KEMP, PAIGE L., M.A. Adult Age Differences In The Forward Testing Effect. (2021)

Directed by Dr. Christopher Wahlheim. 44 pp.

Testing previously learned information can enhance subsequent learning of new information. This benefit was originally attributed to contextual segregation, but recent work suggests that testing can promote integrative encoding of competing information. This present study examined the extent to which such testing benefits older and younger adults by promoting such integration. Subjects studied two lists of associated word pairs that either repeated across lists, appeared only in the second list, or included the same cue with a changed response, and then completed a cued-recall test for List 2 responses. Between the lists, subjects completed a cued recall test with feedback or without feedback for some of the List 1 pairings that would later change. Participants indicated when they recollected that responses changed across lists and reported List 1 responses for those instances. Integrative encoding was inferred from instances of change recollection. Both groups showed a testing benefit on memory for changed pairs, but younger adults showed overall higher performance. More changes were recollected for tested than non-tested items, and change recollection was associated with higher List 2 recall. Feedback increase change recollection, but the benefits were offset by greater interference when change was not recollected. Critically, older adults recollected change less than younger adults and benefitted less when they recollected change. Together, these results suggest that testing promoted integrative encoding for both age groups, but older adults enjoyed these benefits on fewer trials and to a lesser degree.

ADULT AGE DIFFERENCES

IN THE FORWARD TESTING EFFECT

by

Paige L. Kemp

A THESIS

Submitted to

the Faculty of The Graduate School at

The University of North Carolina at Greensboro

in Partial Fulfillment

of the Requirements for the Degree

Master of Arts

Greensboro

2021

Approved by

Dr. Christopher Wahlheim

Committee Chair

DEDICATION

This thesis is dedicated to both my parents. To my father, Allan Kemp, who has always supported me financially, emotionally, and psychologically with his motivating words of encouragement. I will forever be grateful for you always pushing me forward and making me realize my true potential both personally and academically. To my mother, Julie Kemp, for being a source of inspiration and for giving me endless love and support. Thank you for believing in me and for taking part in my journey, I appreciate it very much.

A special feeling of gratitude is also given to my loving Nan, although thousands of miles away, continuously supported me by sending uplifting thought cards. Likewise, due recognition is given to my brother, Jordan, and my soon-to-be sister-in-law, Ellie-Rose Last, for their constant encouragement to accomplish this thesis project, and my wonderful nephew, Albie for keeping my spirits up with all the innocence.

I also dedicate this thesis to my close friends here at UNCG, who have encouraged me all along. In particular, I am indebted to Sydney Garlitch for her sisterly advice and guidance to piece this document together, and AJ Garlitch who always kept me well fed and sane during this process.

And lastly, I dedicate this work and give special thanks to my partner, Talli Sharp for being there for me and listening to the countless hours of mindless ramblings that facilitated the formation of this research project. Your unfaltering support, patience, and calm re-assurance have been hugely helpful. Importantly, thank you for giving me perspective and for making me laugh every single day.

APPROVAL PAGE

This thesis written by Paige L. Kemp has been approved by the following committee of the Faculty of The Graduate School at The University of North Carolina at Greensboro.

Committee Chair	Dr. Christopher Wahlheim
Committee Members	Dr. Robert Wiley
	Dr. Dayna Touron

3/31/2021

Date of Acceptance by Committee

3/27/2021

Date of Final Oral Examination

ACKNOWLEDGEMENTS

I would like to sincerely thank my mentor, Dr. Christopher Wahlheim, who provided me with valuable advice, constructive criticism, and meticulous insights to better this research project. I would also like to express my thanks to the rest of my committee, Dr. Robert Wiley, and Dr. Dayna Touron, for giving me their time and attention in reading my thesis. I thank them for their insightful questions and good-natured support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I: INTRODUCTION

Episodic memory function is commonly known to decline with increasing age (Kausler, 1991; Salthouse, 1991). Some accounts propose that these memory deficits reflect older adults being more susceptible to proactive interference (Hasher & Zacks, 1988; Jacoby et al., 2001; Kausler, 1994). Proactive interference occurs when previously learned information impairs memory for new information (for a review, see Anderson & Neely, 1996). For example, a strong memory for a friend's maiden name could impair memory for their new legal married name. As susceptibility to proactive interference can have a dramatic impact on recall for subsequent information (Underwood, 1957), research has focused on identifying methods to minimize its deleterious effects. Testing previously studied information before the presentation of new information can reducing proactive interference on new learning in both younger and older adults (Pastötter & Bäuml, 2014). Although this forward testing effect has been attributed to a mechanism that segregates contexts, younger adult studies suggest that testing can also promoting integrative encoding (for a review and meta-analysis, see Chan et al., 2018). However, no studies to our knowledge have examined if age-related differences in interference susceptibility can be diminished using interpolated testing to promote integrative encoding.

The overarching aim of the experiment reported here was to address this lacuna by comparing interpolated testing effects in older and younger adults under conditions for which integrative encoding could counteract proactive interference. The specific goals were to 1) verify the role of an integrative mechanism in the effects of testing on proactive interference, and 2) to examine whether there were any age-related deficits that limited older adults' ability to leverage this mechanism in the service of new learning. To achieve this, I used a variant of the classic A-B, A-D paired-associate learning paradigm, wherein subjects learned two competing responses (B, D) paired with the same cue (A) and then completed a cued-recall test that assessed their ability to recall the most recent response (D). To determine how interpolated testing interacted with proactive effects of learning earlier responses, I compared memory for the recent response between conditions that either included or did not include interpolated testing in between the two lists. This comparison allowed me to directly test hypotheses derived from two competing theories of interpolated testing effects: segregation and integration accounts. With respect to the second aim, I compared how interpolated testing influenced the likelihood that younger and older adults would utilize an integrative-encoding mechanism, and how this was associated with memory for the most recent response. Before describing the present experiment, I review relevant literatures on the role of interference susceptibility in age-related memory deficits, interpolated testing effects on memory interference, and the benefits of integrative encoding for counteracting interference.

Age Differences in Memory Interference

As stated above, one account of older adults' deficits in episodic memory (for a review, see Park & Festini, 2017) stems from their greater susceptibility to memory interference (Biss et al., 2012; Hasher & Zacks, 1988;). Age differences in interference susceptibility have been extensively studied in retroactive interference (RI) paradigms, in which memory for previous information is impaired by the learning of new information (for reviews, see Kane & Hasher, 1995; Kausler, 1994). The prevailing finding in RI literature is that older adults are more susceptible to RI, as they recall fewer critical word pairs in comparison to younger adults. However, there are a few exceptions to this trend as some studies demonstrate that older adults do not show more RI than younger adults (e.g., Garlitch & Wahlheim, 2020; Wimer & Wigdor, 1958). These discrepancies may arise from methodological differences, such as length of retrieval lag between-list, degree of semantic association in the material set, and the number of items in the lists.

Age-related declines in performance have also been shown in proactive interference (PI) paradigms, whereby older adults' memory for new information is disproportionally impaired following the prior study of relevant information (Bowles & Salthouse, 2003; Lustig et al., 2001). For instance, studies show that older adults recall fewer target items and produce more intrusion errors under conditions of PI than younger adults (Jacoby et al., 2001; Winocur & Moscovitch, 1983). However, more than in the RI literature, there seems to be a lack of consistency in the findings regarding age-differences in PI, as other studies have shown that older adults do not exhibit greater susceptibility to PI (for a review, see Fernandes & Grady, 2008). As with studies of RI, this may reflect method variance across studies. Collectively, these inconsistencies in age-related deficits in PI suggests that more research is needed to understand the underlying mechanisms of these effects.

Both encoding and retrieval accounts have been forwarded to explain the deleterious effects of PI for younger adults and its amplified effects for older adults when they occur. Encoding-based accounts propose that the prior learning of non-target information disrupts the future encoding of target information. Specifically, it is assumed that after the initial encoding of non-target information, attentional resources decrease and memory load increases, thus reducing how well later lists are encoded (Kliegl et al., 2015; but for a general critique of resource theories, see Navon, 1984). Concerning cognitive aging, previous work has shown that older adults are less likely to engage in effortful memory strategies that are associated with superior performance (Hess, 2014). For instance, it has been demonstrated that older adults are less effective at off-loading items from the focus of attention to storage (Greene et al., 2020) and they tend not to engage in self-initiated elaborative encoding processes unless they are instructed to do so (Kausler, 1982). By this view, it seems plausible to assume that under conditions of PI, older adults will show greater memory deficits as the additional lists will diminish both the quantity and quality of their later memory operations.

Other researchers have attributed PI to retrieval factors (Bennet, 1975; Ikier et al., 2008; Miller et al., 1986). The temporal discrimination theory postulates that PI occurs because individuals are unable to remember the sequential order in which information appeared. For example, the presentation of the additional study list makes it more difficult for participants to discriminate between target items (most recent information) and nontarget items (earlier studied information) during the final test. From a cognitive aging perspective, older adults may be less likely to temporally discriminate target items from nontarget items during the final test due to age-related deficits in source monitoring. Relative to younger adults, older adults typically perform worse in tasks that require item and source recognition (for a review, see Tree & Perfect, 2004). Consequently, this suggests that age-differences in the retrieval of both item and source-based information may be responsible for the increased susceptibility to memory interference observed in older adults.

However, this view has been contradicted by studies indicating that age differences in PI continue to exist even in contexts where competition between target and nontarget responses has been eliminated. For instance, during modified-modified free recall (MMFR) procedures, participants are provided with the first word of each pair (stimuli cue) and instructed to recall the second word of each pair (associated response) from both of the paired-associate lists (Barnes & Underwood, 1959). Siegel (2014) recently employed this procedure and found that in comparison to younger adults, older adults showed significantly greater amounts of proactive interference. This implies that age differences in PI cannot entirely be accounted for by temporal discrimination theory.

The common theme across the mechanisms summarized above suggest that the build-up of PI can partially be attributed to response competition between present and previous material. Based on these accounts, it is natural to inquire whether strong contextual separation between both lists may attenuate or even eliminate interference effects. Indeed, research has shown that when subjects create effective retrieval cues during encoding this promotes list discrimination and enhances target item accessibility at test (Capaldi & Neath, 1995; Watkins & Watkins, 1975). However, although reducing cue-overload can be an effective way to release the build-up of PI, the mechanism underlying this reduction may not be ubiquitous across paradigms. For instance, in reference to the MMFR procedure discussed above, list segregation may be particularly difficult to accomplish in paired-associate paradigms in which many of the items share the same retrieval cues. In these instances, some theoretical models propose that integrating related items during encoding may be a more optimal alternative for preserving temporal order (Kuhl et al., 2010; Wahlheim, 2015). Consequently, this demonstrates that many mechanisms can potentially attenuate interference effects. Given that most of the literature has focused on segregation mechanisms, a critical next step will be to uncover the contributions of response integration mechanisms on reducing PI. Moreover, to have a complete understanding future research should explore the variables that determine whether integration or segregation processes are used to avoid interference.

