Summer 8-1-2021

Evaluating the Contributions of Recollection and Familiarity on Testing and Guessing Benefits in Recognition Memory

Kendal Smith

EVALUATING THE CONTRIBUTIONS OF RECOLLECTION AND FAMILIARITY ON TESTING AND GUESSING BENEFITS IN RECOGNITION MEMORY

by

Kendal A. Smith

A Dissertation
Submitted to the Graduate School,
the College of Education and Human Sciences
and the School of Psychology
at The University of Southern Mississippi
in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

Approved by:

Dr. Mark Huff, Committee Chair
Dr. Alen Hajnal
Dr. Hans Stadthagen
Dr. Lin Agler

August 2021

Published by the Graduate School

THE UNIVERSITY OF SOUTHERN MISSISSIPPI

ABSTRACT

Completing an interpolated memory test or trying to guess non-studied information following study has yielded powerful memory benefits relative to restudy on a final memory test. Across repetitions of testing and guessing, participants may form an expectancy of an upcoming test type, and consequently, adjust their encoding of information in anticipation of the upcoming test. Research has shown that after several task repetitions, participants form an expectancy of the upcoming task type and will encode materials to match task constraints (Huff, Yates, and Balota, 2018). It is uncertain to what extent these expectancy processes aid in facilitating recollection of specific details of studied items or improves familiarity. My dissertation evaluated the contribution of expectancy processes involved in testing and guessing effects on memory by estimating recollection and familiarity processes using the remember/know procedure. Recollection and familiarity processes were estimated under conditions in which expectancy processes were eliminated due to random presentation of restudy, testing and guessing tasks (Experiment 1), or encouraged by having participants repeat restudy, testing, or guessing tasks either 6 (Experiment 2), or 18 (Experiment 3) times consecutively. Testing and guessing benefits were greatest following consecutive task repetitions indicating the presence of task expectancies. Additionally, task expectancies affected recollection of list items similarly to overall correct recognition, whereas expectancy effects on familiarity with critical items was consistent with overall false recognition. Thus, expectancy processes in memory reflect a combination of familiarity- and recollection-based processes. Discussion focuses on repeated testing and guessing as potential strategies to facilitate student performance in educational settings.

ACKNOWLEDGMENTS

First and foremost, I would like to acknowledge my advisor Dr. Mark Huff for his support and extensive contributions to this project. I am thankful for everything that I have learned from working closely with him, and I am truly grateful to have been given the opportunity.

Thank you, as well, to my fellow graduate students both in the Memory, Attentional Control, and Aging lab, especially Nicholas Maxwell, Matthew Gretz, and Jacob Namias. Thank you all for your contributions to my research and for your emotional and intellectual support. Additionally, I would like to thank several other graduate students in the Brain and Behavior program here at the University of Southern Mississippi that supported this project through participating in pilot versions of this study.

I would also like to acknowledge the support and contributions from all of my contributions. Their support overseeing the progression of this project made the completion of my degree possible.

DEDICATION

This dissertation project is dedicated to the memory of Dr. Stanley Kuczaj II, as well as to all of the former members of the Marine Mammal Behavior and Cognition Laboratory. Without your influence as a researcher and most importantly a mentor, I would have never started down this path. Thank you for believing in my potential, and I hope that you are proud of everything that I have accomplished thus far. Thank you to all of the former members of the Marine Mammal Behavior and Cognition Laboratory and especially to the Dolphin League. The emotional and professional support that I have received from every individual has inspired me to continue down this path and finally cross the finish line, despite encountering so many setbacks along the way.

I would also like to dedicate this dissertation project to Michael Vines. You were the foundation during the adversity I have faced. Thank you so much for supporting me in this journey and holding me accountable throughout this entire process. Words can't describe how much I appreciate all that you have done to keep me focused and sane.

Finally, this dissertation project is dedicated to my son, Luca, and to my future daughter. Luca, you are the bravest, most independent, and most headstrong person I know. I hope that you are always this stubborn, and I know that no matter what you set your mind to, you will always achieve it. To my future daughter, I cannot wait to meet you, and I'm so excited that you crossed the finish line with me. I hope you know that no matter what adversity you face in your life and no matter what obstacles stand in your way, you can do anything you set your mind to. I love you both more than you can imagine.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER I - INTRODUCTION

A commonly researched topic for Cognitive Psychologists is how memory, or an individual's ability to encode, store, and retrieve previously stored information, can be improved. Researchers have found enhanced memory benefits for study techniques that promote elaborative or "deep" processing (Craik, 2002; Craik & Lockhart, 1972), study materials that are spaced over time versus massed (Glenberg, 1979; Greene, 1989), and materials that are perceived as distinctive versus non-distinctive (Huff, Bodner, & Fawcett, 2015; Hunt, 2006), including when information is emotionally charged (Schmidt, Patnaik, & Kensinger, 2011). Importantly, techniques that are employed after initial study, such as engaging in retrieval practice or attempting to guess related materials, have also been fruitful in improving later memory (Huff, Balota, & Hutchison, 2016; Roediger & Karpicke, 2006; Rowland, 2014). What is less clear, however, is *how* testing and guessing benefits and the expectancies that may arise when such tasks are repeated, may affect qualitative processes such as by enhancing recollection or familiarity. The purpose of the current study is to provide a close examination of task-expectancy processes following testing and guessing tasks by measuring their effects on recognition and recollection/familiarity processes using the classic remember/know procedure (Tulving, 1985; Yonelinas, 2002).

Test-Expectancy Processes

The effects of expectancy processes on human behavior are well-documented, particularly in the medical sciences. The placebo effect, or behavioral changes that can occur following the administration of a biologically inert substance, is often accounted for by an expectancy-based process (see Stewart-Williams & Podd, 2004, for review).

Though placebo responses are often discussed in the context of medical interventions, it is reasonable to suspect that similar expectancy-based processes may also affect memory performance, particularly situations in which individuals are aware of how memory will be assessed. For instance, knowledge that an upcoming test contains multiple-choice questions which involve discrimination-based retrieval processes may lead to different processing at study than essay questions which require more organization-based/generative retrieval processes. This knowledge may therefore lead to qualitative adjustments in study-based processes in anticipation of an upcoming test.

Despite an intuitive view that individuals may adjust their cognitive processing of study materials in response to a future test, relatively little research has been conducted on how expectancy processes can shape later performance (see Lundeberg & Fox, 1991; Finley & Benjamin, 2012, for reviews). Expectancies likely play a crucial role in how one strategically encodes and processes study materials in terms of both quantitative and qualitative types of processing. Quantitative changes refer to the amount of effort expended towards study in anticipation of an upcoming test (e.g., increased encoding time, repetitions, etc.), whereas qualitative changes refer to adjustments in encoding strategies to process different kinds of information or to organize information more effectively (Neely & Cho, 2014; Tversky, 1973). As an example of a qualitative change, Einstein & Hunt (1980) found that relational processing at study was more likely to promote organizational retrieval strategies, which are beneficial on a recall test, versus item-specific processing, which is more likely to promote discrimination processes that benefit recognition. Qualitative expectancy-driven effects may therefore maximize transfer-appropriate processing in which the effectiveness of a given encoding strategy is contributed to its similarity with the processes that are instantiated at retrieval (Blaxton, 1989; Morris, Bransford, & Franks, 1977; Roediger, Weldon, & Challis, 1989).

Finley and Benjamin (2012) have argued that a disordinal test-expectancy interaction would provide evidence that participants adjust their encoding strategy based on their test expectancies. This interaction refers to a cross-over pattern in which memory performance is greater under conditions in which an individual completes an expected versus an unexpected test. Several test-expectancy studies have similarly predicted a disordinal interaction. Balota and Neely (1980; see too, Neely & Balota, 1981) told their participants whether to expect a recall or recognition test after study of a list of words followed by six practice study-test cycles in which they would complete the expected test type. Using a within-subject design, participants would then complete a final study-test cycle in which the test completed either matched the practiced test (i.e., expected) or mismatched the practice test (i.e., unexpected). Inconsistent with a disordinal interaction, expectation of an upcoming recall test led to improved memory performance both when the final test was an expected recall test or an unexpected recognition test, demonstrating that the expectation of a recall test may produce a qualitative encoding difference that benefits both test types.

Although Balota and Neely (1980) and Neely and Balota (1981) did not find a disordinal interaction, these patterns have been reported in other studies. Postman and Jenkins (1948) reported a disordinal-expectancy effect where those who expected recall tests performed better on expected recall tests and worse on unexpected recognition tests, and vice versa when recognition tests were expected and recall tests were unexpected. Rather than manipulating task expectancy within-subjects as Balota and Neely (1980), Postman and Jenkins (1948) manipulated task expectancy between-subjects. Additionally, study lists differed between the two experiments. Whereas Balota and Neely (1980) had participants repeat six study-test cycles, each using a different set of words, Postman and Jenkins' (1948) repeated the same lists of words five times at study. Thus, repetition of the same words may have contributed to the disordinal pattern. Collectively, the results of both studies indicate that test-expectancy processes develop over repeated study/test cycles, and these expectancies facilitate memory performance, though these benefits may not necessarily produce a disordinal pattern.

Relatedly, Finley and Benjamin (2012) presented participants with 4 study-test cycles of either a cued-recall or a free-recall test for sets of word pairs, to induce an expectancy of a specific test format. Participants were then either given an expected test format or an unexpected test format followed by a self-report questionnaire to determine the influence of their expectancies on their test results. A disordinal-expectancy effect for the final test scores was found and importantly, self-report questionnaires indicated that participants who expected a free-recall test focused on individual studied words versus those who expected a cued-recall test focused on associating word pairs together—a qualitative processing difference. In a subsequent experiment, expectancies were also found to affect the amount of time participants spent studying items. Encoding time was greater for unrelated (vs. related) word pairs that were more difficult to recall when participants anticipated a cued-recall test, suggesting that participants invested additional efforts in generating associations between words that were semantically unrelated. Taken together, the results from these experiments indicated that participants performed better when tested using the expected test types after completing multiple study cycles and adjusted their encoding strategies based on these expectancies.

More recently, disordinal patterns have also been found when specific expectancy processing, such as the processing of semantic or orthographic features, either match or mismatch processing utilized on a final cued-recall test. Cho and Neely (2017) had participants complete four study/test cycles with either a semantically related cue word (e.g. LEG for target word ARM) or an orthographic cue word (e.g., A_M for target word ARM). A final study/test cycle was then completed in which the final test either matched or mismatched the practice tests. A disordinal interaction was again found, demonstrating that expectancy-based processes can develop based on associations between cues and targets with test type held constant. While the literature suggests that expectancy processes develop for anticipated tests, there is also evidence that expectancy processes can develop for upcoming tasks that do not rely upon the explicit use of memory, such as guessing—a discussion with which I will now turn.

Expectancy Processes in Guessing

Research supports that guessing, or having individuals produce information from memory that they have little or no confidence in the accuracy of, can improve retention for information used to make the guess. Pressley, Tanenbaum, McDaniel, and Wood (1990) found that memory for a presented lecture was greater when students were required to answer a set of pre-questions prior to viewing the lecture than when participants did not complete initial questions. The use of pre-questions in this context is noteworthy because participants were not yet knowledgeable of the lecture material and were often incorrect when answering pre-questions. Despite these initial errors, completion of these questions facilitated memory for the lecture. Although guessing through pre-questions benefits memory for subsequent information, the improvements are typically isolated to the specific details inquired by the pre-questions and are accompanied by a memory cost to information that is not pre-tested (see Hamaker, 1986 for review and meta-analysis). Thus, initial guessing does not result in a global memory benefit for all information that is presented subsequently, but only the information that is initially queried.

In addition to guessing processes that occur prior to study, guessing has also been shown to promote memory during study of cue-target pairs. Kornell, Hays, and Bjork (2009; see too Hays, Kornell, & Bjork, 2013; Zawadzka & Hanczakowski, 2018) presented participants with cue-target word pairs that were weakly related (e.g., *whale – mammal*). The pairs were studied as either intact pairs or with the target missing with instructions to guess the word in a within-subject design. Given the weak associates, participants were rarely correct in guessing the target (only 4-5% of targets were correctly guessed). However, participants were always presented with corrective feedback by viewing the intact cue-target pair. Thus, participants always studied an intact pair, but whether the pair was initially guessed or not was manipulated. On a final cued-recall test, correct recall was greater following initial guessing of the cue-target pair versus studying an intact pair, demonstrating that initial guessing paired with corrective feedback facilitated later memory. Subsequent studies have shown that these guessing benefits are reliable, occurring when cue-target pair types were manipulated between-subjects (Kornell et al., 2009), when the final cued-recall test was completed after short and long retention intervals (5 min vs. 24 hours; Yan, Yu, Garcia, & Bjork, 2014), and when guessing was used to learn novel foreign-language pairs (Potts & Shanks, 2014).

Though guessing has been shown to enhance memory for word pairs when accompanied by corrective feedback, it has also been shown to benefit retention for information that is used to derive the initial guess. Huff, Balota, and Hutchison (2016) presented participants with lists of words that were either categorically related, weakly related through ad hoc categories (i.e., things made of wood, things that are black; Hunt & Einstein, 1981, van Overschelde, Rawson, & Dunlosky, 2004), or were completely unrelated. Immediately following study, participants completed one of four between-subject tasks: Completed an arithmetic filler task, restudied the same list of words in a different order, completed a free-recall test of the list of words, or attempted to guess a set of critical words that were related to the studied lists but not actually presented themselves. Participants completed six study-task repetitions which were then followed by a final recognition test. The restudy and free recall tasks were included as standard retrieval-practice comparison groups to gauge the effectiveness of guessing (e.g., Roediger & Karpicke, 2006; Rowland, 2014). On the final test, correct recognition was lowest following the filler and restudy tasks, but greatest following both the recall and guessing tasks—a replication of the retrieval-practice effect and importantly, a demonstration that attempting to guess related information that was not studied could similarly improve recognition. Furthermore, testing and guessing benefits were found across list types, suggesting that the association of the materials may not be critical to whether tasks completed after study improve memory.

Importantly, in a subsequent experiment, Huff et al. (2016) examined whether benefits of testing and guessing over restudy and a filler task were due to the act of completing the task itself, or instead, reflected expectancy-based processes that developed due to successive task repetitions that occurred across the six study-task repetitions. The authors reasoned that task repetitions may have encouraged participants to adjust their study strategies in anticipation for the upcoming test. Thus, the guessing and testing benefits found initially could reflect expectancy processes that may have influenced encoding processes rather than by the completion of the tasks themselves. To eliminate the likelihood that participants were engaging in task-expectancy processes, Huff et al. (2016) manipulated task type within-subjects (restudy, recall, or guessing) and presented the instructions for the task randomly and only after the study list was presented. Under these conditions, participants would have no knowledge of what task would be completed until after study and would therefore be more likely to process all lists similarly without the aid of expectancies. The authors found that guessing and retrieval-practice benefits on the recognition test were eliminated, suggesting that task effects found previously may have been due to participants' anticipatory processes which qualitatively affected study.

