Summer 8-25-2013

The Efficiency of Retrieval Practice as a Function of Spacing and Intrinsic Value in Young and Older Adults

Geoffrey B. Maddox
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychology

Dissertation Examination Committee:
David A. Balota, Chair
Lisa Connor
Janet M. Duchek
Mark McDaniel
Henry L. Roediger, III
James Wertsch

The Efficiency of Retrieval Practice as a Function of Spacing and Intrinsic Value in Young and Older Adults

by

Geoffrey Brandon Maddox

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2013

St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

.

List of Appendices

Acknowledgements

First, I thank my primary advisor and committee chair, Dave Balota, for his guidance and support throughout my time as a graduate student, as well as Jan Duchek for serving as a secondary advisor throughout the past six years. I also thank the other members of my dissertation committee for their helpful input: Lisa Connor, Mark McDaniel, Roddy Roediger, and Jim Wertsch. I am grateful for the helpful feedback and discussions I had with past and present members of the Cognitive Psychology Lab. Finally, I thank Haliday Douglas, my family, and friends who have been extremely supportive and encouraging over the past six years.

This research was supported by NIA Training Grant AG00030.

Please address correspondence to Geoffrey B. Maddox, Rhodes College, 110 Clough Hall, Memphis, TN 38112 or via email maddoxg@rhodes.edu.

ABSTRACT OF THE DISSERTATION

The Efficiency of Retrieval Practice as a Function of Spacing and Intrinsic Value in Young and Older Adults

By

Geoffrey Brandon Maddox

Doctor of Philosophy in Psychology

Washington University in St. Louis, 2013

Professor David A. Balota, Chair

Two powerful methods of improving memory in young and older adults are spacing and testing. The spacing effect refers to the observation that learning material with intervening material between study events, compared to no intervening material between study events, improves long-term memory. The testing effect refers to the finding that retrieving information from memory (via testing) improves memory over merely restudying the information. A combination of the two methods is referred to as spaced retrieval practice, which is the focus of the present dissertation. There were three specific questions addressed. First, how is the function relating continued retrieval practice and long-term memory modulated by the intervening spacing interval (i.e., lag)? Second, how does this function differ between young and healthy older adults given age-related changes in working memory capacity and forgetting rate across short delays? Third, to what extent does the individual's motivation to learn specific information influence the benefit of spaced retrieval practice?

To address the first two questions, Experiment 1 examined the benefit of continued retrieval practice during the acquisition phase on a final cued recall test as a function of lag, retention interval, and age. Participants studied word pairs (e.g., QUEEN – lady) during an

initial acquisition phase and were tested on those pairs one, three or five times (e.g., QUEEN - ?????) which were separated by short or long lags (i.e., 1 vs. 3 intervening items). Following either a short or long retention interval, participants completed a final cued recall test. The results revealed that continued testing in the short lag condition led to consistent increases in retention, whereas continued testing in the long lag condition led to increasingly smaller benefits in retention for both age groups. Analysis of final test response latency revealed a different pattern than that observed in accuracy such that young adults benefited from continued testing in the long lag condition but not the short lag condition, and older adults benefited from continued testing in the short lag condition but not the long lag condition.

Experiment 2 extended the first experiment by examining the benefit of massed (testing without intervening material) versus spaced (Lag 4) retrieval practice across age groups. Young adults benefited from continued testing in both the massed and spaced conditions, whereas older adults showed a selective benefit of continued testing in the spaced condition. Again, analysis of response latency on the final test revealed a different pattern of results such that young adults benefited uniquely from continued testing in the spaced condition but not the massed condition, and older adults benefited from continued testing in both conditions.

In pursuit of the third aim regarding the role of participant motivation on the spacing effect, Experiment 3 examined the benefit of retrieval practice for paired associates assigned either a low point value or a high point value. Participants were asked to earn as many points as possible by successfully retrieving items on the final test. Results revealed the predicted benefit of lag and point value on final test accuracy for both young and older adults with no interaction between these two factors. These results suggest that the manipulation of point value effectively

modulated participant motivation to learn and retain the paired associates similarly across massed and spaced retrieval conditions.

Emphasis on retention (i.e., conditional final test performance) revealed a pattern of results that diverged from past studies such that young and older adults benefited similarly from spaced retrieval when differences in acquisition performance were minimized (Experiment 1). Moreover, age-related differences in refreshing (Experiment 2) and attention (Experiment 3) were implicated as contributing factors to final test performance above and beyond age differences in acquisition accuracy. Discussion focuses on the role of desirable difficulty in producing the benefits of lag, spacing and testing, along with methodological insights into different measures of memory integrity (response latency and accuracy).

Introduction

Healthy aging is marked by broad declines in episodic memory and other cognitive abilities (Arking, 1998; Balota, Dolan & Duchek, 2000; Salthouse, 1996). However, recent evidence suggests there may be many individual differences in the aging process (e.g., Hertzog, Kramer, Wilson & Lindenberger, 2009). Given the substantial increase in our aging population, there is a clear need to identify ways of improving memory that are effective across diverse populations and variable aging trajectories. One technique that is effective across a wide array of contexts and populations, *spaced retrieval*, combines the mnemonic benefits of spacing and testing (see Balota, Duchek, & Logan, 2007, and Cepeda et al., 2006 for reviews). When using this technique, individuals initially learn material and are repeatedly tested on that material with intervening time and/or items between learning and testing events. This technique typically produces better long term retention than conditions in which material is learned and repeatedly tested without any intervening time or items. In fact, spaced retrieval has effectively improved memory in healthy aging individuals (e.g., Balota, Duchek, & Paullin, 1989), Alzheimer's disease individuals (e.g., Balota, Duchek, Sergent-Marshall, & Roediger, 2006; Camp, Foss, Stevens, & O'Hanlon, 1996) and amnesiacs (Schacter, Rich & Stamp, 1985).

Given the effectiveness of spaced retrieval, it is important to better understand why this technique improves memory and how it can be used most effectively. Thus, the current study addresses three specific questions related to the benefits of spaced retrieval practice. First, how can the efficiency of retrieval practice be maximized by varying the spacing interval (i.e., lag) that occurs between retrieval attempts? In other words, how does the relationship between long-term memory performance and additional retrieval practice (e.g., 1 vs. 3 vs. 5 tests) differ as a function of lag? Second, how does the efficiency of retrieval practice differ across healthy young

and older adults given age-related differences in working memory capacity (e.g., McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010; Park et al., 1996) and forgetting rate across short delays (e.g., Giambra & Arenberg, 1993). This aim is motivated by theories that hypothesize a critical role for working memory and forgetting as important mechanisms underlying the benefits of spaced retrieval. Third, to what extent does participant motivation modulate the benefit of retrieval practice for specific items that have relatively high value to an individual? This question is particularly important with respect to memory in natural settings given that individuals must regularly make decisions about which information is particularly important to retain.

Before introducing the current experiments, I will first provide a brief overview of theories of spaced retrieval practice. Next, the implications of age-related changes in episodic and working memory for the efficiency of spaced retrieval will be considered. Finally, recent studies that have examined value-directed encoding in young and older adults will be reviewed to provide a context for how motivation may modulate the efficiency of spaced retrieval practice.

Spaced Retrieval Practice

Spaced retrieval combines the mnemonic benefits of the spacing and testing effects. First, the spacing effect refers to situations in which separating study events with intervening time or material improves long-term retention relative to situations in which study events occur consecutively (Ebbinghaus, 1885) or in relatively close proximity (Melton, 1967; 1970). The benefit of spacing is typically characterized by a nonmonotonic function such that performance initially increases as the lag between repetitions of an item increases and then slowly decreases after some optimal interval size. This finding has been termed the *lag effect* (see Cepeda et al., 2006 for a review). However, it is important to note that the benefits of lag and spacing depend

on the specific retention interval employed. This is most clearly seen in the Lag by Retention Interval interaction, which reflects improved performance for massed items relative to spaced items on an immediate memory test but a significant benefit for spaced over massed items on a delayed memory test (e.g., Balota, Duchek, & Paullin, 1989; Glenberg, 1976; Peterson, Wampler, Kirkpatrick, & Saltzman, 1963). Additionally, the Lag by Retention Interval interaction has been observed in response latency using a naming task (e.g., Spieler & Balota, 1996). Thus, there appears to be greater forgetting across time despite higher initial performance for massed items compared to spaced items, and this occurs in tasks where effortful, recollection-based retrieval processes are necessary and in tasks in which effortful recollection-based retrieval processes appear to be minimized.

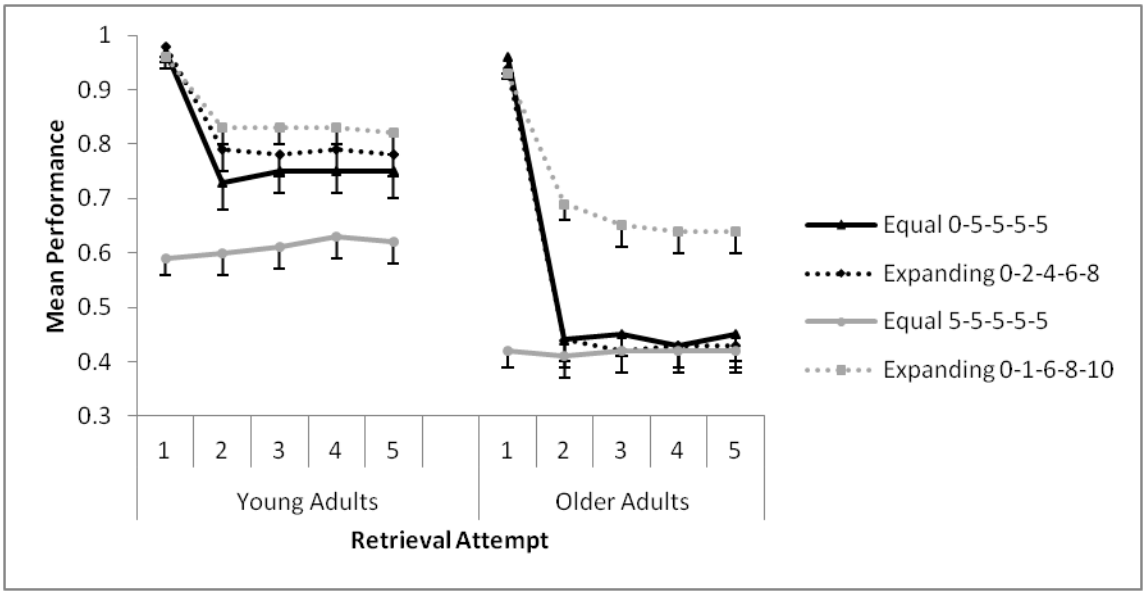
Turning to the testing effect, long-term retention is often better when material is initially learned and then tested compared with situations in which material is initially learned and then restudied (e.g., Hogan & Kintsch, 1971; Roediger & Karpicke, 2006a, 2006b). According to one theoretical account of the testing effect, retrieval practice during the learning phase leads to greater elaboration of the memory trace and increases the number of routes that can later be used to retrieve the material from memory (e.g., Bjork, 1975; McDaniel & Masson, 1985). Indeed, evidence suggests that increasing the overlap in processes used during encoding and retrieval should increase final test performance (transfer appropriate processing; Morris, Bransford & Franks, 1977).

Given the benefits of spacing and testing, researchers have become interested in optimally combining these techniques to maximize long-term memory performance. Past studies examining the benefits of spacing and testing have most often compared forms of spaced retrieval to identify the most effective spacing technique. The three conditions that are typically

compared include *massed retrieval* in which study and test trials occur consecutively without any intervening items, *equal spaced retrieval* in which the study and test trials are separated by an equal number of intervening items, and *expanding retrieval* in which the lag separating study and test trials gradually increases with each successive retrieval attempt. Early studies revealed a benefit of both spaced retrieval conditions over massed retrieval, as well as an additional benefit of expanding over equal spaced retrieval (e.g., Landauer & Bjork, 1978). These results suggested that it was important to maintain high levels of retrieval success across retrieval attempts during learning while shaping the memory trace to be retrieved after longer and longer intervals. However, subsequent studies have produced mixed patterns of data regarding various forms of spaced retrieval. In some studies, expanding and equal spaced retrieval produced equivalent benefit over massed retrieval (e.g., Cull, 2000), whereas in other studies, equal spaced retrieval actually produced a greater benefit than expanding retrieval (e.g., Carpenter & DeLosh, 2005; Karpicke & Roediger, 2007). Critically, these mixed findings have been observed with both young and healthy older adult samples (e.g., Balota et al., 2006; Logan & Balota, 2008).

To better understand the factors that contributed to the mixed benefits of expanding and equal spaced retrieval, Maddox, Balota, Coane and Duchek (2011) examined the influence of early spacing intervals and subsequent changes in lag on final cued recall performance. Across two experiments, Maddox et al. compared two expanding (0-2-4-6-8 and 0-1-6-8-10) and two equal spaced retrieval schedules (0-5-5-5-5 and 5-5-5-5-5), which all produced significant benefits in memory performance over massed retrieval. As seen in Figure 1, young adults benefited during acquisition retrieval practice from the three spacing schedules that included an immediate retrieval attempt compared to the equal spaced retrieval condition in which the first retrieval attempt occurred after five items. In contrast, older adults benefited from the spacing

Figure 1. Acquisition phase performance from Maddox et al. (2011) for young and older adults as a function of spacing condition and retrieval attempt. Error bars are 1 standard error below mean performance.



schedule in which retrieval occurred immediately and then again after a single item interval relative to the other three spacing schedules. These benefits observed in acquisition accuracy were observed 45 minutes later on the final cued recall test. Maddox et al. suggested that age-related differences in working memory capacity (e.g., McCabe et al., 2010; Park et al., 1996; Salthouse, 1991) and forgetting rate (Giambra & Arenberg, 1993) may modulate the extent to which items are maintained across early spacing intervals and successfully retrieved on subsequent retrieval attempts. In turn, the extent to which material is retrieved following various lags during acquisition should depend on the individual's working memory capacity (as indicated by working memory tasks such as Computation Span; Conway et al., 2005). Thus, although memory performance may be similar across individuals with varying working memory capacities following short lags, individuals with low working memory capacity (WMC) may experience relatively greater retrieval difficulty than individuals with high WMC following long lags. When too long of a lag is included between learning and retrieval events, low WMC individuals may experience increased retrieval failure relative to high WMC individuals. Given the reduced WMC for older adults compared to young adults, a shorter lag may be particularly beneficial for the former group to maximize retrieval success.

Another observation noted by Maddox et al. (2011) was that performance remained relatively stable across subsequent spacing events for both age groups after an initial spacing interval had been introduced (see Figure 1). This is not surprising since feedback was not provided, and so one would not expect performance to increase without feedback across retrieval attempts. Of course, it could be the case that performance would decline across retrieval attempts if too much spacing or intervening material was included, and indeed, this was the case between the first and second retrieval attempts in most conditions (see Figure 1). However, one may

question the extent to which continuing to test items following this initial forgetting provides additional benefits to long-term memory performance given the relatively stable performance across later retrieval attempts. I shall now turn to a brief review of the literature on the benefits of continued testing without feedback.

Evidence for the benefits of continued testing without feedback

Although the typical approach to studying the benefits of spaced retrieval is to hold constant the number of retrieval attempts and manipulate the form of spacing (e.g., Maddox et al., 2011), an alternative approach is to vary the number of retrieval attempts without altering the lag between each attempt. Typically, studies that have examined the benefit of continued testing after a single initial test have included feedback following each test trial. Results from one recent study indicated that long-term memory was maximized when material was tested three times separated by relatively long lags following the initial learning event (Rawson & Dunlosky, 2011). When more than three retrieval events have been included during the learning phase, results suggest that the benefit of testing with feedback becomes increasingly smaller with each additional retrieval attempt (e.g., Pyc & Rawson, 2009).

One concern with these studies, however, is that a dropout procedure was used in which an item was dropped from testing once it had been tested a predetermined number of times. As a result, repetitions of items that were to receive continued testing were separated by fewer and fewer items. As the lag separating repetitions decreases, retrieval should become increasingly easy across later retrieval attempts. This change in retrieval difficulty during acquisition is relevant to Bjork's (1994) proposition that the benefit of retrieval practice will be greater when retrieval is difficult compared to instances in which retrieval is easy (presuming both situations lead to successful retrieval), a concept termed *desirable difficulty.* In this light, it is possible that

the use of the contracting spaced interval may have contributed in part to the increasingly smaller benefit in long-term memory performance observed as a function of continued testing in the Pyc and Rawson (2009) and Rawson and Dunlosky (2011) studies.

Turning to studies that did not provide feedback following each retrieval attempt, the extant literature is limited in three ways with respect to the questions addressed in the dissertation. First, most studies have included more encoding time in the single test condition than in the multiple test condition (i.e., 3 study trials and 1 test trial vs. 1 study trial and 3 test trials; Hogan & Kintsch, 1971; Roediger & Karpicke, 2006b; Tulving, 1967). Although the number of exposures to material was equated across conditions, the inclusion of varying amounts of study time may mask the benefit of continued testing. Specifically, increasing initial study time should result in improved performance on test trials. As a result, a larger proportion of studied items should benefit from testing in the single test condition than in the multiple test condition (e.g., the bifurcation model; Halamish & Bjork, 2011, and Kornell, Bjork, & Garcia, 2011). In turn, the observed benefit of continued testing relative to taking a single test may be reduced due to differences between conditions in terms of initial retrieval success.

Second, a limited number of retrieval attempts were typically examined (i.e., 1 vs. 3 retrieval attempts). Thus, the function relating the number of retrieval attempts to the final test performance has typically been limited in range. Third, repeated testing occurred after variable spacing intervals that did not allow an assessment of how lag modulates the benefit of single test versus multiple test conditions. Despite these limitations, two studies provide some insight into the benefits of continued testing.

Wheeler and Roediger (1992) examined memory for a series of 60 pictures when participants were given one forced recall test or three forced recall tests prior to a one-week

retention interval. Each test required participants to recall as many items as possible and then guess until a total of 60 items had been generated before completing the subsequent test. In this sense, even if a specific item was recalled on all three tests, there was intervening time between each of the item's recall events. Results on the final memory test revealed significantly better performance for participants who took three tests compared with those who took one test. However, it is unclear whether continuing to test beyond three retrieval trials would produce an additional increase in memory performance, and it is similarly unclear the extent to which the lag between retrieval attempts modulated the benefit of continued testing.

In addition to the Wheeler and Roediger (1992) study, Karpicke and Roediger (2007) compared two single-retrieval attempt conditions in which the study and test trials were separated by one or five items (Lag 1 and Lag 5, respectively) with two multiple retrieval attempt conditions with the same initial lag (i.e., 3 attempts spaced across learning). Critically for the current study, Karpicke and Roediger reported a significant benefit on a final memory test after a 10 minute delay for three tests over one test, and this benefit was larger on a two day delayed test. Although no analyses were provided with respect to the influence of lag on the benefits of single versus multiple testing, mean performance in each condition was reported, and the data suggest that the benefit of continued testing depends on the lag between study and testing events. As seen in Table 1, performance on the 10 minute delayed test was best after taking three tests spaced with increasingly longer intervals (i.e., expanding retrieval, 1-5-9), equivalent between the single test, Lag 1 condition and taking three tests spaced with equal intervals (i.e., equal spaced retrieval, 5-5-5), and worst in the single test, Lag 5 condition. In contrast, results from the two day delayed test revealed performance that was best for items in the equal spaced retrieval condition, equivalent for the expanding retrieval and Lag 5 conditions,

Table 1. Final recall performance (*M*, *S.E*) from Karpicke and Roediger (2007; Experiment 1).

	10 Minute Delay	2 Day Delay
Expanding (1-5-9)	.71 (.05)	.33 (.05)
Equal (5-5-5)	.62 (.07)	.45 (.05)
Lag 1	.65 (.05)	.22 (.04)
Lag 5	.57 (.06)	.30 (.04)

and worst for the Lag 1 condition. Thus, the benefit of continued testing likely depends on both lag and retention interval, which is consistent with the Lag by Retention Interval interaction discussed earlier.

In sum, results from these two studies suggest that three spaced tests during the acquisition phase produce a long-term memory benefit over a single test for young adults. However, there are three outstanding issues that are important to consider and will be the focus of the dissertation. First, one limitation of the past studies is the relatively limited number of retrieval attempts included during the acquisition phase. Thus, Experiment 1 included conditions with one, three, and five retrieval attempts to examine the function relating long term memory and the influence of repeated retrieval. It is possible that continued testing may lead to decreasing benefits in terms of final retention similar to the results from studies in which feedback was provided (e.g., Pyc & Rawson, 2009; Rohrer et al., 2005). Alternatively, continued testing may produce additional increases in long-term retention when the spacing interval remains constant between retrieval attempts rather than contracting like those in previous studies that provided feedback. In this sense, the retrieval difficulty of later retrieval attempts may be relatively more beneficial when the spacing interval remains constant versus contracting.

Second, it is unclear how the lag separating each retrieval attempt modulates the relationship between continued testing and long-term retention. With respect to Bjork's (1994) concept of desirable difficulty, test trials following longer lags should produce better memory than test trials following shorter lags (so long as retrieval is successful in both cases) due to increased forgetting and retrieval difficulty in the long lag condition. With respect to the current study, continued testing should not produce additional benefit in long-term retention beyond a single retrieval attempt in a short lag condition based on the desirable difficulty account. In

contrast, continued retrieval should improve long-term retention when retrieval attempts are separated by a relatively long lag. Of course, separating study and testing events with a long lag may make early retrieval attempts more effective in improving memory than later retrieval attempts such that the relative increase in performance with each additional test decreases.

Third, the influence of the lag between retrieval attempts and the benefit of continued testing appears to be modulated by the retention interval. Again, past studies have provided evidence of a Lag by Retention Interval interaction in which the lag effect typically increases as the retention interval increases. With this in mind, the benefits of lag and continued testing should increase with longer retention intervals. Thus, it is critical to examine the relative contributions of lag and testing to performance after short and long delays.

To address these outstanding issues, the dissertation utilizes an approach that deviates from past studies. As noted earlier, past research regarding the benefits of spaced retrieval has produced mixed results regarding the optimal spacing schedule. This problem has typically been addressed in subsequent studies by manipulating the factors thought to influence the benefits of various schedules (e.g., Maddox et al., 2011). However, in many cases, the differences in acquisition performance produced by various spacing techniques may have obscured differences in retention that resulted from those techniques (e.g., Balota et al., 2006; Karpicke & Roediger, 2007). Consider the seminal spaced retrieval study reported by Landauer and Bjork (1978) in which the benefit of expanding over equal spaced retrieval decreased across the acquisition phase retrieval trials as the expanding retrieval interval grew increasingly larger (19%, 13%, and 7%, respectively). Importantly, the difference between conditions was further reduced on the final test (5%) which followed a 25 minute retention interval. Despite the evidence reported by Landauer and Bjork suggesting that expanding retrieval was the optimum form of spaced

retrieval practice, the decline in performance across the retention interval occurred at different rates for equal spaced and expanding retrieval practice conditions (see also Balota et al., 2006; Karpicke & Roediger, 2007). Thus, the influence of spaced retrieval on long-term retention for items that could be retrieved during the acquisition phase was obfuscated by different levels of acquisition performance and the chosen retention interval (as seen in Karpicke & Roediger's (2007) data, which are presented in Table 1). Indeed, Karpicke and Bauernschmidt (2011) recently made a similar observation about the differences in acquisition versus differences in retention produced by various spacing manipulations. Thus, one may examine the extent to which retrieved items are retained (or forgotten) across the retention interval. In the dissertation, the focus is on items that were initially recalled correctly during the acquisition phase (conditional performance) in order to directly examine the influence of age, lag and testing on final recall performance, while minimizing the effects of these variables on acquisition accuracy per se.

With respect to the dissertation, consider a comparison between young and older adults in the long lag condition of a spaced retrieval task. Although older adults may have overall lower performance than young adults during acquisition due to poorer episodic memory abilities, the influence of the long lag on retention when items are successfully retrieved during learning may be similar across age groups. Alternatively, analysis of conditional final test performance in the current experiments may still reveal that the benefits of continued testing change as a function of age as well as retention interval given the results reported by Maddox et al. (2011) indicating different optimal spacing schedules across young and older adults. These factors will be considered next.

Aging, Spacing, and Testing

As discussed earlier, Maddox et al. (2011) compared the benefits of various spaced retrieval schedules on final test performance and found that the optimal spacing schedules differed between young and older adults. Specifically, young adults benefited similarly from retrieval schedules that included an immediate retrieval attempt regardless of the subsequent lag, whereas older adults only benefited from the retrieval schedule that paired an immediate retrieval attempt with the shortest possible spacing interval (see Figure 1 above). Maddox et al. suggested that older adults needed this combination of retrieval attempts due to steeper forgetting across short delays compared to young adults (Giambra & Arenberg, 1993) or age-related differences in working memory (e.g., McCabe et al., 2010). It was also the case that young adults appeared to benefit from the single immediate retrieval attempt, whereas older adults did not benefit, e.g., compare the 0-1-6-8-10, 0-2-4-6-8, and 0-5-5-5-5 conditions in Figure 1. Maddox et al. noted that this finding was consistent with research indicating age differences in the benefit of refreshing, which refers to the sustained activation of an item from an immediate retrieval attempt (Johnson, Reeder, Raye and Mitchell, 2002).

As indicated by Maddox et al. (2011), past research suggests that age-related changes in WMC or forgetting may modulate the spaced retrieval benefits in older adults' long term memory performance (see also Coane, 2013; Meyer & Logan, 2013; Tse, Balota & Roediger, 2010). Specifically, older adults may show greater benefit in continued testing when retrieval attempts are separated by a short lag than a long lag at least in nonconditionalized recall performance. This stands in contrast to the earlier prediction in which young adults were predicted to benefit from a longer lag relative to a shorter lag. If the long lag (but not short lag) condition is too long to produce retrieval success, then older adults will not benefit from the additional retrieval attempts during acquisition. Hence, an interesting Age by Lag cross-over

interaction is predicted based on differential forgetting across age groups and lag sizes. More importantly, in terms of conditionalized final test performance, one may still expect an Age by Lag interaction in the function relating lag to retention across age groups. Older adults may experience a relatively greater difference in retrieval difficulty between lag conditions than young adults. Thus, a larger lag effect would be observed for older adults than young adults.

Beyond these direct manipulations of spacing and testing in studying age-related changes in performance, it is also important to consider the extent to which young and older adults have control over which items should be retained across intervening lags. Hence, I will now turn to an emerging literature on value-directed memory encoding.