The Forward Testing Effect

Over the last decade, many researchers have suggested that interpolated testing may be an effective way to reduce the effects of PI. Past research has established the *backward testing effect*, in which taking a memory test on previously learned information can improve the long-term retention of such information (for a review, see Roediger & Butler, 2011). However, recent work has also reported the benefits of testing on the recall of subsequent learning, which is referred to as the *forward testing effect (*for a review and meta-analysis, see Adesope et al., 2017; Chan et al., 2018; Roediger & Karpicke, 2006; Szpunar et al., 2013; Yang et al., 2018). Szpunar et al. (2008) studied the *forward testing effect* using a multi-list paradigm, wherein students studied five word lists in anticipation of a final cumulative recall test. The critical manipulation occurred between lists, in which subjects engaged in one of the following: freely recalled as many words as possible from the list they just studied (test group), restudied the list (restudy group), or completed math problems (control group). Students were then tested on their recall for the fifth list as well as taking a cumulative final test. Relative to the restudy and control groups, the interim test group produced fewer prior-list intrusions and correctly recalled twice as many words on the test of the fifth list.

To explain this pattern of results, Szpunar et al. (2008) proposed that in contrast to restudying, the immediate testing of Lists 1–4 enhances both the short term and long term retention of the content such that participants are better able to segregate the list contexts during the study phase, which improved their ability to isolate retrieval to a specific list on the final test (also see, Jang & Huber, 2008). Specifically, it was assumed that an interpolated testing enhances the discriminability of each list by reducing cue overload and improving source monitoring, leading to a more precise memory search on the final test (Johnson et al., 1993; Watkins & Watkins, 1975). This interpretation aligns with many other studies in the literature that emphasize the critical role of interpolated testing on contextual list segregation and subsequent memory recall (for reviews, see Pastötter & Bäuml, 2014). Collectively, this suggests that interpolated testing could be an effective method to reduce memory interference in both younger and older adults. Despite the obvious relevance of this paradigm for examining how testing ameliorates subsequent memory recall in older adults, to date, research addressing this phenomenon is largely scarce.

Evidence from retroactive learning paradigms suggests that retrieval practice is a potential candidate to improve memory in older adults. In one study of this kind, Meyer and Logan (2013) used prose passages to examine testing effects in younger and older adults at an immediate (5 minutes) and longer delay (2 days). During the task that occurred between the initial study period and final cued-recall test, participants took a recognition test for half of the articles they previously studied, and they restudied the other half. Evidence from the final cued-recall test showed that older adults showed the same amount of testing benefits as younger adults at both short and long delays. The notion that testing produces equivalent benefits for both age groups was further provided by Coane (2013). This study employed a similar method to Meyer and Logan's (2013), but a deep processing condition was included, whereby during initial encoding participants were asked to seek similarities between the word pairs and to generate a mental image that connected both pairs. The results showed that across both age groups retrieval practice and deep processing resulted in enhanced memory relative to restudy, with retrieval practice yielding the greatest benefits overall. This suggested that retrieval practice was a superior technique for repairing older adults' episodic memory deficit.

A recent investigation on proactive effects of learning was conducted by Pastötter and Bäuml (2019) who inferred the late developmental trajectory of the forward testing effect in adulthood in a cross-sectional study including four age-decade groups (40s, 50s, 60s, and 70s). Using a variant of Szpunar et al.'s (2008) method, participants studied three lists of items for a final cumulative recall test. After the presentation of List 1 and List 2, participants were asked to count backwards for 30 s to minimize any long-term recency effects. Immediately following the distraction task, participants in the testing group completed a recall test, in which they were given 60 s to verbally free recall words from the immediately preceding list. Those in the re-study group, however, were re-presented with the words they just studied. To examine how these interpolated tasks affected PI on the final list, all participants were tested on the words studied in List 3. The critical finding was that interpolated testing of Lists 1 and 2 enhanced List 3 recall and reduced the number of prior-list intrusions on the final test relative to interpolated restudy for all age groups. These results suggest that retrieval practice of earlier information produces comparable benefits for both middle-aged and older adults on later memory for new information.

Consistent with list-segregation accounts, Pastötter and Bäuml (2019) attributed these findings to the notion that testing enhances list discrimination which induces context change and reduces memory overload. Again, although there is no direct empirical evidence to support these interpretations, findings from other studies help to augment their assumptions (Jang & Huber, 2008; Szpunar et al. 2008; Pastötter & Bäuml, 2010; Pastötter et al., 2012). Collectively, these findings summarized above provide convincing evidence that the act of retrieval encouraged by testing may be an effective instrument to diminish PI and enhance new learning in older adults.

<u>Theoretical Perspectives Underlying the Forward Testing Effect</u>

As described earlier, one prominent account of the forward testing effect is that the primary mechanism by which testing counteracts PI is by segregating competing sources of information (for reviews, see Pastötter & Bäuml, 2014). However, this account fails to consider how the mechanisms underlying backward and forward testing effects can interact to enhance new learning. For instance, it neglects to articulate how retrieval practice enhances the retention and accessibility of learned information, which can potentially impact how subsequent information is encoded in relation to prior information. Integrating the competing sources of information into one complex memory representation may be particularly useful in paired-associate paradigms, as the shared retrieval cue can be used to activate both responses during the final test. In accord with this idea, Jacoby, Wahlheim, and colleagues recognized the interaction between backward and forward testing in an earlier study and proposed an integrative account to understand how proactive facilitation can occur (Jacoby et al., 2015; Wahlheim & Jacoby, 2013).

To make inferences about the consequences of integrative encoding, Jacoby, Wahlheim, and colleagues proposed the Memory-for-Change (MFC) framework. The MFC (Jacoby et al., 2015; Wahlheim & Jacoby, 2013) builds on Hintzman's (2010) recursive reminding hypothesis and Jacoby's (1991) dual-process theory to explain how remindings can influence episodic memory performance. With respect to the current study, this framework could be used to postulate that testing enhances the accessibility of prior material which may facilitate its subsequent integration with new material. For example, when participants attempt to learn new items that share some features with the previous items (retrieval cue), perceiving a current stimulus may trigger an involuntary or voluntary recollection of a prior stimulus. Consequently, bringing earlier stimuli into the same mental context as the current stimuli may increase the probability of consciously detecting change between items. This change detection should enhance relational processing among these words to the extent that both items can be encoded as one configural representation that also includes memory for the retrieval event (i.e., reminding). At test, recollecting change should be associated with enhanced memory performance (proactive facilitation) because this retrieval process enables access to integrated traces containing both responses and information about their relative temporal order. However, when earlier-detected changes are not recollected this should result in proactive interference because the retrieval practice of the prior responses makes those responses more accessible and more likely to compete with target responses from List 2 at final recall.

Jacoby, Wahlheim, and colleagues provided support for the role of change detection in the integration of responses using variations of the classic A–B, A–D paired-associate learning paradigm (Wahlheim & Jacoby, 2013; Jacoby et al., 2015). Of particular relevance to the current study, Wahlheim (2015) leveraged the MFC framework to determine whether interpolated testing could counteract proactive interference by promoting integrative encoding between earlier and more recently learned information. In that study, participants studied two lists including word pairs with repeated pairs appearing in both lists (A-B, A-B), control pairs appearing only in the second list (A-B, C-D), and changed pairs appearing with the same cue in both lists but with different responses (A-B, A-D). During the interpolated task that occurred between lists, the List 1 items in all except the A-B, C-D condition were divided into conditions such that participants were tested on their recall of the B responses when given the A cues for half the items, while the remaining items were restudied. On the final cued-recall test, participants were asked to recall the responses presented in List 2 and then indicated on a slider how certain they were that a pair had changed between lists. This change recollection measure assayed the extent to which participants detected that a pair had changed and formed an integrated memory trace during the encoding of List 2. The findings showed that for the A-B, A-D condition, PI occurred for pairs that were restudied but not for pairs that were tested in the interpolated phase. The absence of PI in overall performance for the A-B, A-D tested pairs might be attributed to the finding that change recollection occurred more frequently for these items than the restudied items. This association was interpreted as correlational evidence for the causal hypothesis that interpolated retrieval would increase integrative encoding opportunities by making more List 1 responses accessible during the study of List 2. In other words, it suggests that testing facilitates the integration of competing information and promotes later recollection of the attendant representations.

As indicated by the study above, change recollection seems to be paramount for determining whether testing counteracts PI. An objective of the present study was to investigate whether older adults have a deficit in change recollection that limits the benefits they can obtain from interpolated testing. A recent study by Wahlheim (2014) provided evidence suggesting that aging is associated with deficits in change recollection and such deficits partially explained their enhanced susceptibility to proactive interference (also see Hay & Jacoby, 1999; Jacoby et al., 2001). Wahlheim (2014) explained that older adults were poorer at detecting change resulting in fewer integrated traces to be recollected on the final test. Consequently, due to the lack of response integration between targets and competitors this meant that proactive interference was less likely to be eliminated during the final test. Therefore, given that older adults are less efficient at recollecting change and managing response competition, it is reasonable to assume that they are in greater need of procedures to help them counteract the effects of PI.

Despite some evidence suggesting that older adults benefit from testing (Coane, 2013; Meyer & Logan, 2013; Pastötter & Bäuml, 2019), to our knowledge, no studies have examined age-related differences between older and younger adults in proactive facilitation testing effects and whether these differences can partially be explained using an integrative account. One possible reason why this has not been explored is that it goes against recommendations proposed by earlier studies that encourage participants to suppress the nontarget responses during encoding and retrieval. These segregation accounts assume that encouraging contacting between competing information during encoding increases the potential for interference at retrieval by de-segregating list contexts. I addressed this lacuna in the present study by analyzing the proactive effects of interpolated testing on change recollection and associated memory for new information in both younger and older adults using an experimental design also admissible to segregation accounts. Specifically, a variant of the classic A-B, A-D paradigm was used to test competing predictions from integration and segregation accounts simultaneously. By selecting a paradigm that can tease apart integration versus segregation, we can obtain important insights into the underlying mechanisms driving the forward testing effect as well as explore how the efficacy of these mechanisms differ as a function of age.

The Present Study

The primary aim of the present study was to examine the effects of interpolated testing on memory for subsequently learned information, particularly the ability of testing to reduce PI in older and younger adults. The MFC framework assumes that change recollection, following earlier change detection, is required for proactive facilitation (rather than proactive interference) to occur for changed items. However, when change is detected and not recollected this impairs memory performance for changed items, by increasing proactive interference. The MFC framework predicts that interpolated testing promotes response integration by enhancing the accessibility of prior material, which increases the probability of change detection across the cue-pairings and provides an opportunity to integrate the related materials. Considered in combination with well-established age-related episodic memory deficits, the MFC framework also predicts that that older adults should benefit less from interpolated testing than younger adults because their recollection deficits impair their ability to detect and later recollect changes, even after the accessibility of initial responses is enhanced by retrieval practice during interpolated testing.

Important for theory, this prediction runs counter to earlier findings showing comparable forward testing effects in older and younger adults under experimental conditions that support the use of segregation. In contrast to previous studies examing age differences in the efficacy of interpolated testing on new learning, this study will employ study design that lends to the use of integration to counteract proactive interference. If the findings from this experiment reveal that interpolated testing does not produce equivalent testing benefits for younger and older adults, it will more precisely characterize the extent to which testing can improve memory updating in adults of varying ages. Further, if adult age differences in the forward testing effect are associated with differences in the engagement of integrative encoding, it stresses the importance of identifying methods to improve integration.

This current study was designed to build on the previous work by Jacoby, Wahlheim, and colleagues, namely by exploring the effects of interpolated testing on change recollection and its effects on List 2 recall. Specifically, an A–B, A–D paradigm was employed to examine the possibility that retrieval practice for List 1 pairs before the subsequent study of List 2 enhances recall for subsequent learning. Consistent with previous work using this paradigm, this experiment used three items types: pairs that repeated across lists (A–B, A–B); pairs that only appeared in List 2 (C–D); and pairs for which responses changed from List 1 to List 2 (A–B, A–D). The primary manipulation of interest in this study was the type of interpolated task that occurred for some of the A-B, A-D items between List 1 and List 2.