More recently, Huff, Yates, and Balota (2018) further investigated these expectancy processes by examining how task repetitions and the presentation of instructions before and after study lists affected testing and guessing benefits. The authors found that presenting tasks randomly (i.e., unexpectedly) failed to produce guessing and retrieval-practice effects compared to restudy both when task instructions were presented after study (Experiment 1) and when instructions were presented before study so participants would be aware of the upcoming task type (Experiment 2). Guessing and retrieval-practice effects were also not found when participants were given task instructions before study and repeated the task 3 times (Experiment 3); however, these benefits emerged when participants completed 6 task repetitions consecutively (Experiments 4A and 4B). Collectively, these patterns suggest that task benefits are tied to task repetitions that may increase participants' expectancies for the upcoming task.

Although Huff et al. (2018) found evidence for testing and guessing expectancy effects on final recognition, it is unclear what memorial processes specifically were affected. In particular, task-expectancy processes could affect recollective type processes, which are based on an individual's conscious retrieval, or could affect familiarity-based processes, which are based on an individual's assessment of processing fluency in the absence of remembering contextual qualitative details. Therefore, the purpose of the current study was to examine how task expectancy processes that likely build through task repetition may affect recollection and familiarity memorial phenomenology.

Recollection and Familiarity

The dual-process model of recognition memory posits that recognition decisions are made using either recollection- or familiarity-based memory processes (Mandler, 1980; see Yonelinas, 2002 for review). Recollection refers to the conscious retrieval of previously encountered information which is accompanied by situational and contextual details of the past encounter. In contrast, familiarity refers to a vague feeling of remembering previously encountered information which is based on stimulus similarity and/or processing fluency (Higham & Vokey, 2004; Jacoby, 1984). Recognition that is based on recollection provides context-rich information whereas recognition based on familiarity may be relatively context-free.

Although there are many unanswered questions about the mechanism(s) behind recollection and familiarity, researchers generally agree on several points. First, familiarity and recollection do not appear to operate sequentially in which one process becomes activated after the first ends. Instead, recollection and familiarity appear to operate in parallel (Tulving, 1985). Second, familiarity appears to be a faster and more automatic retrieval process whereas recollection is a more controlled search process (Yonelinas, 2002). Third, familiarity has been demonstrated to decay more rapidly than recollection, although there is disagreement as to when the decaying process begins varying between immediately after encoding (Eichembaum, Otto, & Cohen., 1994) to several weeks after encoding (Mandler, 1980). In contrast, recollection remains relatively stable over short intervals. After much longer delays, recollection and familiarity begin to exhibit similar levels of decay. Finally, there is evidence that recollection and familiarity are independent processes. Most neuroanatomical dual-process models identify different brain regions underlying recollection and familiarity processes (Aggleton & Brown, 1999). Factors affecting and dividing attention at encoding more drastically decrease recollection judgments than familiarity judgments, although both processes are ultimately affected. Thus, familiarity processes are less attention-demanding versus recollection processes.

Recollection and familiarity have been found to display different response patterns across a variety of manipulations, suggesting that they provide independent contributions to memory. For instance, deep processing tasks have been shown to affect recollection greater than familiarity, as have studying pairs of words with strong semantic associations rather than more shallow rhyme associations (Gardiner, 1988). Additionally, recollection judgments appear to be affected more by semantic encoding, whereas familiarity judgments appear to be affected more by perceptual encoding (Atkinson & Westcourt, 1975; Yonelinas, 2002). Conceptual manipulations (e.g., generation vs. reading, deep vs. shallow processing, etc.) have been shown to affect both recollection and familiarity, though the effects are generally much larger for recollection estimates. Additionally, retrieval manipulations, such as dividing attention at test have also been shown to affect both processes, again with recollection judgments being more sensitive than familiarity judgments.

A common method of estimating the contributions of recollection and familiarity on recognition is through the remember/know procedure (Tulving, 1985). In this procedure, participants are instructed to introspect on their retrievals and identify whether they recognize a previously studied item based on recollection or familiarity. Specifically, participants are instructed to indicate whether they "remember" or "know" a memory item by qualifying the retrieval phenomenology that accompany items believed to be studied. Standard instructions ask participants to respond with a "remember" judgment if retrieval of an item is accompanied by vivid contextual details of the item's presentation, whereas participants are to respond with a "know" judgment if retrieval of an item is familiar but contextual details are absent (Perfect, Mayes, Downes & Va Eijk, 1996; Rajaram, 1993). Although the remember/know paradigm relies on self-report, studies have shown that responses using the remember/know paradigm are not due to memory confidence (Gardiner & Gregg, 1997: Rajaram, 1993), and remember/know responses accurately gauge recollection and familiarity processes (Higham & Vokey, 2004).

As Yonelinas (2002) suggested, know responses in the remember/know procedure may actually underestimate the familiarity of tested items, as participants are instructed to respond know if they remember the item but cannot recollect specific details rather than to simply respond if the item is familiar to them. A method to correct for underestimation is to apply the independent remember/know correction (Yonelinas & Jacoby, 1995). Using this computation, remember responses are assumed to be a direct probability of recollection of the items (remember = R). Since know responses reflect the absence of recollection (know = $F(1-R)$), the probability of a familiarity response is equivalent to the probability that the item received a know response because it was not recollected (F = know/$(1-R)$).

Given the separation of recollection and familiarity, a critical question that has yet to be addressed is how expectancy processes, especially those that result from repeated testing or guessing tasks, may affect one or both of these memory processes. To ensure a comprehensive examination of the effects of task expectancy on memory phenomenology, Experiment 1 of my dissertation extended the work of Huff et al. (2018; Experiment 1), in which task type was manipulated within-subjects and randomized with task instructions presented only after each list was studied. This procedure was expected to eliminate expectancy processes and should therefore produce a null effect across task types in overall recognition. In Experiment 2, participants completed the same tasks manipulated within-subjects with the exception that tasks were blocked together by repeating the same task across six study/task cycles followed by a recognition test. This procedure replicated Huff et al. (2018; Experiment 4) in which expectancy effects were encouraged through task repetitions and guessing and recall produced improvements in correct recognition over restudy. Finally, Experiment 3 manipulated task type between-subjects where participants used either restudy, recall, or guessing tasks to study 18 lists. Across experiments, the primary research question of interest was how expectancy processes, or the lack of expectancy processes, affect recollection and familiarity estimates—a novel contribution. To compute recollection and familiarity estimates, participants were required to assign a remember or know judgement to test items recognized as studied, as used in prior work (e.g., Tulving, 1985; Yonelinas & Jacoby, 1995). These effects were examined on both correct recognition of studied list items and a set of related lures (i.e., critical items) to assess false recognition. These recognition item types were included to evaluate overall recognition accuracy when both correct and false recognition were examined.

.

CHAPTER II – EXPERIMENT 1

Experiment 1 was designed to replicate and extend the work of Huff et al. (2018) to evaluate whether the null task effect found previously on correct and false recognition when tasks are completed randomly affect recollection and familiarity estimates. To evaluate the generality of testing and guessing effects on a variety of materials, and to be consistent with previous literature, participants studied lists containing words of varying relatedness to non-studied critical items. Specifically, participants studied either strongly related categorized lists, weakly related lists taken from ad hoc categories (e.g., things made of wood), or unrelated lists. Both correct recognition of the studied list items and false recognition of the non-studied critical items were examined, as categorized lists that are strongly related are likely to elicit higher rates of false recognition for the critical items compared to weakly related lists or unrelated lists. Experiment 1 therefore has two hypotheses. Based on Huff et al., no differences on overall correct recognition of list items or false recognition of critical items are expected across restudy, recall, and guessing task conditions, demonstrating a null effect of task type. However, a task type by list type interaction was not expected, as guessing should reduce false recognition similarly for categorized and ad hoc lists (**H_1**). Despite the null task effect, task-related differences are expected for recollection and familiarity estimates. Specifically, a tradeoff between recollection and familiarity was expected in which recollection will be higher for testing and guessing conditions relative to restudy, whereas restudy is expected to show higher familiarity estimates than testing and guessing. False recognition for critical items is expected to be higher for familiarity across list and task types, whereas recollection should be much lower (**H_2**). This pattern is based on findings reported by Chan and

McDermott (2007) who reported higher recollection estimates for testing than restudy conditions and higher familiarity estimates for restudy conditions than retrieval-practice on correct recognition. If guessing operates similarly as testing, guessing is expected to increase recollection estimates relative to restudy. These patterns are expected on both correct and false recognition.

Methods

Participants

Seventy-eight University of Southern Mississippi undergraduates participated for partial course credit. Participants reported fluency in the English language and normal or corrected-to-normal vision. The sample was primarily female (82%) and reported a mean of 13.18 years of education (*SD* = 1.39, Range = 12-16) and a mean age of 19.40 years (*SD* = 2.32, Range = 18-33). A sensitivity analysis using G*Power3 (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that the sample size had adequate power (.80) to predict medium effect sizes or larger (Cohen's *d* = .40) for main effects and interactions. Additionally, this sample size is consistent with previous studies in the literature examining task effects (Huff et al., 2016; 2018).

Materials

A total of 36 word lists created by Huff et al. (2018) served as study materials. Of these lists, 12 were from strongly related categories (e.g., birds, vegetables, spices, etc.), 12 from weakly related ad hoc categories (e.g., things that are green, things made of wood, liquids, etc.), and 12 that were unrelated. Each list contained 20 items. The Battig and Montague (1969) and van Overschelde et al. (2004) categorical word norms were used to generate categorized and ad hoc lists. Unrelated list words were randomly generated and matched to ad hoc lists based on frequency of occurrence in the English language and word length using the SUBTLEX norms (Brysbaert & New, 2009) in the English Lexicon Project (Balota et al., 2007) and on concreteness using the MRC Psycholinguistic database (Coltheart, 1981). For categorized and ad hoc lists, the top 5 exemplars in the norms were not presented at study and instead were utilized as critical items on the recognition test (see Huff, Meade, & Hutchison, 2011, for a similar procedure). Five randomly generated unrelated words served as critical items for unrelated lists. The 36 lists were subdivided into two sets of 18 lists (in which one set was studied and the other set used as control items) and counterbalanced across participants. Each set was further subdivided into three blocks of six lists, each containing two lists of each type. Three study/test blocks were created to reduce testing fatigue. Lists in each block were once randomized with the qualification that no list type is presented consecutively. The order of the blocks was then counterbalanced across participants.

Three recognition tests were constructed corresponding to each of the three blocks of six lists. Each test contained 180 items. Of these items, 60 served as studied list items (10 from each studied list), 30 as critical lures (5 from each of the studied lists), 60 as studied item controls (10 from each non-studied list from the non-studied set), and 30 as critical lure controls (5 from each of the non-studied lists). Recognition items were presented in a newly randomized order for each participant.

Procedure

The procedure was a replication of Huff et al. (2018, Experiment 1), with the exception that participants were tested on 18 total lists (vs. 9) and responded R/K/N (“remember”/ “know”/ “new”) for recognition test items. Participants were tested individually on a computer using E-Prime 3.0 Software (Psychology Software Tools, Pittsburg, PA). Participants were randomly assigned to one of the counterbalanced versions. They were then presented with an instruction screen that informed them that they will be presented with lists containing 20 study words, each presented for 3 seconds. They were then informed that immediately following each list, they will randomly complete one of three tasks: A restudy task, a free recall task, and a guess task. The *restudy task* presented participants with the same word list again with items presented for the same duration but in a different order. The *recall task* required participants to freely recall the words in any order for 60 s. The *guess task* gave participants 60 s to attempt to guess the five critical words that were related to words presented on the studied list but not actually presented. Participants were required to provide at least one guess, but participants could advance to the next list before the 60 s deadline if they completed the task early. All instructions were presented on a single instruction screen which further informed participants that tasks completed after study will occur randomly, and critically, task instructions will only be provided *after* each study list is presented to limit expectancy effects during encoding. Participants were then presented with the first block of six lists which contained two lists each of the restudy, recall, and guess tasks.

Following the sixth list in each block, participants were given a 180-item recognition test in which individual words were presented on the computer screen and participants were asked to make either R (remember), K (know), or N (new) responses. Instructions were taken from Rajaram (1993) as these instructions for the remember/know procedure have shown to be most effective at communicating to participants the concepts of recollection and familiarity (see Appendix A for full instructions). Participants were instructed to make R responses if they could consciously recollect the item from the study list and if the word was accompanied by a specific memory of the item's appearance on the list, an association with the item, or a specific image of the item. Participants were instructed to make K responses if they recognized the item as studied but could not consciously recollect the actual occurrence of the item or any experience associated with the item's presentation on the list. Participants were instructed to make N responses if they did not recognize the test item.

After the first block, participants repeated the same procedure for two additional blocks. An R/K/N recognition test followed each block for a total of 3 recognition tests. Immediately following the third recognition test, participants completed a demographics questionnaire to assess age and education level followed by a full debriefing. Each experimental session lasted approximately 85 min. All participants were tested in-lab with an experimenter present.

Results

For all analyses reported, a $p < .05$ significance level was used unless noted otherwise. Effect size estimates were provided using partial-eta squared (η_p^2) for analyses of variance (ANOVAs) and Cohen's *d* for *t*-tests. Remember/know responses were adjusted using the independent remember/know method (Yonelinas & Jacoby, 1995) to correct for potential underestimation of familiarity (F = know/(1-R)). Mean raw "know" responses for Experiment 1 are reported in Appendix Table B1.

Interpolated Recall and Guessing

Proportion of correct recall was calculated by dividing the number of correctly recalled non-repeated items recalled by the total of list items studied. Table 1 reports mean proportions of correct and false recall on the interpolated recall test and mean proportions of critical items that were correctly guessed on the interpolated guess task. Starting with correct recall, a repeated-measures ANOVA found a significant difference across list types, $F(2, 158) = 41.46$, $MSE = .37$, $\eta_p^2 = .34$, in which correct recall was higher in both categorized and ad hoc lists relative to unrelated lists (.39 vs. .27), $t(79) = 7.65$, $SEM = .02$, $d = 0.91$, and (.39 vs. . 27), $t(77) = 8.61$, $SEM = .01$, $d = 0.95$, respectively, but recall of categorized and ad hoc lists was equivalent, $t < 1$.

Table 1

Mean (SD) Proportion of Correct Recall of List Items and Correct Guessing of Critical Items by list type in Experiment 1 (interpolated task with instructions following list presentations).

	Interpolated Task Type	
Item Type	Recall	Guess
List Items		
Categorized	.39 (.15)	-
Ad Hoc	.39 (.14)	-
Unrelated	.27 (.11)	-
Critical Items		
Categorized	.05 (.09)	.19 (.15)
Ad Hoc	.01 (.03)	.11 (.12)
Unrelated	.00 (.00)	.00 (.00)

Table 1 continued

Extra-List Intrusions per List		
Categorized	.51 (.61)	-
Ad Hoc	.43 (.61)	-
Unrelated	.62 (.93)	-

Critical item intrusions were similarly analyzed, though their recall was rare. No difference was found across list types, $F(2, 158) = 3.39$, $MSE = .003$, $p = .10$. Mean numbers of extra-list intrusions per list did not differ across list types, $F(2, 158) = 1.85$, $MSE = .66$, $p = .16$.