Value-directed Encoding in Young and Older Adults

As noted in the preceding sections, there are numerous exogenous factors (experimenter controlled) that are predicted to influence the benefit of continued testing (e.g., lag, retention interval, number of tests). It is also the case that endogenous factors (participant controlled) may influence the way in which retrieval practice benefits long-term retention. One such factor is the learner's motivation to fully encode, retain and retrieve specific material. Given the wide range of stimuli that individuals are exposed to in everyday life, it is critical that attention is selectively directed toward important information that will be useful at a later time. One way of manipulating the importance of material in an experimental setting is to assign point values to material during encoding that will later be awarded when that material is successfully retrieved (e.g., Castel, 2008; Castel, Balota, & McCabe, 2009; McGillivray & Castel, 2011). Typically, results reveal a smaller age difference in memory performance for high point value material than for low point value material.

In one recent study (Castel et al., 2011), participants were presented with multiple lists of 12 items (with each item assigned a point value from 1 to 12) followed by recall after each list. As expected, overall memory performance was significantly higher for young adults than for the young-older and old-older adult groups. With respect to selecting high value items for encoding, there was no difference between young adults and young-older adults, but old-older adults were significantly less likely to successfully encode and retrieve high point value items than the other two groups. Castel et al. argued that overall memory performance and the ability to select high value items capture two separate mechanisms that represent memory capacity and metacognitive abilities, respectively. As Castel et al. emphasized, selecting and maintaining the goal of encoding high-value items places a demand on attention and requires additional inhibition of irrelevant items (i.e., low point value). The decline in sensitivity to point value in the oldest adults is consistent with the age-related changes in either or both of these processes (e.g., Balota & Faust, 2001; Hasher & Zacks, 1988).

Given findings in the extant literature related to value-directed encoding and the allocation of attention during learning, high-value items may benefit more from retrieval practice than low-value items if enhanced encoding increases the probability of successful retrieval following varying lags. It is also the case that an interesting Age by Lag by Point Value interaction may be observed based on Bjork's (1994) desirable difficulty account. Specifically, high point value items should be better encoded and more easily retrieved during acquisition by young adults than low point value items following a long lag. Thus, successfully retrieved low-value items are likely to be remembered better than high-value items on the final test. Similar to earlier age-related predictions regarding the benefit of lag and continued testing, older adults are expected to show the same pattern as young adults in terms of retention (i.e., conditional final

test performance for items successfully retrieved during acquisition). However, the disparity between point values may be greater for older adults than young adults given the relatively reduced attentional control of the former group (e.g., Balota and Faust, 2001). In turn, this disparity may lead to more difficult retrieval following a long lag compared to young adults. Thus, an Age by Lag by Point Value interaction may be found in final test performance.

Current Study

It is clear from the review of the literature that spacing and testing both improve memory performance for young and healthy older adults. However, it is less clear the extent to which continuing to test material after initial retrieval success provides additional benefit in long-term memory performance and whether the benefit of additional testing is modulated by the lag between retrieval attempts. Moreover, it is unclear if young and older adults will produce the same function relating additional tests and long-term memory performance, and indeed there are theoretically motivated reasons to predict age-related differences in these functions. With these issues in mind, the current experiments were aimed at three goals.

The first goal was to examine the efficiency of continued spaced retrieval practice in terms of final test performance. It is currently unclear how much additional retrieval practice benefits long term recall (i.e., is taking 5 tests better than taking 3 tests?) and the extent to which the efficiency of continued retrieval practice is modulated by the lag that occurs between retrieval attempts (c.f. Rawson & Dunlosky, 2011). Again, one would expect continued testing to produce a benefit in final test performance in the long lag condition but not the short lag condition based on the concept of desirable difficulty (Bjork, 1994). Of course, it will be critical to also obtain some estimate of retrieval difficulty during acquisition performance. The present

study will use response latency for correctly retrieved items during acquisition as a metric for retrieval difficulty.

The second goal was to assess how the efficiency of continued testing and the influence of lag may differ between young and older adults. If the benefit of reminding is modulated by the difficulty of retrieval on an item's second presentation, as Bjork (1994) suggested, then older adults may be more sensitive to the lag and number of retrieval attempts included within the current study provided age-related changes in working memory capacity (e.g., McCabe et al., 2010; Park et al., 1996) and forgetting rate at short delays (e.g., Giambra & Arenberg, 1993). Importantly, the present study directly measured working memory in both young and older adults to determine if age or working memory changes predict the benefits of testing and spacing. Specifically, individuals with lower working memory capacity may find longer lags relatively more difficult than short lags, and as a result these individuals are predicted to benefit more from this manipulation than individuals with high working memory capacity in terms of long-term retention (i.e., conditional final test performance).

The third goal was to examine how participant motivation modulates the benefits of retrieval practice. Specifically, high point values may induce more elaborative encoding or rehearsal which then increases retrieval success over low point value items (especially following longer lags). In turn, retrieval of low value items should be more difficult than retrieval of high value items. Thus, it is predicted that retention (i.e., conditional performance for items that were successfully retrieved acquisition) will be better for low-value items than high-value items and that this difference will be greater for older adults than young adults. Such a pattern may be modulated by age-related differences in working memory capacity (e.g., McCabe et al., 2010)

and forgetting (Balota, Dolan, & Duchek, 2000) that would result in more desirable retrieval difficulty for older adults than young adults.

Methodological Considerations

The current study deviated from the majority of past spaced retrieval research in two critical ways. First, as noted earlier, analysis of nonconditional final test performance leaves open the possibility that differences produced by various spaced retrieval techniques during acquisition are confounded with the influence of the techniques on retention of material. In other words, acquisition performance may differ across conditions (e.g., age group or lag) in a way that obscures the influence of retrieval practice on retention and final test performance, because the manipulation of interest (i.e., retrieval practice) occurs unequally during encoding. Thus, the current study emphasizes conditional performance on the final test for items that were successfully retrieved on the final test trial during the acquisition phase when assessing the benefit of continued testing.

Although emphasizing conditional analyses in the dissertation provides leverage in better understanding the influence of the critical manipulations on final test performance while minimizing the influence of differences in performance during the acquisition phase, one may be concerned that conditional analyses will introduce item selection effects that may undesirably influence the final test performance results. Specifically, as discussed with respect to the critical predictions, older adults should have more difficulty retrieving items than young adults given age-related differences in episodic memory (see Balota, Dolan & Duchek, 2000) and retrieval difficulty should be greater following a longer lag than a shorter lag. In these instances of greater retrieval difficulty, successful retrieval is more likely to occur for items that are inherently easy to remember than for difficult items. In contrast, in instances where retrieval is

easier (e.g., short lag; young adults), successfully retrieved items should be more variable in terms of retrieval difficulty. In other words, the easy-to-remember items will be retrieved along with some of the more difficult-to-remember items. In this sense, conditional analysis may yield relatively greater amounts of retrieval difficulty in the short lag condition than in the long lag condition, which would decrease the likelihood of obtaining a significant lag effect.

Second, the vast amount of work on memory emphasizes accuracy, which is indeed appropriate. However, one may also be interested in the speed to retrieve information. For example, consider the case of an emergency physician who is not only responsible for accessing medical knowledge with high accuracy but is also responsible for accessing that information quickly. Although one might expect the influence of lag and testing to be the same on measures of accuracy and retrieval speed, it is possible that one may find dissociation between these measures. For example, retrieval speed may be more sensitive to repeated testing after accuracy has approached asymptote and no longer benefits from this manipulation. As shown below, it is clear that retrieval speed does not show the same pattern of effects as accuracy. Moreover, by measuring response latency during the acquisition phase, one has an objective measure of retrieval difficulty, which as discussed above, is relevant to the desirable difficulty account of spacing effects.

Experiment 1

The first experiment addressed age-related differences in the influence of continued testing when separated by short versus long lags. As discussed earlier, the benefit of continued testing after an initial spaced retrieval attempt may be greater following a long lag than a short lag especially at a longer retention interval (e.g., Karpicke & Roediger, 2007; Wheeler & Roediger, 1992). Moreover, this relationship may be modulated by age.

Method

Participants. Young adults were undergraduates at Washington University in St. Louis and received partial course credit or monetary remuneration ($15 or $20 for short and long retention intervals, respectively) for their participation. Older adults were healthy, community dwelling adults and received monetary remuneration for their participation ($20). Both age and years of education were significantly different between age groups (*p*s < .005; see Table 2). One-half of the participants in each age group was assigned to the short retention interval, and the other half was assigned to the long retention interval condition. Within each age group, there were no differences in age or years of education between the short and long retention interval conditions (*p*s > .15; see Table 2). An additional group of young adults (n = 1 and n = 3 for short and long retention intervals, respectively) and older adults (n = 5 and n = 4 for short and long retention intervals, respectively) were excluded from analysis due to low performance on the final test (i.e., nonconditional mean accuracy less than 5%), and an additional two young adults in the long retention interval condition were excluded for not completing the second experimental session. Table 2 also includes mean performance on each working memory measure (i.e., Letter Number Sequencing task, Computation Span Task), a standardized working memory composite score, and the Shipley vocabulary test. Although each of these measures will be described in the subsequent Materials and Design Section, it is important to note that there was a main effect of age group in each of these measures (*p*s < .001).

Materials and Design. A 2 (Age) x 2 (Retention Interval) x 2 (Lag, 1 vs. 3) x 3 (Number of Tests, 1 vs. 3 vs. 5) mixed-factor design was used with Age and Retention Interval as between-participant factors and Lag and Number of Tests as within-participant factors. The

Table 2. Mean (*S.D.*) age (in years), education (in years) and working memory performance as a function of age and retention interval for Experiment 1 (top panel), Experiment 2 (middle panel), and Experiment 3 (bottom panel).

		Young		Older	
		Short RI	Long RI	Short RI	Long RI
Exper. 1	n	49	49	42	42
	Age	20.33 (2.46)	20.76 (2.89)	73.81 (5.20)	75.66 (7.49)
	Education	14.39 (1.92)	14.48 (1.53)	15.79 (2.58)	15.21 (2.93)
	LNS	56.34 (19.74)	50.57 (18.40)	27.36 (15.32)	28.31 (14.80)
	CompSpan	8.02 (3.46)	7.47 (3.16)	4.95 (2.52)	4.29 (2.41)
	WMC Composite	.60 (.78)	.38 (.75)	-.53 (.62)	-.61 (.61)
	Shipley	33.71 (2.64)	32.96 (3.05)	35.24 (3.55)	35.95 (3.43)
Exper. 2	n	24	--	24	--
	Age	19.00 (1.02)	--	70.08 (6.19)	--
	Education	13.75 (1.70)	--	16.54 (2.21)	--
	LNS	37.88 (14.81)	--	29.50 (14.96)	--
	CompSpan	5.67 (2.12)	--	4.38 (2.60)	--
	WMC Composite	.27 (.70)	--	-.27 (.71)	--
	Shipley	30.25 (3.49)		35.92 (2.55)	
Exper. 3	n	30	30	24	24
	Age	19.53 (1.25)	20.93 (6.93)	74.25 (5.42)	71.92 (4.63)
	Education	13.03 (2.39)	13.04 (2.25)	16.87 (3.42)	15.88 (2.44)
	LNS	68.30 (31.98)	61.93 (21.99)	36.96 (19.12)	33.39 (10.50)
	CompSpan	7.41 (3.37)	7.03 (2.34)	5.25 (2.83)	4.92 (2.06)
	WMC Composite	.59 (1.11)	.34 (.60)	-.43 (.66)	-.58 (.46)
	Shipley	33.44 (2.93)	32.73 (3.43)	35.74 (2.63)	35.38 (2.67)

LNS = Letter Number Sequencing Task, CompSpan = Computation Span Task, WMC Composite = standardized working memory composite score, Shipley = Shipley Vocabulary Score.

short retention interval was five minutes for both young and older adults. Because of large age-related differences in long-term retention and in an attempt to minimize differences due to scaling of final test performance, the long retention interval was one hour for older adults and one day for young adults. The lag between study and test trials was either a single trial (Lag 1) or three trials (Lag 3), and items were tested one time (1 test), three times (3 test), or five times (5 test) without feedback.

Memory Task. Fifty-six low associate word pairs (e.g., APPLE-evil) were selected from the University of South Florida Free Association Norms (Nelson, McEvoy, & Schreiber, 1998) and have been used in prior spaced retrieval studies (e.g., Maddox et al., 2011). Word pairs shared some features that made them more easily associable (e.g., WHISKEY-water) or could be used to form a sentence (e.g., HORSE-jumped). These critical word pairs were divided into seven sets of eight pairs which were counterbalanced across lists such that each pair occurred equally often in each of the within-participants conditions. Stimuli were statistically equated across sets for word length, frequency, orthographic neighborhood and phonological neighborhood (Balota et al., 2007), and pairs were equated on backward associative strength across stimulus sets (*p*s > .10).

A continuous paired associate task was used for the acquisition phase of the memory task (see Figure 2 for an example). The average serial list position was equated across all conditions (*p*s > .70) as was the average serial list position for the first test (*p*s > .70) and the last test (*p*s > .70) across the six testing conditions. In total, the acquisition phase was comprised of 218 trials consisting of 192 trials for the critical conditions, 18 filler trials, and eight trials that were equally split between primacy and recency buffer items. Of the 192 critical condition trials, 48 trials were encoding trials (e.g., HORSE-jumped) and 144 trials were retrieval practice trials (e.g.,

Figure 2. Partial schedule for three levels of retrieval attempts in Lag 1 (e.g., APPLE – evil) and Lag 3 (e.g., HORSE-jumped) conditions.

1 Retrieval	3 Retrievals	5 Retrievals
--	--	HORSE -- jumped
--	--	--
--	--	--
--	--	--
--	--	HORSE -- ?????
--	HORSE -- jumped	--
--	--	--
--	--	APPLE -- evil
--	--	HORSE -- ?????
--	HORSE -- ?????	APPLE -- ?????
HORSE -- jumped	APPLE -- evil	--
APPLE -- evil	--	APPLE -- ?????
--	APPLE -- ?????	HORSE -- ?????
APPLE -- ?????	HORSE -- ?????	APPLE -- ?????
HORSE -- ?????	APPLE -- ?????	--
--	--	APPLE -- ?????
--	APPLE -- ?????	HORSE -- ?????
--	HORSE -- ?????	APPLE -- ?????
--	--	--
--	--	--
--	--	HORSE -- ?????

HORSE-?????). Filler trials were included to ensure that average serial list position was equated across the critical conditions. The final cued recall test presented the cue for each critical pair (e.g., HORSE-?????). In both the acquisition and final test phases, participants were asked to speak their answers aloud. The experimenter indicated when the participant produced the response via keypress and then typed the participant's response on a second screen.

Computation Span Task. Adapted from the computation span task (CompSpan; Conway et al., 2005), participants were presented with a series of math problems one at a time. Participants read aloud the math problem and indicated whether the solution was correct or incorrect via keyboard button press. Additionally, participants were instructed to remember the middle digit in each equation. For example, if a participant was presented the problem "4 + 9 = 12," the participant would read the problem aloud, respond that the solution was incorrect and then remember the middle digit, "9." Following each series of math problems, participants recalled the center digit of each math problem in the order the digits were presented. The adapted CompSpan task (ELSEM, Storandt, Balota, & Salthouse, 2009) started with the lowest level (i.e., 3 trials in which 1 math problem was presented followed by 3 trials in which 2 math problems were presented, and so on). The task ended when the participant (a) failed to successfully recall the complete sequence of digits on two of the three trials for a given level of CompSpan or (b) after completing the highest CompSpan level (i.e., trials consisting of 7 math problems). Total score was the sum of all correct trials for each of the levels successfully completed.

Letter Number Sequencing Task. The Letter Number Sequencing task (LNS; Wechsler Adult Intelligence Scale-III; Wechsler, 1997) required that participants remember sequences of alternating letters and numbers that ranged between three and twelve total stimuli. At the end of

each sequence, participants were asked to recall the digits numerically followed by the letters alphabetically. Participants completed two trials of each length (for a total of 20 trials). Trials were presented in a fixed-random order with the requirement that one trial of each sequence length occurred in the first half and the second half of the task. Total score was the sum of the sequence lengths for each trial in which all stimuli were correctly recalled. These two measures of working memory capacity (CompSpan and LNS) were selected based on the results from a latent semantic analysis suggesting that LNS, in addition to standard laboratory working memory measures such as the CompSpan, was the best predictor of working memory (Shelton, Elliott, Hill, Calamia, & Gouvier, 2009).

Shipley Vocabulary Test. The Shipley Institute Living Scale vocabulary test (Zachary, 1986) consisted of 40 multiple-choice question trials. Each trial consisted of a target word and four multiple-choice words from which the participant was to choose the word that was most similar in meaning to the target. The sum of total correct answers served as the total score.

Procedure. Participants first reviewed the consent document and then proceeded to complete the acquisition phase of the memory task. On retrieval trials, participants spoke their answers aloud, and the experimenter typed the response immediately upon hearing their response[1]. Immediately following the acquisition phase, participants completed five minutes of a distractor trivia task in which trivia questions were presented at a rate of one question every 10 seconds. Participants were instructed: "Please speak the correct answer aloud. If you do not know the correct answer, make a reasonable guess and then wait for the next question." The procedure following the distractor task differed as a function of retention interval group. For participants in the short retention interval condition, the final cued recall test for the memory task occurred immediately following the trivia task. Again, participants were presented with the cue

word (e.g., HORSE-????) for each of the critical pairs one at time and made an oral response that the experimenter entered into the computer. Participants then completed CompSpan, LNS, the Shipley vocabulary test, and a demographics questionnaire before being debriefed and dismissed. Participants in the long retention interval condition proceeded with CompSpan, LNS, Shipley vocabulary and demographics questionnaire following the trivia task. After completing all other tasks, older adults completed the final cued recall test for the memory task before being debriefed and dismissed. Young adults were dismissed following completion of the other tasks and were asked to return 24 hours later to complete the final cued recall test and receive their debriefing.

Results

Acquisition Performance. Acquisition data were collapsed across retention interval groups, because the acquisition phase was the same for all participants. Indeed, analyses failed to yield any significant differences as a function of retention interval group in acquisition phase accuracy (*p*s > .25). Because there were empty cells for some participants, for purposes of the ANOVA, missing data were estimated by a triangulation procedure in which the relationship in performance between conditions at the group level was used in relation to individual performance at the participant level to provide an estimate for the missing data (see Appendix A for full procedure). It is noteworthy that the pattern of results was similar when analyses were conducted only on participants who had observations for all cells.

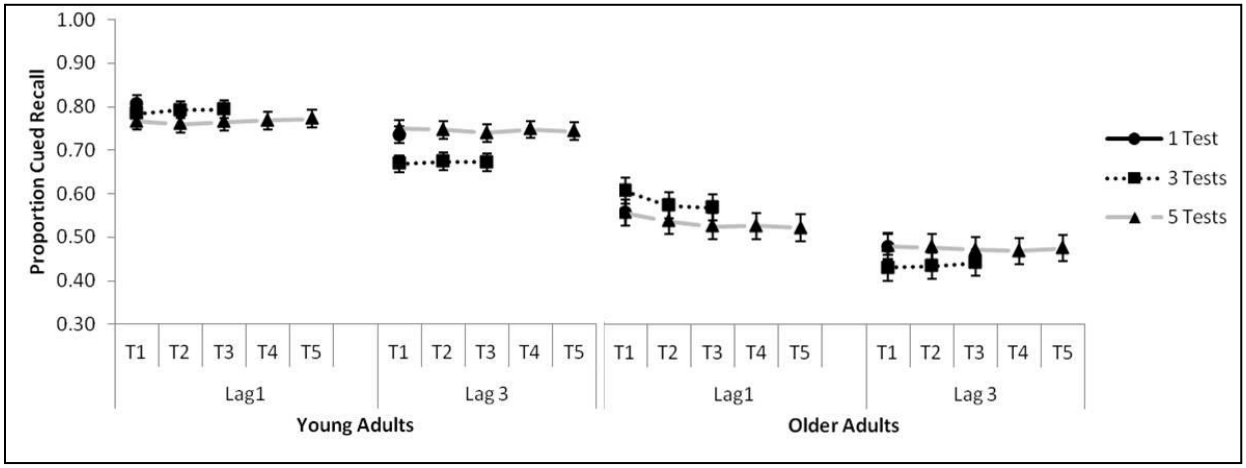
Although there are numerous ways to examine acquisition performance, the current set of analyses focuses on the first and last retrieval attempts in each of the multiple retrieval attempt conditions. Such an analysis directly examines the critical questions of the current study. Specifically, this approach allows for an assessment of stability in accuracy across retrieval

attempts in each lag condition and the extent to which response latency differentially speeds across attempts as a function of lag.

Memory Accuracy. Mean proportion correct recall for young and older adults is shown in Figure 3 as a function of lag, number of tests, and test number. There are three observations to note in this figure. First, young adult performance was higher than older adult performance. Second, performance in the Lag 1 condition was higher than in the Lag 3 condition, and the difference in performance between lag conditions was greater for older adults than young adults. Third, performance remained relatively stable across retrieval attempts in both lag conditions and age groups.

Performance on the first and last test in each of the multiple-retrieval attempt conditions was submitted to a 2 (Age) x 2 (Lag) x 2 (Number of Tests, 3 tests vs. 5 tests) x 2 (Test Number, First vs. Last) mixed-factor ANOVA. Results revealed main effects of Age and Lag, *p*s < .001, as well as significant Lag x Number of Tests, $F(1, 180) = 21.35$, $p < .001$, $\eta^2_p = .11$, and Lag x Test Number interactions, $F(1, 180) = 4.72$, $p = .031$, $\eta^2_p = .03$. The significant Lag x Number of Tests interaction reflected a reversal in the direction of mean performance between the three and five test conditions when separated by a single item lag (*M* = .69 vs. .65, respectively; *p* = .029) relative to a three item lag (*M* = .55 vs. .61, respectively; *p* < .001). The significant Lag x Test Number interaction reflected a small, but significant decrease in performance between the first and last tests in the Lag 1 condition (*M* = .68 vs. .66, respectively; *p* = .045) and no difference in performance between tests in the Lag 3 condition (*M*s = .58; *p* > .90). More importantly, results revealed a significant Age x Lag x Test Number interaction, $F(1, 180) = 11.72$, $p = .001$, $\eta^2_p = .06$. Follow-up *t* tests revealed a single significant difference between the first and final test in the Lag 1 condition for older adults (*p*s = .006). There was no difference in

Figure 3. Mean proportion cued recall during the acquisition phase in Experiment 1 as a function of age, lag, number of tests, and test number (e.g., T1 = first retrieval attempt, T2 = second retrieval attempt, and so on). Error bars represent ± 1 S.E.M.

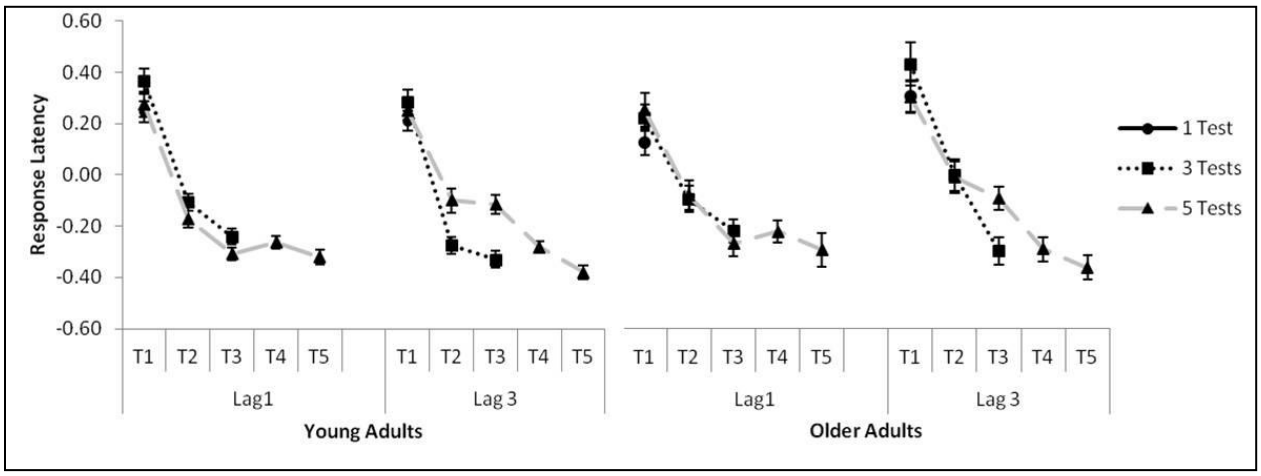


performance between the first and final tests in the Lag 3 condition for the older adults, or either the Lag 1 or Lag 3 conditions for the young adults (*p*s > .20). Thus, it appears that older adults produce some forgetting across repeated tests for items that are initially retrieved at a short lag. This may be due to the fact that success at the initial retrieval event is not as strong of an indicator of encoding quality following the short lag compared to the long lag. If one can maintain the item across the longer lag for the initial retrieval event, then the item is sufficiently well encoded to be produced across the remaining retrieval events. Young adults are not susceptible to this forgetting.

Standardized Response Latency. In the present and all subsequent analyses of response latency, all latencies beyond three *SD*s from the mean were excluded from analysis (< 1%). In addition, because older adults are overall slower than young adults, and this difference in speed can compromise the interpretation of interactions (e.g., Salthouse, 1991), the response latencies were standardized separately for each participant (see Faust, Balota, Spieler & Ferraro, 1999). Specifically, a mean and standard deviation were calculated for each participant's data, and each response latency was then transformed to a *z* score based on that participant's mean and standard deviation.

Standardized mean response latency on correct trials is shown in Figure 4 as a function of age, lag, number of tests condition, and test number. There are three observations to note in the figure. First, response latency decreased across retrieval attempts. Second, the decrease in response latency between the first and last tests was larger for the Lag 3 condition than the Lag 1 condition. Third, the difference between lag conditions in speeding across retrieval attempts was larger for older adults than young adults.

Figure 4. Mean standardized response latency during the acquisition phase in Experiment 1 as a function of age, lag, number of tests, and test number (e.g., T1 = first retrieval attempt, T2 = second retrieval attempt, and so on). Error bars represent ± 1 S.E.M.