During the interpolated phase, all participants received a cued-recall memory test, in which a cue was provided along with a word-fragment for the response, and participants were instructed to produce a response that corresponded to the words studied in List 1. The purpose of including fragments was to give older adults more support during interpolated retrieval. Prior work has shown that forward testing effects are largely determined by the accuracy of List 1 responses during the interpolated phase; therefore, to give older adults the best opportunity to benefit from retrieval practice, environment support was provided. To add to that benefit, a testing condition that provided corrective feedback after participants provided their response was also included. The rationale for including feedback was that if older adults still could not reach the level of correct recall performance as that of the young, the feedback would help to further repair the episodic memory deficits because it provides another opportunity to encode the correct response. In addition to these two conditions, there was also a condition in which changed items were not tested during the interpolated phase. This was done to compare interpolated testing effects against a condition that would encourage maximal context segregation between lists.

Immediately following the presentation of List 2 responses, a final cued recall test was provided, which assessed memory for the most recently studied response. During this test phase, participants were also asked to indicate whether they recollected change and if so, were told to recall the response that had appeared in List 1 (i.e., change recollection measure). Consistent with Wahlheim's study (2015), this change recollection measure indexed the extent to which competing responses were integrated during the encoding of List 2. A change detection measure was not included during the study of List 2 responses because encouraging the retrieval of List 1 responses may have contaminated any differences produced by the two interpolated testing conditions. Nonetheless, it is still possible to infer differences in change detection during List 2 study from the change recollection measure at test, as prior work has illustrated how the two measures converge. For example, manipulations designed to increase change detection also increased the likelihood of recollecting change at test, and failures to report a list 1 response during the List 2 change detection measure, rarely resulted in the recollection of a List 1 response at test (Wahlheim & Jacoby, 2013).

To replicate the results of Wahlheim (2015), I expected that overall recall performance would not differ between untested A-B, A-D items and A-B, C-D items. Further, I expected that subjects would profit from interpolated testing, such that recall performance would be higher for A-B, A-D tested items than untested A-B, A-D items, and this effect will be associated with increased change recollection. In other words, interpolated testing will enhance memory for changed responses by promoting integrative encoding, as measured via change recollection. However, in circumstances where change was not recollected, it was expected that intrusion errors would be the highest for A-B, A-D items that were tested as compared to untested A-B, A-D items. Lastly, it was predicted that the provision of corrective feedback during interpolated testing would further improve memory performance for changed items by increasing change recollection, but the associated memory benefits will also be offset by greater intrusions when change has not been recollected.

As described above, older adults' show impairments in episodic memory tasks that require self-initiated context-reinstatement (for a review, see Craik & Rose, 2012) and so it was expected that older adults would detect and recollect fewer changes than younger adults. As a result of this, I speculated that on average older adults would accurately recall List 2 responses in the A-B, A-D condition less frequently than younger adults, and this would be associated with corresponding deficits in change recollection (which I used as a proxy of change detection here). As in earlier studies (e.g., Garlitch & Wahlheim, 2020), change recollection was defined as instances when subjects indicated that changes had occurred and could remember the original response. Further, although older adults may still benefit from the increased accessibility of original information via interpolated testing, I expected that those benefits would be smaller than for younger adults, due to older adults' impaired ability to recollect integrated representations (Wahlheim, 2014). That is, older adults would yield fewer differences in memory performance between the changed items that were tested (A–B, A–D Test /A-B, A-D Test with Feedback) vs. changed items that were not tested (A–B, A–D) relative to younger adults.

Extant literature has indicated for an item to benefit from retrieval practice, a subject must be able to successfully retrieve that item from memory, and the likelihood of doing so steadily declines with age. Therefore, to determine the extent to which age-related differences in proactive testing effects are due to change recollection deficits above and beyond encoding deficits it is critical to minimize age differences in original learning. In an attempt to equate performance between age groups for responses that later changed between lists (i.e, List 1 A-B, A-D items), during the interpolated phase subjects were sometimes provided with feedback following their interpolated tested response. It was hypothesized that older adults would benefit from feedback such that performance for changed items that were tested with feedback would be higher than changed items that were tested without feedback. Furthermore, it was hypothesized that providing feedback would also enhance subjects' ability to detect change for items that were not recalled initially, leading to greater change recollection and enhanced overall recall of List 2 responses on the final recall test. Nevertheless, because older adults are known to have recollection deficits it is also plausible that they would detect fewer changes regardless of whether feedback has been provided, thus contributing to their lower List 2 recall benefits than younger adults.

CHAPTER II: METHOD

In what follows, I report how I determined the projected sample size, all data exclusions, all manipulations, and all measures that I used (Simmons et al., 2012). The research reported here was approved by the Institutional Review Board at The University of North Carolina at Greensboro (UNCG).

Participants

My final sample consisted of 60 younger adults (44 women, 15 men, 1 gender diverse *M* = 18.73, *SD* = 1.36, *range* = 18-26 years) from UNCG, 35 older adults (21 women, 14 men, *M* = 71.09, *SD* = 5.71, *range* = 65-84 years) from the Greensboro and the surrounding areas. Two younger adults were replaced due to experimenter error (total of 62 younger adult subjects tested). The younger adult sample had an average of 12.72 years of Education (*range* = 12-17, *SD* = 1.15), and the older adults sample had an average of 16.60 years of Education (*range* = 12-19, *SD* = 1.19). Of the younger adult sample 35% identified as African American (21), 33% Caucasian (20), 12% Hispanic or Latino (7), 12% as more than one race (7), 5% Asian or Pacific Islander (3), and 3% Other (2). The racial composition of the older adults was the following: 83% identified as Caucasian (29), 9% African American (3), 3% Hispanic or Latino (1), 3% as more than one race (1), and 3% Asian or Pacific Islander (1). I screened for cognitively healthy older adults by administering the Short Blessed Test (SBT; Katzman et al., 1983) over the phone and the Mini-Mental State Exam (MMSE; Folstein et al., 1975) in person. Subjects were only included in the analyses if they had a SBT weighted error score $\leq$ 4 and a MMSE score $\geq$ 25 as these signify normal cognitive function (Carpenter et al., 2011; Ismail et al., 2010). All older adults had a visual acuity score $\geq$ 20/50 on the Snellen Eye Test (Hetherington, 1954).

For compensation, younger adults received course credit, and older adults received $15 per hour. Vocabulary scores on the Shipley Institute of Living Scale (Shipley, 1986) were significantly lower for young adults (*M* = 26.72, *SD* = 4.18) than for older adults (*M* = 35.97, *SD* = 3.18), t(93) = -11.34. Scores on the Digit Symbol Substitution Task (DSST; Wechsler, 1981) were significantly lower for older adults (*M* = 56.77, *SD* = 11.53) than younger adults (*M* = 68.40, *SD* = 9.82), t(93) = 5.22. Table 1 displays all cognitive ability scores.

Table 1. Descriptive Statistics of Education, Vocabulary, and Cognitive Measures.

	Age			
	Younger		Older	
Task	Mean (SD)	Range	Mean (SD)	Range
Education (years)	12.72 (1.15)	12 – 17	16.60 (1.91)	12 – 19
SBT (error score)	--	--	0.8 (1.21)	0 – 4
MMSE	28.37 (1.55)	24-30	28.49 (1.36)	25 – 30
Vocabulary (out of 40)	26.72 (4.18)	17-36	35.97 (3.18)	25-40
DSST (in 90s)	68.40 (9.82)	42-87	56.77 (11.53)	33-79
DSST (out of 9)	7.52 (1.90)	2 - 9	6.31 (2.47)	1 – 9

The original plan was to collect a sample that allowed each format of the experiment to appear equally often across subjects (five formats) and was larger enough to determine the presence of a forward testing effect by examining the interaction between Age and Item Type for recall performance. In light of this, a power analysis was conducted using G*Power Version 3.1.9.4 (Faul et al., 2009) to determine what sample size was required. For the statistical test, I selected ANOVA: Repeated measures, within-between interaction. For the input parameters, I set the number of groups to 2 (to reflect the levels of the between subject factor – 2 Age groups), the number of measurements to 5 (to reflect the levels of the within subject factor – 5 Item Types), the correlation among repeated measures was set to .5, and the no sphericity correction was 1. The analysis indicated that with these parameters, a total sample size of 122 would be sufficient to detect a small effect size for the interaction ($\eta_p^2 = .010$) according to Cohen (1988). However, in light of the current pandemic crisis this sample size was not obtained because data collection had to be stopped to protect the safety of others. As stated above, I collected usable data from 95 subjects (60 younger adults and 35 older adults) with the aspiration to finish collecting data from older adults in a safer climate. I conducted a sensitivity analysis using G*Power (Version 3.1.9.2; Faul et al., 2009) and I changed the correlation among repeated measures parameter to .504 as this was the smallest bi-variate correlation observed among the Item Types in the dataset. The analysis indicated that with these parameters and with *N* = 95 subjects, I could still detect a moderate effect size for the interaction (η^2 = .011).

Design and Materials

This experiment employed a 5 (Item type: A-B, A-B; A-B, A-D No Test; A-B, A-D Test with No Feedback; A-B, A-D Test with Feedback; A-B, C-D) × 2 (Age: Younger; Older) mixed design. In subsequent analyses, I will abbreviate the A-B, A-D test conditions by referring to their respective interpolated tasks (i.e., Test and Test^{FB}). Item Type was manipulated within-subjects, and age was a between-subjects' variable.

A total of 80 three-word sets were taken from Jacoby (1996) and Nelson et al. (1998) and were divided into groups of 75 critical and 5 buffer items. The 75 critical items were further divided into five groups of 15 items. The buffer items appeared in the primacy and recency positions of the study lists and as fillers in the interpolated test phase. They also served as practice items on the final test. Each set contained one cue word (e.g., sea) and two semantically related responses (e.g., shore, shell). All responses were orthographically related and were the same length because they were used to complete the same word fragment (e.g., sh _ _ _) for the interpolated task that appeared between List 1 and List 2. Note, that the number of underscores in the word fragments (e.g., sh _ _ _) always reflected the number of missing letters associated with the word of interest (e.g., shore, shell). The average lengths of cues (*M* = 5.40, *SD* = 1.58, range 3–9 letters) and responses (*M* = 4.72, *SD* = 1.21, range 3–8 letters) were matched across groups. The average log frequency of cues (*M* = 9.36, *SD* = 1.79, *range* = 5-14) and responses (*M* = 9.35, *SD* = 1.68, *range* = 4-14) in the Hyperspace Analog to Language corpus (HAL; Lund & Burgess, 1996) were also matched across groups. To obtain the log HAL frequency ratings I used the English Lexicon Project database (Balota et al., 2007). For counterbalancing, the groups were rotated through conditions across experimental formats such that each appeared equally often across subjects, resulting in five experimental formats.

The experiment consisted of four phases in the following order: List 1, an interpolated phase, List 2, and a cued recall test (for a schematic, Table 2). For List 1, List 2, and the cued recall test phase, the non-buffer stimuli appeared in a randomized order. However, stimuli in the interpolated phase appeared in a fixed random order with the restriction that no item type appeared more than three times consecutively. To control for serial position effects during the interpolated phase, the average list position for each item type was equated.

Table 2. General Schematic Design for Experiment (Critical Items Only).