Proportions of critical items that were correctly guessed were similarly computed by taking the total correctly guessed critical items by the total number of possible critical items. Guessing rates differed across list types, $F(2, 158) = 73.83$, $MSE = .74$, $\eta_p^2 = .48$, and were higher for categorized and ad hoc lists than unrelated lists (.19 vs. .00), $t(77) = 11.45$, $SEM = .02$, $d = 1.79$ and (.11 vs. .00), $t(77) = 8.66$, $SEM = .01$, $d = 1.30$, respectively, and higher on categorized than ad hoc lists (.19 vs. .11), $t(77) = 4.53$, $SEM = .02$, $d = 0.59$. Thus, lists with stronger semantic consistency produced higher rates of correct recall and correct guessing of critical items.

Correct Recognition

Correct recognition was first analyzed by comparing overall "old" responses to correctly studied list items (i.e., hits) collapsed across recollection and familiarity (see Table 2). A 3 (List Type: Ad hoc vs. Categorized vs. Unrelated) × 3 (Task Type: Recall vs. Guess vs. Restudy) repeated-measures ANOVA yielded a main effect of list type, *F*(2, 154) = 42.91, *MSE* = .90, η_p^2 = .36, in which correct recognition was higher for categorized than unrelated lists (.81 vs. .69), *t*(77) = 9.13, *SEM* = .01, *d* = 0.88 and higher for ad hoc than unrelated lists (.76 vs. .69), *t*(77) = 4.56, *SEM* = .01, *d* = 0.50. Correct recognition was also higher in categorized than ad hoc lists (.81 vs. .76), *t*(77) = 4.83, *SEM* = .01, *d* = 0.40. A main effect of task type was also found, *F*(2, 154) = 14.65, *MSE* = .29, η_p^2 = .16. Post hoc comparisons revealed a reversed retrieval-practice effect in which correct recognition of list items was higher for the restudy task compared to both the recall task (.80 vs. .75), *t*(77) = 3.51, *SEM* = .01, *d* = 0.38, and the guess task (.80 vs. .72), *t*(77) = 5.92, *SEM* = .01, *d* = 0.64. Correct recognition of list items was equivalent between the recall and guess tasks (.75 vs. .72), *t*(77) = 1.56, *SEM* = .01, *p* = .22. There was no significant list type × task type interaction, *F*(4, 308) = 1.95, *MSE* = .03, *p* = .10. Thus, consistent with predictions, recall testing and guessing did not produce a correct recognition advantage over restudy, and instead, restudy produced the highest rate of recognition across list types.

Table 2

Mean (SD) Recognition Results of List Items and Critical Items for Remember (R) and Know (K) Responses Combined for Categorized, Ad hoc, and Unrelated Lists as a Function of Interpolated Task Type in Experiment 1 (Randomized Task with Instructions Following List Presentation).

	Interpolated Task		
	Restudy	Recall	Guess
List Items			
Categorized	.84 (.15)	.80 (.16)	.80 (.14)
Controls		.18 (.15)	
Ad hoc	.80 (.16)	.77 (.19)	.70 (.17)
Controls		.20 (.14)	
Unrelated	.74 (.19)	.67 (.19)	.66 (.18)
Controls		.17 (.14)	
Task Average	**.80 (.12)**	**.75 (.14)**	**.73 (.13)**
Critical Items			
Categorized	.45 (.22)	.58 (.25)	.40 (.22)
Controls		.20 (.15)	
Ad hoc	.47 (.29)	.42 (.25)	.35 (.22)
Controls		.22 (.16)	
Unrelated	.22 (.19)	.20 (.18)	.24 (.17)

Table 2 continued

Controls		.19 (.14)	
Task Average	**.38 (.19)**	**.40 (.19)**	**.32 (.19)**

Recollection estimates. Table B2 reports mean recollection estimates across list and task types for studied list items. A 3(List Type: Ad hoc vs. Categorized vs. Unrelated) × 3(Task Type: Recall vs. Guess vs. Restudy) repeated-measures ANOVA revealed a significant effect of list type, $F(2, 152) = 16.65$, $MSE = .56$, $\eta_p^2 = .18$, in which recollection estimates were higher for categorized than unrelated lists (.55 vs. .45), $t(77) = 5.51$, $SEM = .02$, $d = 0.43$ and for ad hoc than unrelated lists (.51 vs. .45), $t(77) = 2.92$, $SEM = .02$, $d = 0.27$. Recollection estimates for categorized lists were higher than ad hoc lists, (.55 vs. .51), $t(77) = 3.11$, $SEM = .01$, $d = 0.17$. A significant effect of task type was also found, $F(2, 152) = 8.50$, $MSE = .75$, $\eta_p^2 = .10$, in which recollection estimates were higher in the restudy task relative to the guess task (.53 vs. .47), $t(77) = 4.05$, $SEM = .02$, $d = 0.25$, and in the recall task relative to the guess task (.52 vs. .47), $t(77) = 2.99$, $SEM = .02$, $d = 0.21$. There was no difference in recollection estimates between the restudy and recall tasks (.53 vs. .52), $t(77) = 1.08$, $SEM = .03$, $p = .29$. The list type × task type interaction was not significant, $F(4, 304) = 1.07$, $MSE = .02$, $p = .24$. As expected, lists with stronger semantic association produced higher recollection of list items. Recollection of list items suffered a cost in the guess task relative to the recall and restudy tasks.

Familiarity estimates. Table B2 reports mean familiarity estimates across list and task types for studied list items. The same ANOVA revealed a significant effect of list type, $F(2, 152) = 13.86$, $MSE = .61$, $\eta_p^2 = .15$, in which familiarity was higher for categorized lists than unrelated lists (.41 vs. .32), $t(77) = 5.11$, $SEM = .02$, $d = 0.30$, and for ad hoc than unrelated lists (.39 vs. .32), $t(77) = 3.91$, $SEM = .02$, $d = 0.25$. Familiarity for categorized and ad hoc lists was equivalent (.41 vs. .39), $t(77) = 1.29$, $SEM = .02$, $p = .20$. No significant effect of task type was found, $F < 1$. Lists with stronger relatedness increased familiarity estimates for list items.

A significant interaction was found, $F(4, 304) = 3.03$, $MSE = .14$, $\eta_p^2 = .04$. For categorized lists, familiarity estimates did not differ across the three task types, all *t*s < 1; however, for ad hoc lists, familiarity estimates were higher in the recall than guess task (.42 vs. .35), $t(77) = 2.20$, $SEM = .03$, $d = 0.21$. Familiarity estimates were equivalent between the restudy and guess tasks (.39 vs. .35), $t(77) = 1.38$, $SEM = .04$, $p = .17$, and the recall and restudy tasks (.42 vs. .39), $t < 1$. For unrelated lists, familiarity estimates were higher in the restudy task than the recall task (.35 vs. .27), $t(77) = 2.76$, $SEM = .03$, $d = 0.26$, and marginally higher in the guess than the recall task (.32 vs. .27), $t(77) = 1.85$, $SEM = .03$, $p = .07$, $d = 0.18$, but were equivalent between the restudy and guess tasks (.35 vs. .32), $t < 1$. Thus, when list items are strongly related, participants likely relied on list relatedness rather than on familiarity processes when identifying list items. Recall improved familiarity of list items only when compared to the guess task for ad hoc lists. However, when lists were unrelated, restudy improved familiarity of list items only compared to testing.

False Recognition

False recognition was first analyzed by comparing overall incorrect “old” responses to non-presented critical lures (i.e., false alarms; see Table 2). A 3 (List Type)

× 3 (Task Type) repeated-measures ANOVA found a main effect of list type, $F(2, 154) = 105.85$, $MSE = 4.23$, $\eta_p^2 = .58$, in which false recognition was higher for categorized than unrelated lists (.48 vs. .22), $t(78) = 13.81$, $SEM = .02$, $d = 0.33$ and higher for ad hoc than unrelated lists (.41 vs. .22), $t(78) = 10.29$, $SEM = .02$, $d = 0.99$. False recognition was higher in categorized lists than ad hoc lists, (.48 vs. .41), $t(78) = 3.45$, $SEM = .02$, $d = 0.33$. A main effect of task type was also found, $F(2, 154) = 13.93$, $MSE = .33$, $\eta_p^2 = .15$, in which false recognition was lower in the guess task relative to both the restudy task (.33 vs. .38), $t(78) = 3.79$, $SEM = .01$, $d = 0.28$, and the recall task (.33 vs. .40), $t(78) = 5.16$, $SEM = .01$, $d = 0.39$. There was no difference in false recognition between the restudy and recall tasks (.38 vs. .40), $t(78) = 1.10$, $SEM = .01$, $p = .27$. As lists became more semantically related, false recognition of critical items increased. As expected, when asked to guess the critical items from study lists, participants were able to monitor and correctly reject these critical items at test significantly better versus the test and restudy tasks.

A significant list type × task type interaction was found, $F(4, 308) = 16.90$, $MSE = .34$, $\eta_p^2 = .18$. For categorized lists, false recognition was lowest in the guess task relative to both the restudy task (.40 vs. .45), $t(77) = 2.62$, $SEM = .02$, $d = 0.23$, and the recall task, (.40 vs. .58), $t(77) = 6.48$, $SEM = .03$, $d = 0.76$. False recognition was also lower in the restudy than recall task, (.45 vs. .58), $t(77) = 4.82$, $SEM = .03$, $d = 0.55$. For ad hoc lists, false recognition was lower for the guess task compared to both the restudy task (.35 vs. .47), $t(77) = 4.53$, $SEM = .03$, $d = 0.47$ and the recall task (.35 vs. .42), $t(77) = 3.55$, $SEM = .02$, $d = 0.30$; false recognition in the restudy task was marginally higher than the recall task (.47 vs. .42), $t(77) = 1.77$, $SEM = .03$, $p = .08$, $d = 0.18$. For unrelated items, false recognition was only lower in the recall versus the guess task (.24 vs. .20), $t(77) = 2.31$, $SEM = .02$, $d = 0.23$, with all other comparisons non-significant, $t < 1$. Regardless of strength of associations, when list items were semantically related, guessing the critical items at study significantly reduced false recognition.

Recollection estimates. Table B2 reports mean recollection estimates for critical items across list and task type. A significant effect of list type was found, $F(2, 154) = 38.55$, $MSE = 1.32$, $\eta_p^2 = .33$ in which recollection estimates were higher for categorized than unrelated lists (.22 vs. .08), $t(78) = 7.42$, $SEM = .02$, $d = 0.89$. Recollection estimates were also higher for ad hoc than unrelated lists (.19 vs. .08), $t(78) = 6.18$, $SEM = .02$, $d = 0.72$. Finally, recollection estimates were higher for categorized than ad hoc lists, (.22 vs. .19), $t(78) = 2.36$, $SEM = .01$, $d = 0.16$. An effect of task type, $F(2, 154) = 11.14$, $MSE = .15$, $\eta_p^2 = .13$, indicated that estimates were lower in the guess task compared to both the recall task (.14 vs. .19), $t(78) = 4.24$, $SEM = .01$, $d = 0.34$, and the restudy task (.14 vs. .18), $t(78) = 3.69$, $SEM = .01$, $d = 0.27$. There was no difference in recollection estimates between the restudy and recall tasks (.18 vs. .19), $t < 1$.

A significant interaction was found, $F(4, 308) = 13.20$, $MSE = .16$, $\eta_p^2 = .15$. For categorized lists, recollection estimates were lower in the guess than recall task (.18 vs. .29), $t(77) = 5.37$, $SEM = .02$, $d = 0.51$, and marginally lower in the guess than restudy task (.18 vs. .21), $t(77) = 1.90$, $SEM = .02$, $p = .06$, $d = 0.15$. Recollection estimates were also lower in the recall than restudy tasks (.29 vs. .21), $t(77) = 3.56$, $SEM = .02$, $d = 0.35$. For ad hoc lists, estimates were similarly lower in the guess task than both the restudy task (.14 vs. .24), $t(77) = 4.67$, $SEM = .02$, $d = 0.46$, and the recall task (.14 vs. .20), $t(77) = 3.06$, $SEM = .02$, $d = 0.31$, and lower in the recall task relative to the restudy task (.20

vs. .24), *t*(77) = 2.15, *SEM* = .02, *d* = 0.17. For unrelated lists, recollection estimates were higher in the guess task compared to the recall task (.10 vs. .07), *t*(77) = 2.23, *SEM* = .01, *d* = 0.26, and marginally higher than the restudy task (.10 vs. .08), *t*(77) = 1.79, *SEM* = .01, *d* = 0.17. Recollection estimates between the recall and the restudy tasks were equivalent, *t* < 1. Overall, recollection estimates followed a similar pattern as overall false recognition, with semantic relatedness of lists increasing recollection of critical items. Likewise, guessing at study significantly decreased recollection of critical items at test. Similar to overall false recognition patterns, guessing at study significantly decreased recollection of critical items for related lists, whereas for unrelated lists, testing at study significantly decreased recollection of critical items.

Familiarity estimates. Table B2 reports mean familiarity estimates for critical items. An effect of list type was found, *F*(2, 154) = 39.20, *MSE* = 1.26, η_p^2 = .33, in which familiarity estimates were higher in categorized than unrelated lists (.24 vs. .10), *t*(77) = 7.58, *SEM* = .02, *d* = 0.79, and for ad hoc than unrelated lists (.21 vs. .10), *t*(77) = 6.16, *SEM* = .02, *d* = 0.67. Familiarity estimates were also higher in categorized lists than ad hoc lists, (.24 vs. .21), *t*(77) = 2.74, *SEM* = .01, *d* = 0.14. A significant effect of task type was also found, *F*(2, 154) = 6.18, *MSE* = .12, η_p^2 = .07, in which familiarity estimates were lower in the guess task than both the recall task (.16 vs. .20), *t*(77) = 3.29, *SEM* = .01, *d* = 0.23, and the restudy task (.16 vs. .19), *t*(77) = 2.57, *SEM* = .01, *d* = 0.18. There was no difference in familiarity estimates between the restudy and recall tasks (.19 vs. .20), *t* < 1.