Again, data from the first and last test taken in each condition were submitted to a 2 (Age) x 2 (Lag) x 2 (Number of Tests, 3 tests vs. 5 tests) x 2 (Test Number, First vs. Last) mixed-factor ANOVA. Results revealed main effects of Number of Tests, $F(1, 180) = 4.17$, $p = .043$, $\eta^2_p = .02$, which reflected a small difference in response latency between the three and five test conditions ($M = .03$ vs. -.03, respectively), and Test Number, $F(1, 180) = 473.19$, $p < .001$, $\eta^2_p = .72$, which reflected the speeding of response latency between the first and last tests ($M = .30$ vs. -.31, respectively). Additionally, the Lag x Test Number interaction was significant, $F(1, 180) = 6.94$, $p = .009$, $\eta^2_p = .04$, and further qualified by a significant Age x Lag x Test Number interaction, $F(1, 180) = 4.65$, $p = .032$, $\eta^2_p = .03$.

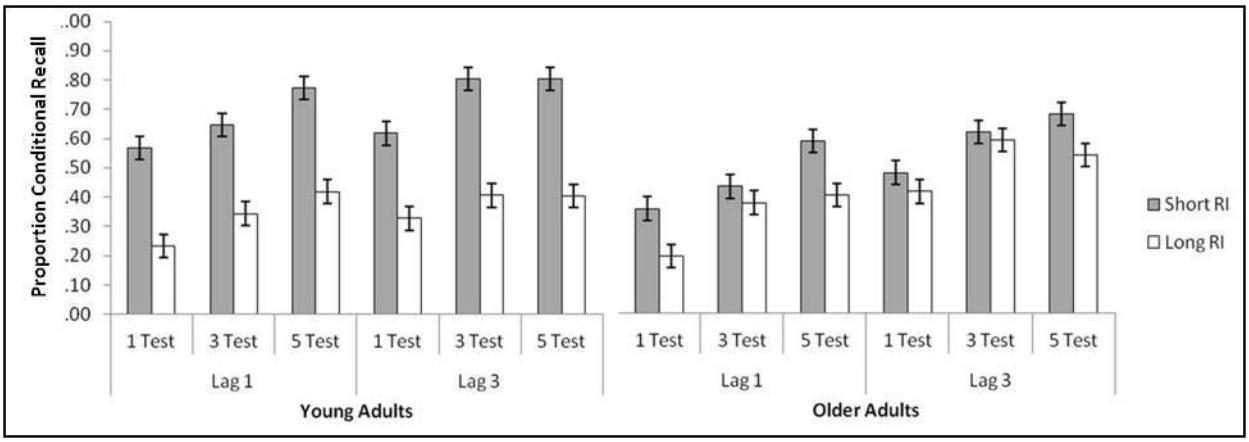
To examine the three-way interaction, separate 2 (Lag) x 2 (Test Number) ANOVAs were conducted for young and older adults. Analysis of young adult performance revealed significant effects of Lag and Test Number, ps $< .05$, but no interaction, whereas older adult performance revealed a significant effect of Test Number, $p < .001$, and a significant Lag x Test Number interaction, $F(1,83) = 8.32$, $p = .005$, $\eta^2_p = .09$. The significant Lag x Test Number interaction reflected significantly slower response latency on the first test for the Lag 3 condition than the Lag 1 condition ($M = .37$ vs. .24, respectively, $p = .034$) and a numerical reversal of this pattern on the final test ($M = -.33$ vs. -.25, respectively, $p = .154$). This finding is consistent with Bjork's (1994) desirable difficulty account in which items that are retrieved with relatively more difficulty will be strengthened to a greater extent than items retrieved with less difficulty. In the current results, the time it took to initially retrieve items provides evidence that Lag 3 produced a relatively more difficult retrieval event than Lag 1. As a result, the trace may be strengthened to a greater extent and consequently retrieved faster on the final retrieval attempt in the Lag 3 condition compared to the Lag 1 condition.

Final Test Phase Performance. As noted earlier, the current analyses emphasize changes in conditional recall performance. In the following analyses, final test performance is examined for items that the participant correctly retrieved on the last retrieval attempt during the acquisition phase[2]. These analyses are important for investigating the influence of age, lag and continued testing on the retention of material, since if an item is not correctly recalled during the acquisition phase it does not receive retrieval practice. Although analysis of conditional final test performance generally accords with nonconditional final test performance, there is one noteworthy exception considered in Appendix B. In addition to analysis of retention, the present results also include measures of response latency.

Conditional Memory Accuracy. Mean proportion conditional recall is shown in Figure 5 as a function of age, retention interval, lag, and number of tests. There are three observations to note in this figure. First, retention was greater for young adults than older adults after a short retention interval but was comparable across age groups following the long retention interval. This confirms that retention performance can be matched across young and older adults at the long retention interval by increasing the retention interval more for young adults than for older adults. Second, continued retrieval during the acquisition phase led to increased retention for young and older adults when tests were spaced by a single item regardless of retention interval. A similar increase in retention was observed across age groups and retention intervals in the longer, Lag 3 condition when pairs were tested three times versus one time, but no additional benefit was observed in the five test condition relative to the three test condition. Third, older adults produced a larger lag effect than young adults at both the short and long retention intervals.

These observations were supported by the results of a 2 (Age) x 2 (Retention Interval) x 2

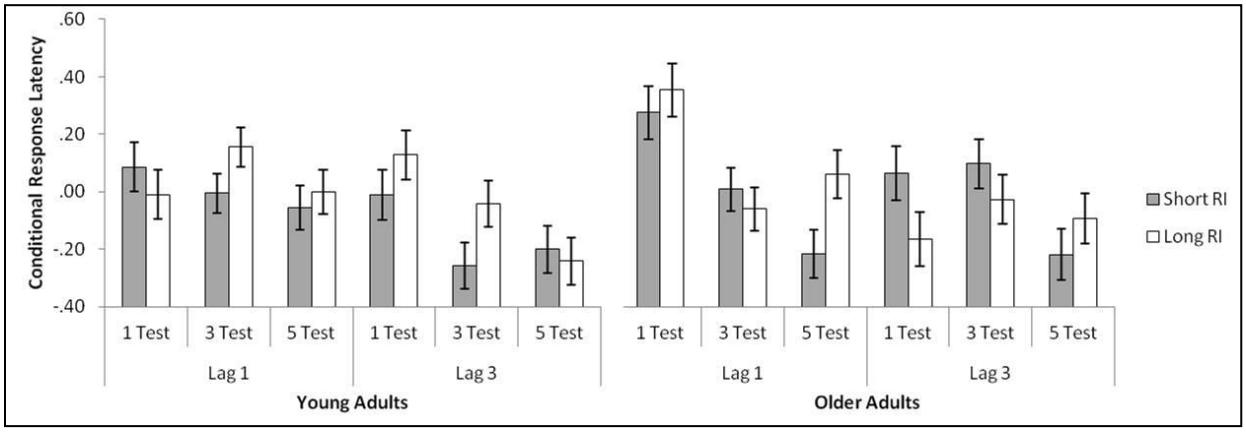
Figure 5. Mean proportion conditional cued recall on the final test in Experiment 1 as a function of age, retention interval (RI), lag, and number of tests. Error bars represent ± 1 S.E.M.



(Lag) x 3 (Number of Tests) mixed-factor ANOVA. The main effects of retention interval, lag, and number of tests were significant (*p*s < .05), in addition to a marginal effect of age ($p = .062$). The effects of age and retention interval were qualified by a significant two-way interaction, $F(1, 178) = 17.69$, $p < .001$, $\eta^2_p = .09$, which reflected a significant age difference in retention following a short retention interval ($p < .001$) but statistically equivalent performance between groups following a long retention interval ($p = .134$). The Lag x Number of Tests interaction was also significant, $F(2, 356) = 4.52$, $p = .012$, $\eta^2_p = .03$. As shown in Figure 5, this interaction reflected significant increases in performance in the Lag 1 condition as the number of tests during acquisition increased, ($M = .34$, .45, and .55 for 1-test, 3-tests, and 5-tests conditions, respectively; *p*s < .001), whereas performance in the Lag 3 condition increased from the one-test to three-test condition ($M = .46$ and $M = .61$, $p < .001$) but did not increase further with five tests ($M = .61$, $p > .90$). Finally, the Age x Lag interaction was significant, $F(1, 178) = 12.65$, $p < .001$, $\eta^2_p = .07$, which reflected a larger lag effect for older adults (.16) than for young adults (.06).

Conditional Standardized Response Latency. Mean conditional standardized response latency on the final cued recall tests is presented in Figure 6 as a function of age, retention interval, lag and number of tests taken during acquisition. There are three observations to note in the figure. First, overall response latency decreased across number of tests taken during encoding. Second, young adult response latency was relatively similar across conditions in the Lag 1 condition, but appeared to produce a testing effect in the Lag 3 condition. Third, older adults produced substantial benefit of testing in the Lag 1 condition, especially following the short retention interval.

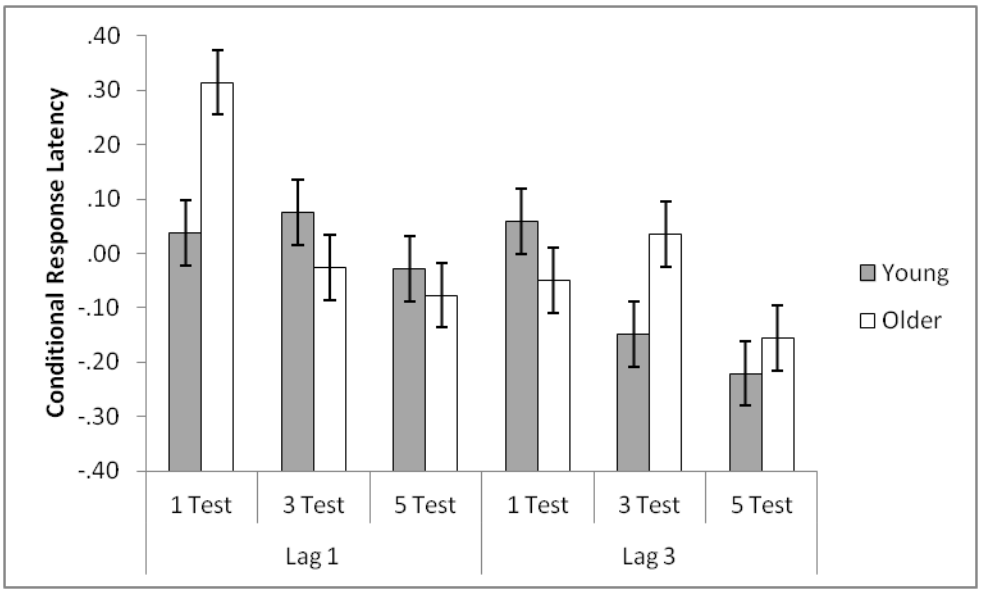
Figure 6. Mean conditional standardized response latency on the final test in Experiment 1 as a function of age, retention interval (RI), lag, and number of tests. Error bars represent ±1 S.E.M.



Results from a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 3 (Number of Tests) mixed-factor ANOVA revealed main effects of age, retention interval, and number of tests, *p*s < .05. These effects were further qualified by two three-way interactions. First, the Age x Lag x Number of Tests interaction was significant, $F(2, 356) = 7.73$, $p = .001$, $\eta^2_p = .04$. As seen in Figure 7, young adults' response latency was facilitated by conditions that placed the greatest demand on retrieval processing (i.e., Lag 3-3 test and Lag 3-5 tests). In contrast, older adults' response latency was facilitated in the Lag 1 condition when increasing from one to three tests but showed no changes across other conditions. To further explore the three-way interaction, separate Lag x Number of Tests ANOVAs were conducted for each age group.

Analysis of young adult performance revealed main effects of Lag and Number of Tests, *p*s < .05 , and a marginally significant Lag x Number of Tests interaction, $F(2, 194) = 2.63$, $p = .074$, $\eta^2_p = .03$. To explore the marginal interaction, response latencies for each lag condition were submitted to separate ANOVAs. Analysis of Lag 1 response latencies failed to yield an effect of Number of Tests, $p > .40$. However, examination of Lag 3 response latencies yielded a significant effect of Number of Tests, $F(2, 194) = 5.50$, $p = .005$, $\eta^2_p = .05$. Bonferroni corrected comparisons revealed significantly slower response latency in the single test condition compared to the multiple test conditions, *p*s < .05, and no difference between the multiple test conditions, $p > .999$. Analysis of older adult performance revealed main effects of Lag and Number of Tests, *p*s < .05, and a significant Lag x Number of Tests interaction, $F(2, 166) = 5.17$, $p = .007$, $\eta^2_p = .06$. Again, response latencies for each lag condition were submitted to separate ANOVAs. Analysis of Lag 1 response latencies revealed a significant effect of Number of Tests, $F(2, 166) = 9.91$, $p < .001$, $\eta^2_p = .10$. Follow-up comparisons with Bonferroni correction revealed significantly slower response latency in the single test condition compared to the multiple test

Figure 7. Age x Lag x Number of Tests interaction in conditional final test response latency. Error bars represent ± 1 S.E.M.



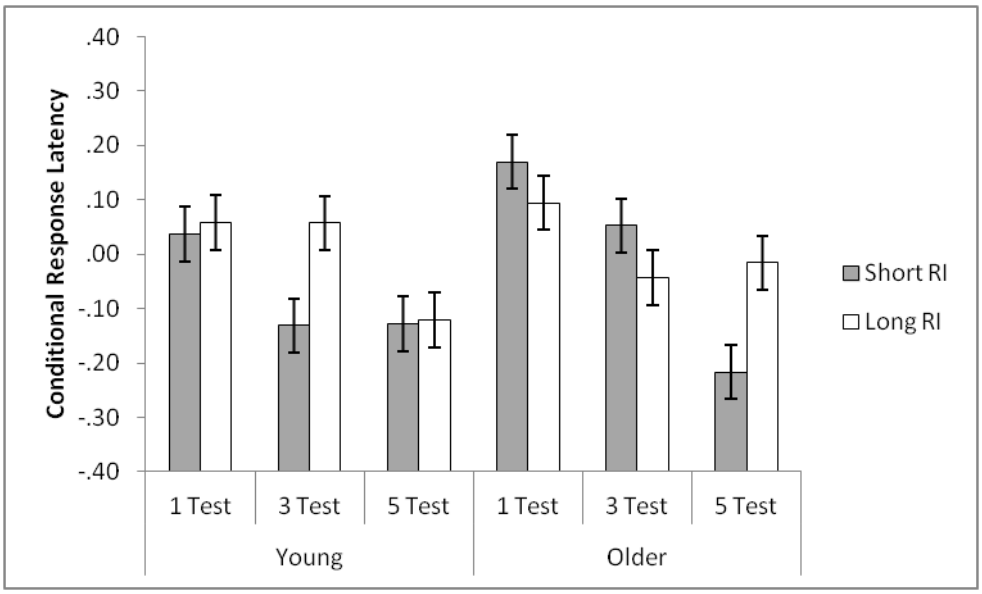
conditions, *p*s < .005, and no difference between the multiple test conditions, $p > .999$. Examination of Lag 3 performance failed to yield a significant effect of Number of Tests, $p > .15$.

There was also a significant Age x Retention Interval x Number of Tests interaction, $F(2, 356) = 3.48$, $p = .032$, $\eta^2_p = .02$ (displayed in Figure 8). This pattern appears to reflect a differential influence of additional tests for young and older adults as a function of retention interval. Although the interaction reached statistical significance, it is unclear why this pattern would be found and so no further follow-up analyses will be presented. Regarding the working memory measures, I will postpone the discussion of these measures until Experiment 2 is also presented to minimize redundancy.

Discussion

Given the emphasis on conditional final test performance, it is important to consider predictions for this variable based on Bjork's (1994) concept of desirable difficulty. Continued retrieval should increase retention more for items in the Lag 3 condition compared to the Lag 1 condition. Moreover, the function relating lag and continued testing to retention should be similar across age groups or may even reflect a stronger effect in the older adult group due to relatively more desirable difficulty for older adults compared to young adults as a result of age-related differences in cognition (e.g.. Giambra & Arenberg, 1993; McCabe et al., 2010). Clearly the results from Experiment 1 are inconsistent with these predictions. Results from Experiment 1 revealed significant increases in retention with each increase in testing in the Lag 1 condition but an isolated benefit in final retention with an initial increase in testing in the Lag 3 condition (i.e., increased retention when increasing from 1 to 3 tests but no additional benefit when increasing to 5 tests). Moreover, the results failed to reveal a benefit in retention for older adults

Figure 8. Age x Retention Interval (RI) x Number of Tests interaction in conditional final test response latency. Error bars represent ± 1 S.E.M



over young adults following both retention intervals. Of course, these predictions address only one way in which desirable difficulty may influence the long-term memory benefits of lag and continued testing for young and older adults. Alternative ways in which desirable difficulty may influence long-term memory performance will be considered in the General Discussion.

Unlike previous studies that have typically limited analysis to accuracy, the current study examined both accuracy and response latency. Critically, response latency results diverged from conditional final test accuracy performance. With respect to final test response latency, there were the expected main effects of number of tests, lag and retention interval. Additionally, the significant Age x Lag x Number of Tests interaction reflected a facilitation effect in young adult standardized response latency for combinations of lag and test that placed the greatest demand on retrieval processing (i.e., Lag 3-3 test and Lag 3-5 tests) and a facilitation effect in older adult standardized response latency for minor increases in retrieval processing (i.e., increasing from 1 to 3 tests in the Lag 1 condition).

In sum, the number of tests included during the acquisition phase and the lag by which tests were spaced both influenced final test performance. In addition, the current results revealed dissociation between measures of final test conditional accuracy and response latency. Of course, comparing two spacing intervals will not always produce a pattern of results similar to that observed in Experiment 1 given the nonmonotonic function relating lag with final test performance noted earlier (see Cepeda et al., 2006). Indeed, continued testing with an ineffective lag, namely massed retrieval, should produce little to no benefit in final test performance. Thus, Experiment 2 was motivated to extend the results from Experiment 1 through an examination of massed versus spaced retrieval practice. This extension is also

motivated by the results from Maddox et al. (2011) which indicated that massed performance may produce a specific benefit for young adults, compared to older adults.

Experiment 2

Although the results from Experiment 1 were intriguing, they were inconsistent with the original prediction that retrieval practice would produce a larger benefit in final test performance for items separated by the longer lag than the shorter lag. Moreover, analysis of conditional final test accuracy and response latency revealed different patterns of results suggesting that both measures are useful for understanding the manipulations examined in the current paradigm. In order to further examine these differences, Experiment 2 compared the benefit of continued testing when tests were massed versus spaced (i.e., the spacing effect; Lag 0 and Lag 4, respectively). The use of massed retrieval in Experiment 2 made the lag conditions more distinct than the lag conditions included in Experiment 1 and more importantly offered an interesting condition in which continued testing should produce minimal increases in memory performance over a single massed attempt.

A second important prediction for Experiment 2 stems from the observation noted in the Introduction that young adults may benefit from immediate refreshing more than older adults (e.g., Johnson et al., 2002; Maddox et al., 2011). Past evidence (Johnson et al., 2002) suggests that older adults are slower to retrieve items on massed retrieval attempts and benefit less on a later recognition test for these items than young adults. Hence, one might expect young adults' long-term memory performance to benefit more from continued massed retrieval than older adults' performance. Continued testing may also differentially influence response latency as a function of lag and age group. Specifically, Experiment 1 results indicated that older adults' response latency benefited from increased retrieval practice in the short lag condition, whereas

young adults' retrieval fluency benefited from increased retrieval practice in the long lag condition. Thus, older adult response latency may be facilitated from continued testing in the massed condition even when no benefits are obtained in terms of overall retention, and young adults may show a selective benefit in response latency with continued testing in the spaced retrieval condition.

Given the specific interest in examining the relationship between spacing, testing and final test performance, Experiment 2 included two levels of testing (1 vs. 3 tests) and one retention interval (5 minutes). These changes in methodology have the added benefit of reducing the overall list length and increasing older adult performance above the near-floor performance in the long retention interval condition observed in Experiment 1.

Method

Participants. Young adults were undergraduates at Washington University in St. Louis and received partial course credit or monetary remuneration ($10) for their participation. Older adults were healthy, community dwelling adults and received monetary compensation ($15) for their participation. Age and years of education differed significantly between age groups (*p*s < .001; see Table 2). Analysis of the two working memory measures revealed marginally significant differences between age groups in both the CompSpan, $t(46) = 1.94$, $p = .058$, and the LNS, $t(46) = 1.89$, $p = .066$. However, the working memory composite score differed significantly between age groups, $t(46) = 2.64$, $p = .011$. Finally, there was a significant difference between groups in Shipley vocabulary scores, $t(46) = 6.42$, $p < .001$.

Design. A 2 (Age) x 2 (Spacing, 0 vs. 4) x 2 (Number of Tests; 1 vs. 3) mixed-factor design was used in Experiment 2. Age was a between-participants factor. Lag and Number of

Tests were within-participants factors. Retention interval was five minutes for both age groups.

Materials.

Memory Task. A subset of thirty-two low associate word pairs was selected from Experiment 1. Word pairs were divided into four sets of eight pairs, and these sets were counterbalanced across lists such that each pair occurred once in each of the within-participants conditions. A continuous paired associate task was again used for the acquisition phase of the memory task. The average serial list position was equated across all trials (*p*s > .90) as well as the average serial list position for the first test (*p*s > .80) and the last test (*p*s > .75) across the two testing and two spacing conditions. In total, the acquisition phase consisted of 139 trials of which 96 trials were included for critical conditions, 35 were filler trials, and the remaining eight trials were equally split between primacy and recency buffer items. Of the 96 critical condition trials, 32 trials were encoding trials and 64 trials were retrieval practice trials. Filler trials were included to ensure that average serial list position was equated across the critical conditions.

In addition to the memory task, participants completed the Shipley vocabulary test and the two working memory tasks (CompSpan and LNS) used in Experiment 1.

Procedure. The procedure was the same as the procedure used in Experiment 1 with the following two exceptions: (a) only single test and three test conditions were included; (b) a single five minute retention interval was used.

Results

Acquisition Performance.

Again, the current set of analyses emphasizes performance on the first and last retrieval attempts in each of the multiple retrieval attempt conditions as a way of assessing the stability of retrieval accuracy across testing events and the degree to which response latency decreases across test events as a function of lag.

Memory Accuracy. Mean proportion correct recall for young and older adults is shown in Figure 9 as a function of lag, number of tests, and test number. Accuracy data from the first and last retrieval attempt in the three test conditions were submitted to a 2 (Age) x 2 (Lag) x 2 (Test Number) mixed-factor ANOVA. Results revealed main effects of Age and Lag, *p*s < .005, that were further qualified by a significant Age x Lag interaction, $F(1, 46) = 5.86$, $p = .020$, $\eta^2_p = .11$. The interaction revealed statistically equivalent performance across age groups in the Lag 0 condition, $p = .173$, but a significant difference in Lag 4 performance between young (.55) and older adults (.39), $p = .003$.

Standardized Response Latency. Mean standardized response latency for young and older adults is shown in Figure 10 as a function of lag, number of tests, and test number. Again, response latency data from the first and last retrieval attempt in the three test condition were submitted to a 2 (Age) x 2 (Lag) x 2 (Test Number) mixed-factor ANOVA. All main effects were significant, *p*s < .05, in addition to a significant Lag x Test Number interaction, $F(1, 46) = 10.87$, $p = .002$, $\eta^2_p = .19$. Moreover, the three-way interaction was significant, $F(1, 46) = 6.17$, $p = .017$, $\eta^2_p = .12$. Separate analysis of young adult response latency revealed main effects of Lag and Test Number, *p*s < .001, with no interaction, $p > .50$. Analysis of older adult response latency also revealed significant main effects of Lag and Test Number, *p*s < .001, and a significant Lag x Test Number interaction, $F(1, 23) = 14.12$, $p = .001$, $\eta^2_p = .38$. As shown in Figure 10, response latency for older adults decreased from the first to second retrieval attempt in the Lag 0 condition ($p = .002$) but remained stable from the second to third retrieval attempt ($p > .40$). In contrast, response latency significantly decreased across all retrieval attempts in the Lag 4 condition, *p*s < .05

Final Test Phase Performance. Similar to Experiment 1, conditional probability of

Figure 9. Mean proportion cued recall during the acquisition phase in Experiment 2 as a function of age, lag, number of tests, and test number (i.e., T1 = first retrieval attempt, T2 = second retrieval attempt, T3 = third retrieval attempt). Error bars represent ± 1 S.E.M.

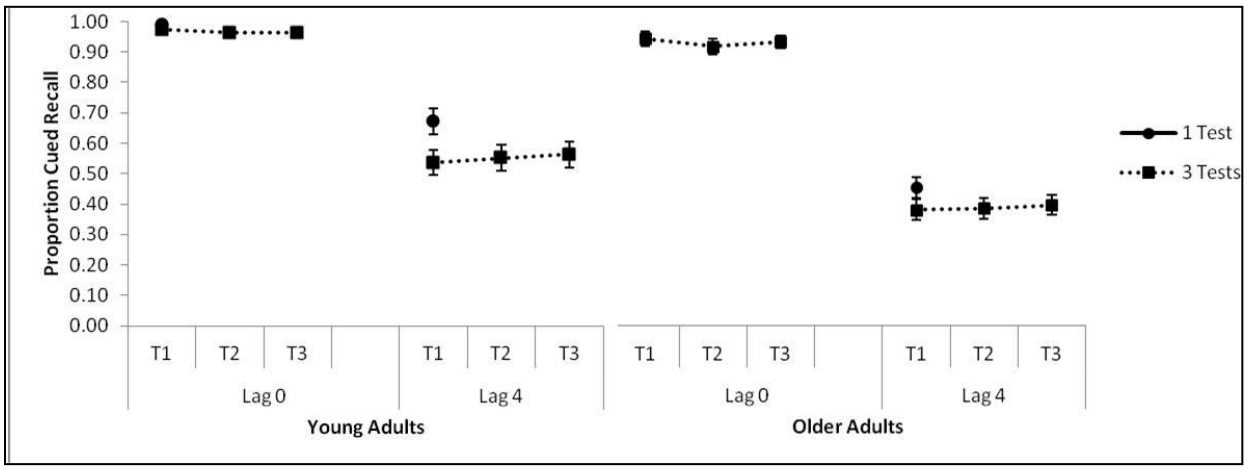
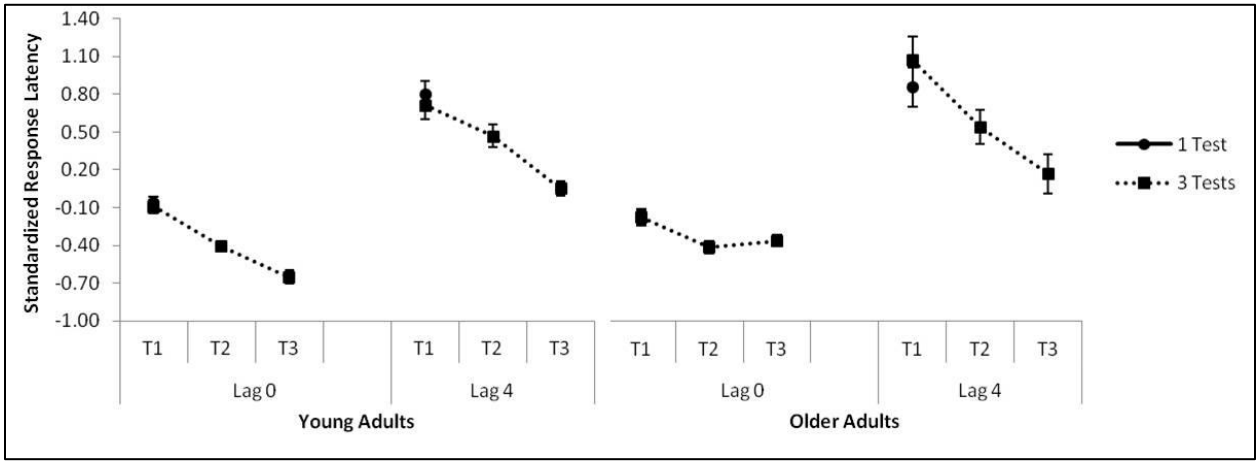


Figure 10. Mean standardized response latency during the acquisition phase in Experiment 2 as a function of age, lag, number of tests, and test number (i.e., T1 = first retrieval attempt, T2 = second retrieval attempt, T3 = third retrieval attempt). Error bars represent ± 1 S.E.M.