Item Type	List 1 (60 items)	Interpolated Task (30 items)	List 2 (75 items)	List 2 Test (75 items)
A-B, A-D	15 A-B items	None	15 A-D items	15 (A - ?)
	15 A-B items	15 test (A - ?)	15 A-D items	15 test (A - ?)
	15 A-B items	15 test with feedback (A - ?; B)	15 A-D items	15 test with feedback (A - ?)
A-B, A-B	15 A-B items	None	15 A-B items	15 (A - ?)
A-B, C-D	None	None	15 C-D items	15 (C - ?)

List 1 consisted of 64 word pairs, including 4 buffer items (two primacy and two recency) that were evenly divided across four item type conditions (A-B, A-B; A-B, A-D; A-B, A-D Test; A-B, A-D Test^{FB}). Two of these buffer items were later used as filler pairs in the interpolated task, but all four were used in List 2 and served as practice pairs at the final test. The interpolated phase consisted of 32 word pairs (30 critical, 2 filler) that were evenly distributed across the A-B, A-D Test and A-B, A-D Test^{FB} item types (15 critical, 1 filler per condition). As stated previously, the interpolated test items consisted of a cue word (e.g., sea) and a fragment (e.g., sh_ _ _ that could either be completed by *shore* or *shell*). List 2 consisted of 80 word pairs, including 5 buffer items (two primacy and three recency) that were evenly distributed across all five item type conditions. Of the 75 critical items in List 2, A-B, A-B word pairs from List 1 were repeated exactly, but all three A-B, A-D word pairs included repeated cues (A) and changed responses (B, D). Additionally, C-D pairs only appeared in List 2 as they were control pairs against which to assess proactive effects of memory. At test, the five buffer pairs from List 2 were used as practice items before the actual test, which included all 75 critical pairs.

<u>Procedure</u>

Table 3 shows an overview of the procedure. Subjects were tested individually with an experimenter present. The experimental stimuli were presented on computers using E-Prime software (Version 2, Psychology Software Tools, Inc). In all phases of the experiment, stimuli appeared in white Arial size 24 pt lowercase font on a black background. In List 1, word pairs appeared individually for 5000 ms each followed by a 500 ms interstimulus interval (ISI). Subjects were told to read word pairs aloud and to study them for an upcoming memory test. They did not receive any instructions on how to encode the pairs nor did they receive any details on the format for the upcoming test.

Table 3. Overview of the Procedure.

Item Type	List 1	Interpolated	List 2	List 2 Test
A-B, A-D	chalk - write		chalk - white	chalk - ?
A-B, A-D Test	knee - bend	knee – b_n_	knee - bone	knee - ?
A-B, A-D Test^{FB}	sea - shore	sea – sh_ _ shore	sea - shell	sea - ?

During the interpolated phase, subjects were presented with two types of items (i.e, Test Item or Test^{FB} Items). All items appeared as the cue paired with a fragment that could be completed with the response from List 1 (e.g., sea – sh_ _ _). Subjects were told to read the cue and recall the List 1 response aloud while the stimulus was still on the screen. In the Test condition, items appeared as cue-fragment pairs for 7000 ms and were followed by a 500 ms ISI; subjects were not provided with feedback. In the Test^{FB} condition, items appeared as cue-fragment pairs for 5000 ms and then feedback in the form of List 1 response (e.g., shore) appeared below the cue for 2000 ms (7000 ms total). Subjects were told to read the feedback silently. All responses were recorded by an experimenter, and subjects were instructed to stay silent when they could not remember the response. Discouraging subjects from guessing was purposefully done to limit the number of extra-experimental intrusions on the final cued recall test.

During List 2, word pairs were presented in the same way as List 1. Subjects were told to read the word pairs aloud and to study them for a subsequent memory test. They were also told that the upcoming list would contain pairs they had seen before, stimulus cues they had seen before paired with a new response, and pairs they had never seen before. They were instructed to mentally track whether the pairs on List 2 were the same as List 1, changed from List 1, or were new to List 2. Encouraging subjects to notice the discrepancies between cue-response pairings across lists was meant to foster change detection and thus integrative encoding, albeit without measuring them directly.

On the final cued recall test, a cue-question mark pair (e.g., sea - ?) appeared, and subjects were asked to type the response that appeared with the cue in List 2 (e.g., shell). After entering their response, subjects were asked to indicate whether the response that was paired with the cue changed from List 1 to List 2. Subjects were told to press the “1” key to indicate that the response changed and the “0” key to indicate that the response did not change. When subjects indicated that the response changed, they were then asked to type the response that appeared in List 1. When subjects indicated that a pair did not change the program advanced to the next item. If subjects could not remember the target word, they could omit responses by pressing “enter” on the keyboard.

After completing the experiment, all subjects completed the Shipley vocabulary test (Shipley, 1986), the MMSE, and the DSST in that order.

CHAPTER III: RESULTS

Exclusion Criteria

As stated previously, subjects were excluded from the analyses when they failed to achieve scores on the MMSE and Snellen Eye Test that indicated normative status. Subjects were also excluded from analyses when errors were caused by the experimenter. For instance, when the experimenter failed to articulate to the subject how to report answers during the interpolated phase, how the instructions for the List 2 paired-associates differed from List 1, and how to appropriately report answers on the final test with respect to the change recollection measure. To verify when these instances occurred, I referred to the notes on the subject tracking log that were provided by the research assistant running the session.

Analysis Plan

All analyses were conducted using R software (R Core Team, 2019). Before fitting any models, I used the matrix function from the Base R package to define and re-code "contrast" coefficients using a forward difference coding scheme to specify levels of the categorical variable that were required for testing the apriori hypotheses. After defining the contrasts, I fitted a logistic mixed effects models using the glmer from the lme4 package (Bates et al., 2015). Random intercepts by-participants and by-items were included as well as random slopes for Item-Type by-participants and for Age Group-by items, while the predictor variables Age Group and Item Type were included as fixed effects. After fitting the models, I performed hypothesis tests using the Anova function of the car package (Weisberg & Fox, 2011), and pairwise comparisons using the emmeans package (Lenth, 2018). Population estimates for each measure in each cell generated from the mixed effects models were extracted from the emmeans output and visualized along with 95% confidence intervals.

Interpolated Recall

For my first set of analyses, I examined potential age differences in List 1 Recall performance in the interpolated phase (see Figure 1). As stated earlier, during the interpolated test phase subjects were provided with a cue word along with a fragmented List 1 response to strengthen the encoding of List 1 responses. This was done for two reasons, firstly, prior work has indicated that testing benefits for an item are most evident following successful retrieval during interim testing: and secondly, any age-related differences in memory for List 1 responses should be attenuated due to the enhanced level of environmental support, thereby providing older adults with greater opportunity to benefit from retrieval practice. These study design features were proven to be effective, as List 1 Recall performance in the interpolated phase was high and did not differ between younger (M = .82, CI = [.81, .84]) and older (M = .79, CI = [.77, .82]) adults, χ^2 (1) = 1.35, p = .24. There was also no difference between the Test (M = .82, CI = [.80, .84]) and Test^{FB} (M = .80, CI = [.78, .82]) Item Types, χ^2 (1) = 1.24, p = .27, which was expected given that feedback was only provided after subjects submitted their response, thus not influencing their performance. Lastly, there was a no significant Age × Item Type interaction χ^2 (1) = .35, p = .55. Based on these findings, we can conclude that any differences between Age groups and Item Types on the final cued recall test were not due performance differences during the interpolated test.

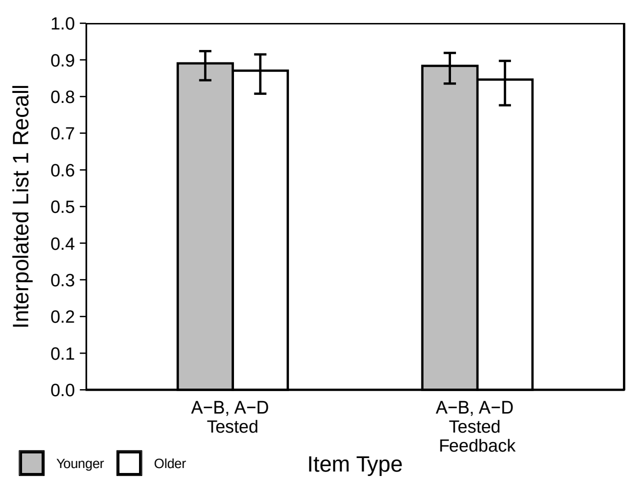

Figure 1. Interpolated List 1 Recall as a Function of Item Type and Age.

The bar plots display List 1 Recall for the two Item Types tested during the Interpolated Task (A-B, A-D Test and A-B, A- D Test^{FB}) for both younger (grey bars) and older adults (white bars). Error bars are bootstrap 95% confidence intervals

Final Cued Recall Test

List 2 Recall

To assess the effects of the interpolated task manipulation on new learning, I examined overall List 2 Recall as a function of Age and Item Type (Figure 2). As expected, a model with Age and Item Type as fixed effects (A-B, C-D; A-B, A-D No Test; A-B, A-D Test; A-B, A- D Test^{FB}) revealed a significant effect of Age, χ^2 (1) = 10.99, *p* < .001, showing that overall memory performance was higher for young than for older adults, *z* = 3.31, *p* < .001. This is presumably because older adults have a compromised ability to create and retrieve associative bindings. There was also a significant Item Type effect, χ^2 (3) = 94.38, *p* < .001 but the interaction between Age and Item Type was not significant, χ^2 (3) = 2.34, *p* = .50. A follow-up analysis showed that interpolated testing on changed items led to an increase in overall memory performance, as List 2 Recall for A-B, A-D Test Items was higher than A-B, A-D No Test Items, illustrating the presence of a significant forward testing effect. List 2 Recall accuracy between A-B, A-D Test Items and A-B, A-D Test^{FB} Items did not differ, *z* = -0.37, *p* = .71, indicating that feedback did not provide any additional recall benefits for List 2 changed responses than testing without feedback. Lastly, there was no evidence for proactive interference in overall recall of List 2 responses for untested changed items, as memory performance did not differ between A-B, C-D Items and A-B, A-D No Test Items, *z* = -0.49, *p* = .62.

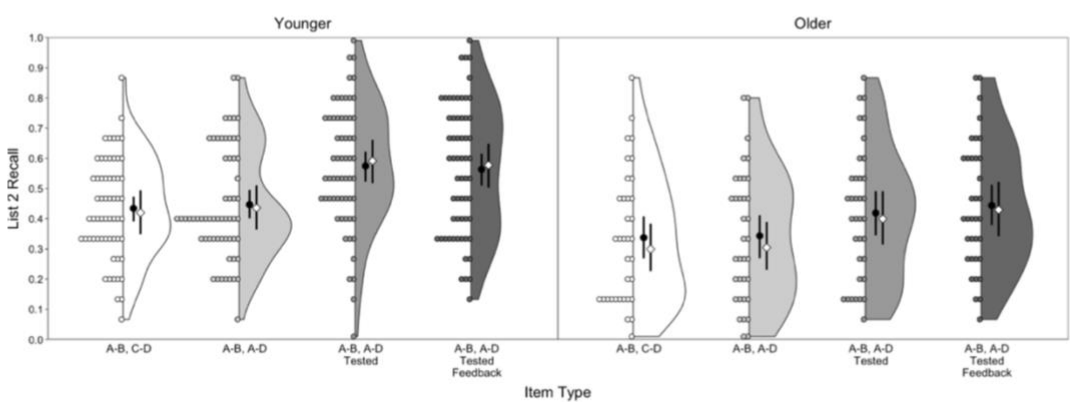

Figure 2. List 2 Recall as a Function of Item Type and Age.