Effects of list type and task type were qualified by a significant interaction, *F*(4, 308) = 5.62, *MSE* =.10, η_p^2 = .07. For categorized lists, familiarity estimates were lower in the guess than the recall task (.20 vs. .31), *t*(77) = 3.66, *SEM* = .03, *d* = 0.41, but not relative to the restudy task (.20 vs. .22), *t*(77) = 1.07, *SEM* = .02, *p* = .29. Familiarity estimates were also lower in the restudy than recall task (.22 vs. .31), *t*(77) = 3.59, *SEM* = .02, *d* = 0.33. For ad hoc lists, familiarity estimates were lower in the guess task relative to the restudy task (.18 vs. .23), *t*(77) = 2.07, *SEM* = .03, *d* = 0.21, but did not differ between the guess and recall tasks (.18 vs. .23), *t*(77) = 1.27, *SEM* = .02, *p* = .21, or the restudy and recall tasks (.23 vs. .20), *t*(77) = 1.00, *SEM* = 03, *p* = .32. For unrelated lists, familiarity estimates in the recall task were only marginally lower than the restudy task (.09 vs. .12), *t*(77) = 1.76, *SEM* = .01, *p* = .08, *d* = 0.21, with all other comparisons non-significant, *t*s < 1.43. Overall, familiarity estimates followed a similar pattern as recollection estimates, with list type relatedness increasing familiarity of critical items. Guessing also reduced familiarity of non-presented critical items at test compared to the other tasks. For categorized lists, guessing reduced familiarity of critical items compared to only testing, but not to the restudy task. However, for ad hoc lists, guessing only reduced familiarity of critical items compared to the restudy task, but not to testing. When lists were unrelated, testing only reduced familiarity of critical items compared to the restudy task.

Discussion

Experiment 1 replicated the findings of Huff et al. (2018; Experiment 1), in which recall and guessing benefits were eliminated when tasks were unpredictable through randomization of task types and presentation of task instructions after study. A reversed retrieval-practice effect was again found, where overall correct recognition was higher for restudy lists than recall and guessing lists. As predicted in the first hypothesis, no effect of correct recognition on list type was found, and there was no interaction of list type with task type. To evaluate the mechanisms behind task-related differences, Experiment 1 also assessed recollection and familiarity estimates. As predicted by the second hypothesis, eliminating task expectancies differentially affected recollection and familiarity estimates, where recollection estimates were similarly higher on the restudy task, but only relative to the guess task. Regarding list type, when lists were strongly related, participants appeared to rely on list item relationships to inform their recognition decisions, rather than recollection of those items. Familiarity for list items that were weakly related demonstrated a benefit of the recall task, whereas familiarity for lists with no associations demonstrated a benefit of the restudy and guess task.

Turning to false recognition of critical items, false recognition was found to be higher for strongly related categorized lists than ad hoc and unrelated lists, and these patterns were sensitive to task type. As predicted in $\mathbf{H_1}$, guessing was found to reduce false recognition, but only when critical lures were more likely to be identified in categorized and ad hoc lists. This interaction pattern replicated Huff et al. (2018) in which guessing also reduced false recognition on categorized and ad hoc lists. False recognition processes were further examined using recollection and familiarity estimates. Unlike with correct recognition, eliminating task expectancies produced similar false recognition patterns for recollection and familiarity estimates. Recollection and familiarity estimates closely matched the patterns of overall false recognition, where both were highest for categorized lists. As predicted by $\mathbf{H_2}$, recollection and familiarity estimates were similarly affected by task type, where guessing reduced false recognition of critical lures. As semantic relatedness of study lists increased, recollection and familiarity estimates of critical items were reduced by the guessing task, demonstrating a memorial benefit of guessing at test for reducing false recognition of related information. However, guessing did not effectively reduce recollection or familiarity estimates for critical lures for information that was completely unrelated, partially supporting $\mathbf{H_2}$.

CHAPTER III - EXPERIMENT 2

As demonstrated in the literature, because testing and guessing benefits over restudy do not occur when task expectancies are eliminated, it is reasonable that the benefits of testing and guessing may be dependent upon the individual's expectancies for completing those tasks (e.g., Huff et al., 2018). The remaining experiments therefore examine testing and guessing effects on correct and false recognition when participants were encouraged to develop task expectancies. Based on Huff et al. (2016), task expectancies appear to occur when tasks are repeated consecutively which increases participants' awareness of the upcoming task type. Indeed, the magnitude of testing and guessing benefits on recognition have been shown to be positively related to the number of task repetitions participants receive (Huff et al., 2018 Experiments 3 and 4). The purpose of Experiment 2 was therefore to encourage task-expectancy processes through task repetitions and gauge their effects on recollection and familiarity estimates. Huff et al. (2016; 2018) found that blocking tasks together such that they were completed repetitively increased task-related effects such that interpolated recall and guessing tasks produced both correct and false recognition benefits over restudy. Therefore, Experiment 2 will attempt to replicate Huff et al. (2018; Experiment 4) with the addition of remember and know judgments to assess the contributions of recollection and familiarity on task effects for correct recognition and false recognition. Based on previous findings, I expected (**H_1**) that blocking recall and guessing tasks should produce an increase in correct recognition and a decrease in false recognition relative to restudy—a *mirror effect* pattern (Glanzer & Adams, 1990). Regarding correct recognition, recollection estimates should increase for words presented in the free recall and guessing blocks, whereas familiarity estimates should not differ by task type. For false recognition, familiarity estimates are expected to be higher for critical items than recollection estimates, but familiarity estimates are not expected to differ by task type (**H_2**). This prediction is based on Chan and McDermott (2007) in which initial testing was found to increase recollection estimates and reduced familiarity estimates on final recognition.

Methods

Participants

Eighty-four students from The University of Southern Mississippi served as research participants for Experiment 2. Participants were either compensated with partial course credit or a $10 gift card. Participants reported fluency in the English language and normal or corrected-to-normal vision. Participants were primarily female (72%), reported a mean of 13.98 years of education (*SD* = 1.91, Range = 12-19), and a mean age of 21.20 years (*SD* = 4.59, Range = 18-48). The G*Power3 (Faul et al., 2007) sensitivity analysis indicated that the sample size had adequate power (.80) to predict medium effect sizes or larger (Cohen's *d* = .40) for main effects and interactions.

Materials and Procedure

Materials and procedures from Experiment 1, including the R/K/N recognition test instructions, were again used with the following exceptions. First, participants now completed the same study task repeatedly over a block of 6 lists. Of the 18 total lists, 6 were blocked together for the guess task, 6 for the recall task, and 6 for the restudy task. Each block contained 2 lists of each of the 3 list types (categorized, ad hoc, and unrelated). As in Experiment 1, a 180-item recognition test was created for each block that contained 60 studied list items (10 selected from each studied list), 30 critical lures (5 from each of the studied lists), 60 studied item controls (10 selected from each non-studied list from the non-studied set), and 30 critical lure controls (5 from each of the non-studied lists). The lists used in each task block were once randomized and arranged into 3 versions that were counterbalanced across participants, totaling 6 different versions of the experiment. Task orderings across the three blocks (i.e., guess, recall, restudy; restudy, guess, recall, etc.) were similarly counterbalanced across participants. Additionally, task instructions were provided at the beginning of each task block, and these instructions were repeated *before* the presentation of each study list to enhance expectancy processes on memory.

Results

Similar analyses were conducted as in Experiment 1, including collapsing across test blocks for analyses and computing recollection and familiarity estimates. Mean raw "know" responses for Experiment 2 are reported in Appendix Table B3.

Interpolated Recall and Guessing

Starting with mean proportions of correct recall, a significant difference was found across list types, $F(2, 166) = 86.86$, $MSE = .47$, $\eta_p^2 = .51$ (Table 3). Like Experiment 1, correct recall was higher in both categorized and ad hoc lists relative to unrelated lists (.46 vs. .32), $t(84) = 13.43$, $SEM = .01$, $d = 1.40$ and (.44 vs. .32), $t(84) = 9.74$, $SEM = .01$, $d = 1.09$, respectively, and correct recall of categorized items was higher than ad hoc items (.46 vs. .44), $t(84) = 2.10$, $SEM = .02$, $d = 0.18$.

Table 3

Mean (SD) Proportion of Correct Recall of List Items and Correct Guessing of Critical Items by List Type in Experiment 2 (Task Blocked in 6 Repetitions with Instructions Prior to List Presentation).

	Interpolated Task Type	
Item Type	Recall	Guess
List Items		
Categorized	.46 (.10)	-
Ad Hoc	.44 (.12)	-
Unrelated	.32 (.10)	-
Critical Items		
Categorized	.10 (.09)	.22 (.18)
Ad Hoc	.08 (.12)	.10 (.08)
Unrelated	.00 (.00)	.00 (.01)
Extra-List Intrusions per List		
Categorized	.31 (.49)	-
Ad Hoc	.34 (.60)	-
Unrelated	.49 (.83)	-

Critical item intrusions also differed across list types, $F(2, 166) = 4.43$, $MSE = .03$, $\eta_p^2 = .29$, and were higher for categorized than unrelated lists (.10 vs. .00), $t(84) = 3.55$, $SEM = .03$, $d = 1.57$, and higher for ad hoc than unrelated lists (.08 vs. .00), $t(84) = 2.17$, *SEM* = .04, *d* = 0.94. Critical item intrusions did not differ between categorized and ad hoc lists (.10 vs. .08), $t < 1$.

Mean number of extra-list intrusions per list were similarly analyzed and were found to differ across list types, $F(2, 166) = 3.70$, $MSE = .77$, $\eta_p^2 = .04$. Intrusions were lower for categorized than unrelated lists (.31 vs. .49), $t(84) = 2.69$, $SEM = .06$, $d = 0.26$, but equivalent to ad hoc lists (.31 vs .34), $t < 1$. Intrusions on unrelated lists were marginally higher than ad hoc lists (.49 vs .34), $t(84) = 1.87$, $SEM = .08$, $p = .07$, $d = 0.21$.

Correct guessing rates also differed across lists, $F(2, 166) = 86.74$, $MSE = .99$, $\eta_p^2 = .51$. Correct guessing was higher for both categorized and ad hoc lists relative to unrelated lists (.22 vs. .00), $t(84) = 11.01$, $SEM = .02$, $d = 1.73$, and (.09 vs. .00), $t(84) = 9.96$, $SEM = .01$, $d = 1.58$, respectively, and were higher for categorized lists than ad hoc lists (.22 vs. .09), $t(84) = 6.77$, $SEM = .02$, $d = 0.93$. Thus, consistent with Experiment 1, lists that were semantically related produced an increase in correct recall, an increase in false recall, and an increase in correct guessing of critical items.

Correct Recognition

Correct recognition was first analyzed by comparing overall “old” responses collapsed across recollection and familiarity (see Table 4). A 3(List Type) × 3(Task Type) repeated-measures ANOVA found an effect of list type, $F(2, 166) = 33.44$, $MSE = .71$, $\eta_p^2 = .29$, in which correct recognition was higher for both categorized lists and ad hoc lists compared to unrelated lists (.79 vs. .69), $t(83) = 8.18$, $SEM = .01$, $d = 0.59$, and (.76 vs. .69), $t(83) = 5.22$, $SEM = .01$, $d = 0.41$, respectively. Correct recognition was also higher for categorized than ad hoc lists (.79 vs. .76), $t(83) = 2.66$, $SEM = .01$, $d = 0.19$.

Unlike Experiment 1, there was no main effect of task type, $F < 1$, and no significant list type × task type interaction, $F(4, 332) = 1.46$, $MSE = .03$, $p = .22$, indicating that correct recognition in the recall and guessing tasks were equivalent to the restudy task

Table 4

Mean (SD) Recognition Results of List Items and Critical Items for Remember (R) and Know (K) Responses Combined for Categorized, Ad hoc, and Unrelated Lists as a Function of Blocked Task Type in Experiment 2 (Task Blocked in 6 Repetitions with Instructions Presented Prior to List Presentation).

	Interpolated Task		
	Restudy	Recall	Guess
List Items			
Categorized	.77 (.22)	.81 (.20)	.79 (.18)
Controls		.20 (.17)	
Ad hoc	.76 (.22)	.75 (.19)	.76 (.19)
Controls		.20 (.17)	
Unrelated	.68 (.22)	.67 (.22)	.72 (.24)
Controls		.19 (.16)	
Task Average	**.74 (.19)**	**.74 (.17)**	**.76 (.18)**
Critical Items			
Categorized	.49 (.28)	.41 (.28)	.43 (.25)
Controls		.23 (.20)	
Ad hoc	.39 (.29)	.31 (.25)	.34 (.26)

Table 4 continued

Controls		.22 (.18)	
Unrelated	.24 (.25)	.18 (.21)	.22 (.22)
Controls		.19 (.18)	
Task Average	**.38 (.19)**	**.40 (.19)**	**.32 (.19)**

Recollection estimates. Appendix Table B3 reports mean recollection estimates across list and task types. A 3(List Type) × 3(Task Type) repeated-measures ANOVA revealed a significant effect of list type, $F(2, 166) = 34.08$, $MSE = .90$, $\eta_p^2 = .29$, in which recollection estimates were higher for categorized than unrelated lists (.53 vs. .41), $t(83) = 7.80$, $SEM = .01$, $d = 0.53$, and higher for ad hoc than unrelated lists (.50 vs. .41), $t(83) = 6.23$, $SEM = .01$, $d = 0.41$. Recollection estimates were only marginally higher for categorized than ad hoc lists, (.53 vs. .50), $t(83) = 1.94$, $SEM = .02$, $d = 0.13$, $p = .06$. There was no effect of task type, $F(2, 166) = 1.95$, $MSE = .11$, $p = .15$, but a reliable list type × task type interaction, $F(4, 332) = 2.40$, $MSE = .06$, $\eta_p^2 = .03$. For categorized lists, recollection judgements were higher for the recall than restudy task (.56 vs. .49), $t(83) = 2.58$, $SEM = .03$, $d = 0.25$, but all other task comparisons were equivalent, *t*s < 1. For ad hoc lists, no task effects were found, *t*s < 1. For unrelated lists, recollection judgements were higher in the guess than the restudy task (.44 vs. .38, $t(83) = 2.32$, $SEM = .03$, $d = 0.23$, but all other task comparisons were equivalent, *t*s < 1. As in Experiment 1, recollection estimates of list items improved as list items became more related. For categorized lists, testing at study improved recollection of list items only compared restudy. However, for unrelated lists, guessing increased recollection estimates relative to restudy.

Familiarity Estimates. Table B4 reports mean familiarity estimates as a function of list and task type. Using the same ANOVA as above, a marginal effect of list type was found, $F(2, 164) = 2.61$, $MSE = .13$, $\eta_p^2 = .03$, in which familiarity estimates were higher for categorized than unrelated lists (.42 vs. .38), $t(83) = 2.28$, $SEM = .02$, $d = 0.13$ and higher for ad hoc than unrelated lists (.41 vs. .38), $t(83) = 1.57$, $SEM = .02$, $d = 0.10$. Familiarity estimates for categorized and ad hoc lists were equivalent (.42 vs. .41), $t < 1$. There was no significant effect of task type for familiarity estimates, $F < 1$, and the interaction was not significant, $F < 1$. Increasing task repetitions therefore did not appear to affect familiarity estimates for list items.