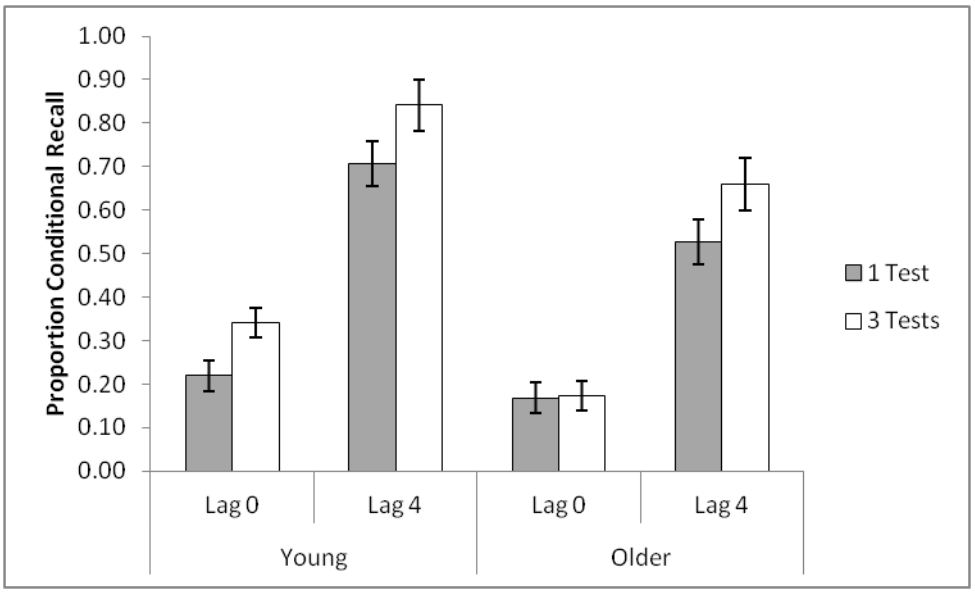


cued recall on the final test given the item was recalled successfully on the final retrieval attempt during the acquisition phase is reported prior to response latency analyses. Analysis of nonconditional final test performance generally produced the same pattern of performance as those analyses reported for conditional final test performance (see Appendix C for analysis of nonconditional final test performance).

Conditional Memory Accuracy. Figure 11 displays mean proportion conditional recall as a function of lag and number of tests for young and older adults. As displayed in Figure 11, performance was higher for young adults than older adults (M = .53 vs. M = .38, respectively), Lag 4 items were remembered better than Lag 0 items (M = .68 vs. M = .23, respectively), and taking three tests led to better retention than taking a single test (M = .50 vs. M = .41, respectively).

Conditional probability was submitted to a 2 (Age) x 2 (Lag) x 2 (Number of Tests) mixed-factor ANOVA. There were significant effects of age, $F(1, 46) = 15.44$, $p < .001$, $\eta^2_p = .25$, lag, $F(1, 46) = 272.33$, $p < .001$, $\eta^2_p = .86$, and number of tests, $F(1, 46) = 8.03$, $p = .007$, $\eta^2_p = .15$. Although the three-way interaction was clearly not significant ($F < 1.00$, $p = .340$), separate analyses were conducted to examine the benefit of additional testing for each lag condition given *a priori* predictions based on age differences in refreshing discussed above (cf. Johnson et al., 2002; Maddox et al., 2011). Analysis of Lag 0 performance revealed main effects of age and number of tests (ps < .05) which were further qualified by a significant Age x Number of Tests interaction, $F(1, 46) = 4.20$, $p = .046$, $\eta^2_p = .08$. This interaction reflected a significant increase in performance when testing was increased from one to three tests for young adults (p = .005) but not for older adults ($p > .90$). With regard to Lag 4 performance, the ANOVA revealed main effects of Age and Number of Tests (ps < .05) but no interaction, $p > .95$. As predicted,

Figure 11. Mean proportion conditional cued recall on the final test in Experiment 2 as a function of age, lag, and number of tests. Error bars represent ± 1 S.E.M.



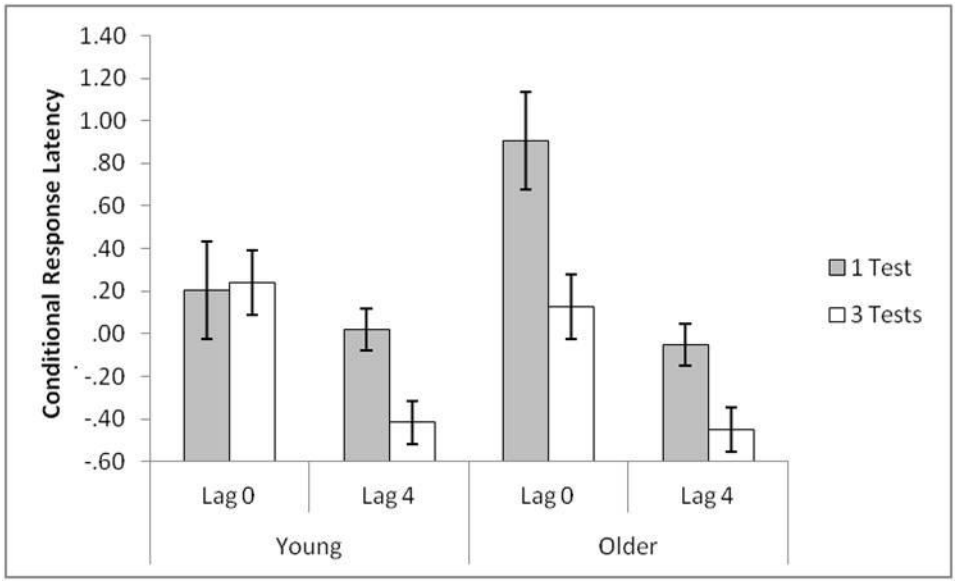
these results indicate that young adults benefited from repeated testing when refreshing was engaged in the Lag 0 condition, but older adults did not produce this benefit.

Conditional Standardized Response latency. Mean conditional standardized response latency was submitted to a 2 (Age) x 2 (Lag) x 2 (Number of Tests) mixed-factor ANOVA which yielded main effects of Lag, $F(1, 46) = 33.49$, $p < .001$, $\eta^2_p = .42$, and Number of Tests, $F(1, 46) = 15.40$, $p < .001$, $\eta^2_p = .25$. As shown in Figure 12, standardized response latency was faster in the Lag 4 condition ($M = -.24$) than in the Lag 0 condition ($M = .31$), and was faster in the three test condition ($M = -.14$) than in the one test condition ($M = .21$). Again, the three-way interaction ($F = 2.69$, $p = .108$) was examined given *a priori* predictions regarding refreshing and age differences in final test response latency as a function of spacing condition and number of tests in Experiment 1. A separate analysis of Lag 0 standardized response latency revealed a marginally significant Age x Number of Tests interaction ($F = 3.81$, $p = .057$). Follow-up comparisons revealed a significant difference in response latency between the single and three test conditions for older adults ($p = .035$) but not for young adults ($p > .85$). Analysis of Lag 4 standardized response latency failed to yield a significant Age x Number of Tests interaction ($p > .85$). Therefore, in contrast to the accuracy data, the response latency data reflect a larger benefit from testing in the Lag 0 condition for older adults compared to young adults. Again, these results indicate it is important to examine response latency in conjunction with accuracy.

Discussion

The results from Experiment 2 are clear. First, the final test results again indicate that conditional accuracy and response latency are distinct measures of memory performance and must be considered together when making inferences about the effects of variables within this paradigm. Second, as predicted, both age groups benefited from spaced retrieval and continued

Figure 12. Mean conditional standardized response latency on the final test in Experiment 2 as a function of age, lag, and number of tests. Error bars represent ± 1 S.E.M.



testing in the Lag 4 condition. However, only young adults benefited in terms of conditional accuracy from continued testing in the Lag 0 condition, which is consistent with previously reported age differences in refreshing described above. Third, older adult response latency benefited from additional massed and spaced testing, whereas young adult retrieval fluency only benefited when continued retrieval occurred in the Lag 4 condition. Although motivated *a priori*, these later comparisons were marginal and should be interpreted with caution.

Cross-experiment Comparison

When considering the results from both Experiments 1 and 2, there are several consistent findings that elucidate the relationship between experimenter-controlled variables and the long-term benefits of continued testing. As shown in Figure 13, conditional accuracy increased with increased testing for both age groups across lag conditions with one exception. As noted earlier, older adults did not show a benefit in retention from increased massed testing which may reflect age-related differences in refreshing (c.f. Johnson et al., 2002; Maddox et al., 2011). With respect to response latency, first consider young adult performance in Figure 14. Across both experiments, young adults showed a selective benefit in response latency (i.e., faster response latency) from repeated testing in the long lag conditions but not the short lag conditions. In contrast, older adults produced the largest benefit in response latency with increased testing in the short lag conditions compared to the long lag conditions. Clearly, the response latency data showed different influences of testing on retrieval fluency as a function of lag and age.

In order to directly test the consistency in this pattern across experiments, response latency data from the short retention interval in Experiment 1 and data from Experiment 2 were entered into a 2 (Age) x 2 (Experiment) x 2 (Lag: Short, Long) x 2 (Number of Tests: 1, 3) mixed-factor ANOVA. Results revealed a significant Age x Lag x Number of Tests interaction,

Figure 13. Mean proportion conditional cued recall on the final test following a 5 minute retention interval for the single test and three test conditions in Experiments 1 and 2 as a function of age and lag. Error bars represent ± 1 S.E.M.

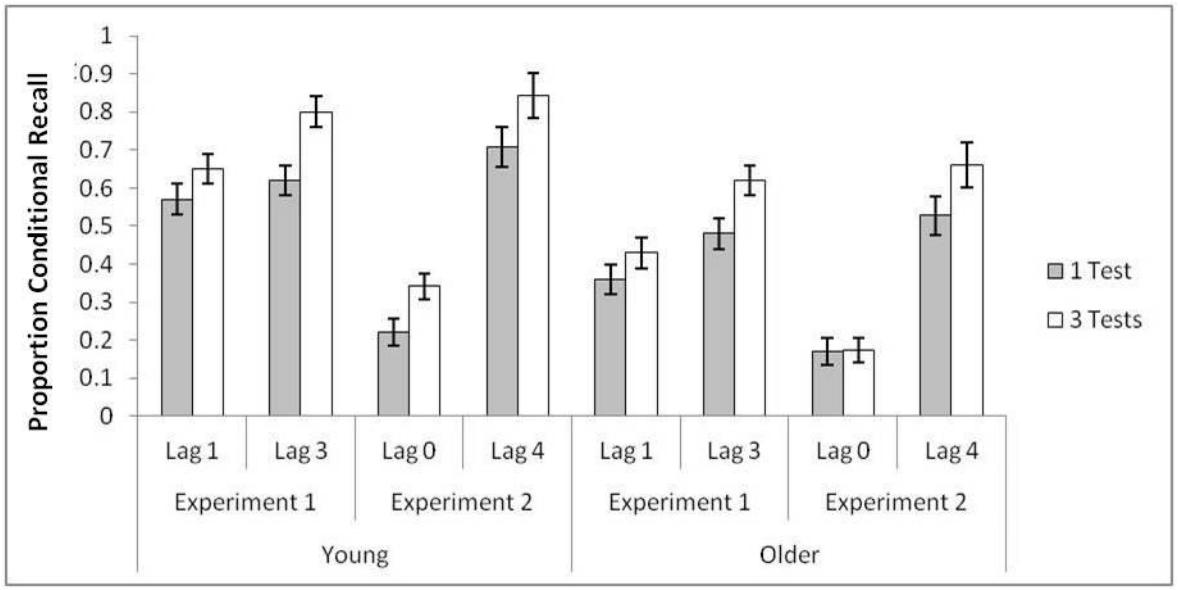
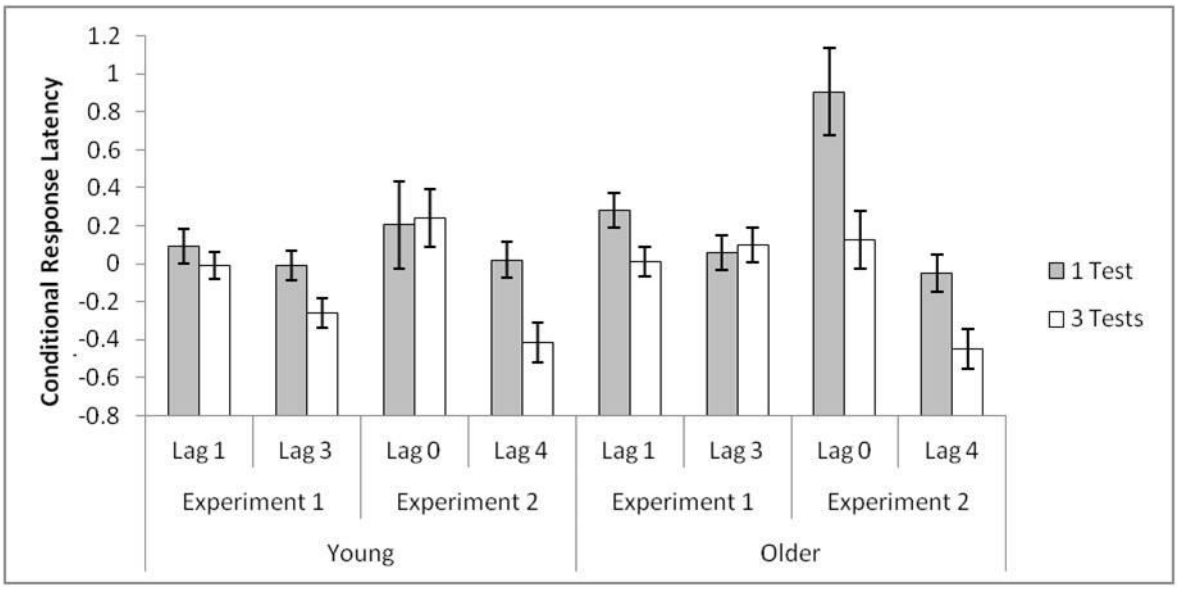


Figure 14. Mean conditional standardized response latency on the final test following a 5 minute retention interval for the single test and three test conditions in Experiments 1 and 2 as a function of age and lag. Error bars represent ± 1 S.E.M.



$F(1, 135) = 6.56$, $p = .012$, $\eta^2_p = .05$, in which Experiment did not contribute additional variance, as reflected by the nonsignificant four-way interaction, $F < 1.00$, $p > .40$.

Working Memory Capacity and Retrieval Practice

A secondary interest in the current study was to examine the relationship between working memory and the benefits of continued spaced retrieval practice. First consider Experiment 1. CompSpan and LNS scores were standardized and averaged to form a composite working memory score for each participant because the two tasks were significantly correlated when controlling for age ($r = .35$, $p < .001$). Initial correlation analyses between the working memory composite and conditional final test performance yielded no systematic effects. Thus, a median split was used to create low and high working memory groups based on the composite score. Each of the acquisition and final test analyses reported above were replicated when collapsing across age groups. Working Memory Group served as a between-participants factor and age (in years) served as a covariate. Given that only the short retention interval was consistent across age groups the following analyses focused on the short retention interval data[4]. Significant results that involve working memory group are reported separately for each experiment.

The results from Experiment 1 acquisition phase analyses revealed a marginal effect of WMC Group in accuracy, $F(1, 88) = 2.81$, $p = .097$, $\eta^2_p = .03$, which reflected overall better performance by the high working memory group than the low working memory group ($M = .66$ vs. .60, respectively). With respect to standardized response latency, the results revealed a significant WMC Group x Lag interaction, $F(1, 88) = 9.97$, $p = .002$, $\eta^2_p = .10$, which reflected a larger difference in response latency between lag conditions for the low WMC group than the high WMC group ($M_{diff} = .17$ vs. .09). This latter effect is consistent with the notion that various

lags are qualitatively different in terms of retrieval difficulty as a function of WMC. In terms of final test performance, there were no effects of WMC on conditional performance.

Results from Experiment 2 were analyzed similarly to those from Experiment 1 such that WMC group (established using a median split of the standardized composite score) served as a between-participants factor and age (in years) served as a covariate. Acquisition phase analyses failed to yield any significant effects involving WMC Group in accuracy or response latency. Similarly, there were no significant effects involving WMC Group in conditional final test performance. One possibility for this lack of finding is that the working memory measures may have yielded unreliable estimates of WMC. Indeed, performance on the LNS was not significantly correlated with performance on the CompSpan when controlling for age in Experiment 2 ($r = .03$, $p > .80$). Alternatively, the comparison of massed versus spaced retrieval may have limited the ability to detect differences between working memory groups if both low and high WMC groups could easily retrieve items immediately following the encoding trial and had comparable difficulty in retrieving items following a relatively longer spacing interval (compared with the spacing intervals used in Experiment 1). The failure to obtain differences during acquisition is consistent with this latter possibility.

In summary, the results from Experiments 1 and 2 suggest that working memory capacity may modulate retrieval success during the acquisition phase when the lags used to separate learning and testing events fall within a sensitive range. However, when the lags used are more extreme (i.e., either no spacing interval or too long of a spacing interval as in Experiment 2), differences in the benefits of retrieval practice may be minimized between WMC groups. Analysis of conditional final test performance failed to yield any significant differences between WMC groups. This finding suggests that WMC is a more critical consideration when selecting

the lags that will separate learning and testing events as a means of maximizing acquisition performance across groups and that retention of material across time may be similar across various levels of WMC (see Appendix D for analysis of nonconditional final test performance).

Experiment 3

The first and second experiments addressed the influence of experimenter controlled factors on the benefits of testing. Specifically, lag and the number of tests included during the acquisition phase were both manipulated to examine predictions based on the desirable difficulty hypothesis (Bjork, 1994). Beyond these experimenter-controlled factors, the participant's motivation to learn material may also modulate the benefits of testing during the acquisition phase. For example, being highly motivated to learn material may make subsequent retrieval of that information during acquisition easier than retrieval of material deemed relatively less important to learn. In turn, less difficult retrieval for material deemed highly important may actually lead to *reduced* retention compared to more difficult retrieval of material encoded under low motivation. In order to address the third critical question of the dissertation, Experiment 3 further examined the benefit of retrieval practice as a function of age and lag, while also manipulating participant motivation at encoding.

The participant's motivation to learn information was manipulated on a trial-by-trial basis using a value-directed encoding procedure. Past studies (e.g., Castel, 2008; Castel, Balota, & McCabe, 2009; McGillivray & Castel, 2011) that have investigated the effects of value-directed encoding have typically used a list learning paradigm in which items were presented one at a time along with an associated point value (e.g., each word in a 12 item list was assigned a value ranging from 1 to 12 points). Words have generally been presented at a relatively fast rate (e.g., 1 word per second), and memory was tested after a short retention interval (e.g., most often

immediately following the study phase). The results from this procedure typically indicate that point value modulates overall memory performance and that the age-related difference in episodic memory is most clearly observed for low point value items (e.g., Castel, 2008).

Experiment 3 extended the standard point value paradigm in three different ways. Most importantly, Experiment 3 compared the benefit of massed and spaced retrieval practice for high and low-value word pairs. Moreover, the final test was administered shortly after the acquisition phase (i.e., 30 second retention interval) or following a longer delay (i.e., 15 minute retention interval). These extensions allow for an examination of the extent to which value-directed encoding differentially influences acquisition of material versus retention of material. Finally, Experiment 3 utilized paired associates and a slower presentation rate (4.5 seconds per pair).

The slower presentation rate used in the current paradigm, selected to accommodate the acquisition of relatively more demanding stimuli (i.e., paired associates), may be a critical departure from previous studies that have examined value-directed encoding for single words. Specifically, a fairly quick presentation rate like those used in the past may lead participants to adopt a strategy in which high-value items are attended to and low-value items receive little encoding. Indeed, Castel et al. (2011) reported a factor analysis which revealed two classes of items. Specifically, one factor captured high-value items and the other factor captured low-value items. In contrast, a slower presentation rate like the one used in the current experiment may encourage participants to use a strategy in which sufficient encoding is attempted for all items and only subsequent rehearsal is modulated by point value.

Given the prediction that point value will exert a greater influence on rehearsal of the stimulus than initial encoding of the stimulus, extending the value-directed encoding procedure to a spaced retrieval paradigm in which some repetitions are massed affords an interesting and

important comparison. Specifically, past research suggests that massed items suffer from reduced processing and rehearsal on the item's second presentation relative to spaced items (e.g., Johnston & Uhl, 1976; Rundus, 1971; Shaughnessy, Zimmerman, & Underwood, 1972). In turn, this deficient processing contributes in part to the spacing effect obtained when both massed and spaced items are presented in the same list. When items are assigned point values, low-value pairs are predicted to suffer to a greater extent than high-value pairs from deficient processing on the pair's second massed presentation. Consequently, low-value massed items are expected to show a greater deficit in terms of long-term retention than high-value massed items. With regard to the spaced retrieval condition, if high-value word pairs are rehearsed to a greater extent than low-value pairs, then retrieval practice should be easier and more successful for the former class of stimuli. Thus, according to Bjork's (1994) desirable difficulty account, low point value pairs retrieved after a lag should be better remembered on the final test compared to high point value pairs, because these items should be more difficult to retrieve during the acquisition phase, since they are less likely to be maintained during the lag. As in the previous experiments, conditional recall should capture the consequences of different levels of desirable difficulty.

Finally, it is important to consider the shift from single items in prior value-directed encoding studies to paired associates in the current paradigm. Past research indicates that older adults have particular difficulty with paired associate stimuli relative to young adults (e.g., Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000). This difficulty can be seen in the preceding experiments in terms of acquisition accuracy and nonconditional performance on the final cued recall test (see Appendices B and C). To provide a second measure of episodic memory performance in the current paradigm, a recognition test was administered following the cued recall test. The recognition test consisted of a series of word pairs all of which were

comprised of previously studied stimuli. Pairs were divided equally between intact pairs (i.e., the first and second words were paired exactly as they were presented during encoding) and rearranged pairs (i.e., words were re-paired with other studied, incorrect words). As a result, the extent to which lag and point value modulate recognition performance that relies heavily on retrieving associations formed during the acquisition phase can be examined. Specifically, if lag and point value lead to increased processing of the relationship between cue and target during the acquisition phase, incorrectly endorsing rearranged pairs as "intact" (i.e., false alarms) should be relatively low for "strong" items (i.e., high point value and/or spaced retrieval) given that retrieving either of the two original word pairs would provide sufficient evidence to correctly reject a lure. This may be especially true for young adults compared with older adults who have particular difficulty with associative encoding and retrieval (e.g., Naveh-Benjamin, 2000). However, one may predict the opposite pattern of results if the critical manipulations (i.e., point value and spacing interval) exert a broader influence on processing (e.g., increased rehearsal of target items at the expense of fully encoding the relationship between cue and target). In this instance, each component of the lure item (the cue and target) will be highly familiar which should in turn lead to a higher probability of false alarms to these items than to lures comprised of two relatively "weak" items (i.e., massed items and low point value items). Such a pattern of results is more likely to be observed in older adult performance given past research indicating their use of relatively ineffective encoding strategies (e.g., Hertzog, Price & Dunlosky, 2012) and overall difficulty with associative memory (e.g., Naveh-Benjamin, 2000) compared to young adults.

Similar to Experiments 1 and 2, the influence of working memory capacity in the current paradigm will also be examined. Although the previous experiments yielded little evidence of an

influence of working memory, it is predicted that individuals with low working memory capacity will benefit more from the manipulation of point values during encoding than individuals with high working memory capacity. Much like older adults, these individuals may not be able to retrieve items successfully following a long lag without specifically directing attention and effort to maintaining high-value items across the interval at the cost of low-value items. In turn, when retrieval is successful following a longer lag, it will be with relatively greater difficulty for older adults than young adults and for individuals with low WMC compared to those with high WMC. Thus, one may predict a larger lag effect in conditional recall for the former group relative to the latter group with respect to both the age and WMC comparisons.

Method

Participants. Young and older adults were recruited from the same population used in Experiments 1 and 2. Age and years of education differed significantly between age groups (*p*s < .001; see Table 2). Young adults were given course credit or monetary remuneration for their participation ($15), and older adults were monetarily compensated for their time ($15). Analysis revealed significant differences in performance between age groups for each working memory measure, the working memory composite, and Shipley vocabulary test, *p*s < .005.

Design. A 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 2 (Point Value) mixed-factor design was used in the current experiment. Age and Retention Interval were between-participant factors. Lag and Point Value were within-participant factors. Retention interval was either 30 seconds or 15 minutes for both young and older adults. Similar to Experiment 2, retrieval occurred either immediately following the encoding trial (i.e., massed retrieval) or following a short lag (i.e., Lag 4), and each word pair was identified on its encoding trial as being worth

either 3 points (low-value) or 6 points (high-value). These point values were selected so that the high-value condition was worth double the number of points as the low-value condition.