The split violin plots display List 2 Recall for each of the four Item Types for both younger (left panel) and older adults (right panel). The right-side of the violin depicts the shape of the frequency distribution of List 2 Recall and the dots on the left-side of depicts the individual values that correspond to the distributions. Black points inside the plots represent overall performance for each Item Type across both younger and older adults. White diamonds inside the plots represent the model-estimated probabilities for each Item Type across both younger and older adults.

Following my research proposal plan, after fitting the model, a series of planned contrasts were conducted to test the a priori hypotheses of interest. To test the prediction that there would be age differences in the proactive effects of memory for A-B, A-D items, I compared the mean accuracy A-B, C-D Items in List 2 to A-B, A-D No Test Items between age groups. This contrast was not significant $z = -0.23$, $p = .82$. Next, I tested for age differences in the effects of interpolated testing on memory for changed responses by comparing the mean accuracy for List 2 for A-B, A-D items that were tested and were not tested between age groups. This contrast was not significant $z = -1.18$, $p = .24$. Lastly, I examined whether there were age differences in the effects of interpolated testing with feedback on memory for changed responses relative to interpolated testing without feedback by comparing the mean accuracy for List 2 for A-B, A-D Test Items to A-B, A-D Test^{FB} Items between age groups. This contrast was not significant $z = 1.03$, $p = .30$. In sum, contrary to expectations, these findings indicate that younger and older adults showed similar levels of memory performance for items that could have led to proactive interference (i.e., A-B, A-D items) relative to themselves; there were no age differences in the magnitude of the forward testing effect; and the provision of feedback during interpolated testing did not lead to memory performance differences between age groups.

List 1 Intrusions

Another way to identify proactive interference effects is by analyzing the number of List 1 Intrusions errors generated during the retrieval of List 2. To explore the effects of the interpolated task manipulation on proactive interference, I examined overall List 1 Recall on the final cued recall test as a function of Age and Item Type (Figure 3). Note that intrusions for A-D, C-D items were not included in the model, since they did not have a corresponding List 1 pairing to impinge subsequent recall. Consequently, this means that A-B, A-D items that were not tested during the interpolated task served as the baseline measure here relative to A-B, A-D items that were tested with or without feedback. Using a comparable model to that fit to the List 2 Recall data, the model indicated no significant effect of Age, χ^2 (1) = 0.17, p = .68, showing that both age groups had a comparable overall intrusions rate. There was a significant Item Type effect, χ^2 (3) = 245.26, p < .001, but the interaction between Age and Item Type was not significant, χ^2 (3) = 3.75, p = .29. A follow-up analysis showed that intrusion errors did not differ between A-B, A-D No Test Items and A-B, A-D Test Items, z = -0.71, p = .48, showing that interpolated testing (without feedback) did not increase the likelihood of List 1 Intrusions. However, when feedback was provided during interpolated testing this negatively impact memory performance, as List 1 Intrusions were higher for A-B, A-D Test Items than A-B, A-D Test^{FB} Items, z = -2.57, p = .01.

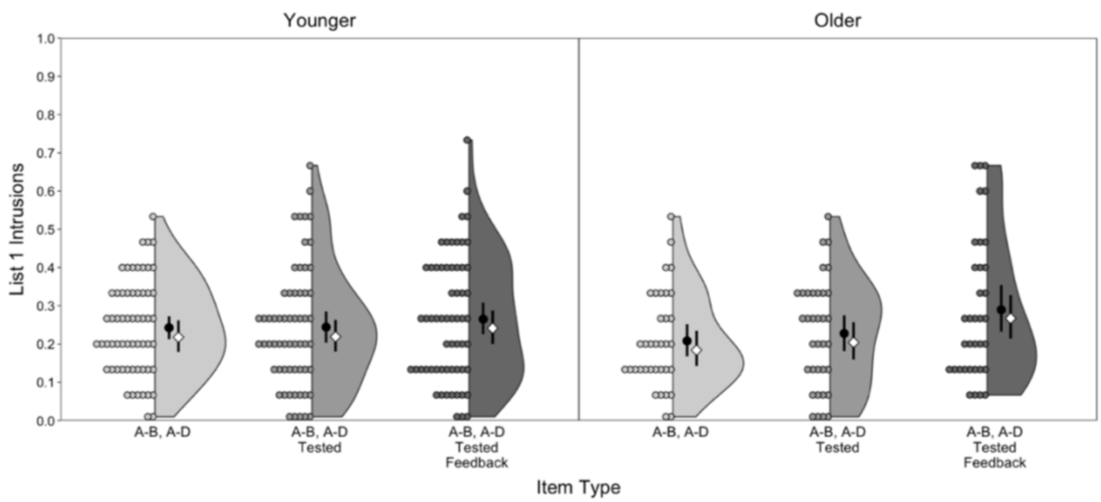

Figure 3. List 1 Intrusions as a Function of Item Type and Age.

The split violin plots display List 1 Intrusions for each of the four Item Types for both younger (left panel) and older adults (right panel). The right-side of the violin depicts the shape of the frequency distribution of List 1 Intrusions and the dots on the left-side of depicts the individual values that correspond to the distributions. Black points inside the plots represent overall performance for each Item Type. White diamonds inside the plots represent the model-estimated probabilities for each Item Type.

After fitting the model, a series of planned contrasts were conducted to test a priori hypothesis. Compatible with predictions for correct recall for List 2 responses, it was anticipated that interpolated testing would significantly benefit younger adults more than older adults, such that fewer intrusions would be generated. To examine age differences in the effects of the interpolated task manipulation on overall List 1 intrusions, I compared the average number of List 1 Intrusions during List 2 Recall for A-B, A-D Items that were tested and not tested between age groups. This contrast was not significant $z = 0.63$, $p = .53$, showing that interpolated testing (without feedback) did not led to differences in the number of intrusions across age groups. The second contrast examined whether the provision of feedback during interpolated testing resulted in differences in intrusion rates between younger and older adults. To do this, I compared mean number of List 1 Intrusions during List 2 Recall for A-B, A-D Test Items to A-B, A-D $Test^{FB}$ Items between age groups. This contrast was not significant $z = 1.22$, $p = .22$, showing that there were no age differences in the effects of feedback on overall List 1 Intrusions. Together, inconsistent with predictions, interpolated testing did not benefit younger adults more so than older adults, in the form of reducing overall intrusions errors; and the provision of feedback did not change this pattern.

Change Classifications on the Final Cued Recall Test

Next, I computed change classification probabilities for A-B, A- D items (Table 4) as a first step towards examining the role of interpolated testing on integrative encoding. Consistent with earlier studies of proactive effects of memory, there were three types of change classification. *Change Recollection* refers to instances when items were correctly classified as changed and List 1 responses were correctly recalled. Extant literature suggests that experimental manipulations have comparable consequences for measures of change detection during List 2 and change recollection on a later test (Jacoby et al., 2013; Wahlheim & Jacoby, 2013). As a result of this, this measure was used to analyze differences in change detection resulting from the interpolated task manipulation and to assess differences in change recollection that I predicted would be associated with differences in memory accuracy for changed responses. *Change Remembered* refers to instances when items were classified as changed and List 1 responses were not correctly recalled. *Change Not Remembered* refers to instances when items were not classified as changed.

Table 4. Change Classification Probabilities as a Function of Item Type and Age.

		Classification		
Item Type	Age	Change Recollected	Change Remembered	Change Not Remembered
A-B, A-D	Younger	.27 [.22, .31]	.07 [.06, .10]	.66 [.61, .71]
	Older	.27 [.20, .33]	.18 [.14, .24]	.55 [.49, .61]
A-B, A-D Test	Younger	.51 [46, .57]	.07 [.05, .10]	.42 [.36, .47]
	Older	.43 [.34, .52]	.17 [.11, .24]	.40 [.33, .47]
A-B, A-D $Test^{FB}$	Younger	.56 [.51, 62]	.05 [.04, .07]	.39 [.34, .44]
	Older	.47 [.39, .55]	.15 [.11, .20]	.38 [.31, .45]

Note. Bootstrap 95% confidence intervals are displayed in brackets

To examine potential age-related differences in the effects of interpolated testing on integrative encoding, as measured through change recollection, I compared estimates across A-B, A-D Item Types for Change Recollection by fitting a 2 (Age) × 3 (Item Type) model. There was no significant effect of Age $\chi^2(1) = 2.33$, $p = .13$, but there was a significant effect of Item Type, $\chi^2(2) = 276.85$, $p < .001$, which was qualified by a significant Age × Item Type interaction, $\chi^2(2) = 7.48$, $p = .02$. Follow-up comparisons provide evidence suggesting that testing List 1 responses in the interpolated phase promoted the integration of competing responses for changed item types, as change recollection was significantly higher for A-B, A-D Test Items than A-B, A-D No Test Items, $z = -12.19$, $p < .001$. The provision of feedback during interpolated testing was shown to further enhance the probability of integrative encoding during List 2, as evinced by the higher rates of change recollection for A-B, A-D $Test^{FB}$ Items than A-B, A-D Test Items, $z = 2.71$, $p < .01$.

After fitting the model, two planned contrasts were conducted to test the a priori hypotheses. First, I examined age differences in the effects of the interpolated task manipulation on change recollection rates for changed responses by comparing change recollection rates for A-B, A-D

No Test Items and A-B, A-D Test Items between age groups. As expected, this contrast was significant $z = -2.36$, $p < .01$, showing that the difference in change recollection between untested and tested changed items was greater for younger adults than older adults. In other words, interpolated testing led to a significantly higher probability of integrative encoding for younger adults than older adults. The second planned contrast examined age differences in the effects of feedback on change recollection for A-B, A-D Item types by comparing rates of change recollection for A-B, A-D Test Items to A-B, A-D Test^{FB} Items between age groups. This contrast was not significant $z = -0.11$, $p = .91$, indicating that the difference in change recollection rates between A-B, A-D items tested with and without feedback was similar for both age groups. Thus, meaning that one age group did not benefit from testing with feedback more than the other, in terms of recollecting change.

List 2 Recall and List 1 Intrusions Conditionalized on Change Classifications

CONDITIONAL LIST 2 RECALL

In my next set of analyses, I examined List 2 Recall conditionalized on the change classification measures at test for each A-B, A-D condition. As described earlier, the MFC framework makes the clear prediction that change recollection should be associated with proactive facilitation and the absence of change recollection should be associated with interference effects. To verify these associations, I fit a model with Age, Item Type (A-B, A-D No Test; A-B, A-D Test; A-B, A- D Test^{FB}), and Change Recollection (Change Recollected; Change Not Recollected) as factors to List 2 Recall. Note that I collapsed Change Remembered and Change Not Remembered into Change Not Recollected, this is because prior studies have shown that proactive facilitation effects only really occur when change is recollected.

Figure 4 displays List 2 correct recall conditionalized on change classifications. As expected, the model indicated a significant effect of Age, $\chi^2(1) = 17.50$, $p < .001$, showing that when change was recollected List 2 Recall performance was significantly higher for younger adults than older adults $z = 4.97$, $p < .001$. Furthermore, in accord with predictions from the MFC framework, the model indicated a significant effect of Change Recollection, $\chi^2(1) = 889.09$, $p < .001$, showing that conditional List 2 Recall accuracy for changed items was significantly higher when change was recollected in comparison to when it was not, z = 28.91, p < .001. There was also a significant interaction between Age × Change Recollection, χ^2 (1) = 30.99, p < .001, and a significant Item Type effect, χ^2 (2) = 7.57, p = .02, which was qualified by an Item Type × Change Recollection interaction, χ^2 (2) = 9.64, p < .01. No other interactions were significant.