False Recognition

False recognition was analyzed by comparing overall incorrect “old” responses to non-presented critical items collapsed across recollection and familiarity (see Table 4). An effect of list type was found, $F(2, 166) = 111.76$, $MSE = 3.37$, $\eta_p^2 = .57$, in which false recognition was higher for categorized than unrelated lists (.44 vs. .21), $t(83) = 12.20$, $SEM = .02$, $d = 1.09$, and higher for ad hoc than unrelated lists (.35 vs. .21), $t(83) = 9.93$, $SEM = .01$, $d = 0.68$. False recognition was also higher for categorized than ad hoc lists, (.44 vs. .35), $t(83) = 7.15$, $SEM = .01$, $d = 0.40$. A main effect of task type was also found, $F(2, 166) = 7.66$, $MSE = .33$, $\eta_p^2 = .08$, in which false recognition was lower in the guess than restudy task (.33 vs. .37), $t(83) = 2.38$, $SEM = .02$, $d = 0.18$, but equivalent to the recall task (.33 vs. .30), $t(83) = 1.67$, $SEM = .02$, $p = .10$. False recognition was lower in the recall than restudy task (.30 vs. .37), $t(83) = 3.56$, $SEM = .02$, $d = 0.30$. There was no significant list type × task type interaction, $F < 1$. Like Experiment 1, guessing reduced false recognition of critical items, but only relative to the restudy task and not the recall task.

Recollection estimates. Table B4 reports mean recollection estimates across list and task types. A significant effect of list type was found, $F(2, 166) = 66.72$, $MSE = 1.33$, $\eta_p^2 = .45$, in which recollection estimates were higher for both categorized and ad hoc lists compared to unrelated lists (.21 vs. .06), $t(83) = 9.27$, $SEM = .02$, $d = 1.05$, and (.15 vs. .06), $t(83) = 8.64$, $SEM = .01$, $d = 0.76$, respectively. Recollection was also higher for critical items in categorized than ad hoc lists, (.21 vs. .15), $t(83) = 5.21$, $SEM = .01$, $d = 0.37$. There was no significant effect of task type, $F(2, 166) = 2.18$, $MSE = .06$, $p = .12$. The interaction was not significant, $F < 1$.

Familiarity estimates. Table B4 reports mean familiarity estimates across list and task types. Like recollection estimates, an effect of list type was found, $F(2, 164) = 37.00$, $MSE = .85$, $\eta_p^2 = .31$, in which familiarity was higher for categorized than unrelated lists, (.25 vs. .13), $t(83) = 7.16$, $SEM = .02$, $d = 0.61$. Familiarity estimates for ad hoc lists were higher relative to unrelated lists (.19 vs. .13), $t(83) = 5.91$, $SEM = .01$, $d = 0.32$. Finally, familiarity estimates were higher for categorized than ad hoc lists (.25 vs. .19), $t(83) = 3.98$, $SEM = .01$, $d = 0.27$. Unlike recollection estimates however, a significant effect of task type was found, $F(2, 164) = 4.48$, $MSE = .13$, $\eta_p^2 = .05$, in which familiarity estimates were lower in the guess task relative to the restudy task (.18 vs. .22), $t(83) = 2.81$, $SEM = .01$, $d = 0.19$, and in the recall task relative to the restudy task (.22 vs. .18), $t(83) = 2.30$, $SEM = .02$, $d = 0.19$. Familiarity estimates were equivalent between the recall and guess tasks (.18 vs. .18), $t < 1$. Finally, the interaction was not significant, F(4, 328)= 1.97, *MSE* = .04, *p* =.10. Recollection and familiarity estimates of critical items were affected by list relatedness, with more related list items increasing recollection of non-presented critical items. However, only familiarity estimates of critical items was affected by task type, with both guessing and testing decreasing false recognition of critical items at test compared to restudy.

Discussion

The purpose of Experiment 2 was to assess whether task expectancy processes that were encouraged through instructions presented before study and task repetitions would affect recognition and recollection/familiarity estimates. Starting with overall correct recognition, we found correct recognition to be higher for categorized than ad hoc and unrelated lists. However, correct recognition was equivalent across testing, guessing, and restudy tasks. Although a task effect for overall correct recognition was not found as expected in $\mathbf{H_1}$, inducing task expectancies eliminated the benefit of the restudy task that is demonstrated when there are no task expectancies in Experiment 1. Therefore, it is possible that task repetitions did result in an expectancy effect which eliminated the reversed retrieval-practice effect found previously. As predicted in $\mathbf{H_2}$, task effects on recollection estimates were contingent on list type, where testing increased recollection of list items for highly related categorized lists, while guessing increased recollection for unrelated lists. Separately, familiarity estimates demonstrated the same list type pattern in which correct recognition was highest for categorized and ad hoc lists compared to unrelated lists. Therefore, as predicted in $\mathbf{H_2}$, enhancing task-expectancy effects did not appear to affect familiarity estimates for correct recognition. When task expectancies are encouraged, participants appeared to rely more on recollective processes to identify list items.

Turning to false recognition of critical items, false recognition was higher in t categorized lists than both ad hoc and unrelated lists. As predicted by **H_1**, testing and guessing significantly decreased false recognition compared to the restudy task. As predicted by **H_2**, recollection estimates were affected by list type but not task type. Familiarity estimates were also similarly affected by list type, with familiarity of critical items being higher the more related the list items were. Further supporting **H_2**, similar to overall false recognition, familiarity estimates were lower in testing and guessing tasks compared to the restudy task, demonstrating that when task expectancies were enhanced, testing and guessing produced a memorial benefit by decreasing false memories of critical lures. Whereas recollective processes appeared to affect correct recognition of list items, familiarity processes appeared to affect false recognition of critical lures more than recollective processes.

The results of my second experiment failed to replicate task effects with 6 list repetitions found in Huff et al. (2018). It is possible that 6 list repetitions were insufficient to induce expectancy processes that were hypothesized to induce task effects. Additionally, a within-subjects design could produce carry-over effects from one task block to another, limiting the ability to detect task effects on recognition memory. Therefore, task effects were examined by increasing task repetitions to 18 to enhance task expectancies and manipulating task type between-subjects to eliminate potential for carry-over effects.

CHAPTER IV – EXPERIMENT 3

In Experiment 3, testing and guessing effects were further evaluated under conditions designed to increase task-expectancy processes. Specifically, participants used either the restudy, recall, or guessing tasks repeatedly to study a total of 18 lists, a three-fold increase relative to Experiment 2. Additionally, task type was manipulated between-subjects rather than within-subjects to eliminate any potential carry-over effects which may have persisted across blocks and affected task-expectancy processes in Experiment 2. Huff et al. (2018; Experiments 3 and 4) reported that the magnitude of testing and guessing benefits increased as a function of task repetitions. Therefore, the goal of Experiment 3 was to further enhance task-expectancy processes through 18 task repetitions and determine their effects on recollection and familiarity estimates. Based on previous findings (Huff et al., 2016; 2018), the additional repetitions should further enhance expectancy effects, which would result in a mirror effect of recognition in the testing and guessing groups on categorized and ad hoc lists ($\mathbf{H_1}$). Regarding recollection/familiarity estimates, recollection of list items is expected to increase for participants in the guessing and recall groups, even more so than Experiment 2 due to increased task repetitions. Familiarity of list items is also expected to increase for participants in the guessing and free recall groups, but not at the same rate as recollection. False recognition familiarity judgements are expected to be higher across tasks when compared to recollection judgments, with familiarity judgments being highest for false recognition in the restudy task ($\mathbf{H_2}$).

Methods

Participants

Eighty participants were taken from The University of Southern Mississippi undergraduate research participant pool or recruited locally from the greater Hattiesburg, MS community. Participants were randomly assigned to task groups (restudy = 27, recall = 27, guess = 26). Due to the COVID-19 pandemic's disruption of in-person data collection, participants completed the study online and were either recruited through the institutional research pool or contacted directly. Participants were compensated with either partial course credit (for students) or a $10 gift card (for community members). Participants reported fluency in the English language and normal or corrected-to-normal vision. Participants were primarily female (75%) and reported a mean of 14.3 years of education (*SD* = 2.29, Range = 12-22) and a mean age of 22.67 years (*SD* = 7.51, Range = 18-51). A G*Power3 (Faul et al., 2007) sensitivity analysis indicated that the sample size had adequate power (.80) to predict medium effect sizes or larger (Cohen's *d* = .66) for main effects and interactions.

Materials and Procedure

All materials and procedures from Experiments 1 and 2, including the R/K/N recognition test instructions, were again used with the following exceptions. First, participants now completed the same study task repeatedly for all 3 blocks, totaling 18 study/task lists. As in Experiment 2, each of the 3 blocks contained 2 lists of each of the list types (categorized, ad hoc, and unrelated). We utilized the same 180-item recognition test from Experiment 2 for each block. The lists used in each task block were once randomized and arranged into 3 versions that were counterbalanced across participants, totaling 6 different versions of the experiment. Task instructions were again provided at the beginning of each task.

Results

Similar analyses were conducted as in the two previous experiments, including collapsing across test blocks for analyses and computing recollection and familiarity estimates. Mean raw “know” responses for Experiment 3 are reported in Appendix Table B5.

Recall and Guessing Tasks

Proportions of correct recall were found to differ across the three list types, $F(2, 52) = 56.28$, $MSE = .17$, $\eta_p^2 = .68$ (see Table 5). Like Experiments 1 and 2, correct recall was higher for both categorized and ad hoc lists versus unrelated lists (.54 vs. .38), $t(26) = 8.65$, $SEM = .02$, $d = 1.59$ and (.50 vs. .38), $t(26) = 7.63$, $SEM = .02$, $d = 1.19$, respectively. Correct recall of categorized lists was higher than ad hoc lists (.54 vs. .50), $t(26) = 2.83$, $SEM = .01$, $d = 0.44$.

Table 5

Mean (SD) Proportion of Correct Recall of List Items and Correct Guessing of Critical Items by List Type in Experiment 3 (Task Repeated for 18 lists).

	Interpolated Task Type	
Item Type	Recall	Guess
List Items		
Categorized	.54 (.09)	-
Ad Hoc	.50 (.09)	-

Table 5 continued

Unrelated	**.38 (.11)**	**-**
Critical Items		
Categorized	**.14 (.07)**	**.22 (.16)**
Ad Hoc	**.07 (.06)**	**.12 (.11)**
Unrelated	**.01 (.02)**	**.00 (.00)**
Extra-List Intrusions per List		
Categorized	.34 (.34)	-
Ad Hoc	.35 (.37)	-
Unrelated	.27 (.26)	-

False recall of critical items also differed across lists, $F(2, 20) = 15.03$, $MSE = .05$, $\eta_p^2 = 0.60$. False recall was higher for categorized than unrelated lists (.14 vs. .01), $t(10) = 5.93$, $SEM = .02$, $d = 2.60$, and higher for ad hoc than unrelated lists (.07 vs. .01), $t(10) = 3.46$, $SEM = .02$, $d = 1.39$. False recall was also higher for categorized than ad hoc lists (.14 vs. .07), t (10) = 2.28, $SEM = .03$, $d = 1.07$. Mean number of extra-list intrusions were similarly analyzed, but no differences were found across list types, $F < 1$.

Correct guessing rates also differed across lists, $F(2, 50) = 35.87$, $MSE = .31$, $\eta_p^2 = .59$. Correct guessing was higher for both categorized and ad hoc lists relative to unrelated lists (.22 vs. .00), $t(25) = 6.94$, $SEM = .03$, $d = 1.94$, and (.12 vs. .00), $t(25) = 5.60$, $SEM = .02$, $d = 1.54$, respectively, and was higher for categorized than ad hoc lists (.22 vs. .12), $t(25) = 4.26$, $SEM = .02$, $d = 0.73$. Thus, consistent with Experiments 1 and

2, lists that were semantically related produced an increase in correct recall, an increase in false recall, and an increase in correct guessing of critical items.

Correct Recognition

Appendix Table B6 displays proportions of “old” responses to studied list items, non-studied list item controls, critical items, and non-studied critical item controls. Overall recognition rates across recollection and familiarity estimates were analyzed first. Because task type was manipulated between subjects, recognition proportions were first adjusted using a hits minus false alarms correction for both correct recognition (hits for studied list items minus false alarms for non-studied list items) and false recognition (hits for critical items minus false alarms for non-studied critical items). This correction was used to control for potential response biases that may be due to task-type differences (see Huff et al., 2018 for a similar procedure).

A 3(List Type) × 3(Task Type) mixed ANOVA found an effect of list type, $F(2, 154) = 9.51$, $MSE = .10$, $\eta_p^2 = .11$, in which correct recognition was higher for categorized than unrelated lists (.61 vs. .54), $t(79) = 3.86$, $SEM = .02$, $d = 0.29$. No difference was found for correct recognition between ad hoc and unrelated lists (.56 vs. .54), $t(79) = 1.61$, $SEM = .01$, $p = .11$. Correct recognition was higher for categorized than ad hoc lists, (.61 vs. .56), $t(79) = 2.93$, $SEM = .02$, $d = 0.19$. An effect of task type was also found, F (2, 77) = 8.45, MSE =1.07, $\eta_p^2 = .18$, in which correct recognition was higher in the recall group than the restudy group (.62 vs. .44), $t(52) = 2.93$, $SEM = .06$, $d = 0.80$, and higher in the guess group than the restudy group (.66 vs. .44) $t(51) = 3.61$, $SEM = .06$, $d = 1.02$. Correct recognition was equivalent between the recall and guess groups (.62 vs. .66), $t < 1$. The list type × task type interaction was not reliable, $F(4, 154)$

= 1.73, *MSE* = .02, *p* = .15. Therefore, consistent with predictions, extensive testing and guessing task repetitions resulted in large increases in correct recognition and this pattern was equivalent across list types.

*Recollection estimates***.** Appendix Table B6 reports mean recollection estimates across list and task types. A 3(List Type) × 3(Task Type) mixed ANOVA revealed a significant main effect of list type, *F*(2, 154) = 18.57, *MSE* = .20, η_p^2 =.20, in which recollection estimates were higher for both categorized lists and ad hoc lists relative to unrelated lists (.49 vs. .40), *t*(79) = 5.27, *SEM* = .02, *d* = 0.37 and (.47 vs. .40), *t*(79) = 4.97, *SEM* = .02, *d* = 0.32, respectively. Recollection estimates did not differ between categorized and ad hoc lists (.49 vs. .47), *t*(79) = 1.08, *SEM* = .02, *p* = .28. Consistent with Experiment 2, there was no effect of task type, *F* < 1, but a marginal list type × task type interaction was found, *F*(4, 154) = 2.23, *MSE* = .02, η_p^2 = .06, *p* = .07.