Materials

Memory Task. A continuous paired associate task similar to those used in Experiments 1 and 2 was used for the acquisition phase of the memory task. Pilot testing indicated that the acquisition phase should be split in half. It appeared that when all encoding occurred in a single phase, no effect of point value was observed, likely due to near floor performance. This may be due to the extra load of interpreting the point value information during encoding. Thus, participants completed the memory task two times with novel word pairs in each list.

Each list included 64 trials. Of the 64 trials, there were 16 primacy and recency trials, 16 filler trials, and 32 critical trials. Critical trials were equally divided between each of the four Lag x Point Value conditions. For each critical word pair, the first presentation was an encoding trial which included the associated point value for the pair directly beneath the stimulus, and the second presentation was the pair's retrieval trial without any indicator of point value. Thus, the influence of the point value presumably should modulate the likelihood of maintaining the item across the lag.

Following cued recall during each final test phase, participants completed a recognition test in which half of the pairs in each condition were presented intact and the other half were rearranged (see Table 3). Rearranged pairs were constructed within each condition. For example, two of the four word pairs in a given condition were intact target pairs and the remaining two pairs were rearranged such that the targets were repaired with the incorrect cue. Participants were asked to respond whether the pair presented on the screen appeared exactly as it had been presented during the acquisition phase (i.e., the pair was *intact*) or if the words had

Table 3. Example of Intact versus Rearranged Recognition Test with Correct Recognition Response.

Study List	Recognition Test	Recognition Response
HORSE – jumped	HORSE – jumped	Intact
APPLE – evil	APPLE – title	Rearranged
KING – title	KING – evil	Rearranged
MARKET - shelf	MARKET – shelf	Intact

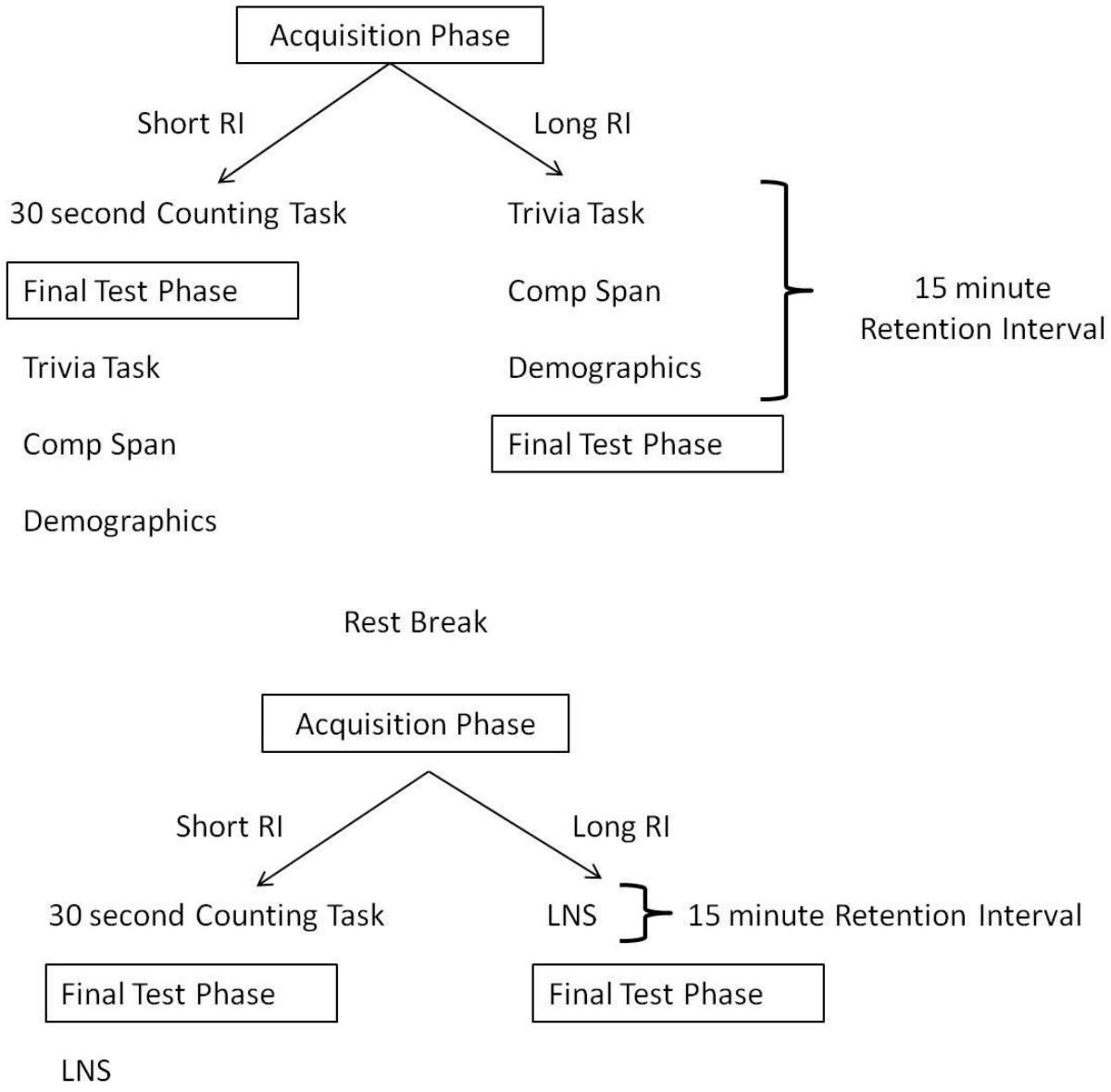
been mismatched from different pairs (i.e., the pair was *rearranged*). Recognition trials were randomly ordered and appeared individually on the screen until participants responded. Importantly, the Intact versus Rearranged recognition test was added to the paradigm following the first two counterbalancing sets of older adult participants in the long retention interval condition. Thus, only 10 of the 24 older adults in the long retention interval condition completed this task. Otherwise, all other participants took both the cued recall and recognition tests.

In addition to the memory task, participants completed the trivia distractor task, Shipley vocabulary test and the two working memory tasks (CompSpan and LNS) used in Experiments 1 and 2.

Procedure. Because the memory task consisted of two separate lists, the first and second halves of the experimental session will be considered separately. A schematic of the experimental session is presented in Figure 15.

As displayed in Figure 15, all participants were presented with the first acquisition list during which they studied word pairs and responded aloud on retrieval trials. Again, the experimenter entered responses immediately upon vocalization by the participant. Following the first acquisition phase, the procedure differed based on retention interval condition. Participants in the short retention interval condition completed a brief, 30 second counting task in which they counted backwards by sevens from a three digit number before completing the first final test phase of the memory task (consisting of both the final cued recall and recognition tests). Then participants completed a battery of tasks which included the trivia question distractor task, the CompSpan (Conway et al., 2005) and a demographics questionnaire. Participants in the long retention interval condition completed the trivia task, CompSpan, and the demographics questionnaire between the initial acquisition phase and final test phase for the first list of word

Figure 15. Schematic of the experimental session for short retention interval (30 seconds) and long retention interval (15 minute) participants in Experiment 3. Each phase of the value-directed encoding task appears within an additional border.



pairs which produced a 15 minute retention interval. Participants in both the short retention interval and long retention interval conditions were given the opportunity to take a break before beginning the second half of the experiment.

Following the brief break, all participants were presented with the second acquisition phase that included all novel word pairs. Participants in the short retention interval condition again completed the 30 second counting task between the acquisition phase and final test phase. Following the final test phase, participants completed the LNS task. Participants in the long retention interval condition completed the LNS task between the acquisition phase and final test phases (which again produced a 15 minute retention interval).

Results

Acquisition phase performance and conditional final test performance will be considered separately. Nonconditional final test performance is reported in Appendix E.

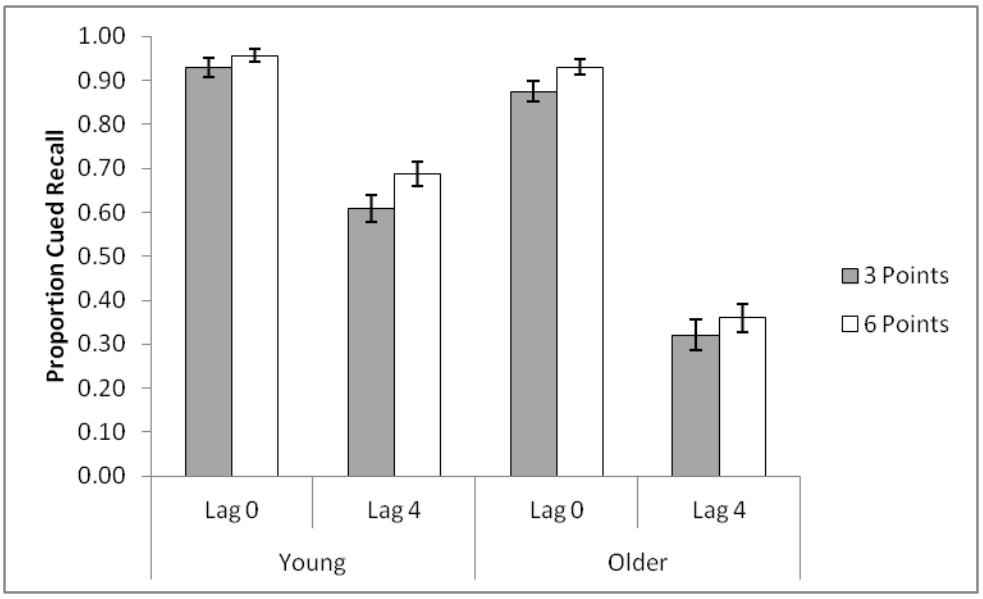
Acquisition Performance.

Performance was collapsed across retention interval groups given that the acquisition phase was the same for both groups. Indeed, analyses failed to yield any significant differences in performance as a function of retention interval condition (*p*s > .25).

Memory Accuracy. Mean proportion cued recall for young and older adults is shown in Figure 16 as a function of lag and point value. There are three observations to note in this figure. First, performance was better in the Lag 0 condition than in the Lag 4 condition, and young adults performed better than older adults. Second, the difference between lag conditions was larger for older adults than young adults. Third and most critically, the value-directed encoding manipulation produced better performance for high-value pairs compared to low-value pairs.

The above observations were supported by a 2 (Age) x 2 (Lag) x 2 (Point Value)

Figure 16. Mean proportion cued recall during the acquisition phase in Experiment 3 as a function of age, lag and point value. Error bars represent ± 1 S.E.M.



mixed-factor ANOVA. Results revealed main effects of age, lag and point value, *p*s < .001, as well as a significant Age x Lag interaction, $F(1, 104) = 51.25$, $p < .001$, $\eta_p^2 = .33$, which reflected a larger difference between lag conditions for older adults ($M_{diff} = .56$) than young adults ($M_{diff} = .30$).

Standardized Response Latency. Mean standardized response latency is presented in Figure 17 as a function of age, lag and point value. There are again three critical observations to note in this figure. First, young adult response latency was faster than older adult response latency. Second, Lag 0 response latency was faster than Lag 4 response latency. Third, the difference in response latency as a function of lag was larger for older adults than for young adults.

Results from a 2 (Age) x 2 (Lag) x 2 (Point Value) mixed-factor ANOVA revealed main effects of age and lag, *p*s < .001, and a significant Age x Lag interaction, $F(1, 103) = 4.32$, $p = .040$, $\eta_p^2 = .04$. This interaction reflected a larger difference in response latency between lag conditions for older adults ($M = 1.11$) than young adults ($M = .88$). In contrast to the accuracy data, there was no influence of point value on response latency during acquisition ($M = .13$ and .16 for high-value and low-value conditions, respectively).

Final Test Phase Performance.

Conditional Memory Accuracy. Mean proportion conditional recall is shown in Figure 18 as a function of age, retention interval, lag, and point value. There are three observations to note in this figure. First, retention was better for young adults than older adults. Second, retention was better for items that received spaced practice during learning (i.e., Lag 4) relative to massed practice. Third, high point value items appear to be retained better than low point value items, and this effect appears particularly salient for young adults.

Figure 17. Mean standardized response latency during the acquisition phase in Experiment 3 as a function of age, lag and point value. Error bars represent ± 1 S.E.M.

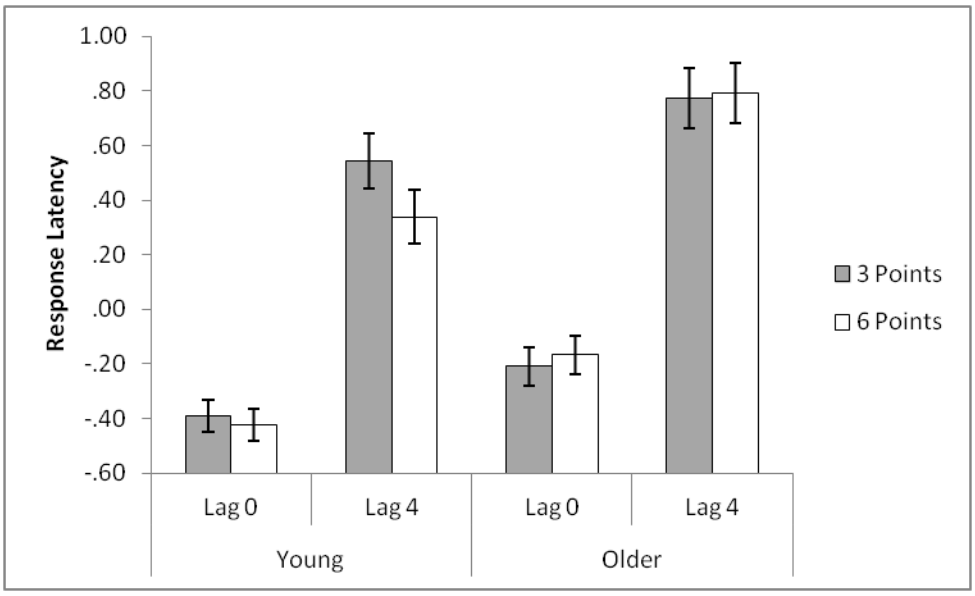
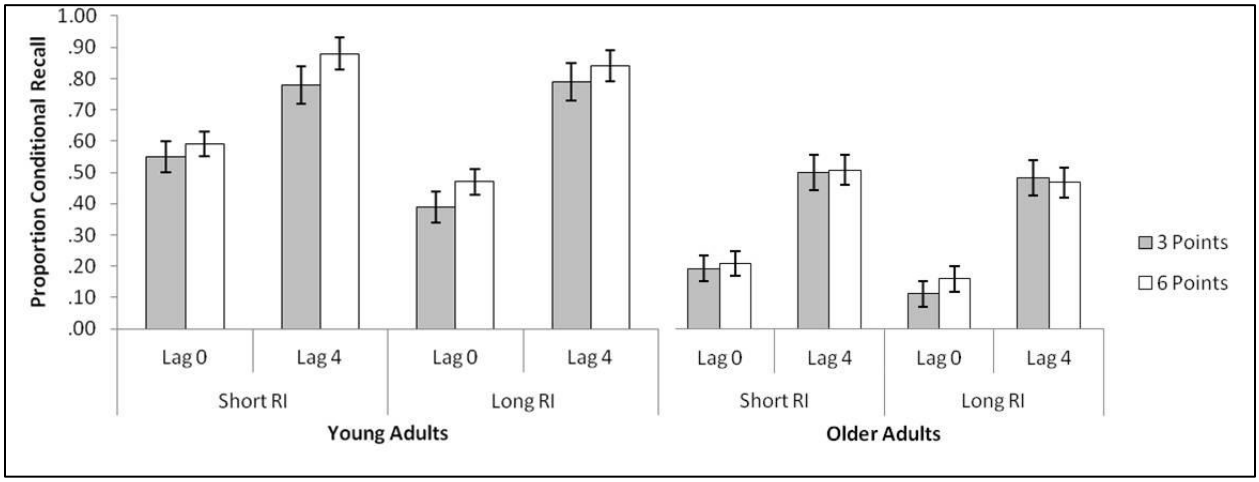


Figure 18. Mean proportion conditional cued recall on the final test in Experiment 3 as a function of age, retention interval, lag, and point value. Error bars represent ± 1 S.E.M.



Results of a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 2 (Point Value) mixed-factor ANOVA revealed main effects of Age and Lag, *p*s < .001, which reflected greater retention by young adults compared to older adults (*M* = .66 vs. .41, respectively) and greater retention in the Lag 4 condition compared to the Lag 0 condition (*M* = .71 vs. .36, respectively). The effect of Point Value was marginally significant (*p* = .061) which reflected greater retention in the high-value condition than in the low-value condition (*M* = .56 vs. .51, respectively). As shown in Figure 19, it appears as if the young adults produced an effect of point value in conditional recall performance, but older adults did not produce a similar effect. Although the interaction between age and point value was clearly not reliable, $F = 1.38$, $p = .244$, follow-up analyses confirmed that the young adults produced a reliable effect of point value, $t(59) = 2.63$, $p = .011$, with little effect in older adults, $p > .65$.

Conditional Standardized Response Latency. Mean conditional standardized response latency on the final cued recall test is presented in Figure 20 as a function of age, retention interval, lag and point value. There are three critical observations to note in the figure. First, a lag effect was obtained in young adult response latency but not older adult response latency. Second, mean response latency was similar across point values in the Lag 4 condition following both retention intervals for young and older adults. Third, mean response latency in the Lag 0 condition was faster for high-value items following the short retention interval but faster for low-value items following the long retention interval.

Results from a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 2 (Point Value) mixed-factor ANOVA revealed main effects of age, retention interval, and lag, *p*s < .005. In addition to these main effects, there was a reliable Age x Lag interaction, $F(1, 104) = 24.24$, $p < .001$, $\eta_p^2 = .19$, which indicated a lag effect in response latency for young adults ($M = .42$, $p < .001$) but no lag

Figure 19. Mean conditional performance on the final cued recall test in Experiment 3 as a function of age and point value. Error bars represent ± 1 S.E.M.

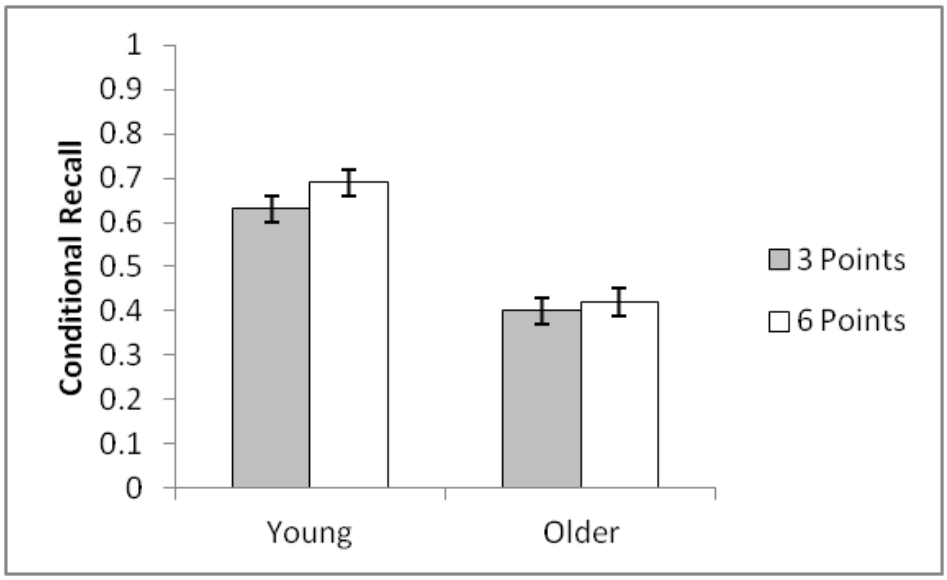
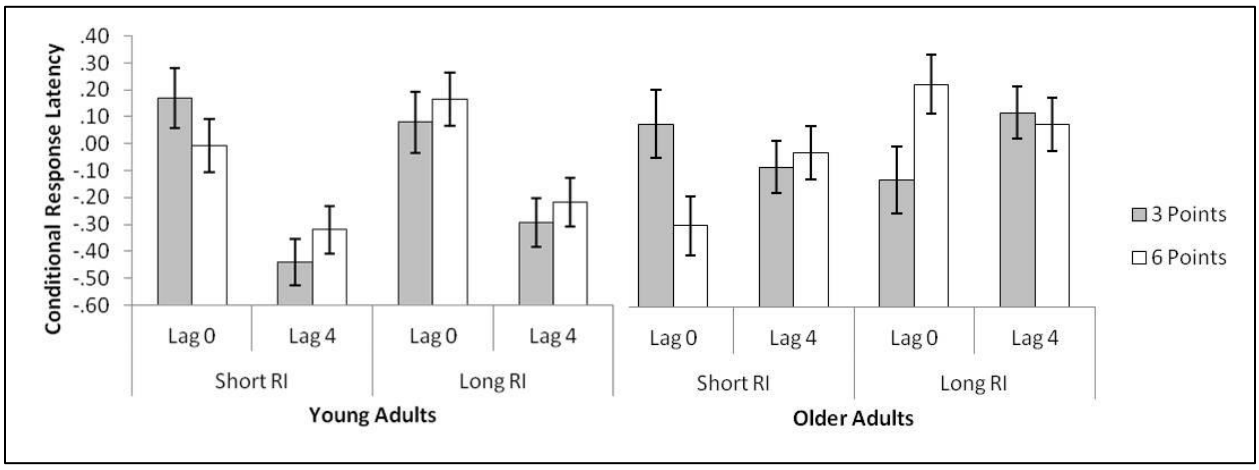


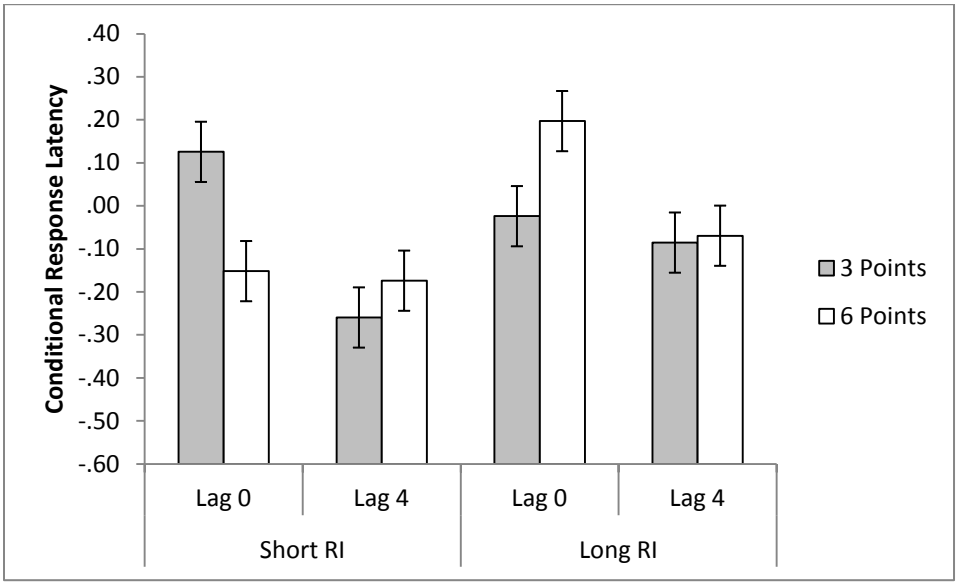
Figure 20. Mean conditional standardized response latency on the final cued recall test in Experiment 3 as a function of age, retention interval, lag, and point value. Error bars represent ±1 S.E.M.



effect for older adults (M = .05, p > .60).

The Retention Interval x Point Value interaction was also significant, p = .036, which was further qualified by a significant Retention Interval x Lag x Point Value interaction, F(1, 104) = 4.75, p = .031, η_p^2 =.04. To further explore the three-way interaction (see Figure 21) separate Retention Interval x Point Value ANOVAs were conducted for each lag condition. Analysis of Lag 0 response latency revealed a significant Retention Interval x Point Value interaction, F(1, 106) = 7.10, p = .009, η_p^2 = .06, which reflected faster response latency for high versus low value items following a short retention interval that reversed following the longer retention interval. Follow-up comparisons in the Lag 0 condition revealed a significant difference between high and low point value conditions in standardized response latency following the short retention interval, t(53) = 2.18, p = .034, and a marginally significant difference between point value conditions following the long retention interval, t(53) = 1.61, p = .114. In contrast, analysis of Lag 4 response latency revealed a significant effect of retention interval, F(1, 106) = 4.94, p = .028, η_p^2 = .04, but no influence of point value (ps > .45). Although this interaction was unexpected, it is possible that this pattern reflects differences in acquisition retrieval difficulty. Pairs that were successfully retrieved following a four item interval may have a well-established retrieval route that is easily accessed on the final test regardless of point value. In the massed condition, however, high point values may influence the quantity of rehearsal but not necessarily the quality of rehearsal (e.g., shallow versus elaborative processing). In turn, increased rehearsal for high-value, massed items may result in increased accessibility shortly after learning that decreases across time. Although this is an intriguing pattern of results, the interaction may be spurious and the complexity of this interaction clearly demands replication in the future.

Figure 21. Retention Interval x Lag x Point Value interaction in Experiment 3 response latency. Error bars represent ± 1 S.E.M.



Intact vs. Rearranged Recognition Accuracy. Given that the recognition test always occurred following the cued recall test, one may be concerned that recognition performance was influenced by performance on the prior test. Although this may be the case for young adults provided that analysis of their cued recall performance revealed an effect of point value as well as a level of nonconditional performance above floor (see Appendix E), older adult performance failed to reveal a significant effect of point value, and more importantly, older adult nonconditional final cued recall performance was near floor. Thus, the influence of the cued recall test on the subsequent recognition test performance should be minimal for this group of participants.

Mean proportion hits (i.e., correctly calling an intact pair "intact") and false alarms (i.e., incorrectly calling a rearranged pair "intact") are presented in Table 4 in addition to measures of recognition discriminability (d'; a measure of accuracy) and criterion (C; a measure of response bias) as a function of age, retention interval, lag, and point value. Analyses will emphasize signal-detection measures (d' and C), because analysis of a corrected recognition measure (Proportion Hits minus Proportion False Alarms) was generally consistent with the d' analysis.