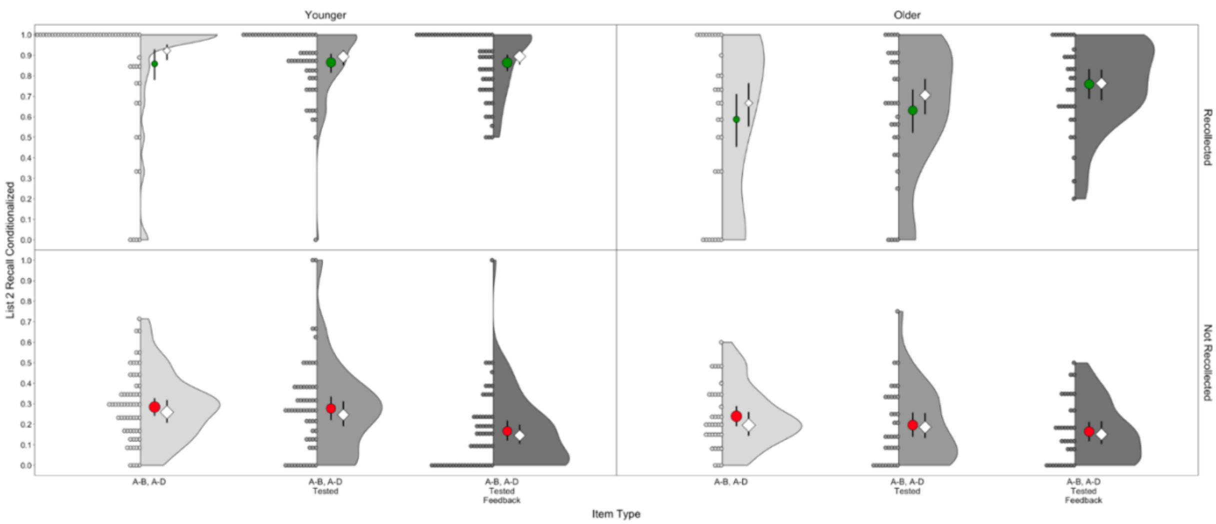

Figure 4. Conditional List 2 Recall as a Function of Item Type, Age, and Change Recollected.

The split violin plots display List 2 Recall conditionalized on change recollected (top panels) and not recollected (bottom panel) for both younger (left panels) and older adults (right panels). The right-side of the violin depicts the shape of the frequency distribution of conditionalized List 2 and the dots on the left-side of depict the individual values that correspond to the distributions. Green points (top panels) represent the conditionalized correct recall for each A-B, A-D item type for instances where subjects indicated change and were able to recall the List 1 response (change recollection). Red points (bottom panels) represent the conditionalized correct recall for each A-B, A-D item type for instances where change was not recollected. White diamonds represent the estimated model means for conditional List 2 Recall respectively for each change classification. The size of the points reflect the proportion of observations included in each cell.

Similar to above, after fitting the model, two planned contrasts were conducted to examine age difference. The first contrast examined whether there were age differences in the interactive effects of interpolated testing and change recollection on memory accuracy. To do this, I compared the mean accuracy for Change Recollected A-B, A-D No Test Items to Change Recollected A-B, A-D Test Items between age groups. This contrast was not significant, z = 1.14, p = .25, showing that the difference in List 2 Recall conditionalized on change recollection between non-tested and tested A-B, A-D items was similar for both age groups. The second contrast assessed age differences in the interactive effects of feedback and change recollection on memory accuracy, by comparing the mean accuracy for Change Recollected A-B, A-D Test Items to Change Recollected A-B, A-D Test^{FB} Items between age groups. This contrast was not significant, $z = -0.26$, $p = .79$, indicating that the difference in List 2 Recall conditionalized on change recollection between A-B, A-D item tested with and without feedback was similar for both age groups.

Conditional List 1 Intrusions

Next, I examined List 1 Intrusions conditionalized on whether change was indicated or not at test for each A-B, A-D condition. As was described in the Introduction, the MFC framework proposes that when change is not recollected this should be associated with more intrusion errors on final cued recall, because the lack of access to an integrated memory trace containing temporal order memory should promote interference. Further, within the context of this study it was hypothesized that interpolated testing would be associated with more intrusions errors because the act of retrieving an earlier response strengthens its representation, thereby increasing its influence as a source of interference. This pattern was expected to be augmented for conditions when feedback was provided during interpolated testing due to the additional rehearsal opportunity for the original cue-response pairings. To test these predictions, I fit a model with Age, Item Type (A-B, A-D No Test; A-B, A-D Test; A-B, A- D Test^{FB}), and Change Remembered (Change Remembered; Change Not Remembered) as factors to conditional List 1 Intrusions. The rationale for excluding the Change Recollected classification from the model was due to prior work showing that intrusions rarely occur when the earlier response is recall correctly, suggesting that examinations with this classification measure would be trivial.

Figure 5 displays List 1 Intrusions conditionalized on change classification. The model indicated a significant Item Type, $\chi^2(2) = 135.10$, $p < .001$, showing lower probabilities of List 1 Intrusions for A-B, A-D No Test Items than A-B, A-D Test Items, $z = -6.45$, $p < .001$, and for A-B, A-D Test Items than A-B, A-D Test^{FB} Items, $z = -4.22$, $p < .001$. Congruent with the MFC account, this suggests that interpolated testing inflated the likelihood of intrusions, and this was particularly apparent when feedback was included during the interpolated phase. There was also a significant effect of Change Remembered, $\chi^2(1) = 6.24$, $p < .01$, showing lower probabilities of List 1 Intrusions when changes were remembered than when they were not. No other effects or interactions were significant. These results suggest that remembering change was sufficient to reduce the probabilities of intrusions during final recall, but interpolated testing increased the likelihood of making intrusions, especially when feedback was provided.

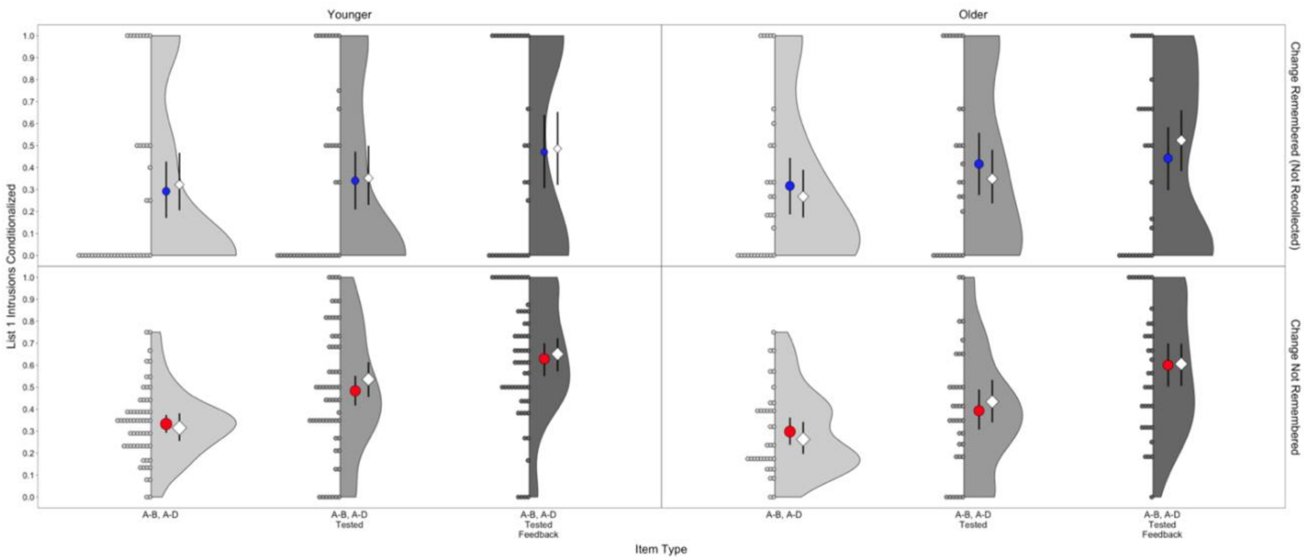

Figure 5. Conditional List 1 Intrusions as a Function of Item Type, Age, and Change Not Remembered.

The split violin plots display List 1 Intrusions conditionalized on change remembered (top panels) and not remembered (bottom panel) for both younger (left panels) and older adults (right panels). The right-side of the violin depicts the shape of the frequency distribution of conditionalized List 1 Intrusions and the dots on the left-side of depict the individual values that correspond to the distributions. Blue points (top panels) represent the conditionalized intrusion rate for instances where subjects indicated change but were unable to recall the List 1 response (change recollection). Red points (bottom panels) represent the conditionalized intrusion rate for instances where change was not remembered. White diamonds represent the estimated model means for conditional List 2 Recall respectively for each change classification. The size of the points reflect the proportion of observations included in each cell.

Similar to that of conditionalized List 2 Recall, after fitting the model a series of planned contrasts were conducted. The first contrast examined age differences in the interactive effects of interpolated testing and change classification on List 1 Intrusions, by comparing the average number of List 1 Intrusions for Change Not Remembered A-B, A-D No Test Items to Change Not Remembered A-B, A-D Test Items between age groups. This contrast was not significant, $z = 0.67$, $p = .50$, showing that the difference in List 1 Intrusions conditionalized on change not remembered between non-tested and tested A-B, A-D items was similar for both age groups. The second planned contrast assessed age differences in the interactive effects of feedback and change classification on List 1 Intrusions. To do this, I compared the average number of List 1 Intrusions for Change Not Remembered A-B, A-D Test Items and Change Not Remembered A-B, A-D Test^{FB} Items. This contrast was not significant, $z = -0.08$, $p = .94$, indicating that the difference in List 1 Intrusions conditionalized on change not remembered between A-B, A-D items tested with and without feedback was similar for both age groups.

CHAPTER IV: DISCUSSION

The present experiment was the first to examine whether age-related deficits under conditions that could lead to proactive interference can be alleviated using interpolated testing to promote integrative encoding. Similar to prior work using an analogous paradigm, there was no evidence for proactive interference in overall performance, presumably due to the high semantic and orthographic overlap across pairings facilitating cue-dependent retrieval of earlier memories during new learning. Moreover, consistent with literature, a comparable forward testing effect was observed for both age groups, with retrieval practice of earlier information (A-B pairs) enhancing subsequent memory for new information (A-D pairs), but older adults were shown to have poorer overall performance. In addition, the present results implicate that the proactive effects of testing can partially be attributed to an integration encoding mechanism. Integrated encoding was measured by change recollection which indexed subject's ability to recall List 1 responses at test due to detecting change and integrating both responses during the study of List 2. Change recollection was higher for tested items (A-B, A-D Tested) than untested items (A-B, A-D), and change recollection was associated with enhanced List 2 Recall and the absence of change recollection was associated with impaired List 2 Recall. Older adults recollected fewer changes and benefitted less from recollected changes than younger adults. This partly accounted for older adults' poorer overall recall. Collectively, the obtained results imply that interpolated testing promoted integrative encoding for both age groups, but older adults enjoyed these benefits on fewer trials and to a lesser degree. In what follows, I will discuss the theoretical and applied implications of these findings.