For the restudy task, recollection estimates were significantly higher on categorized than unrelated lists (.48 vs. .39), *t*(26) = 3.47, *SEM* = .03, *d* = 0.34, and marginally higher for ad hoc lists than unrelated lists (.44 vs. .39), *t*(26) = 2.05, *SEM* = .02, *d* = 0.21, *p* = .051. Recollection estimates were equivalent between categorized and ad hoc lists for the restudy group (.48 vs. .44), *t*(26) = 1.59, *SEM* = .03, *p* = .12. For the recall group, recollection estimates were higher for both categorized and ad hoc lists compared to unrelated lists, (.51 vs. .38), *t*(26) = 3.59, *SEM* = .04, *d* = 0.56 and (.46 vs. .38), *t*(26) = 3.09, *SEM* = .03, *d* = 0.38, respectively. Recollection estimates were also higher for categorized lists than ad hoc lists, (.51 vs. .46), *t*(26) = 2.05, *SEM* = .02, *d* = 0.21. For the guess group, recollection estimates were marginally higher for categorized lists than unrelated lists (.47 vs. .42), *t*(25) = 2.01, *SEM* = .03, *d* = .22, *p* = .06, and marginally higher for ad hoc lists compared to categorized lists (.52 vs. .47), $t(25) = 1.75$, $SEM = .03$, $d = 0.21$, $p = .09$. Recollection estimates were significantly higher for ad hoc than unrelated lists, (.52 vs. .42), $t(25) = 3.37$, $SEM = .03$, $d = 0.48$[1]. Consistent with the previous experiments, recollection estimates of list items was higher for lists that were more strongly related. This pattern was especially true for the recall group. However, for list items that converged on a broader category, guessing the critical items improved recollection of list items at test.

Familiarity estimates. Appendix Table B6 reports mean familiarity estimates as a function of list and task type. An effect of list type was found, $F(2, 154) = 47.08$, $MSE = 1.34$, $\eta_p^2 = .38$, in which familiarity estimates were higher for categorized and ad hoc lists than unrelated lists (.56 vs. .33), $t(79) = 9.08$, $SEM = .03$, $d = 0.95$ and (.55 vs. .33), $t(79) = 8.11$, $SEM = .03$, $d = 0.91$, respectively. Familiarity estimates for categorized and ad hoc lists were equivalent (.56 vs. .55), $t < 1$. There was no significant effect of task type for familiarity estimates, $F(2, 77) = 1.11$, $MSE = .14$, $p = .33$, and the list type × task type interaction was not significant, $F(2, 154) = 1.68$, $MSE = 05$, $p = .16$. Once again, enhancing task-expectancy effects did not appear to affect familiarity of recognized list items.

False Recognition

Analyses for critical item false recognition were similarly calculated as for correct recognition (i.e., using corrected scores; see Table B6). A main effect of list type was

[1] Given the interest in task type differences, this interaction was also investigated by examining task effects within each list type. However, no task-type effects were found within each list type, *t*s < 1.06, *p*s > .24, indicating that the interaction was driven by list-type differences.

found, *F*(2, 154) = 45.57, *MSE* = .77, η_p^2 = .37, in which false recognition was higher for categorized and ad hoc than unrelated lists (.21 vs. .01), *t*(79) = 8.43, *SEM* = .02, *d* = 1.43 and (.11 vs. .01), *t*(79) = 5.44, *SEM* = .02, *d* = 0.91, respectively. Additionally, false recognition was higher for categorized than ad hoc lists, (.21 vs. .11), *t*(79) = 4.77, *SEM* = .02, *d* = 0.68. There was no effect of task type, *F*(2, 77) = 1.84, *MSE* = .04, *p* = .17, and the interaction was not significant, *F* (4, 154) = 1.43, *MSE* = .02, *p* = .23. Therefore, enhanced task type repetition did not appear to affect overall false recognition of critical items.

Recollection estimates. Table B6 reports mean recollection estimates as a function of list and task type. An effect of list type was found, *F*(2, 154) = 47.59, *MSE* = .34, η_p^2 = .38, in which recollection estimates were higher for both categorized and ad hoc than unrelated lists, (.22 vs. .09), *t*(79) = 8.05, *SEM* = .02, *d* = 0.87 and (.16 vs. .09), *t*(79) = 6.53, *SEM* = .01, *d* = 0.56, respectively. Recollection estimates were also higher for categorized than ad hoc lists, (.22 vs. .16), *t*(79) = 4.70, *SEM* = .01, *d* = 0.37 There was also a significant effect of task type, *F*(2, 77) = 4.23, *MSE* = .20, η_p^2 = .10. Recollection estimates of critical items were marginally lower in the recall than the restudy group, (.14 vs. .21), *t*(52) = 1.87, *SEM* = .04, *d* = 0.52, *p* = .07, but lower in the guess than the restudy group, (.11 vs. .21), *t*(51) = 2.84, *SEM* = .03, *d* = 0.79. Recollection estimates of critical items were equivalent between the recall and guess groups, (.14 vs. .11), $t < 1$. The interaction was not reliable, $F < 1$. The recall and guess tasks appeared to reduce recollection of non-presented critical items relative to the restudy task.

Familiarity estimates. Table B6 reports mean familiarity estimates across list and task types. Like recollection, an effect of list type was found, *F*(2, 154) = 32.03, *MSE* =

.35, η_p^2 = .29, in which familiarity was higher for critical items from categorized and ad hoc lists than unrelated lists (.26 vs. .13), *t*(79) = 7.04, *SEM* = .02, *d* = 0.76 and (.21 vs. .13), *t*(79) = 4.84, *SEM* = .02, *d* = 0.48, respectively. Familiarity was higher for categorized than ad hoc critical items (.26 vs. .21), *t*(79) = 3.68, *SEM* = .01, *d* = 0.29. A marginal effect of task type was found, *F*(2, 77) = 2.74, *MSE* = .16, η_p^2 = .07, *p* = .07, in which familiarity estimates were lower in the recall task than the restudy task (.16 vs. .25), *t*(52) = 2.21, *SEM* = .04, *d* = 0.60. Familiarity estimates in the recall task were equivalent to the guess task, *t* < 1, and familiarity estimates in the guess task were equivalent to the restudy task (.19 vs. .25), *t*(51) = 1.36, *SEM* = .04, *p* = .18. Finally, the list type × task type interaction was not significant, *F*(4, 154)= 1.01, *MSE* = .01, *p* = .40. In sum, both recollection and familiarity estimates of critical items were affected by list relatedness, with more strongly related lists increasing recollection and familiarity of non-presented critical items. Recollection estimates of critical items were affected by task type, with both guessing and testing at study decreasing false recognition of critical items at test compared to the restudy task. Familiarity estimates were only reduced in the recall task versus the restudy task.

Discussion

Experiment 3 extended the results of Experiment 2 by increasing the number of task repetitions and manipulating tasks between-subjects to enhance expectancy effects on recognition memory. Consistent with our previous experiments, overall correct recognition was highest for categorized lists compared to ad hoc lists and unrelated lists. As predicted by **H_1**, extensive task repetitions led to large task differences, where correct recognition was higher for both the recall and guess groups compared to the restudy groups. List type effects on recollection estimates mirrored overall correct recognition, where recollection of list items was highest for categorized and ad hoc lists compared to unrelated lists. This effect was qualified by a significant interaction with task type, where recollection estimates in the restudy task were highest for both categorized and ad hoc lists compared to unrelated lists and recollection estimates in the recall task were higher for categorized lists compared to ad hoc lists and ad hoc lists were higher than unrelated lists. Interestingly, recollection estimates in the guess task were highest for ad hoc lists compared to categorized lists and categorized lists compared to unrelated lists. Whereas participants appeared to recollect categorical contextual information in the restudy and recall groups, enhanced task repetitions in the guess task group effectively increased recollective processes for more weakly related list materials. Familiarity estimates also mirrored overall correct recognition, where correct recognition in categorized and ad hoc lists were higher than unrelated lists. Contrary to $\mathbf{H_2}$, there was no significant effect of task type, indicating that enhancing expectancy processes did not appear to affect familiarity of list items.

Turning to overall false recognition, consistent with the previous experiments, false recognition of critical items was highest in categorized lists followed by ad hoc lists and then unrelated lists. Contrary to $\mathbf{H_1}$, there was no effect of task type on overall false recognition. List type effects on both recollection and familiarity were consistent with overall false recognition. However, recollective and familiarity processes were differentially affected by task type. Contrary to the predictions of $\mathbf{H_2}$, recollection of critical items was reduced in the recall and guess groups compared to the restudy task, demonstrating that when task expectancies were enhanced, testing and guessing decreased recollection of critical lures. Alternatively, familiarity estimates of critical items was only higher in the restudy group compared to the recall group, partially supporting $\mathbf{H}_2$'s predictions.

CHAPTER V – GENERAL DISCUSSION

The purpose of the current study was to investigate how expectancy processes generated by repeated testing and guessing tasks affect qualitative memory processes by estimating recollection and familiarity when task expectancies were eliminated (Experiment 1), induced by blocking task by 6 repetitions (Experiment 2), and further increased over 18 task repetitions (Experiment 3). Beginning with correct recognition, a consistent effect of list type was found across all three experiments, where correct recognition was highest for categorized lists, which contained strongly related words, second highest for ad hoc lists, which contained words that were loosely related to a broader category, and lowest for unrelated lists, which contained words that were randomly generated. When task order was randomized across list presentations and participants were given task instructions after list presentation, eliminating expectancies of upcoming tasks, a reversed retrieval-practice effect was found, where overall correct recognition was highest for restudy lists, rather than recall and guess lists. These findings replicated those of Huff et al. (2018; Experiment 1) and demonstrated that recall and guessing benefits were eliminated and even reversed, in the absence of task expectancies. Additionally, when task expectancies were induced over 6 task repetitions with task instructions presented prior to list presentation (Experiment 2), although the same list effect was present, where correct recognition was higher for related versus unrelated lists, no task differences were found. One possibility is that, given the reversed retrieval-practice effect in Experiment 1, testing and guessing in Experiment 2 did yield expectancy effects, but these benefits were only sufficient in magnitude to boost correct recognition to the same recognition level as the restudy task. Consistent with this possibility, when task expectancies were increased over 18 task repetitions and manipulated between-subjects (Experiment 3), recall and guess tasks both produced an increase in correct recognition that was greater than 20% over the restudy task.

In addition to analyzing correctly recognized list items, estimates of recollection and familiarity processes were computed. These estimates were extracted by using the remember/know procedure described by Tulving (1985) and corrected for familiarity underestimation using the remember/know correction (Yonelinas & Jacoby, 1995). Recollection and familiarity estimates were differentially affected by task type regarding correct recognition. Starting with recollection, list item effects were similar to overall correct recognition, with recollection highest for categorized lists, then ad hoc lists, then unrelated lists across all task manipulations. When task expectancies were eliminated in Experiment 1, recollection was higher for the restudy and recall lists compared to the guess list. When task expectancies were induced over 6 repetitions, an interaction occurred in which for categorized lists, recollection for recall lists increased over restudy lists and for unrelated lists, recollection for guess lists increased to equivalent levels of restudy. Finally, when task expectancies were enhanced over 18 repetitions, task effects on recollection were minimal, except for the guess group, where recollection was highest for ad hoc list items, followed by categorized then unrelated list items.

Unlike recollection estimates, familiarity was only significantly affected by task type when tasks were randomized and expectancies were eliminated. Specifically, an interaction of task effects and list type on familiarity estimates was found, where categorized lists were not affected by task type, as participants likely relied on list relatedness rather than familiarity processes in identifying list items at test; familiarity for ad hoc list items was highest in the recall task; and familiarity for unrelated list items was highest for the guess and restudy lists compared to the recall lists. In contrast, familiarity estimates were not affected by task expectancies and familiarity only differed as a result of list type effects. These results suggest that participants may rely on recollective processes when deciding list item identification and do not appear to utilize familiarity processes.

Turning to overall false recognition, across all experimental manipulations, false recognition of critical items was highest for categorized lists, followed by ad hoc lists and unrelated lists. When task expectancies were eliminated by randomizing tasks in Experiment 1, false recognition was lowest in the guess task compared to the restudy and recall task; however, guessing only produced the lowest false recognition rate when critical lures could be successfully guessed, such as for categorized and ad hoc lists. The benefit of guessing critical items at study, therefore, appeared to persist across all list types when task expectancies were eliminated. This pattern likely occurred because guessing potentially assisted participants in identifying critical items at study that they could monitor for later at test. When task expectancies were induced across 6 list repetitions in Experiment 2, false recognition was lower in the recall and guess tasks compared to the restudy task. Contrary to our expectations, when task expectancies were enhanced across 18 list repetitions, false recognition did not differ statistically based on task type. We consider possible explanations for this later in the discussion.

Again, as with correct recognition, recollection and familiarity processes were differentially affected by task type. Beginning with estimates of recollection, recollection of critical items across all 3 experiments was affected by list effects consistent with overall false recognition. Additionally, like overall false recognition, the guess task significantly reduced recollection of critical items compared to both the recall and restudy tasks. List effects interacted with task effects, where for categorized lists, recollection of critical items was highest for the recall task, for ad hoc lists, recollection was highest for the restudy task, and for unrelated lists, recollection was highest for the guess task. Recollection estimates of critical items was not affected by task type when expectancies for a task were induced over 6 task repetitions. However, when task expectancies were enhanced over 18 task repetitions, recollection of critical items was highest in the restudy task, followed by the recall task, and lowest in the guess task. This latter finding indicates that attempting to identify critical items reduced the likelihood that participants would recollect critical items as studied during the recognition test.

Turning to familiarity estimates of critical items, list effects across all experimental manipulations were consistent with overall false recognition. Additionally, when task expectancies were eliminated, as with overall false recognition and recollection estimates, familiarity estimates for critical items were reduced in the guess lists compared to the restudy and recall lists. Again, list type and task type interacted, where for categorized lists, familiarity for critical items was highest in the recall lists, for ad hoc lists, familiarity was highest in the restudy lists, and for unrelated lists, familiarity was higher in the restudy lists but only compared to the recall lists. When task expectancies were induced over 6 task repetitions, consistent with overall false recognition, familiarity for critical items was highest in the restudy task. Finally, when expectancies were enhanced over 18 task repetitions, familiarity for critical items was only higher for restudy lists compared to recall lists. These results suggest that false recognition may be attributed more so to familiarity processes rather than recollective processes as was the case with correct recognition.

Overall, as task expectancies were enhanced across the three experiments, so were task effects on correct recognition of studied list items. When expectancies were completely eliminated, we found a reversed retrieval-practice, where the restudy task increased correct recognition compared to the recall and guess tasks (cf. Huff et al., 2018), even though recall and guess tasks commonly increase recognition, as both tasks have been found to facilitate relational processing in recognition memory (Huff & Bodner, 2018). When task expectancies were induced in Experiment 2, task effects were completely eliminated, as the recall and guess tasks increased correct recognition to be comparable with restudy. Finally, when these task expectancies were increased in Experiment 3, recall and guess significantly benefitted recognition memory compared to the restudy task. These findings replicate previous studies and provide further evidence that the benefit of the recall and guess tasks is in fact influenced by participants' expectations of completing the task prior to studying list words.

Task effects on false recognition of critical lures also appeared to be influenced by task expectancies across the course of the study. When expectancies were altogether eliminated, only the guessing task decreased false recognition of critical lures compared to the recall and restudy task across list types. As predicted, when task expectancies were induced over 6 list repetitions, guessing and testing at study benefitted recognition memory by decreasing false memories of critical lures. However, when task expectancies were enhanced over 18 list repetitions, there was no significant effect of task type. A possible explanation for this finding is that the guessing and recall tasks both rely on relational processing, or focusing on the similarities and relatedness of list items, which has been demonstrated to facilitate encoding critical lures and thereby increase false recognition (Coane, Huff & Hutchison, 2016; Huff & Bodner, 2018).