With respect to d', there are several notable observations in this table. Consistent with expectations, d' was higher for young adults than for older adults and was higher following the short retention interval compared to long retention interval. Discriminability was also better for spaced stimuli compared to massed stimuli and for high-value items compared to low-value items. Discriminability (d') was submitted to a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 2 (Point Value) mixed-factor ANOVA. All main effects were significant. As expected, d' was higher for young adults than older adults (M = 1.70 vs. .78, respectively; $p < .001$) and was higher following the short retention interval compared to the long retention interval (M = 1.43 vs.

Table 4. Mean proportions (and standard errors) of hits and false alarms (FAs), with d' and Criterion, as a function of age, retention interval, lag and point value.

		Young				Older			
		Lag 0		Lag 4		Lag 0		Lag 4	
		3 Points	6 Points	3 Points	6 Points	3 Points	6 Points	3 Points	6 Points
Short RI	**Hits**	.45 (.02)	.46 (.02)	.43 (.02)	.45 (.02)	.39 (.02)	.36 (.02)	.35 (.03)	.38 (.02)
	FAs	.05 (.02)	.05 (.02)	.05 (.02)	.04 (.02)	.15 (.02)	.13 (.02)	.10 (.02)	.09 (.02)
	d'	1.82 (.14)	1.74 (.13)	1.68 (.15)	1.94 (.16)	.99 (.16)	1.02 (.15)	1.04 (.17)	1.21 (.18)
	C	1.03 (.07)	.97 (.07)	1.04 (.07)	1.11 (.06)	.80 (.08)	.88 (.07)	1.01 (.08)	.98 (.07)
Long RI	**Hits**	.41 (.02)	.48 (.02)	.42 (.02)	.43 (.02)	.33 (.04)	.34 (.03)	.29 (.04)	.36 (.04)
	FAs	.09 (.02)	.08 (.02)	.07 (.02)	.06 (.02)	.28 (.04)	.23 (.03)	.16 (.03)	.15 (.03)
	d'	1.43 (.14)	1.71 (.13)	1.57 (.15)	1.67 (.16)	.06 (.25)	.45 (.23)	.59 (.26)	.89 (.28)
	C	.95 (.08)	.92 (.07)	1.00 (.07)	1.02 (.06)	.61 (.12)	.66 (.11)	.88 (.12)	.93 (.11)

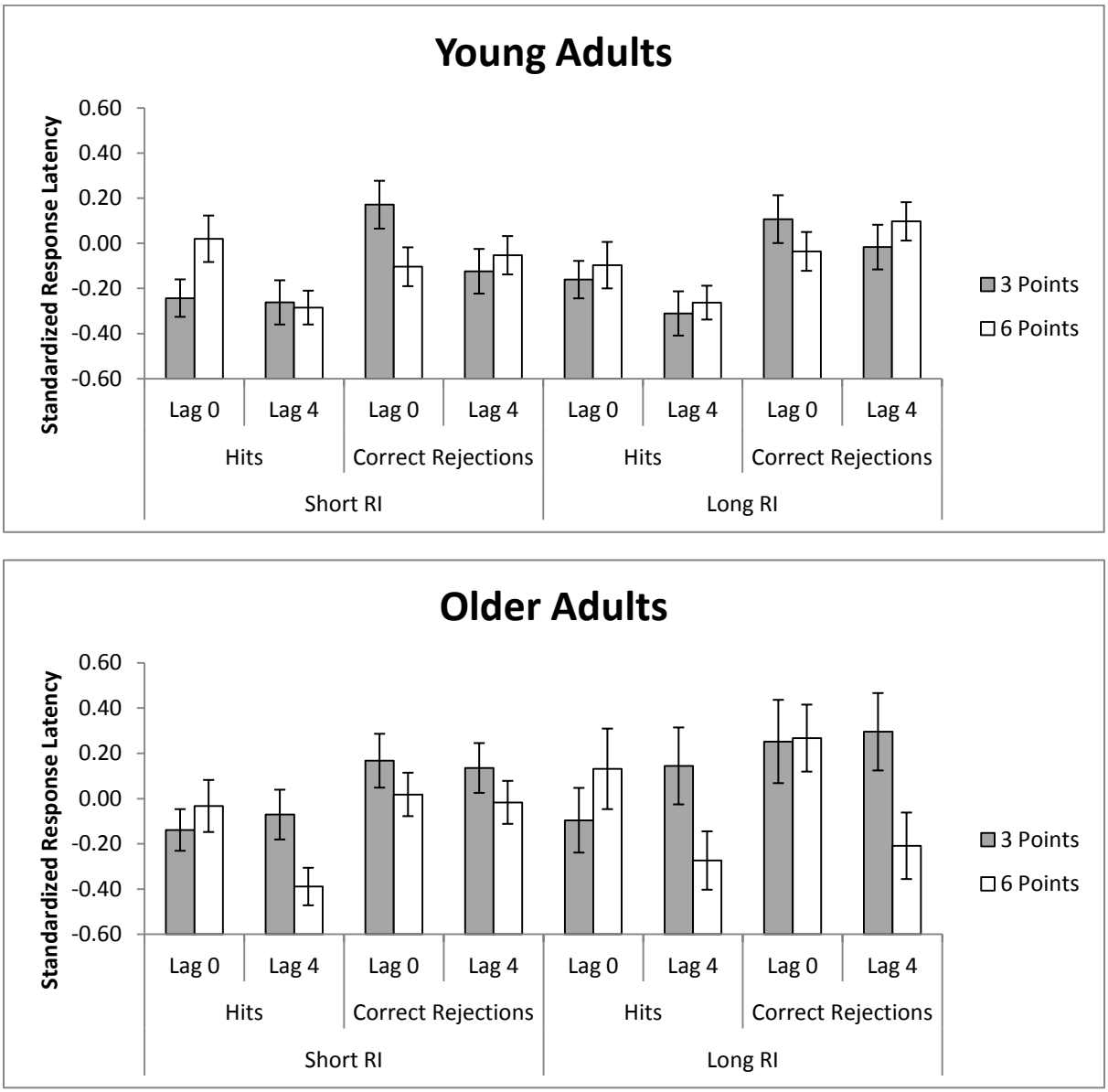
1.05, respectively; $p = .003$). With regard to the influence of spacing on final recognition performance, the main effect of lag, $F(1, 90) = 4.34$, $p = .043$, $\eta_p^2 = .05$, reflected higher d' for pairs in the Lag 4 condition compared to the Lag 0 condition ($M = 1.32$ vs. 1.15, respectively). Finally, d' was higher for high-value pairs ($M = 1.33$) than low-value pairs ($M = 1.15$), $F(1, 90) = 5.90$, $p = .017$, $\eta_p^2 = .06$.

Turning to criterion, there are two observations to note in Table 4. First, older adults set a more liberal criterion (as indicated by lower values) than young adults. Second, criterion was lower for Lag 0 items than for Lag 4 items. Data were submitted to a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 2 (Point Value) mixed-factor ANOVA which revealed main effects of age, $F(1, 90) = 7.99$, $p = .006$, $\eta_p^2 = .08$, and lag, $F(1, 90) = 13.81$, $p < .001$, $\eta_p^2 = .13$.

Intact vs. Rearranged Recognition Response latency. To examine retrieval fluency on the recognition test, response latency from correct trials (i.e., Hits and Correct Rejections) was submitted to a 2 (Age) x 2 (Retention Interval) x 2 (Trial Type) x 2 (Lag) x 2 (Point Value) ANOVA. Mean response latency is presented in Figure 22 as a function of Age, Retention Interval, Trial Type, Lag and Point Value. There are three observations to note in this figure. First, response latency was faster for hits than for correct rejections. Second, although a lag effect was generally observed in response latency across point value conditions for young adults, it appears that a lag effect was only observed in the high point value condition for older adults. Third, it appears that the lag effect observed on Hit trials was driven by high-value items, whereas the lag effect observed on Correct Rejection trials was driven by low-value items.

Results revealed main effects of Age, Retention Interval, Trial Type, and Lag, ps $< .05$. Additionally the Age x Point Value interaction was significant, $F(1, 90) = 4.50$, $p = .037$, $\eta^2_p = .05$, and was further qualified by a significant Age x Lag x Point Value interaction, $F(1, 90) =$

Figure 22. Mean standardized response latency for hits and correct rejections on the Intact versus Rearranged Recognition test as a function of lag and point value for young (top panel) and older adults (bottom panel). Error bars represent ± 1 S.E.M.



6.74, $p = .011$, $\eta_p^2 = .07$. To further explore this interaction (see Figure 23), separate Lag x Point Value ANOVAs were conducted for young and older adults. Analysis of young adult response latency revealed a single main effect of Lag, $F(1, 59) = 5.03$, $p = .029$, $\eta_p^2 = .08$, with no interaction between point value and lag. Analysis of older adult response latency revealed a main effect of point value, $p = .029$, which was further qualified by a significant Lag x Point Value interaction, $F(1, 33) = 5.21$, $p = .029$, $\eta_p^2 = .14$, that reflected a significant effect of point value in the Lag 4 condition, $t(33) = 3.09$, $p = .004$ but no difference in Lag 0 response latency, $p > .80$.

The ANOVA also yielded a reliable Trial Type x Lag x Point Value interaction, $F(1, 90) = 5.17$, $p = .025$, $\eta_p^2 = .05$. To further explore this interaction (see Figure 24), separate Lag x Point Value ANOVAs were conducted for each trial type. Analysis of Hits revealed a main effect of lag, $p = .002$, that was further qualified by a significant Lag x Point Value interaction, $F(1, 93) = 6.08$, $p = .016$, $\eta_p^2 = .06$. There was no difference in response latency across lag conditions for low point value items, $p > .85$, but response latency was significantly faster for high point value items that were separated by Lag 4 compared to Lag 0 (M = -.30 vs. -.02, respectively), $t(93) = 3.90$, $p < .001$. Analysis of Correct Rejections yielded a marginal effect of point value, $F(1, 93) = 3.50$, $p = .065$, $\eta_p^2 = .04$, which reflected faster response latency in the high-value condition compared to the low-value condition (M = -.01 vs. .09, respectively).

Working Memory, Retrieval Practice and Value-Directed Encoding.

A working memory composite score was created in the same way described for Experiments 1 and 2. Again, each of the analyses reported above were replicated with Working Memory Group as a between-participants factor and age (in years) serving as a covariate.

Acquisition Performance. Analysis of acquisition accuracy revealed a main effect of

Figure 23. Age x Lag x Point Value interaction for Intact versus Rearranged recognition response latency in Experiment 3. Error bars represent ± 1 S.E.M.

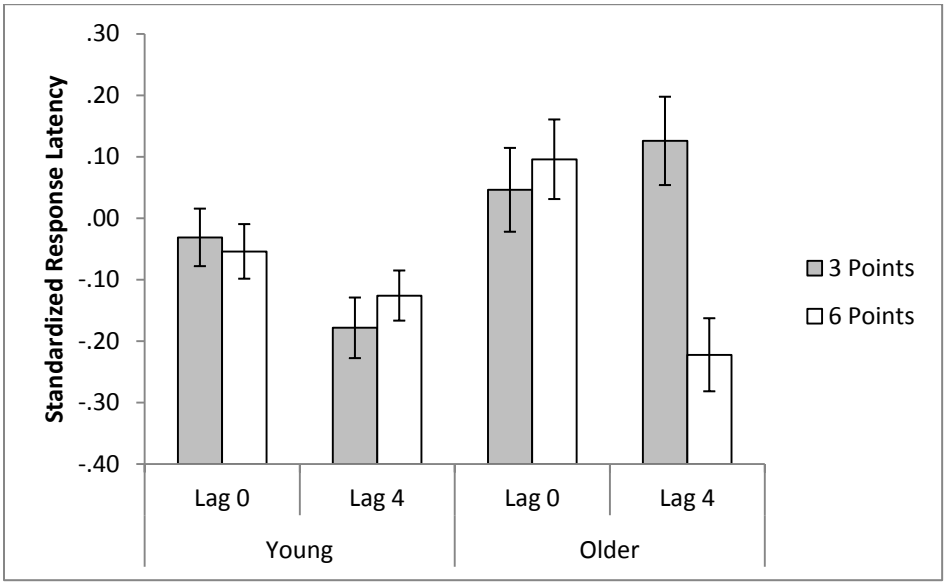
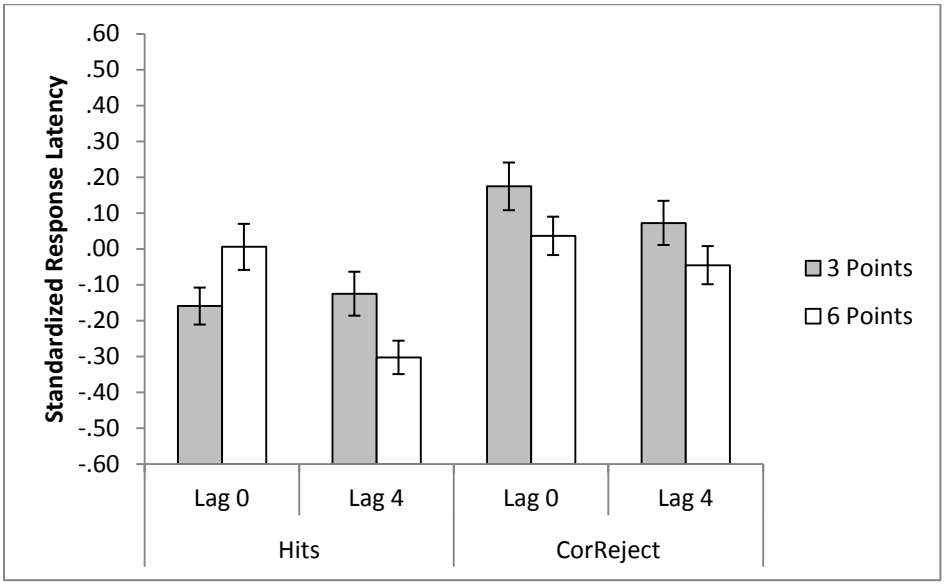


Figure 24. Trial Type x Lag x Point Value interaction for Intact versus Rearranged Recognition response latency in Experiment 3. Error bars represent ± 1 S.E.M.



WMC Group, $F(1, 103) = 5.12$, $p = .026$, $\eta_p^2 = .05$, and a significant WMC Group x Lag interaction, $F(1, 103) = 4.20$, $p = .043$, $\eta_p^2 = .04$. As seen in Figure 25, the significant interaction reflected a larger difference between Lag 0 and Lag 4 performance for the low WMC group compared to the high WMC group ($M = .45$ vs. .38, respectively). Analysis of acquisition response latency failed to yield a significant effect or any significant interactions involving WMC group, ps > .67.

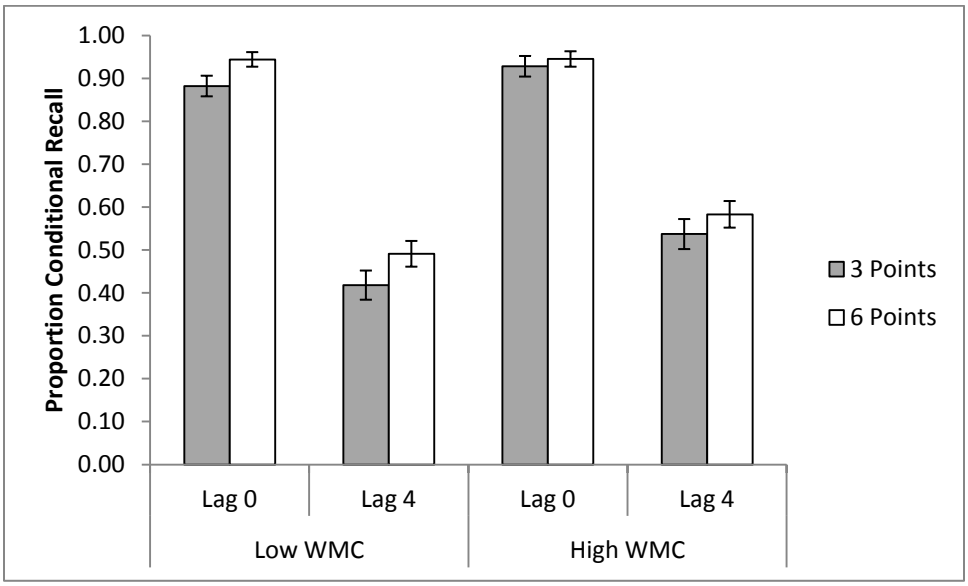
Final Test Performance. Analysis of conditional final test accuracy revealed a marginally significant WMC Group x Retention Interval x Lag interaction, $F(1, 101) = 3.82$, $p = .053$, $\eta_p^2 = .04$. This interaction reflected a similar size spacing effect across retention interval conditions for the high WMC group (Ms = .34) and an increase in the size of the spacing effect from the short to long retention interval for the low WMC group ($M = .27$ and $M = .44$, respectively). Analysis of conditional final test response latency failed to yield a significant effect or any significant interactions involving WMC group, ps > .19.

Intact vs. Rearranged Recognition. Analysis of d' and C failed to reveal any significant main effects of WMC group. However, a 2 (WMC Group) x 2 (Retention Interval) x 2 (Trial Type) x 2 (Lag) x 2 (Point Value) mixed-factor ANOVA of standardized response latency yielded two significant interactions. Both the WMC Group x Retention Interval x Lag x Point Value interaction, $F(1, 87) = 4.40$, $p = .039$, $\eta_p^2 = .05$, and the WMC Group x Retention Interval x Trial Type x Lag interaction were significant, $F(1, 87) = 6.89$, $p = .010$, $\eta_p^2 = .07$. Follow-up analyses for each interaction failed to yield any systematic patterns, and so will not be further discussed.

Discussion

The results from Experiment 3 yielded four critical findings. First, the manipulation of

Figure 25. Acquisition accuracy as a function of WMC group, Lag and Point Value in Experiment 3. Error bars represent ± 1 S.E.M.



the participant's motivation during the acquisition phase via value-directed encoding extends to a paradigm that differed from past studies in several important ways. The current study included word pairs rather than single words and used a slower presentation rate (4.5 seconds rather than 1 second per item). Moreover, cued recall results revealed a benefit of value-directed encoding that persisted across short and longer retention intervals (30 second and 15 minute retention intervals for young adults but not older adults). Most importantly, the benefit of value-directed encoding persists when stimuli receive retrieval practice with varying lags. This is consistent with the possibility that high-value pairs receive enhanced encoding (e.g., more rehearsal) and are consequently retrieved with more success following various lags compared to low-value pairs.

Second, because performance for items was assessed during the learning phase, the extent to which value-directed encoding influences acquisition performance and retention could be examined, via the conditional analyses that targeted final recall performance only for those items that were successfully retrieved during encoding. In fact, current analyses of final test accuracy suggest that manipulation of point value during the acquisition phase differentially influences the encoding of information but exerts a relatively smaller influence on retention as indicated by the marginal effect of point value in conditional cued recall. Of course, this finding is inconsistent with predictions based on different levels of retrieval difficulty that result from the manipulation of point value. Specifically, if higher point value items are more likely to be maintained between initial encoding and retrieval during acquisition then, based on the desirable difficulty hypothesis, one might actually predict that final test conditional performance should be lower for these items. However, this was not the case in the current study. Further consideration of this pattern will be provided below in the General Discussion.

Third, a benefit of value-directed encoding was observed on the intact versus rearranged recognition test across short and long retention intervals for both young and older adults. This finding is important for several reasons. Results from the cued recall test failed to show a benefit of value-directed encoding for older adults which is inconsistent with past research on value-directed encoding (e.g., Castel et al., 2012). Of course, the lack of an effect in cued recall performance in the current paradigm may reflect the relatively more demanding nature of the stimuli compared to previous value-directed encoding studies. Specifically, the shift from single items to paired associates may disproportionately reduce cued recall performance for older adults compared to young adults (e.g., Naveh-Benjamin, 2000), which in turn may limit the ability to detect a value-directed encoding effect in the former group. However, when additional experimental support was provided in the form of a recognition test (see also Craik & McDowd, 1987), older adult performance revealed a significant benefit in d' for high-value items over low-value items.

Of course, one may be concerned that performance on the recognition test was confounded with test order (i.e., the recognition test always followed the cued recall test). However, if a prior cued recall test served to exaggerate the effects obtained in recognition, then a value-directed encoding effect should not have been obtained on the recognition test for older adults. This clearly was not observed in the current results.

Finally, value-directed encoding had a clearer influence on accuracy than on response latency. One possibility for this finding is that the effect observed in accuracy was relatively small compared to other factors (e.g., lag). Thus, one may not expect a large effect of point value on retrieval fluency (i.e., response latency).

General Discussion

This dissertation addressed three questions regarding the benefits of spaced retrieval practice across age groups and retention intervals. First, how does lag modulate the extent to which continued testing improves long-term memory? Second, how does the function relating lag and continued testing to final test performance differ across young and older adults? Third, how does participant motivation modulate the benefit of retrieval practice in young and older adults? Before turning to a discussion of these issues, it is important to note that the present methods diverged from standard approaches in this literature in two important ways.

First, emphasis was placed on analysis of conditional final test performance (rather than overall performance) to minimize encoding confounds when assessing the influence of lag and age on long-term memory. Specifically, as noted earlier, if an item cannot be retrieved during the acquisition phase (e.g., due to age-related declines in episodic memory and/or a long lag during acquisition), then the item will not incur retrieval practice, and so one cannot directly measure the benefit of retrieval practice for those items or for other related items presented in the study list (i.e., test induced facilitation for non-tested material related to the tested material; e.g., Chan, McDermott, & Roediger, 2006; Szpunar, McDermott, & Roediger, 2008). As noted in the Introduction, the use of conditional analyses may lead to item selection effects in final test performance. However, it is comforting that conditional analyses generally accorded with nonconditional final test analyses, which suggests that the influence of item selection effects on the interpretation of conditional final test performance may be of minimal consequence in the current experiments. Nevertheless, subsequent studies may wish to implement a methodology which yields similar levels of performance during the acquisition phase (e.g., near perfect performance) to avoid this concern in the future.

Second, the dissertation diverged from previous spaced retrieval studies in that it examined the influence of age, spacing, and testing on final test response latency as well as conditional accuracy. Although one might expect that measures of accuracy and response latency will produce similar patterns of results as a function of experimental manipulations (e.g., lag and number of retrieval attempts), these two measures may reflect the strength of the memory trace in different ways. If one assumes that the integrity of the memory trace is reflected by a continuous measure of strength, accuracy will reflect discrete states in which the item is or is not above some response threshold. In contrast, response latency may capture variation in trace strength above and beyond the critical threshold at which the item can be correctly recalled, because response latency is a continuous dependent variable. With these methodological extensions in mind, I shall turn to each of the goals of the dissertation by first considering Experiments 1 and 2 before turning to Experiment 3.

Aim 1: Retrieval Practice as a Function of Lag.

The first aim of the dissertation was to examine the function relating continued testing and lag to final test performance. Here I will focus on the results from Experiment 1, and I will discuss the related spacing manipulation in Experiment 2 in a later section. The first experiment included two manipulations to address this aim. First, the lag between study and testing events (Lag 1 vs. Lag 3) was experimentally manipulated during acquisition. Second, continued testing was examined at three levels (1 vs. 3 vs. 5 tests) to more fully capture the function relating continued testing and lag to final test performance.

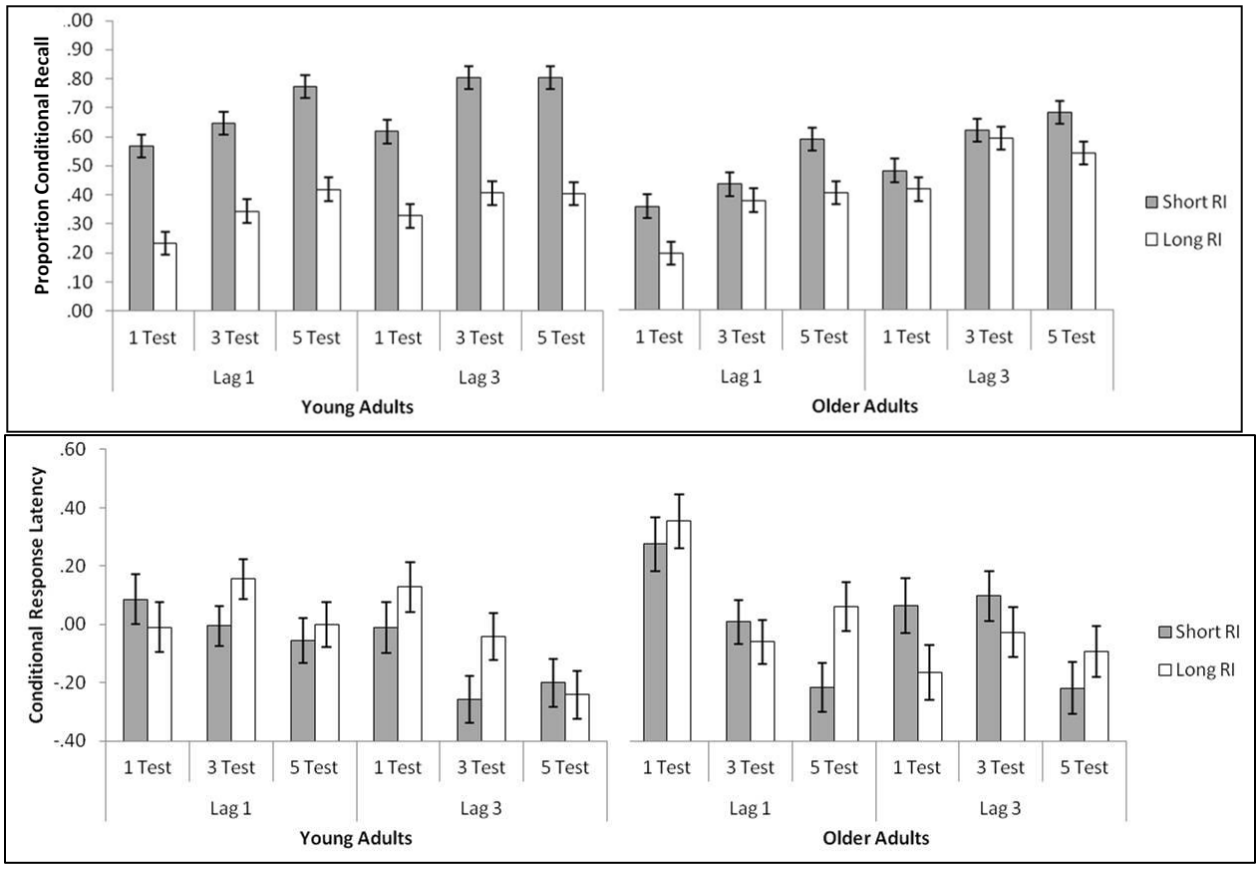
Similar to previous studies (e.g., Karpicke & Roediger, 2007; Wheeler & Roediger, 1992), Experiment 1 revealed a long-term retention benefit with increased testing when comparing a single test condition to a three test condition in both the Lag 1 condition (11%

benefit) and Lag 3 condition (15% benefit). More importantly, the inclusion of an additional level of testing revealed a difference in the function relating continued testing and lag to final test performance (see top panel of Figure 26). Specifically, retention continued to increase with additional retrieval practice in the short lag condition (10% from 3- to 5-tests) but did not increase in the long lag condition (0% from 3- to 5-tests). Thus, the benefits of additional retrieval practice appear to asymptote after three successful retrieval events in the long lag condition but not in the short lag condition. Although this pattern in accuracy occurred for both young and older adults, there were differences across age groups with respect to the influence of testing and lag on response latency. Hence, I now turn to Aim 2 regarding age-related changes in the influence of lag and retrieval practice on long-term retention.