Implications for Theories of Age Differences in Interference Effects

Earlier studies examining aging and interference effects have produced mixed findings, with some studies showing that older adults are more susceptible to interference effects, while others showing no age differences. When age differences are observed, they are generally attributed to mechanisms during encoding, retrieval, or a combination of both. Across all the proposed mechanisms, the theme that emerges is that older adults are less able to differentiate between materials learned at different points in time as well as younger adults. Compatible with this view, research has shown that one way to improve list segregation is through interpolated testing. Recently, Pastötter and Bäuml (2019) found that interpolated testing helped older adults overcome the occurrence of proactive interference in a multi-list learning paradigm, as evidenced by their higher recall for target words and fewer prior-list intrusion errors on subsequently learned material relative to a restudy condition. To account for this forward testing effect, it was proposed that testing between lists induces context change, leading to improved temporal discrimination and enhanced encoding for subsequent materials. Conversely, empirical research has also shown that proactive interference can be reduced in situations that encourage integration. For instance, Wahlheim (2015) found that practicing retrieval of a competing response counteracted proactive interference in part by integrating competing sources of information during encoding. Although such effects have been shown in younger adults, the extent to which interpolated testing influences integrative encoding and subsequent memory recall in older adults remains unknown.

This was addressed in the current study using a paradigm that was conducive to integration and segregation mechanisms, enabling one to directly test hypotheses derived from each account. Segregation accounts assume that inserting retrieval practice between lists enhances temporal discrimination by promoting more distinct contextual representations that enhances participants ability to use list-specific context cues to recall target lists. Integration accounts though, advocate that testing fosters the co-activation of related information, allowing both items and their associated context to be encoded into a configural trace. To disentangle the competing predictions within one study, a A-B, A-D list learning paradigm was employed, wherein subjects learned different responses (B, D) that shared the same retrieval cue (A) and were later tested on their memory for the most recent response (D). If the segregation account holds true here, the responses across the lists will be kept separate, such that the retrieval of one is not dependent on the other. Contrary to this, integration accounts predict that memory performance for the recent response (D term) will be enhanced if an earlier response (B term) can be retrieved, since the two responses are integrated.

Consistent with prior work, the results showed that retrieval practice of earlier learned materials (A-B pairs) enhanced memory for subsequent materials (A-D pairs) for both age groups, but older adults had lower recall performance overall. Beyond replicating the forward testing effect, the results provided evidence in support of a role of an integrative mechanism in the effects of testing on proactive interference. The extent to which new responses were integrated with pre-existing memories due to testing was revealed by greater frequencies of change recollection for tested items relative to non-tested. Analyses of the recall for changed responses conditionalized on change recollection showed that when change was recollected this was associated with higher memory accuracy for the change response compared to when change was not recollected. In other words, List 2 responses were recalled more often when List 1 responses were also recalled. Integration accounts can account for this retrieval dependency in a way that contextual segregation accounts cannot. Contextual segregation accounts that propose that the responses are stored in independent memory representations, meaning that retrieval List 1 of impairs memory for List 2 in part by diminishing contextual differentiation. Whereas integrative encoding accounts propose that when overlapping responses are connected during encoding, the retrieval of a List 1 response facilitates memory for List 2, since they are bound together into one representation.

It was also shown that providing feedback during the interpolated test tended to increase change recollection, this is presumably because feedback serves as an additional study opportunity to strengthen the memory representation for the original cue-response pairing making it more likely to be reactivated later. However, there were downstream consequences for List 2 recall when change was not recollected, as evidenced by greater number of intrusion errors. This is because in the absence of change recollection the enhanced accessibility of prior responses provided by feedback renders them more competitive for automatic retrieval. This finding substantiates the results of Wahlheim (2015) and similar findings from Wahlheim (2014) who found that increasing the number of List 1 repetition facilitates change recollection, but the associated benefits were offset by greater proactive interference when changed was not recollected. Overall, a combination of all these findings highlights that the detection and later recollection of change plays a critical role in the forward testing effects.

However, of critical importance for the current study was to examine whether there were any age-related deficits in change recollection that limited the benefits older adults could enjoy from testing. Based on previous findings illustrating that age differences in susceptibility to proactive interference can partially be explained by a deficit in change recollection (Wahlheim, 2014), I expected that older adults would benefit less from testing than younger adults because their recollection deficits would impair their ability to integrate effectively. As stated above, there were no age-related differences in the proactive effects of testing in overall List 2 recall performance, but it was revealed that older adults were less likely to recollect change and they obtained fewer benefits even when they did recollect change relative to younger adults. That is, the retrieval of a List 1 response did not necessarily lead to successful retrieval of a List 2 response. One potential explanation for why older adults may not have benefitted from change recollection to the same extent as younger adults is because their known age-related bindings deficits (Naveh-Benjamin, 2000) may have compromised their ability to form and recollect integrated traces. Overall, while inconclusive, the present results imply that change recollection could have played a role in age-related differences in overall List 2 recall performance, but this interaction was unable to be detected possibly because the smaller final sample size than was originally planned.

So far, I have emphasized how the data provides some evidence to support the notion that integration plays a significant role in the benefit of testing in insulating against proactive interference. It is worth noting, however, that there could be other factors that contribute to explain the pattern of results observed here. It was revealed that older adults benefitted less from change recollection in comparison to younger adults and so its plausible that another mechanism may be responsible for why older adults still show the same overall benefits relative to themselves than the young do. As mentioned previously, some accounts propose that testing helps differentiate the lists and so it might be that for the times when older adults were successful at recalling from List 1 but unsuccessful at integrating or even noticing the change, it might be because these representations are more separated. Consequently, given that these mechanisms are not mutually exclusive, it could be that there is a mixture of change recollection for some trials and greater differentiation for others that is driving the forward testing effect in older adults. This view aligns with previous literature suggesting that multiple mechanisms may contribute to the forward testing effect (Pastötter & Frings, 2019).

Implications for Theories of Memory Integration

The present results have general implications for the MFC framework (Jacoby et al., 2015; Wahlheim & Jacoby, 2013) that was previously mentioned in the Introduction. The account was originally proposed to explain the role of remindings and memory for change in proactive effects of memory. The MFC framework argues that when current events share an overlapping association with previously encoded information this may trigger the retrieval of past information. The co-activation of old and new memories provides a unique opportunity to register the contextual differences and form an integrated representation. At test, when change is recollected the retrieval of these representations can preserve temporal order and also facilitate memory for the updated information. However, when earlier detected changes are not recollected this can lead to proactive interference because the study-phase retrieval of an earlier response makes it more competitive for reactivation. The current study supports these assumptions by showing that when change was recollected this was associated with facilitation of List 2 recall. Further, the findings here imply that interpolated testing facilitated integrative encoding by increasing the accessibility of the earlier information, thus providing greater opportunities for the detection and later recollection of change.

As mentioned previously, the results reported here build on prior work by Wahlheim (2015) by showing the benefits of interpolated testing on integration encoding in older adults as well as younger adults. However, one notable difference between the two experiments is that this study did not find proactive interference in overall List 2 recall for either age group. One explanation that may contribute to this discrepancy is the use of different materials; whereas Wahlheim (2015) used responses that were neither semantically nor orthographically related to each other, all of the responses used here were orthographically related so that they could complete the same word fragment. As a result, the strong associations among the terms may have encouraged contact between the lists permitting subjects to detect change in List 2 and thus, foster an environment for integration. Future research should aim to examine the role of interpolated testing on new learning using different materials that are conducive to integration encoding.

Empirical evidence to support the idea that details of prior events can be retrieved when they share an overlapping association with present information and can be integrated into a memory trace for later use has also been provided by several neuroimaging studies. For example, Shohamy and Wagner (2008) used functional magnetic resonance imaging (fMRI) to examine the involvement of the hippocampus and midbrain during an acquired equivalence paradigm. In this study, subjects first studied a series of individual associations (F1 – S1; F1 - S2; F2 – S1) and then later were asked to make generalizations across pairs of items (F2 - ?) whose relationship was mediated by an explicitly learned common element (S1). Findings showed that hippocampal and midbrain activation during learning was associated with later generalization performance, suggesting that the hippocampus may be important for relating information across pairings for later flexible use. Similar findings have been documented using associative interference paradigms in which hippocampal activation during *encoding* predicted later inference success (Zeithamova & Preston, 2010). Most relevant to the current study is the finding that interpolated testing facilitates new learning by enhancing the medial prefrontal cortex representations that support memory integration (Ye et al., 2020). Together, the converging results from neural and behavioral studies highlight that integration during encoding plays a critical role in reducing interference effects at retrieval, and interpolated testing serves to foster integrative encoding.

Limitations and Future Directions

The empirical results reported herein should be considered in light of some potential limitations. One concern is that the effects of interpolated testing on integrative encoding are inferred from differences in change recollection at test. As a result, the lack of direct evidence to support the presence of integrative encoding could lead one to argue that other mechanisms are responsible for the testing effects observed here. However, this seems untenable since previous studies have demonstrated that if subjects are unable to retrieve the earlier response during the presentation of the second list, then the probability of showing this retrieval dependency at test is highly rare (Wahlheim & Jacoby, 2013). The reason why the present study excluded an overt change detection measure was to avoid any reactive effects on subsequent encoding and thus prevent contaminating the effects of the interpolated task. Therefore, to verify that the downstream measure of change recollection is indicative of this earlier detection future work should find ways to include an overt detection measure without overriding the effects of the interpolated task.

Another concern with the current experiment is that some of the analyses of final test performance were conditionalized on change recollection to examine the influence of interpolated testing on integrative encoding and subsequent recall. As mentioned in the Method, efforts to collect data from older adults were hampered by the ongoing pandemic of coronavirus disease 2019, such that just over half of the intended sample size was collected. Consequently, this smaller sample size limits the statistical power of the current sample to detect the effects of interest, potentially explaining why an age-related interaction was not observed in overall recall performance. Thus, to overcome this limitation, I plan to continue older adult recruitment until the target sample size is has been collected.

Conclusions

The present experiment was the first to characterize the role of change recollection on the forward testing effect in both younger and older adults. Findings demonstrated that interpolated testing produced comparable benefits for both age groups in part by promoting the integration of competing responses. For instance, interpolated testing increasing change recollection, and change recollection was associated with enhanced memory for subsequent information. This study contributes to a growing body of evidence suggesting that age-related memory deficits created by response competition can be reduced through memory integration processes. It also provides a new perspective to the test-effect literature by showing that testing can counteract proactive interference by integrating competing information for both younger and older adults. Future work should focus on exploring the variables that drive the forward testing effect and investigate the boundary conditions for which testing counteracts proactive interference in older adults.

REFERENCES

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: A meta-analysis of practice testing. *Review of Educational Research*, *87*(3), 659-701.

Anderson, M., & Neely, J. (1996). Interference and Inhibition in Memory Retrieval. *Memory*, 237-313. https:/doi.org/10.1016/b978-012102570-0/50010-0

Barnes, J. M., & Underwood, B. J. (1959). " Fate" of first-list associations in transfer theory. *Journal of experimental psychology*, *58*(2), 97.

Bates, D., Mächler, M., & Dai, B. (2011). *Lme4: Linear Mixed-Effects Models Using S4 Classes at.* Available at: http://lme4.r-forge.-r-project.org/

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1-48.