We also analyzed the contributions of recollection and familiarity processes in recognition memory, as well as how these processes are specifically influenced by task type, a unique contribution to the literature. Regarding correct recognition, we found that recollection of list items followed a similar pattern to overall correct recognition, where task effects became more influential as participant expectations for the task increased. These findings indicate that recollection appears to be significantly affected by task effects induced by participant expectancies and that participants may rely on these recollective processes when making recognition decisions of studied information at test. On the other hand, familiarity was only marginally affected by task type when no expectancies were induced, indicating that task effects do not appear to affect familiarity processes in regard to recognition of studied information.

Recollection and familiarity processes also appeared to differentially contribute to false recognition. As with overall false recognition, when expectancy effects were eliminated, guessing benefitted recollection and familiarity processes by reducing false recognition compared to the recall and restudy tasks. However, expectancy processes influenced recollection and familiarity at different rates. Task expectancies appeared to influence recollection of falsely remembered critical lures with both the recall and guess tasks, where guessing reduced recollection of critical lures even more so than the recall task, but not until 18 task repetitions. In contrast, task expectancies also influenced familiarity estimates of falsely remembered critical lures, but in different ways. When task expectancies were induced, familiarity estimates were affected similarly to overall false recognition, where the recall and guess tasks decreased familiarity for critical items compared to the restudy tasks. However, when tasks were enhanced over 18 repetitions, only recall significantly decreased familiarity of critical lures compared to the restudy control task. Overall, when task expectancies were amplified, the guessing task only reduced recollection of critical lures, not familiarity. As false recognition appears to be a more familiarity-based process, since the guess task does not significantly affect familiarity, it is likely that the task effects of guessing on reducing false recognition may be minimal.

Our assessment of recollective and familiarity-based processes utilized the Tulving (1985) remember/know procedure. The remember/know procedure, however, relies on participant introspection and response, rather than objective measurements of these memorial processes. Therefore, it is possible that participants may have over or under reported their "remember" or "know" responses at test. One way that we accounted for this was by using the remember/know correction (Yonelinas & Jacoby, 1995) for "know" responses to adjust for underreporting of familiarity processes. The criteria for determining a "remember" or a "know" response to a stimuli is arguably quite similar, and it is possible that participants may not have fully understood the distinction between the two responses. However, as we used the same procedure instructions from the original study that have shown to be effective at distinguishing between the two processes in the literature and because our data illustrate substantial differences between the two responses, this does not seem likely.

Additionally, although Experiments 1 and 2 were conducted in a laboratory setting, due to the disruption of data collection from the COVID-19 pandemic, Experiment 3 data were collected online. Therefore, it is possible that participants in Experiment 3 may not have been as focused on completing the study and could have been off-task throughout the course of the study. However, this possibility does seem unlikely as there was a powerful task effect for testing and guessing in Experiment 3. Further, Huff et al. (2018; Experiments 4A and 4B) similarly collected participant data online whereas all previous experiments were collected in-lab. Similar to the present findings, strong recall and guess benefits were found using an online sample, providing evidence that task effects are reliable across online and in-person experimental modalities.

One task related difference worthy of discussion is the difference in the duration of the guess, recall, and restudy tasks participants completed. Although restudy and recall participants were required to either restudy list items or engage in free recall for 60 s, guess participants were able to advance to the next study list provided they generated at least one guess. It is possible that guessing effects could be confounded by potential differences in the time required to complete each task, as participants were more likely to remain on-task while guessing because they could control the speed in which they completed the task. Time spent engaged in tasks was not analyzed in the present experiments; however, Huff et al., (2016; 2018) reported that the mean time participants engaged in the guess task was often less than 30 s. Despite this short duration however, guess participants were still able to produce large expectancy effects that increased correct recognition to the same level as the recall test. This pattern suggests that guessing, which does not involve the explicit retrieval of study list items, may be a more efficient method for promoting retention, at least when the final recognition test is completed relatively shortly following study. Future studies should evaluate the generality of guessing as a study strategy when expectancies are involved as it may produce more efficient yet equally potent benefits as retrieval practice.

The current study demonstrates that the benefits of repeated testing and guessing that lead to improved recognition memory can largely be attributed to the development of expectations for the upcoming task. An obvious application of these findings is that efforts to promote expectancies for upcoming tests in educational settings through detailed knowledge of and repeated practice with upcoming assessments may be effective ways to promote retention. For example, clearly defining the test type for an upcoming test, such as whether students should expect multiple-choice questions or essay questions, would enhance students' expectations for that test and facilitate quantitative and qualitative changes at study to improve overall test scores. Importantly, the current studies indicate that test type repetitions are positively related to the guessing and testing benefits found on subsequent recognition. An interesting prospect for future research is whether similar benefits found using categorized, ad hoc, and unrelated word lists would also be found using textbook or lecture materials which are common in educational settings. Our use of different types of study materials was designed to test the generality of testing and guessing benefits, and given that testing and guessing improvements over restudy occurred across list types, it is reasonable to predict that testing and guessing would produce similar benefits on educational materials. Examining testing and guessing with more externally valid materials will be critical for determining the utility of these task types and the expectancies that can be promoted in practice.

The findings of the current study demonstrate that participants appeared to rely more so on recollective processes over familiarity when making decisions regarding list item identification. Therefore, it is reasonable to presume that increasing recollective processes at study could also improve correct recognition of studied information at test. The testing and guessing appeared to benefit recollection of studied information compared to the restudy task, indicating that similarly to recognition, recollective processes are differentially affected by study task strategies. Therefore, future studies should focus on identifying study strategies that specifically enhance recollective processes, and thereby increasing correct recognition.

Conclusions

The purpose of my dissertation study was to evaluate the contribution of expectancy processes involved in testing and guessing effects on memory by estimating recollection and familiarity processes when expectancy processes were eliminated by randomizing task presentation and when expectancy processes were induced by blocking tasks in either 6 repetitions or 18 repetitions. Eliminating expectancy processes in Experiment 1 resulted in a reversed retrieval-practice effect, where correct recognition was highest in the restudy task. Similarly, restudy benefited recollection of list items. Expectancy for upcoming tasks induced over 6 repetitions in Experiment 2 eliminated the restudy benefit as correct recognition was equivalent between restudy, recall, and guess tasks. Recall and guess also benefitted recollection of list items. Recall and guess benefits on correct recognition were only found when task expectancies were further enhanced over 18 repetitions in Experiment 3. Recollection was differentially affected by task type dependent on list relatedness. In contrast, familiarity with list items was not affected by task effects across all experimental manipulations.

Regarding false recognition, in the absence of task expectancy, the guess task effectively reduced false recognition, recollection, and familiarity with non-presented critical items at test. Expectations of upcoming task resulted in decreased false recognition and familiarity for critical items in both the recall and guess tasks. Recollection, on the other hand, was not affected. However, these task effects were eliminated altogether for overall false recognition, whereas the recall and guess tasks decreased both recollection and familiarity of critical items. Collectively, these experiments indicated that task effects were driven by expectancies of upcoming tasks rather than any inherent benefit of the tasks themselves.

APPENDIX A – Instructions for Remember/Know Procedure

Remember judgments: If your recognition of the word is accompanied by a conscious recollection of its prior occurrence in the study list, please select “R” for remember. The “remembered” word should bring to mind a particular association, image, a personal experience at time of study, or a specific memory about the word’s appearance or position on the list (i.e. words that came before or after).

Know judgements: If you recognize that the word was in the study list, but you cannot consciously recollect anything about the actual occurrence of the word or what happened or was experienced when the word occurred, please select “K” for know. The word you “know” was presented is a word that you are certain of recognizing the word from the study lists, but this word fails to evoke any specific conscious recollection from the study list.

New judgments: If you do not recognize the word at all from the study list, please select “N” for new. The “new” word should have no memory of the occurrence of the word on the study list at all.

To further clarify the difference between Remember and Know, consider the following examples. If someone asks you for your name, you would respond in the “Know” sense, because you would respond without becoming consciously aware of anything about a particular event or experience. However, if someone asks you for the last movie you saw, you would respond in the “Remember” sense, because you would become consciously aware again of some specific aspects of the original experience. If you have any questions or need further clarification regarding these judgments, please ask the researcher.

APPENDIX B –Raw "Remember" and "Know" Responses for Experiments 1-3

Table B1

Mean (SD) Recognition Results of List Items and Critical Items for Separated Raw Remember (R) and Know (K) Responses for Categorized, Ad hoc, and Unrelated Lists as a Function of Interpolated Task Type in Experiment 1 (Randomized Task with Instructions Following List Presentation).

	Interpolated Task					
	Restudy		Recall		Guess	
	R	K	R	K	R	K
List Items						
Categorized	.59 (.28)	.24 (.23)	.54 (.26)	.26 (.22)	.53 (.29)	.27 (.22)
Controls			.06 (.08)	.11 (.10)		
Ad hoc	.53 (.29)	.27 (.24)	.54 (.25)	.23 (.17)	.45 (.24)	.25 (.21)
Controls			.08 (.10)	.13 (.11)		
Unrelated	.48 (.28)	.26 (.21)	.47 (.26)	.20 (.20)	.42 (.24)	.24 (.18)

Table B1 continued

Controls			.07 (.08)	.11 (.09)		
Task Average	**.53 (.25)**	**.26 (.22)**	**.52 (.22)**	**.23 (.20)**	**.47 (.22)**	**.26 (.20)**
Critical Items						
Categorized	.21 (.21)	.24 (.18)	.29 (.24)	.28 (.22)	.18 (.20)	.22 (.22)
Controls			.07 (.10)	.14 (.11)		
Ad hoc	.24 (.24)	.23 (.22)	.20 (.21)	.22 (.18)	.14 (.18)	.20 (.16)
Controls			.08 (.10)	.14 (.11)		
Unrelated	.08 (.12)	.14 (.14)	.08 (.12)	.12 (.14)	.10 (.11)	.14 (.13)
Controls			.06 (.08)	.13 (.12)		
Task Average	**.18 (.16)**	**.20 (.19)**	**.19 (.16)**	**.21 (.19)**	**.14 (.13)**	**.19 (.16)**

Table B2

Mean (SD) Recognition Results of List Items and Critical Items for Separated Remember (R) and Familiarity (F) Estimates for Categorized, Ad hoc, and Unrelated Lists as a Function of Interpolated Task Type in Experiment 1 (Randomized Task with Instructions Following List Presentation).

	Interpolated Task					
	Restudy		Recall		Guess	
	R	F	R	F	R	F
List Items						
Categorized	.59 (.28)	.39 (.38)	.54 (.26)	.41 (.35)	.53 (.29)	.44 (.38)
Controls			.06 (.08)	.09 (.11)		
Ad hoc	.53 (.29)	.39 (.37)	.54 (.25)	.42 (.36)	.45 (.24)	.35 (.31)
Controls			.08 (.10)	.10 (.12)		
Unrelated	.48 (.28)	.35 (.34)	.47 (.26)	.27 (.28)	.42 (.24)	.32 (.29)
Controls			.07 (.08)	.08 (.10)		

Table B2 continued

Task Average	**.53 (.25)**	**.38 (.32)**	**.52 (.22)**	**.37 (.28)**	**.47 (.22)**	**.37 (.28)**
Critical Items						
Categorized	.21 (.21)	.22 (.23)	.29 (.24)	.31 (.31)	.18 (.20)	.20 (.22)
Controls			.07 (.10)	.11 (.12)		
Ad hoc	.24 (.24)	.23 (.27)	.20 (.21)	.20 (.24)	.14 (.18)	.18 (.20)
Controls			.08 (.10)	.11 (.13)		
Unrelated	.08 (.12)	.12 (.16)	.08 (.12)	.09 (.13)	.10 (.11)	.10 (.12)
Controls			.06 (.08)	.09 (.11)		
Task Average	**.18 (.16)**	**.19 (.18)**	**.19 (.16)**	**.20 (.20)**	**.14 (.13)**	**.16 (.15)**

Table B3

Mean (SD) Recognition Results of List Items and Critical Items for Separated Raw Remember (R) and Know (K) Responses for Categorized, Ad hoc, and Unrelated Lists as a Function of Interpolated Task Type in Experiment 2 (Task Blocked in 6 Repetitions with Instructions Prior to List Presentation).

	Interpolated Task					
	Restudy		Recall		Guess	
	R	K	R	K	R	K
List Items						
Categorized	.49 (.29)	.29 (.24)	.56 (.28)	.26 (.21)	.52 (.31)	.29 (.24)
Controls			.06 (.08)	.15 (.14)		
Ad hoc	.49 (.29)	.28 (.21)	.51 (.27)	.25 (.21)	.48 (.27)	.29 (.22)
Controls			.06 (.08)	.15 (.14)		
Unrelated	.38 (.26)	.31 (.21)	.41 (.23)	.27 (.20)	.44 (.27)	.28 (.21)
Controls			.06 (.08)	.14 (.13)		

Table B3 continued

Task Average	.45 (.24)	.29 (.22)	.50 (.23)	.26 (.21)	.48 (.26)	.29 (.22)
Critical Items						
Categorized	.22 (.21)	.28 (.23)	.18 (.20)	.24 (.20)	.23 (.23)	.21 (.17)
Controls			.06 (.08)	.17 (.17)		
Ad hoc	.15 (.18)	.25 (.20)	.14 (.19)	.19 (.19)	.16 (.19)	.19 (.18)
Controls			.07 (.10)	.16 (.13)		
Unrelated	.08 (.12)	.16 (.19)	.05 (.10)	.14 (.16)	.15 (.15)	.16 (.17)
Controls			.05 (.08)	.15 (.15)		
Task Average	**.18 (.16)**	**.23 (.21)**	**.19 (.16)**	**.19 (.18)**	**.14 (.13)**	**.19 (.18)**

Table B4

Mean (SD) Recognition Results of List Items and Critical Items for Separated Remember (R) and Familiarity (F) Responses for Categorized, Ad hoc, and Unrelated Lists as a Function of Interpolated Task Type in Experiment 2 (Task Blocked in 6 Repetitions with Instructions Prior to List Presentation).