Aim 2: Retrieval Practice, Lag and Age

Turning to the second aim of examining age differences in the function relating lag and continued testing to final test performance, there are two intriguing aspects of both Experiment 1 and 2. First consider Experiment 1 which examined the function relating lag and continued testing to final test performance across age groups and retention intervals. As noted earlier, in the top panel of Figure 26, this function was generally similar across age groups and retention interval conditions in terms of conditional accuracy, i.e., both groups appeared to asymptote after three retrieval attempts in the long lag condition but not in the short lag condition. The similarity in functions relating lag and number of tests to conditional accuracy on the final test stands in contrast with past studies that have indicated age differences in the optimal spacing schedules (e.g., Maddox et al, 2011). This again highlights the importance of conditional analyses in the current dissertation and the need to account for differences in acquisition performance across age groups and lag conditions when considering final test performance.

Figure 26. Proportion conditional accuracy (top panel) and standardized response latency (bottom panel) on the final test in Experiment 1 as a function of age, retention interval, lag, and number of tests. Error bars representation ±1 S.E.M.



In contrast to the accuracy data, the function relating continued testing and lag to response latency on the final test differed across young and older adults (see the bottom panel of Figure 26). Young adult response latency did not change as a function of testing in the short lag condition but was facilitated with increased testing in the long lag condition. However, older adult response latency benefited from increased testing in the short lag condition but did not change as a function of number of tests in the long lag condition. Because these patterns were observed for items that were successfully retrieved on the final test, it is important to consider how age-related differences in episodic memory may influence the speed with which these items are accessed.

It is intriguing to consider why older adults benefit more (i.e., in terms of reduced response latency) from continued testing at the short lag, whereas, young adults benefit more from continued testing at the long lag. One possibility is that response latency on the final test may reflect the integrity or strength of the retrieval route used to access the item in memory. In turn, the age differences relating lag and number of tests to final test response latency may reflect the degree to which an effective retrieval route was established and assessed during the learning phase. Specifically, young adults, whose encoding abilities are superior to those of older adults, may initially encode material well enough that enhanced retrieval fluency (as indicated by response latency) is only obtained with repeated practice in the more difficult long lag condition. In this sense, repeated retrieval across a greater number of intervening trials is a better indicator of the efficiency of the retrieval route (i.e., an item can be retrieved following more variable interference or in more variable contexts when retrieved repeatedly in the long lag condition; see Estes, 1955a; 1955b; Glenberg, 1976). In contrast, due to older adults' relatively poorer encoding abilities, successful retrieval of an item following a short lag during encoding suggests

that it is accessible but does not necessarily reflect that a robust retrieval route has been established. Thus, continued practice with the item further strengthens the item in terms of overall accuracy and also helps to establish a more efficient retrieval route as indicated by changes in response latency (both during acquisition and on the final test). Changes in response latency would not necessarily be observed with additional testing in the long lag condition given that the retrieval route initially established for those items was effective enough to ensure successful retrieval following a longer lag. Obviously, this account is post hoc and needs further examination.

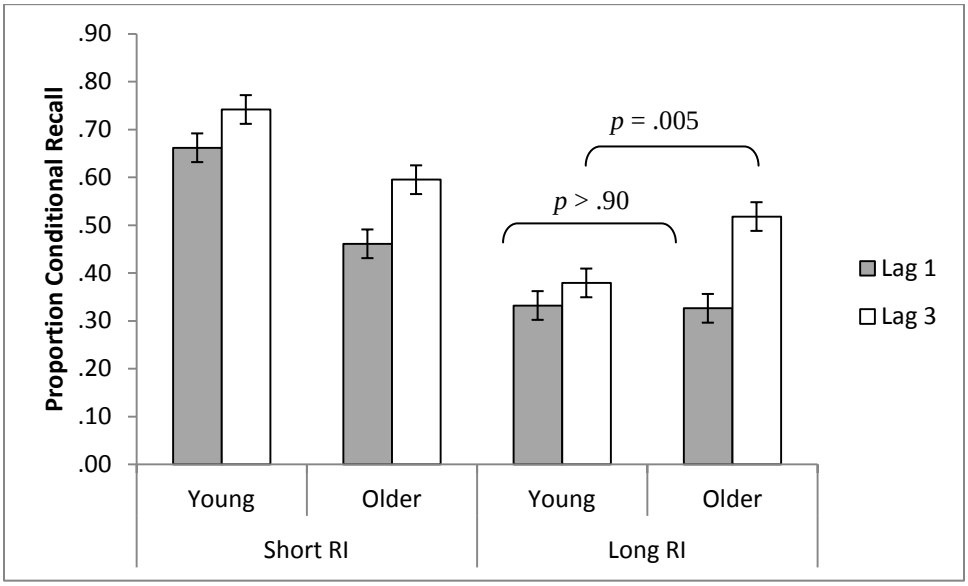
There is a second interesting aspect of Experiment 1 that is important to consider at this point. Specifically, although there was a very similar pattern relating lag and retrieval practice across age groups, there was also a reliable Age x Lag interaction which reflected a larger lag effect in conditional accuracy for older adults than for young adults. This interaction was primarily driven by the long retention interval condition. Although the three-way interaction between Age, Retention Interval, and Lag did not reach significance ($F = 2.64$, $p = .106$), the pattern is noteworthy. As shown in Figure 27, a similar pattern was observed across both age groups following a short retention interval, whereas a larger lag effect was observed for older adults compared to young adults following a long retention interval. Indeed, analysis of short retention interval performance did not yield a reliable Age by Lag interaction, $p > .15$, whereas, the Age by Lag interaction was highly reliable following the long retention interval, $p = .001$.

The pattern of results at the long retention interval in Figure 27 is particularly intriguing. As shown, the young and older adults were equated in the short lag condition, but the older adults actually produced better performance than the young adults in the long lag condition. At this level, it appears that older adults actually benefit more from the Lag 3 condition in terms of

conditional accuracy, when age groups are equated at the Lag 1 condition. It is tempting to conclude that this increased lag effect in older adults may be due to age-related differences in retrieval effort and desirable difficulties (Bjork, 1994). Specifically, because older adults have lower performance during the long lag condition during encoding, those items that did survive may have benefited from more desirable difficulty in the older adult group than the young adult group, and hence, may have produced a stronger long-term trace. Of course, one must be cautious not to overinterpret these results, because the young and older adults had different long retention intervals in Experiment 1. Given the evidence of a nonmonotonic lag by retention interval interaction in the literature (see Crowder, 1976, Chapter 9 for a discussion), these results may be due to idiosyncratic points in this function for the young and older adults. It is also possible that the different retention intervals used across age groups captured different relationships between lag and retention interval. Specifically, research by Cepeda et al. (2009) suggests that the optimal lag between repetitions is a decreasing proportion of the retention interval with the optimal ratio ranging from 1.0 when the final test is administered after short delays to 0.10 at very long delays. With respect to the current experiments, the optimal ratio occurs in the short RI condition for both age groups (i.e., the ratio for Lag 3 items to the RI approaches the .10 ratio, whereas all other ratios are substantially lower than this value). In terms of the age-related differences observed in long retention interval performance between young adults and older adults, it is possible that the older adult ratio is more optimal than the young adult ratio. In either case, this pattern is unique in the aging literature and deserves further exploration.

Turning to Experiment 2, an interesting dissociation between conditional accuracy and response latency was observed when comparing the benefit of continued massed and spaced

Figure 27. Proportion conditional accuracy on the final test in Experiment 1 as a function of age, retention interval and lag. Error bars representation ± 1 S.E.M.



retrieval practice. As a reminder, this experiment investigated massed versus spaced retrieval, instead of spaced conditions with various lags as was examined in Experiment 1. The results indicated that although the spacing effect was of similar magnitude across age groups (as was the case following the short retention interval in Experiment 1), young and older adults benefited from continued testing in different ways as a function of lag. Specifically, only the young adults benefited from additional testing in the massed condition (possibly due to refreshing), whereas, both groups benefited from continued testing in the spaced condition. Interestingly, however, response latency in the massed condition was facilitated with increased testing for older adults but not young adults. Thus, a trade-off in the benefit of refreshing was observed between conditional accuracy and response latency for both age groups such that additional massed retrieval practice benefited young adults in terms of accuracy but not response latency, whereas, additional massed retrieval benefited older adults in terms of response latency but not accuracy.

Although the current results are useful for examining the benefits of spaced retrieval for items that successfully incurred retrieval practice during encoding, an emphasis on conditional analyses overlooks overall differences between age groups and lag conditions in performance during acquisition. Indeed, the benefit of a long lag during acquisition is offset by reduced acquisition performance. Thus, future studies may extend recent work reported by Rawson and Dunlosky (2011) to an older adult population as a means of examining the benefits of criterion level learning and the benefits of continued testing with feedback in this group.

Continued Retrieval Practice and Desirable Difficulty. As discussed in the Introduction, the benefits of various spaced retrieval schedules have been tied to the degree to which a given schedule produces desirably difficult retrieval (e.g., Bjork, 1994) during the learning phase. Based on Bjork's concept of desirable difficulty, it was originally predicted that

continued testing would improve retention in the long lag condition but would produce relatively no improvement in the short lag condition given that longer lags should lead to more difficult retrieval attempts than short lags. Clearly, this pattern of data was not observed in either Experiment 1 or Experiment 2. Of course, this prediction from the desirable difficulty perspective assumes that each retrieval attempt in the Lag 3 condition was more difficult than in the Lag 1 condition. But, how does one measure desirable difficulty? The present experiments afforded a measure of desirable difficulty during encoding, i.e., response latency, and hence can provide some direct evaluation of this prediction. Based on the response latency during acquisition, the initial desirable difficulty prediction may have been incorrect regarding the benefits of additional testing as a function of lag. Specifically, items that are more difficult to retrieve initially should benefit to a greater extent from retrieval than less difficult items in terms of strengthening of the memory trace, and as a result, subsequent retrieval attempts may actually be *faster and easier* in the nominally "more difficult" lag condition relative to the "less difficult" lag condition. Two approaches to examining the role of desirable difficulty in producing the benefit of spaced retrieval will be considered next.

First, one may expect difficulty on the first retrieval attempt to influence long-term retention given past research suggesting a long initial lag produces increased long term memory versus a short initial lag regardless of subsequent form of spacing (i.e., equal spaced vs. expanding retrieval; Karpicke & Roediger, 2007). Analysis of standardized response latency on the first retrieval attempt in each condition of Experiment 1 revealed that the three-test condition produced slower response latency than the one-test condition. Additionally, a significant Age x Lag interaction reflected similar response latency across lag conditions for young adults but faster response latency in the Lag 1 condition than the Lag 3 condition for older adults. Thus,

one would predict a benefit in conditional accuracy for the three-test condition compared to the one-test condition, and that the lag effect should be larger for older adults than young adults.

Turning to the conditional final test results, it appears that conditional accuracy generally accords with predictions based on acquisition response latency. First, the three test conditions produced a benefit over the one test conditions for both Lag 1 and Lag 3. The one exception to this prediction is that taking five tests in the Lag 1 condition produced significantly better performance on the final test than taking three tests. Of course, this increase may reflect the benefit in final test performance obtained from additional retrieval rather than shifts in desirable difficulty. Thus, the influence of repeated exposure to material via testing may compensate for less effective spacing intervals. The second prediction from this perspective is that one should find a larger lag effect for older adults than young adults in Experiment 1, which as described in detail above, was indeed observed. Thus, there is some evidence in support of the desirable difficulty account in the current results.

Alternatively, one may expect that more difficult retrieval attempts soon after encoding an item will lead to overall greater speeding across later retrieval attempts. Thus, response latency on the final retrieval attempt may best capture the variability in retrieval difficulty across continued testing during acquisition.

In contrast with the preceding analysis which emphasized response latency for the first retrieval attempt at various lags, *faster* response latency on the final retrieval attempt during acquisition would indicate larger benefit from difficult, early retrieval attempts. Analysis of standardized response latency on the last retrieval attempt in each multiple-retrieval attempt condition of Experiment 1 revealed faster response latency in the Lag 3 condition compared to the Lag 1 condition and faster response latency in the five-test condition compared with the

three-test condition. With consideration for these differences in acquisition response latency, one would predict a significant lag effect in conditional final test accuracy and increased performance for items tested five times compared to three times. Indeed, final test accuracy was greater for Lag 3 items than Lag 1 items, and performance for items tested five times was greater than performance for items tested three times. However, this latter finding was only observed in the Lag 1 condition. Thus, although there are clearly aspects from the final test performance that are consistent with the predictions from the desirable difficult account (based on response latency data during acquisition), it is clear that other factors likely contribute to performance in the current paradigm. Future studies may wish to explore other mechanisms previously proposed to account for the spacing effect (e.g., encoding variability; Estes, 1955a; 1955b).

In summary, regarding the first and second aims, the present results indicate that continued testing in the short lag condition produces continued increases in terms of long-term accuracy, whereas continued testing in the long lag condition produces an initial increase in retention that approaches asymptote after three tests (e.g., Experiment 1). The results from Experiments 1 and 2 also demonstrate dissociations between accuracy and response latency which highlights the importance of considering multiple measures of memory performance when examining the benefits of lag and continued testing. Finally, the relation between the acquisition response latency and the final test performance are generally consistent with a desirable difficulty account of the combined effects of spacing and testing effect (e.g., Bjork, 1994).

In addition to experimenter-controlled factors that may modulate the benefit of continued testing, the participant's motivation to learn material was also examined in the dissertation. The results from Experiment 3 provide some insight into this third aim of the dissertation.

Aim 3: Retrieval Practice and Participant Motivation

The third aim of the dissertation was to examine the extent to which participant motivation influences the benefits of retrieval practice across age groups and retention intervals. In contrast to past value-directed encoding studies, Experiment 3 utilized paired associates that were assigned a low or high point value in a retrieval practice paradigm. It was predicted that retrieval difficulty during acquisition would be greater for low-value items than high-value items, which would produce a benefit in conditional final test performance for the former class of items over the latter class of items. Results yielded two critical findings. First, a benefit of high point value items was observed in long-term conditional memory performance as reflected in both cued recall performance as well as in intact versus rearranged recognition performance. However, the point-value effect observed in cued recall performance was only significant for young adults, whereas the benefit observed in recognition performance was significant for both age groups.

Second, the manipulation of point values within a retrieval practice paradigm provided the opportunity to examine the influence of value-directed encoding on acquisition performance versus retention of material across multiple retention intervals. Results revealed that point value exerts a stronger influence on the encoding of information, as reflected by the influence on acquisition performance, compared with the relatively weaker influence on final conditional test performance. Initial predictions regarding the influence of point value on final test performance emphasized the potential role of desirable difficulty in modulating the extent to which high versus low valued items were retained across the two retention intervals. Analysis of acquisition response latency failed to yield any differences across point values. Instead, there was a significant Age x Lag interaction which reflected a larger difference between spacing conditions for older adults than for young adults. However, a similar effect was not observed in final test

performance, which suggests that desirable difficulty may not be able to fully account for the effects observed in Experiment 3. Thus, it is important to consider the mechanism originally predicted to modulate retrieval difficulty as a function of point value, namely, deficient processing. As noted earlier, the deficient processing account of the spacing effect (e.g., Johnston & Uhl, 1976; Rundus, 1971; Shaughnessy, Zimmerman, & Underwood, 1972) suggests that spaced items are rehearsed at the detriment of rehearsing massed items when both conditions occur within the same list. For example, in the current paradigm massed items may be tagged as low or high-value which then results in the low value massed item being dropped from subsequent rehearsal. Thus, the influence of difficulty of retrieval during the learning phase on final test performance may be masked by the influence of additional rehearsal for high value, long lag items compared to low value, massed items.

Future studies may wish to extend the current paradigm to situations in which participants control the duration of study, the number of retrieval trials, and the spacing interval between learning and subsequent retrieval attempts (e.g., Maddox & Balota, 2012). The use of such a paradigm may be particularly interesting with respect to Experiment 3 and the benefits of value-directed encoding with older adults. Specifically, older adults failed to produce a point value effect on the more effortful cued recall test in the current paradigm, which may have reflected the influence of dividing attention in which older adults had difficulty with adequately encoding word pairs and processing the point value. Allowing participants to pace their study should provide the opportunity to establish and execute a strategy in which high value items are given more attention, more encoding and more retrieval practice than low value items, which in turn should produce a significant point value effect for older adults in more effortful retrieval tasks. Of course, individuals often do place different priorities on learning information (e.g.,

when an instructor says this is going to be on the exam, or the feedback from a doctor), and so it is important to develop a better understanding how an individual's perceived value modulates later memory performance.

Secondary Aim: Retrieval Practice and Working Memory

A secondary interest in the dissertation was the relationship between working memory capacity and the benefit of retrieval practice in terms of conditional final test performance. Across experiments, it was expected that individuals with low WMC would benefit more from the lag manipulation than individuals with high WMC given between-group differences in the sensitivity to the consequences of various lags (e.g., interference). Indeed, past research suggests that low and high WMC groups show optimum benefit in free recall performance from differing levels of task demand when repetitions are spaced by easy versus difficult tasks (Bui, Maddox, & Balota, 2012). Despite some variability in the patterns of acquisition and final test results across experiments, none of the experiments produced the predicted WMC Group x Lag interaction in final conditional test performance. However, there were some hints of this interaction in nonconditional performance (see Appendices D and F).

One possibility for why the WMC Group x Lag interaction was not observed may reflect the reliability of the WMC measures and/or the reliability of the memory estimates. Clearly, there is evidence that WMC can powerfully predict episodic memory performance (see McCabe et al., 2010). Alternatively, it is quite possible that there really are minimal differences between high and low WMC participants when initial encoding is comparable based on conditional recall analyses. As noted earlier, there is some evidence of an effect of WMC on nonconditional recall performance. However, this confounds the WMC with initial retrieval success.

Conclusions

The present dissertation provided some new insights regarding the influence of age on the benefits of retrieval practice and lag. Because these analyses emphasized the influence of lag and retrieval practice on conditional final test performance (i.e., final test performance for items that were successfully retrieved during learning), age differences during acquisition were minimized. Indeed, results from Experiment 1 revealed a similar function relating lag and continued testing to final test performance across young and older adults. However, there were also some notable differences. For example, young but not older adults benefited from massed retrieval practice, which replicates and extends past findings of an age-related difference in refreshing (e.g., Johnson et al., 2002; Maddox et al., 2011). Finally, the reduced benefit of point value in older adults compared to young adults in Experiment 3 may reflect age-related differences in the capacity to attend to both the encoding task and the point value.

Second, response latency during acquisition and on the final test provided additional constraints on the influence of age, lag and testing on retrieval fluency. Importantly, analyses revealed dissociations between these two measures (i.e., accuracy and response latency). Clearly, future studies should replicate the dissociation between accuracy and response latency and should also replicate the patterns of final test response latency observed in the current studies.

Finally, results from each experiment were generally consistent with the concept of desirable difficulty (e.g., Bjork, 1994). The response latency data during acquisition were particularly informative with these predictions. In this sense, a more difficult retrieval attempt will better enhance the memory trace than a less difficult retrieval attempt presuming both attempts are successful (e.g., Bjork, 1994). However, it is clear that other mechanisms are likely implicated in producing the benefits of spaced retrieval. This was especially true in Experiment

3. Future studies may wish to examine the potential role of other mechanisms previously proposed to account for the benefits of spacing (e.g., encoding variability; Estes, 1955a; 1955b; and deficient processing, e.g., Johnston & Uhl, 1976; Rundus, 1971; Shaughnessy, Zimmerman, & Underwood, 1972).

With these limitations in mind, the current results still afford several useful findings. Notably, the benefits of continued testing approached asymptote in the long lag condition but not the short lag condition, and this relationship between lag, continued testing and final test performance was similar across age groups. Moreover, the multiple dissociations observed between accuracy and response latency on the final test suggest that examining both of these measures is critical for understanding the influence of lag and testing in the current paradigm. Finally, results from Experiment 3 provide initial evidence that manipulation of point value during acquisition exerts more influence on acquisition performance than on retention.

Works Cited

Arking, R.A. (1998). Perspectives on aging. In R.A. Arking, *Biology of Aging* (2nd ed., pp2-36). Sunderland, MA: Sinauer Associates.

Balota, D.A., Dolan, P.O., & Duchek, J.M. (2000). Memory changes in healthy young and older adults. In E. Tulving & F.I.M. Craik (Eds.), *Oxford Handbook of Memory* (pp. 395-410). Oxford: Oxford University Press.

Balota, D.A., Duchek, J.M., & Logan, J.M. (2007). Is expanded retrieval practice a superior form of spaced retrieval? A critical review of the extent literature. In J.S. Nairne, (Ed.), *The foundations of remembering: Essays in honor of Henry L. Roediger III* (pp. 83-105). New York: Psychology Press. doi: 10.3758/MC.38.1.116.

Balota, D.A., Duchek, J.M., & Paullin, R. (1989). Age-related differences in the impact of spacing, lag and retention interval. *Psychology and Aging, 4*, 3-9.

Balota, D.A., Duchek, J.M., Sergent-Marshall, S.D., & Roediger III, H.L. (2006). Does expanded retrieval produce benefits over equal-interval spacing? Explorations of spacing effects in healthy aging and early stage Alzheimer's disease. *Psychology and Aging, 21*, 19-31. doi: 10.1037/0882-7974.21.1.19.

Balota, D.A, & Faust, M. (2001). Attention in Dementia of the Alzheimer's Type. In F. Boller & S. Cappa (Eds.), *Handbook of Neuropsychology*, 2nd Edition, Vol. # 6, Elsevier Science, 51-80.Benjamin, A. S. & Tullis, J. G. (2010). What makes distributed practice effective? *Cognitive Psychology, 61*, 228-247.

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

Bjork, R. A. (1975). Retrieval as a memory modifier. In R. Solso (Ed.), *Information processing and cognition: The Loyola Symposium* (pp. 123-144). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bjork, R.A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A.P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 185-205). Cambridge, MA: MIT Press.

Bui, D.C., Friedman, M.C., McDonough, I.M., & Castel, A.D. (in press). False memory and important: Can we prioritize encoding without consequences? *Memory & Cognition.*

Bui, D.C., Maddox, G.B., & Balota, D.A. (2012). The roles of working memory and intervening task difficulty in determining the benefits of repetition. *Psychonomic Bulletin & Review, 20*, 341-347.

Camp, C.J., Foss, J.W., Stevens, A.B., & O'Hanlon, A.M. (1996). Improving prospective memory task performance in person with Alzheimer's disease. In M. Bandimonte, G.O. Einstein & M.A. McDaniel (Eds.). *Prospective memory: Theory and applications* (pp. 351-367). Mahwah, NJ: USum Associates.

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name-learning. Applied Cognitive Psychology, 19, 619-636.

Castel, A.D. (2008). The adaptive and strategic use of memory by older adults: Evaluative processing and value-directed remembering. In A.S. Benjamin & B.H. Ross (Eds.), *The psychology of learning and motivation* (Vol. 48, pp. 225-270). London: Academic Press.

Castel, A.D., Balota, D.A., & McCabe, D.P. (2009). Memory efficiency and the strategic control of attention at encoding: Impairments of value-directed remembering in Alzheimer's disease. *Neuropsychology, 23*, 297-306.

Castel, A.D., Humphreys, K.L., Lee, S.S., Galvan, A., Balota, D.A., & McCabe, D.P. (2011). The development of memory efficiency and value-directed remembering across the lifespan: A cross-sectional study of memory and selectivity. *Developmental Psychology, 47*, 1553-1564.

Cepeda, N.J., Coburn, N., Rohrer, D., Wixted, J.T., Mozer, M.C., & Pashler, H. Optimizing distributed practice: Theoretical analysis and practical implications. *Experimental Psychology, 55*, 236-246. DOI: 10.1027/1618-3169.56.4.236.

Cepeda N.J., Pashler, H., Vul, E., Wixted, J.T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin, 132*, 354-380.

Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory and Cognition, 24*, 403–416. doi:10.3758/BF03200930

Chan, J.C.K., McDermott, K.B., & Roediger, H.L. (2006). Retrieval-induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General, 34*, 1273-1284.

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

Conway, A.R.A., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12*, 769-786.

Craik, F.I.M., & McDowd, J.M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition, 13*, 474-479. doi: 10.1037/0278-7393.13.3.474

Crowder, R.G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Erlbaum.

Cull, W.L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215-235.

Ebbinghaus, H. (1885). *Uber das Gedachtnis.* New York: Dover.

Estes, W.K. (1955a). Statistical theory of spontaneous recovery and regression. *Psychological Review, 62*, 145-154

Estes, W.K. (1955b). Statistical theory of distributional phenomena in learning. *Psychological Review, 62*, 369-377.

Faust, M.E., Balota, D.A., Spieler, D.H., & Ferraro, F.R. (1999). Individual differences in information processing rate and amount: Implications for group differences in response latency. *Psychological Bulletin, 125*, 777-799.

Giambra, L.M., & Arenberg, D. (1993). Adult age differences in forgetting sentences. *Psychology and Aging, 8*, 451-462.

Glenberg, A.M. (1976). Monotonic and nonmonotonic lag effects in paired-associate and recognition memory paradigms. *Journal of Verbal Learning and Verbal Behavior, 15*, 1-16.

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 801-812.

Hasher, L., & Zacks, R.T. (1988). Working memory, comprehension, and aging: A review and new view. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation, Vol. 22* (pp. 193-225). New York, NY: Academic Press.