Bennett, R. W. (1975). Proactive interference in short-term memory: Fundamental forgetting processes. *Journal of Verbal Learning and Verbal Behavior*, *14*(2), 123-144.

Biss, R. K., Campbell, K. L., & Hasher, L. (2013). Interference from previous distraction disrupts older adults' memory. *Journal of Gerontology: Psychological Sciences*, *68*, 558–561. https:/doi.org/10.1093/geronb/gbs074.

Bowles, R. P., & Salthouse, T. A. (2003). Assessing the age-related effects of proactive interference on working memory tasks using the Rasch model. *Psychology and Aging, 18,* 608–615.

Capaldi, E. J., & Neath, I. (1995). Remembering and forgetting as context discrimination. *Learning & Memory, 2,* 107–132.

Carpenter, C. R., Bassett, E. R., Fischer, G. M., Shirshekan, J., Galvin, J. E., & Morris, J. C. (2011). Four sensitive screening tools to detect cognitive dysfunction in geriatric emergency department patients: Brief Alzheimer's Screen, Short Blessed Test, Ottawa 3DY, and the Caregiver-Completed AD8. *Academic Emergency Medicine*, *18*(4), 374-384.

Chan, J. C., Meissner, C., & Davis, S. (2018). Retrieval Potentiates New Learning: A Theoretical and Meta-Analytic Review. *Psychological Bulletin*, *144*(11), 1111-1146.

Coane, J. H. (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition, 2*(2), 95–100.

Craik, F. I., & Rose, N. S. (2012). Memory encoding and aging: a neurocognitive perspective. *Neuroscience & Biobehavioral Reviews*, *36*(7), 1729-1739.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*(4), 1149–1160. https:/doi.org/10.3758/brm.41.4.1149

Fernandes, M., & Grady, C. (2008). Age Differences in Susceptibility to Memory Interference During Recall of Categorizable but Not Unrelated Word Lists. *Experimental Aging Research*, *34*(4), 297-322. https:/doi.org/10.1080/03610730802273860

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). “Mini-mental state”. A practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research, 12*, 189–198.

Greene, N. R., Naveh-Benjamin, M., & Cowan, N. (2020). Adult age differences in working memory capacity: Spared central storage but deficits in ability to maximize peripheral storage. *Psychology and Aging, 35*(6), 866–880. https://doi.org/10.1037/pag0000476

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: a review and a new view. *The Psychology of Learning and Motivation, 50,* 193–225

Hay, J. F., & Jacoby, L. L. (1999). Separating habit and recollection in young and older adults: Effects of elaborative processing and distinctiveness. Psychology and Aging, 14, 122–134.

Hess, T. M. (2014). Selective engagement of cognitive resources: Motivational influences on older adults’ cognitive functioning. *Perspectives on Psychological Science*, *9*(4), 388-407.

Hetherington, R. (1954). The Snellen chart as a test of visual acuity. *Psychologische Forschung*, *24,* 349-357.

Hintzman, D. L. (2009). How does repetition affect memory? Evidence from judgments of recency. *Memory & Cognition, 38*(1), 102–115. https:/doi.org/10.3758/MC.38.1.102.

Ikier, S., Yang, L., & Hasher, L. (2008). Implicit Proactive Interference, Age, and Automatic versus Controlled Retrieval Strategies. *Psychological Science, 19*(5), 456-461.

Ismail, Z., Rajji, T. K., & Shulman, K. I. (2010). Brief cognitive screening instruments: an update. *International Journal of Geriatric Psychiatry: A journal of the psychiatry of late life and allied sciences*, *25*(2), 111-120.

Jacoby, L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal Of Memory And Language*, *30*(5), 513-541. https:/doi.org/10.1016/0749-596x(91)90025-f

Jacoby, L. L., Debner, J. A., & Hay, J. F. (2001). Proactive interference, accessibility bias, and process dissociations: Valid subjective reports of memory. *Journal of Experimental*

Psychology: Learning, Memory, and Cognition, 27, 686–700. https:/doi.org/10.1037/0278-7393.27.3.686

Jacoby, L., Wahlheim, C., & Yonelinas, A. (2013). The role of detection and recollection of change in list discrimination. *Memory & Cognition*, *41*(5), 638-649. doi: 10.3758/s13421-013-0313-x

Jacoby, L. L., Wahlheim, C. N., & Kelley, C. (2015). Memory consequences of looking back to notice change: Retroactive and proactive facilitation. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 41(5),* 1282-1297. https:/doi.org/10.1037/xlm0000123

Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term memory drives list isolation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 34,* 112–127.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*(1), 3–28. https:/doi.org/10.1037/0033-2909.114.1.3

Kane, M. J., & Hasher, L. (1995). Interference. In G. L. Maddox, R. C. Atchley, J. Grimley Evans, C. E. Finch, D. E. Hultsch, R. A. Kane,... I. C. Siegler (Eds.), The encyclopedia of aging: A comprehensive re-source in gerontology and geriatrics(pp. 514–516). New York, NY: Springer.

Katzman, R., Brown, T., Fuld, P., Peck, A., Schechter, R., & Schimmel, H. (1983). Validation of a short orientation-memory concentration test of cognitive impairment. *American Journal of Psychiatry,* 140, 734-739.

Kausler, D. H. (1982). Experimental psychology and human aging. New York: John Wiley and Sons.

Kausler, D.H. (1991) Experimental Psychology, Cognition, and Human Aging. 2nd Edition, Springer Verlag, New York. https://doi.org/10.1007/978-1-4613-9695-6

Kausler, D. H. (1994). Learning and memory in normal aging. New York, NY: Academic Press.

Kliegl, O., Pastötter, B., & Bäuml, K.-H. T. (2015). The contribution of encoding and retrieval processes to proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1778–1789. https:/doi.org/10.1037/xlm0000096

Kuhl, B. A., Shah, A. T., DuBrow, S., & Wagner, A. D. (2010). Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. *Nature neuroscience*, *13*(4), 501-506.

Lenth, R. (2018). Emmeans: Estimated marginal means, aka, least-squares means. R Package. https://CRAN-Rproject.org/package=emmeans.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130,* 199–207.

Meyer, A. N., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, *28*(1), 142.

Miller, R., Kasprow, W., & Schachtman, T. (1986). Retrieval Variability: Sources and Consequences. *The American Journal of Psychology, 99*(2), 145-218. https:/doi.org/10.2307/1422275

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170.

Navon, D. (1984). Resources—A theoretical soup stone?. *Psychological review*, *91*(2), 216.

Park, D. C., & Festini, S. B. (2017). Theories of memory and aging: A look at the past and a glimpse of the future. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 72,* 82–90.

Pastötter, B., & Bäuml, K. (2010). Amount of post cue encoding predicts amount of directed forgetting. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *36*(1), 54-65. https:/doi.org/10.1037/a0017406

Pastötter, B., & Frings, C. (2019). The forward testing effect is reliable and independent of learners' working memory capacity. *Journal of cognition*, *2*(1).

Pastötter, B., Kliegl, O., & Bäuml, K. (2012). List-method directed forgetting: The forget cue improves both encoding and retrieval of post cue information. *Memory & Cognition*, *40*(6), 861-873. https:/doi.org/10.3758/s13421-012-0206-4

Pastötter, B., & Bäuml, K.-H. T. (2014). Retrieval practice enhances new learning: The forward effect of testing. *Frontiers in Psychology, 5,* Article 286.

Pastötter, B., & Bäuml, K. (2019). Testing enhances subsequent learning in older adults. *Psychology And Aging*, *34*(2), 242-250. https:/doi.org/10.1037/pag0000307

Roediger, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, *15*(1), 20-27. https:/doi.org/10.1016/j.tics.2010.09.003

Roediger, H.L., III, & Karpicke, J.D. (2006). The power of testing memory: Implications for educational practice. Unpublished manuscript, Washington University in St. Louis.

Salthouse, T. A. (1991). *Theoretical perspectives on cognitive aging.* Hillsdale, NJ: Lawrence Erlbaum

Shohamy, D., & Wagner, A. D. (2008). Integrating memories in the human brain: hippocampal-midbrain encoding of overlapping events. *Neuron*, *60*(2), 378-389.

Shipley, W. C. (1986). Shipley Institute of Living Scale. Los Angeles: Western Psychological Services.

Siegel, J. (2014). *Brain Volume Differences and Proactive Interference: A VBM Investigation of Healthy Cognitive Aging* (Doctoral dissertation, University of Pittsburgh).

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2012). A 21-word solution. Dialogue: The Official Newsletter of the Society for Personality and Social Psychology, 26, 4–7.

Szpunar, K., McDermott, K., & Roediger, H. (2008). Testing during study insulates against the build-up of proactive interference. *Journal Of Experimental Psychology: Learning, Memory, And Cognition*, *34*(6), 1392-1399. https:/doi.org/10.1037/a0013082

Szpunar, K. K., Khan, N. Y., & Schacter, D. L. (2013). Interpolated memory tests reduce mind wandering and improve learning of online lectures. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 110*(16), 6313–6317. https:/doi.org/10.1073/pnas.1221764110

Tree, J., & Perfect, T. (2004). A Reexamination of Source Monitoring Deficits in the Elderly: Evidence for Independent Age Deficits of Item and Source Memory. *Brain Impairment*, *5*(2), 138-144. https:/doi.org/10.1375/brim.5.2.138.58251

Underwood, B. (1957). Interference and forgetting. *Psychological Review*, *64*(1), 49-60. https:/doi.org/10.1037/h0044616

Wahlheim, C. N., & Jacoby, L. L. (2013). Remembering change: The critical role of recursive remindings in proactive effects of memory. *Memory & Cognition, 41,* 1–15. https:/doi.org/10.3758/s13421-012- 0246-9

Wahlheim, C. N. (2014). Proactive effects of memory in young and older adults: the role of change recollection. *Memory & cognition*, *42*(6), 950–964. https:/doi.org/10.3758/s13421-014-0411-4

Wahlheim, C. N. (2015). Testing can counteract proactive interference by integrating competing information. *Memory & cognition, 43*(1), 27–38. https:/doi.org/10.3758/s13421-014-0455-5

Watkins, O. C., & Watkins, M. J. (1975). Build-up of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory, 104,* 442–452.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised*. New York, NY: Psychological Corporation.

Weisberg, S., & Fox, J. (2011). *An R Companion to Applied Regression*. (2 ed.) Thousand Oaks: Sage. http://socserv.mcmaster.ca/jfox/ Books/Companion/index.html

Wimer, R. E., & Wigdor, B. T. (1958). Age differences in retention of learning. *Journal of Gerontology*, *13*(3), 291-295.

Winocur, G., & Moscovitch, M. (1983). Paired-associate learning in institutionalized and noninstitutionalized old people: An analysis of interference and context effects. *Journal of Gerontology, 38*(4), 455–464. https://doi.org/10.1093/geronj/38.4.455

Yang, C., Potts, R., & Shanks, D. R. (2018). Enhancing learning and retrieval of new information: a review of the forward testing effect. *NPJ science of learning*, *3*(1), 1-9.

Ye, Z., Shi, L., Li, A., Chen, C., & Xue, G. (2020). Retrieval practice facilitates memory updating by enhancing and differentiating medial prefrontal cortex representations. *Elife*, *9*. https://doi.org/10.7554/elife.57023

Zeithamova, D., & Preston, A. R. (2010). Flexible memories: differential roles for medial temporal lobe and prefrontal cortex in cross-episode binding. *Journal of Neuroscience*, *30*(44), 14676-1468.