	Interpolated Task					
	Restudy		Recall		Guess	
	R	F	R	F	R	F
List Items						
Categorized	.49 (.29)	.40 (.37)	.56 (.28)	.44 (.38)	.52 (.31)	.43 (.38)
Controls			.06 (.08)	.09 (.11)		
Ad hoc	.49 (.29)	.40 (.34)	.51 (.27)	.41 (.35)	.48 (.27)	.43 (.37)
Controls			.06 (.08)	.10 (.15)		
Unrelated	.38 (.26)	.38 (.33)	.41 (.23)	.37 (.32)	.44 (.27)	.39 (.35)
Controls			.06 (.08)	.08 (.10)		
Task Average	**.45 (.24)**	**.29 (.18)**	**.50 (.23)**	**.26 (.17)**	**.48 (.26)**	**.29 (.16)**

Table B4 continued

Critical Items						
Categorized	.22 (.21)	.29 (.31)	.18 (.20)	.23 (.26)	.23 (.23)	.21 (.24)
Controls			.06 (.08)	.11 (.12)		
Ad hoc	.15 (.18)	.22 (.25)	.14 (.19)	.18 (.23)	.16 (.19)	.18 (.26)
Controls			.07 (.10)	.11 (.13)		
Unrelated	.08 (.12)	.13 (.21)	.05 (.10)	.12 (.16)	.15 (.15)	.14 (.19)
Controls			.05 (.08)	.09 (.11)		
Task Average	**.18 (.16)**	**.22 (.23)**	**.19 (.16)**	**.18 (.19)**	**.14 (.13)**	**.18 (.19)**

Table B5

Mean (SD) Final Recognition Proportions for List Items and Critical Items and Recollection/Familiarity Estimates of Categorized, Ad Hoc, and Unrelated Lists as a Function of Interpolated Task Lists in Experiment 3.

	Restudy			Recall			Guess		
	Categorized	Ad Hoc	Unrelated	Categorized	Ad Hoc	Unrelated	Categorized	Ad Hoc	Unrelated
Raw Recognition									
List Items	.74 (.20)	.73 (.18)	.68 (.21)	.82 (.15)	.77 (.18)	.70 (.17)	.83 (.14)	.81 (.13)	.76 (.14)
Controls	.27 (.18)	.29 (.20)	.26 (.20)	.14 (.10)	.16 (.13)	.13 (.12)	.16 (.16)	.16 (.14)	.12 (.13)
Critical Items	.47 (.22)	.42 (.20)	.31 (.23)	.40 (.21)	.28 (.18)	.16 (.14)	.38 (.23)	.29 (.20)	.15 (.14)
Controls	.32 (.21)	.33 (.23)	.29 (.21)	.15 (.12)	.15 (.12)	.15 (.13)	.17 (.14)	.17 (.14)	.14 (.14)
Corrected Recognition									
List Items	**.47 (.27)**	**.43 (.27)**	**.42 (.29)**	**.69 (.18)**	**.60 (.21)**	**.56 (.19)**	**.67 (.20)**	**.65 (.16)**	**.64 (.18)**

Table B5 continued

Task Average		**.44 (.26)**			**.62 (.18)**			**.66 (.16)**	
Critical Items	**.15 (.15)**	**.09 (.13)**	**.02 (.15)**	**.25 (.17)**	**.13 (.10)**	**.00 (.08)**	**.21 (.17)**	**.12 (.14)**	**.01 (.07)**
Task Average		**.09 (.15)**			**.13 (.16)**			**.11 (.16)**	
Recollection/Familiarity Estimates									
List Items Recollection									
List Items	.48 (.29)	.44 (.24)	.39 (.23)	.51 (.26)	.46 (.22)	.38 (.20)	.47 (.26)	.52 (.22)	.42 (.20)
Controls	.14 (.13)	.14 (.13)	.13 (.12)	.05 (.07)	.08 (.10)	.06 (.09)	.05 (.07)	.04 (.05)	.03 (.05)
List Items Familiarity									
List Items	.46 (.27)	.51 (.26)	.32 (.22)	.60 (.28)	.58 (.28)	.32 (.21)	.63 (.24)	.56 (.27)	.35 (.20)
Controls	.16 (.14)	.18 (.15)	.15 (.15)	.09 (.07)	.09 (.09)	.07 (.08)	.12 (.13)	.13 (.12)	.10 (.11)
Critical Items Recollection									
List Items	.26 (.18)	.22 (.16)	.15 (.13)	.22 (.19)	.14 (.13)	.06 (.10)	.17 (.17)	.12 (.12)	.05 (.08)
Controls	.16 (.14)	.14 (.14)	.14 (.14)	.07 (.09)	.07 (.08)	.07 (.10)	.05 (.09)	.04 (.06)	.04 (.06)
Critical Items Familiarity									

List Items	.29 (.19)	.26 (.19)	.20 (.21)	.23 (.16)	.16 (.14)	.10 (.10)	.27 (.19)	.21 (.17)	.10 (.11)
Controls	.19 (.18)	.23 (.20)	.18 (.17)	.08 (.07)	.08 (.09)	.09 (.09)	.13 (.11)	.14 (.12)	.10 (.12)

Table B6

Mean (SD) Final Recognition Proportions for "Know" Responses of Categorized, Ad Hoc, and Unrelated Lists as a Function of Interpolated Task Lists in Experiment 3.

	Restudy			Recall			Guess		
	Categorized	Ad Hoc	Unrelated	Categorized	Ad Hoc	Unrelated	Categorized	Ad Hoc	Unrelated
Know Responses									
List Items	.25 (.21)	.29 (.20)	.28 (.18)	.30 (.24)	.30 (.19)	.30 (.19)	.35 (.25)	.29 (.22)	.34 (.20)
Controls	.13 (.12)	.15 (.13)	.12 (.12)	.08 (.07)	.08 (.08)	.07 (.07)	.11 (.11)	.12 (.11)	.09 (.10)

Table B6 continued

Critical Items	.21 (.14)	.20 (.14)	.16 (.16)	.17 (.12)	.13 (.11)	.09 (.09)	.21 (.15)	.18 (.12)	.09 (.10)
Controls	.16 (.15)	.19 (.16)	.15 (.14)	.07 (.07)	.08 (.08)	.08 (.09)	.12 (.09)	.13 (.11)	.10 (.11)

REFERENCES

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal anterior thalamic axis. *Behavioral and Brain Sciences, 22*(3), 425-444. doi: 10.1017/S0140525X99002034

Atkinson, R. C., & Westcourt, K. T. (1975) Some remarks on a theory of memory. In Rabbitt, P. M. A., & Dornic, S. (Eds.) *Attention and Performance V*. San Francisco: Academic Press.

Balota, D. A., & Neely, J. H. (1980). Test-expectancy and word-frequency effects in recall and recognition. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 576-587. doi: 10.1037/0278-7393.6.5.576

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, B., Neely, J. H., Nelson, D. L, Simpson, G. B., & Trieman, R. (2007). The English lexicon project. *Behavior Research Methods*, *39*(3), 445-459.

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for a transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 657-668. doi: 10.1037/0278-7393.15.4.657

Battig, W. F., & Montague, W. E. (1969). Category norms of verbal items in 56 categories. A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology*, *80*, 1-46. doi: 10.1037/h0027577

Brysbaert, M., & New, B. (2009). Moving beyond Kucera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavioral Research Methods, 41*(4), 977-990. doi: 10.3758/BRM.41.4.977

Chan, J. C. K., & McDermott, K. B. (2007). The testing effect in recognition memory: A dual process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(2), 431-437. doi: 10.1037/0278-7393.33.2.431

Coane, J. H., Huff, M. J. & Hutchison, K. A. (2016). The ironic effect of guessing: Increased false memory for mediated lists in younger and older adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology, and Cognition, 23*(3). 282-303. doi: 10.1080/13825585.2015.1088506

Coltheart, M. (1981). The MRC psycholinguistic database. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 33,* 497-505. doi: 10.1080/14640748108400805

Craik, F. I. (2002) Levels of processing: Past, present, and future? *Memory, 10*(5-6). 305-318. doi: 10.1080/09658210244000135

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 671-684. doi: 10.1016/S0022-537(72)80001-x

Cho, K. W., & Neely, J. H. (2017). The roles of encoding strategies and retrieval practice in test expectancy effects. *Memory, 25*(5), 626-635. doi: 10.1080/09658211.2016.1202983

Eichenbaum, H., Otto, T., & Cohen, N. J. (1994). Two functional components of the hippocampal memory system. *Behavioral & Brain Sciences, 17*(3), 449-517. doi: 10.1017/S0140525X00035391

Einstein, G. O., & Hunt, R. R. (1980) Levels of processing and organization: Additive effects of individual-item and relational processing. *Journal of Experimental Psychology: Human Learning and Memory, 6*(5), 588-598. doi: 10.1037/0278-7393.6.5.588

Faul, F., Erdfelder, E., Lang, A. G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175-191. doi: 10.3758/bf03193146

Finley, J. R. & Benjamin, A. S. (2012). Adaptive and qualitative changes in encoding strategy with experience: Evidence from the test-expectancy paradigm. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(3), 632-652. doi: 10.1037/a0026215

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*(4), 309-313. doi: 10.3758/BF03197041

Gardiner, J. M., & Gregg, V. H. (1997). Recognition memory with little or no remembering: Implications for a detection model. *Psychonomic Bulletin & Review, 4*(4), 474-479. doi: 10.3758/BF03214336

Glanzer, M. & Adams, J. K. (1990). The mirror effect in recognition memory: Data and theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 5-16. doi: 10.1037/0278-7393.16.1.5

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition, 7*(2), 95-112. doi: 10.3758/BF03197590

Greene, R. L. (1989). Spacing effects in memory: Evidence for a two-process account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 371-377. doi: 10.1037//0278-7393.15.3.371

Hamaker, C. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research, 56*(2), 212-242. doi: 10.2307/1170376

Hays, M. J., Kornell, N., & Bjork, R. A. (2013). When and why a failed test potentiates the effectiveness of subsequent study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39,* 290-296. doi: 10.1037/a0028468

Higham, P. A., & Vokey, J. R. (2004) Illusory recollection and dual-process models of recognition memory. *The Quarterly Journal of Experimental Psychology, 57A*(4), 714-744. doi: 10.1080/02724980343000468

Huff, M. J. & Bodner, G. E. (2018). Item-specific and relational processing both improve recall accuracy in the DRM paradigm. *Quarterly Journal of Experimental Psychology*, *72*(6). 1493-1506. doi: 10.1177/1747021818801427

Huff, M. J., Balota, D. A., & Hutchison, K. A. (2016). The costs and benefits of testing and guessing on recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1559-1572. doi: 10.1080/09658211.2018.14679

Huff, M. J., Bodner, G. E., & Fawcett, J. M. (2015). Effects of distinctive encoding on correct and false memory: A meta-analytic review of costs and benefits and their orgins in the DRM paradigm. *Psychonomic Bulletin & Review, 22*(2), 349-365. doi: 10.3758/s13423-014-0648-8

Huff, M. J., Meade, M. L., & Hutchison, K. A. (2011). Age-related differences in guessing on free and forced recall tests. *Memory (Hove, England), 19,* 317-330. doi: 10.1080/09658211.2011.568494

Huff, M. J., Yates, T. J., & Balota, D. A. (2018). Evaluating the contributions of task expectancy in the testing and guessing benefits on recognition memory. *Memory, 26*(8). doi: 10.1080/09658211.2018.1467929

Hunt, R. R. (2006). The concept of distinctiveness in memory research. In Hunt, R. R., & Worthen, J. B. (Eds.), *Distinctiveness and Memory* (3-25). New York: Oxford University Press.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20,* 497-514. doi: 10.1016/S0022-5371(81)90138-9

Jacoby, L. L. (1984) Incidental versus intentional retrieval: Remembering and awareness as separate issues. In Squire, L. R. & Butters, N. (Eds.) *Neuropsychology of memory* (145-156). New York: Guilford Press.

Kornell, N., Hays, M. J., & Bjork, R. A. (2009). Unsuccessful retrieval attempts enhance subsequent learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35,* 989-998. doi: 10.1037/a0015729

Lundeberg, M. A., & Fox, P. W. (1991). Do laboratory findings on test-expectancy generalize to classroom outcomes? *Review of Educational Research, 61,* 94-106. doi: 10.2307/1170668

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*(3), 252-271

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. doi: 10.1016/S0022-5371(77)80016-9

Neely, J. H., & Balota, D. A. (1981). Test-expectancy and semantic-organization effects in recall and recognition. *Memory & Cognition, 9,* 283-300. doi: 10.3758/BF03196962

Neely, J. H., & Cho, K. W. (2014). Testing and retrieval practice effects: Assessing the contributions of encoding and retrieval mechanisms. In Lindsay, D. S., Kelley, C. M., Yonelinas, A. P., & Roediger, H. L. (Eds.), *Remembering: Attributions, processes, and control in human memory: Papers in honour of Larry L. Jacoby* (29-45). New York: Psychology Press.

Perfect, T. J., Mayes, A. R., Downes, J. J., & Van Eijk, R. (1996). Does context discriminate recollection from familiarity in recognition memory? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*(3), 797-813. doi: 10.1080/713755644

Postman, L., & Jenkins, W. O. (1948). An experimental analysis of set in rote learning: The interaction of learning instruction and retention performance. *Journal of Experimental Psychology, 38,* 683-689. doi: 10.1037/h0057311

Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General, 143*, 644-667. doi: 10.1037/a0033194

Pressley, M., Tanenbaum, R., McDaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology, 15*(1). 27-35. doi: 10.1016/0361-476X(90)90003-J

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory and Cognition, 21*(1), 89-102. doi: 10.3758/BF03211168

Roediger, H. L, III, & Karpicke, D. J. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1*, 181-210. doi: 10.1111/j.1745-6916.2006.00012x

Roediger, H. L., III, Weldon, M. S., & Challis, B. H. (1989). Explaining dissociations between implicit and explicit measures of retention: A processing account. In Roediger, H. L., & Craik, F. I. M. (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (3-39). Hillsdale, NJ: Erlbaum.

Rowland, C. A. (2014). The effects of testing versus restudy on retention: A meta analytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432-1463. doi: 10.1037/a0037559

Schmidt, K., Patnaik, P., & Kensinger, E. A. (2011) Emotion’s influence on memory for spatial and temporal context. *Cognition & Emotion, 25*(2), 229-243. doi: 10.1080/02699931.2010.483123

Stewart-Williams, S. & Podd, J. (2004). The placebo effect: Dissolving the expectancy versus conditioning debate. *Psychological Bulletin, 130*(2), 324-340. doi: 10.1037/0033-2909.130.2.324

Tulving, E. (1985). Memory and consciousness. *Canadian Psychology, 26*(1), 1-12. doi: 10.1037/h0080017

Tversky, B. (1973). Encoding processes in recognition and recall. *Cognitive Psychology, 5*, 275-287. doi: 10.1016/0010-0285(73)90037-6

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50,* 289-335. doi: 10.1016/j.jml.2003.10.003

Yan, V. X., Yu, Y., Garcia, M. A., & Bjork, R. A. (2014). Why does guessing incorrectly enhance, rather than impair, retention? *Memory & Cognition*, *42*, 1373-1383. doi: 10.3758/s13421-014-0454-6

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. doi: 10.1006/jmla.2002.2864

Yonelinas, A. P., & Jacoby, L. L. (1995). The relation between remembering and knowing as bases for recognition: Effects of size congruency. *Journal of Memory and Language*, *34*(5), 622-643. doi: 10.1006/jmla.1995.1028

Zawadzka, K., & Hanczakowski, M. (2018). Two routes to memory benefits of guessing. *Journal of Experimental Psychology: Learning, Memory, and Cognition.* doi: 10.1037/xlm0000676