Hertzog, C., Kramer, A.F., Wilson, R.S., & Lindenberger, U. (2009). Enrichment effects on adult cognitive development. Can the functional capacity of older adults be preserved and enhanced? *Psychological Science in the Public Interest, 9,* 1-65.

Hertzog, C., Price, J., & Dunlosky, J. (2012). Age differences in the effects of experimenter-instructed versus self-generated strategy use. *Experimental Aging Research, 83,* 42-62.

Hogan, R.M, & Kintsch, W. (1971). Differential effects of study and test trials on long-term recognition and recall. *Journal of Verbal Learning and Verbal Behavior, 10,* 562-567.

Johnson, M.K., Reeder, J.A., Raye, C.L., & Mitchell, K.J. (2002). Second thoughts versus second looks: An age-related deficit in reflectively refreshing just-active information. *Psychological Science, 13,* 64-67.

Johnston, W.A., & Uhl, C.N. (1976). The contributions of encoding effort and variability to the spacing effect on free recall. *Journal of Experimental Psychology: Human Learning and Memory, 2,* 153-160.

Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37,* 1250-1257.

Karpicke, J.D., & Roediger III, H.L. (2007). Expanding retrieval practice promotes short-term retention, but equally spaced retrieval promotes long-term retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33,* 704-719. doi: 10.1037/0278-7393.33.4.704.

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language,65,* 85-97.

Landauer, T.K., & Bjork, R.A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P.E. Morris, & R.N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.

Logan, J.M. & Balota, D.A. (2008). Expanded vs. equal spaced retrieval practice in healthy young and older adults. *Aging, Cognition, and Neuropsychology,* 15, 257-280. doi: 10.1080/13825580701322171.

Maddox, G.B., & Balota, D.A. (2012). Self control of when and how much to test face-name pairs in a novel spaced retrieval paradigm: An examination of age-related differences. *Aging, Neuropsychology, and Cognition, 19,* 620-643.

Maddox, G.M., Balota, D.A., Coane, J.H., & Duchek, J.M. (2011). The role of forgetting rate in producing a benefit of expanded over equal spaced retrieval in young and older adults. *Psychology and Aging, 26,* 661-670.

McCabe, D.P., Roediger, H.L., McDaniel, M.A., Balota, D.A., Hambrick, D.Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology, 24,* 222-243.

McDaniel, M.A., & Masson, M.E.J. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 371–385.

McGillivray, S., Castel, A.D. (2011). Betting on memory leads to metacognitive improvement in younger and older adults. *Psychology and Aging, 26,* 137-142.

Melton, A.W. (1967). Repetition and retrieval from memory. *Science, 158,* 532.

Melton, A.W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior, 9,* 596-606.

Meyer, A.N., & Logan, J.M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging, 28,* 142-147.

Morris, C.D., Bransford, J.D., & Franks, J.J. (1977). Levels of processing versus transfer-appropriate processing. *Journal of Verbal Learning and Verbal Behavior, 16,* 519-533.
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,* 1170–1187.
Nelson, D.L., McEvoy, C.L., & Schreiber, T.A. (1998). The University of South Florida word association, rhyme, and word fragment norms. http://www.usf.edu/FreeAssociation/.
Park, D.C., Smith, A.D., Lautenschlager, G., Earles, J., Frieske, D., Zwahr, M., & Gaines, C. (1996). Mediators of long-term memory performance across the life span. *Psychology and Aging, 11,* 621-637.
Peterson, L.R., Wampler, R., Kirkpatrick, M., & Saltzman, D. (1963). Effect of spacing presentations on retention of a paired associate over short intervals. Journal of Experimental Psychology, 66(2), 206-209.
Pyc, M.A. & Rawson, K.A. (2009). Testing the retrieval effort hypothesis: Does greater difficulty correctly recalling information lead to higher levels of memory? *Journal of Memory and Language, 60,* 437-447.
Rawson, K.A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: How much is enough? *Journal of Experimental Psychology: General, 140,* 283-302.
Roediger, H. L. & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181-210.
Roediger, H. L., & Karpicke, J. D. (2006b). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17,* 249-255.
Roediger, H.L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory and Cognition, 24(4), 803–814.
Rohrer, D., Taylor, K., Pashler, H., Wixted, J.T., & Cepeda, N.J., (2005). The effect of overlearning on long-term retention. *Applied Cognitive Psychology, 19,* 361-374.
Rundus, D. (1971). Analysis of rehearsal processes in free recall. JEP, 89, 63-77.
Salthouse, T.A. (1991). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science, 2,* 179-183.
Salthouse, T.A. (1996). The Processing Speed Theory of Adult Age Differences in Cognition. *Psychological Review, 103,* 403-428.
Schacter, D.L., Rich, S.A., & Stamp, M.S. (1985). Remediation of memory disorders: Experimental evaluation of the spaced-retrieval technique. *Journal of Clinical and Experimental Neuropsychology, 7,* 70-96.
Shaughnessy, J.J., Zimmerman, J., & Underwood, B.J. (1972). Further evidence on the MP-DP effect in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, *11*, 1-12.
Shelton, J. T., Elliott, E. M., Hill, B. D., Calamia, M.R. & Gouvier, W. D. (2009). A comparison of different assessments of working memory and their prediction of fluid intelligence: The laboratory versus the clinical setting. *Intelligence, 37,* 283-293.
Spieler, D.H., & Balota, D.A. (1996). Characteristics of associative learning in younger and older adults: Evidence from an episodic priming paradigm. *Psychology and Aging, 11,* 607-620.

Storandt, M., Balota, D.A., & Salthouse, T.A. (2009). ELSEM: A computerized battery to assess executive, linguistic, spatial, and memory abilities. (http://www.psych.wustl.edu/coglab).
Szpunar, K.K., McDermott, K.B., & Roediger, H.L. (2008). Testing during study insulates against the build-up of proactive interference. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34,* 1392-1399.
Tse, C.S., Balota, D.A., Roediger, H.L. (2010). The benefits and costs of repeated testing on the learning of face-name pairs in healthy older adults. *Psychology and Aging 25, 833-845.*
Tulving, E. (1967). The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning and Verbal Behavior, 6,* 175-184.
Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—3rd Edition (WAIS-3®)* San Antonio, TX: Harcourt Assessment.
Wheeler, M.A., & Roediger, H.L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3,* 240-245.
Zachary, R.A. (1986). Shipley Institute of Living Scale: Revised Manual. Los Angeles: Western Psychological Services.

Footnotes

1. One may be concerned that having the experimenter type the participant's response may unduly influence the response latency results in two possible ways. First, raw response latency may be slower overall than in past studies in which participants entered their own responses. However, the range of mean response latency within each age group was comparable with the range of mean response latency in the previous experiments in which response latencies were measured by the participant's response (e.g., Balota et al., 2006; Maddox et al., 2011). Second, the experimenter may be biased when entering responses if certain items are better retained than others. This situation is unlikely given that (a) items were counterbalanced across conditions which should ensure that any experimenter biases are equally represented across conditions and (b) accuracy and response latency produce different pattern of results (especially in final test performance) which is inconsistent with the expectation of similar patterns of enhanced accuracy and response latency for items biased by a given experimenter.
2. Across age groups, items that were retrieved on the final test trial in each condition were retrieved successfully on each of the preceding test trials. Specifically, 99.6% of Lag 1 – 3 test items, 100% of Lag 1 – 5 test items, 99.9% of Lag 3 – 3 test items, and 99.9% of Lag 3 – 5 test items were successfully retrieved on all test trials during acquisition.
3. Conditional raw response latency was submitted to a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 3 (Number of Tests) mixed-factor ANOVA. Results replicated those reported for conditional standardized response latency with the addition of a significant four-way interaction, $F(2, 342) = 3.11$, $p = .046$, $\eta^2_p = .02$.
4. Separate analysis of long retention interval data failed to yield any significant results for young adults or older adults.

Appendix A.

In order to estimate the response latency (ms) for missing cells, the relevant conditional mean for participants who had at least one observation per cell was taken in proportion to the grand mean for those same participants. In turn, this proportion was used to estimate a given participant's missing cell(s) by multiplying the [Conditional mean/Grand Mean] proportion for all participants and the participant's grand mean.

The estimated response latency (ms) was then treated as an individual trial response latency for which a *z* score was calculated using the standard formula.

Appendix B

Nonconditional Final Test Phase Performance: Experiment 1

Memory Accuracy. Mean proportion correct cued recall on the final test is presented in Figure 28 as a function of age, retention interval, lag and number of tests. Data were submitted to a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 3 (Number of Tests) mixed-factor ANOVA. All main effects were significant, *p*s < .001. The Age x Retention Interval interaction was significant, $F(1, 178) = 15.37$, $p < .001$, $\eta^2_p = .08$, which reflected a larger difference in performance between retention interval conditions for young adults ($M_{diff} = .27$) than for older adults ($M_{diff} = .05$). Critically, the Lag by Number of Tests interaction was only marginally significant, $F(2, 356) = 2.57$, $p = 078$, $\eta^2_p = .01$. As shown in Figure 28, continuing to test material in both lag conditions led to continued increases in final test performance. This finding stands in contrast with the conditional final test performance in which there was no additional benefit obtained in terms of retention from taking five tests versus three tests in the long lag condition. Of course, one must be concerned with performance during the acquisition phase for these various conditions. In fact, the additional 3% benefit in final test performance obtained from taking five tests versus three tests in the long lag condition is only half the size of the difference between these conditions observed during the acquisition phase (7% benefit, as shown in Figure 3), which suggests that final test performance may simply reflect acquisition phase differences in performance rather than a unique benefit from continued testing when separated by a long lag. This complication in interpreting nonconditional final test performance again underscores the importance of the conditional analyses reported earlier.

Standardized Response latency. Mean standardized response latency is presented in Figure 29 as a function of age, retention interval, lag and number of tests. Data were submitted

to a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 3 (Number of Tests) mixed-factor ANOVA. Age, Lag and Number of Tests were all significant (*p*s < .05). These main effects were further qualified by a three-way interaction between Age, Lag and Number of Tests, $F(2, 356) = 7.94$, $p = .003$, $\eta^2_p = .04$.

Figure 28. Mean nonconditional performance on the final cued recall test in Experiment 1 as a function of age, retention interval (RI), lag, and number of tests. Error bars represent ±1 S.E.M.

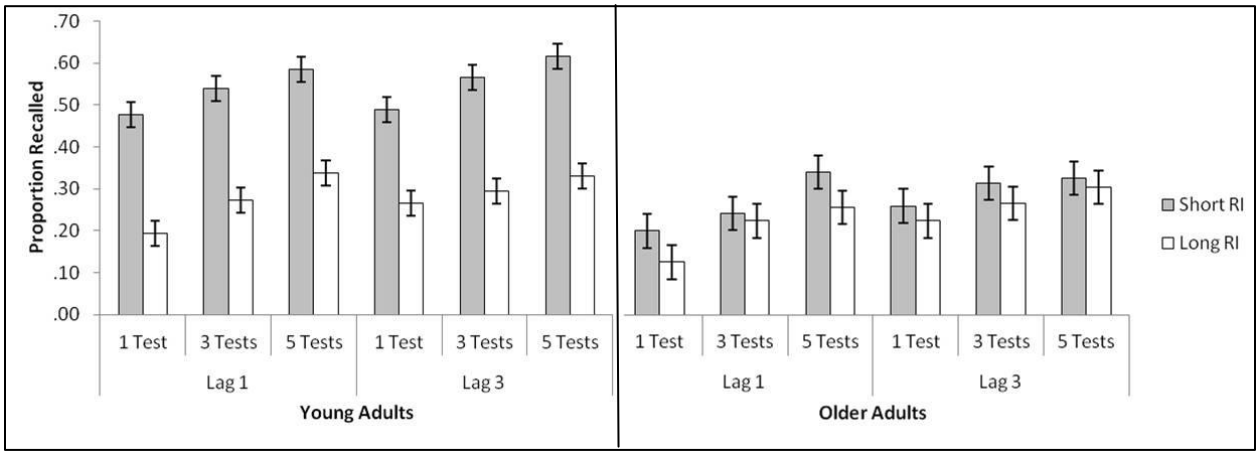
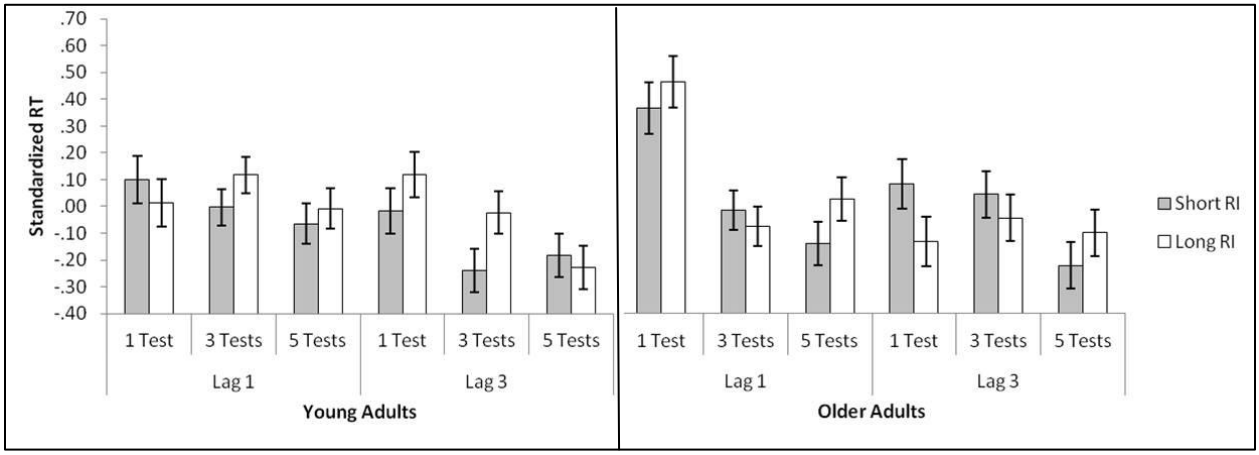


Figure 29. Mean nonconditional standardized response latency on the final cued recall test in Experiment 1 as a function of age, retention interval (RI), lag, and number of tests. Error bars represent ± 1 S.E.M.



Appendix C

Nonconditional Final Test Phase Performance: Experiment 2

Memory Accuracy. Mean proportion correct cued recall on the final test is presented in Figure 30 as a function of age, lag and number of tests. Data were submitted to a 2 (Age) x 2 (Lag) x 2 (Number of Tests) mixed-factor ANOVA. The predicted main effects of Age and Lag were both significant, *p*s < .001. Similarly, Number of Tests was marginally significant, $F(1, 46) = 3.35$, $p = .074$, $\eta^2_p = .07$, such that taking three tests during the learning phase ($M = .32$) produced better performance than having taken one test ($M = .28$). Finally, the Age x Lag interaction was significant, $F(1, 46) = 7.59$, $p = .008$, $\eta^2_p = .14$. As shown in Figure 30, the spacing effect was larger for young adults ($M = .21$) than older adults ($M = .10$).

Standardized Response latency. Mean standardized response latency is presented in Figure 31 as a function of age, lag and number of tests. Lag and Number of Tests were both significant, *p*s < .001. The Age x Number of Tests interaction was significant, $F(1, 46) = 4.51$, $p = .039$, $\eta^2_p = .09$. Finally, the three-way interaction was significant, $F(1, 46) = 5.32$, $p = .026$, $\eta^2_p = .10$.

For young adults, the main effect of lag was significant, $F(1, 23) = 15.08$, $p = .001$, $\eta^2_p = .40$, and the Lag x Number of Tests interaction was marginally significant, $F(1, 23) = 3.48$, $p = .075$, $\eta^2_p = .13$. Standardized response latency was significantly faster for the Lag 4-3 test condition than the other three conditions, *p*s < .005, and there was no difference between the other conditions, *p*s > .15. For older adults, both main effects were significant, *p*s < .001, but the Lag x Number of Tests interaction did not approach significance, $p > .15$.

Figure 30. Mean nonconditional performance on the final cued recall test in Experiment 2 as a function of age, lag, and point value. Error bars represent ± 1 S.E.M.

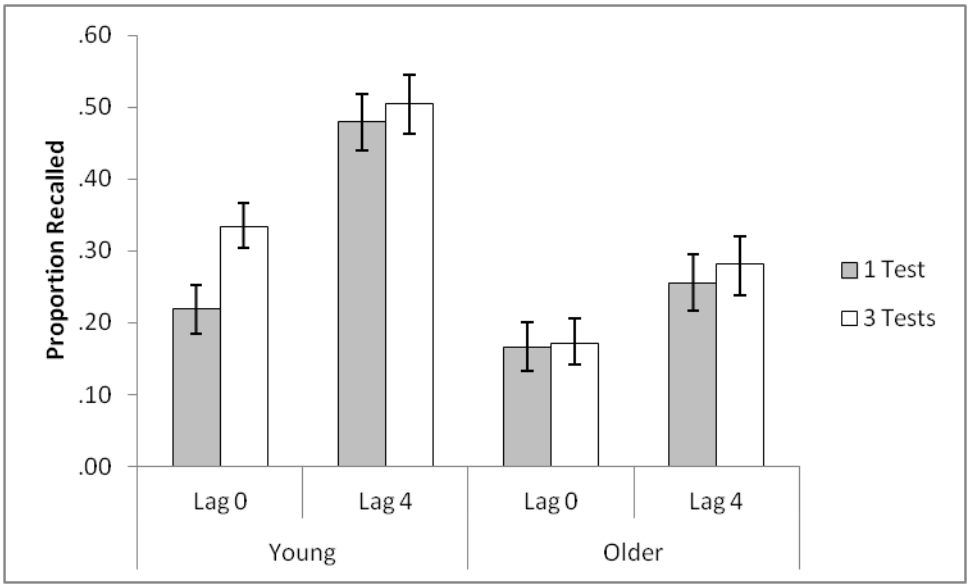
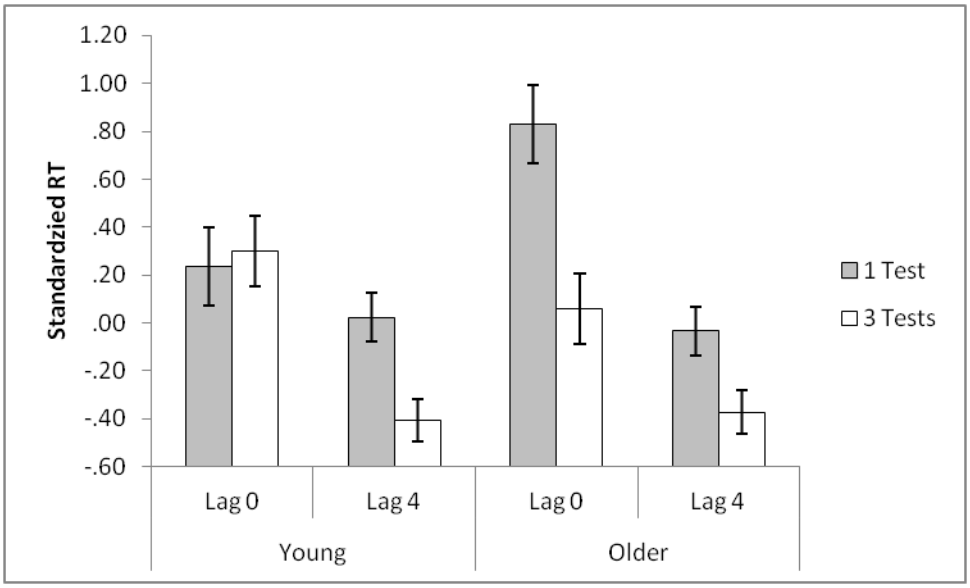


Figure 31. Mean nonconditional standardized response latency on the final cued recall test in Experiment 2 as a function of age, lag, and point value. Error bars represent ± 1 S.E.M.



Appendix D

One may expect that working memory will exert more influence on nonconditional final test performance than conditional performance if the critical processes are implicated in the encoding of information (i.e., once an item is encoded and subsequently retrieved during the acquisition phase it may be retained similarly for individuals with low and high working memory capacity). Analysis of nonconditional accuracy in Experiment 1 failed to yield any significant effects involving WMC Group, but analysis of nonconditional response latency yielded a significant WMC Group x Lag interaction, $F(1, 88) = 5.01$, $p = .028$, $\eta^2_p = .05$. As shown in Figure 32, this interaction reflected a larger lag effect for Low WMC Group than the High WMC Group ($M_{diff} = .26$ vs. .01, respectively).

Turning to Experiment 2, analysis of nonconditional final test accuracy failed to reveal any significant effects involving WMC Group as a factor. With respect to nonconditional final test response latency, however, the WMC Group x Lag interaction was marginally significant, $F(1, 45) = 3.20$, $p = .080$, $\eta^2_p = .07$. As displayed in Figure 33, the difference in response latency between lag conditions was again larger for the Low WMC group compared to the High WMC Group ($M_{diff} = .74$ vs. .38, respectively).

Figure 32. Working Memory Capacity Group x Lag interaction in Experiment 1 nonconditional final test response latency. Error bars represent ± 1 S.E.M.

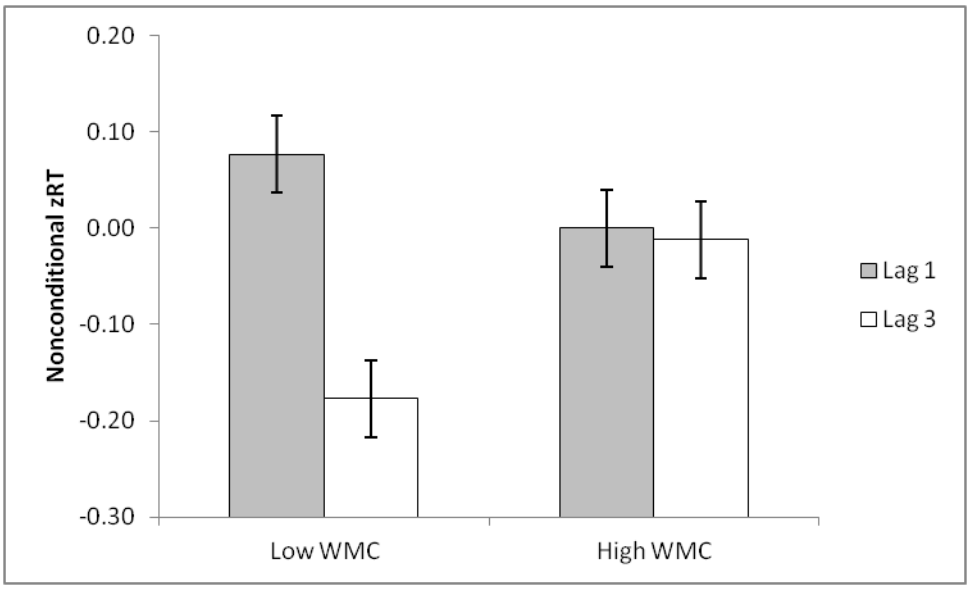
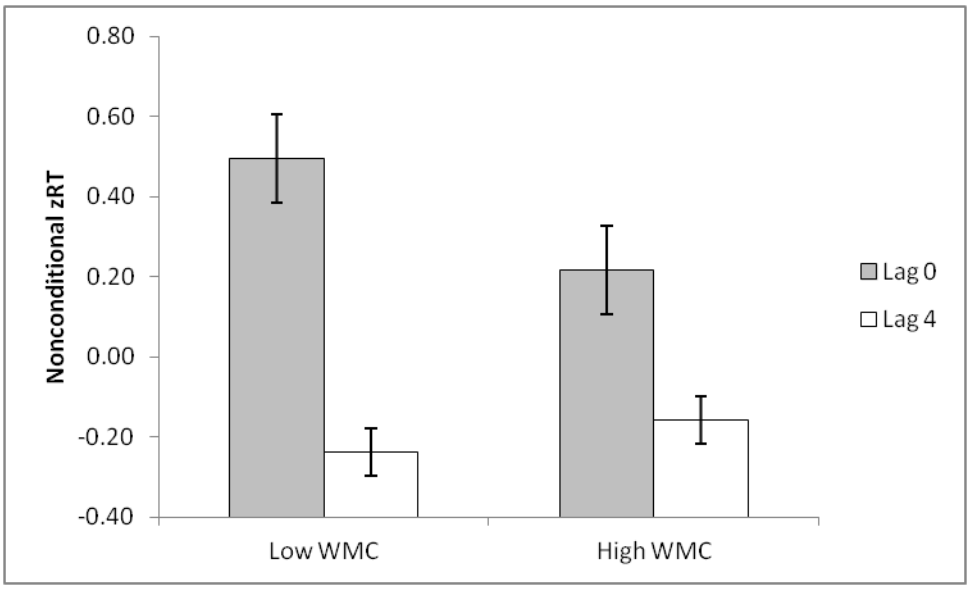


Figure 33. Working Memory Capacity Group x Lag interaction in Experiment 2 nonconditional final test response latency. Error bars represent ±- 1 S.E.M.



Appendix E

Nonconditional Final Test Phase Performance: Experiment 3

Memory Accuracy. Mean proportion correct cued recall on the final test is presented in Figure 34 as a function of age, retention interval, lag and point value. Data were submitted to a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 2 (Point Value) mixed-factor ANOVA. All main effects were significant, *p*s < .05, and were further qualified by two interactions. The Age x Lag interaction, $F(1, 104) = 5.51$, $p = .021$, $\eta_p^2 = .05$, reflected a larger difference between lag conditions for young adults ($M = .20$) than older adults ($M = .08$). However, it is important to note that older adult performance was near floor ($M = .07$ and $M = .15$ for Lag 0 and Lag 4, respectively). The significant Retention Interval x Lag interaction, $F(1, 104) = 4.48$, $p = .037$, $\eta_p^2 = .04$, reflected an increase in the lag effect from the short retention interval ($M = .07$) to the long retention interval ($M = .19$).

Standardized Response Latency. Mean standardized response latency on the final test is presented in Figure 35 as a function of age, retention interval, lag and point value. Data were submitted to a 2 (Age) x 2 (Retention Interval) x 2 (Lag) x 2 (Point Value) mixed-factor ANOVA. Results revealed main effects of age, $F(1, 104) = 5.63$, $p = .020$, $\eta_p^2 = .05$, and lag, $F(1, 104) = 7.32$, $p = .008$, $\eta_p^2 = .07$. Additionally, the Age x Lag interaction was significant, $F(1, 104) = 3.97$, $p = .049$, $\eta_p^2 = .04$. This interaction reflected a larger lag effect for young adults ($M = .28$) than older adults ($M = .04$). Finally, the Retention Interval x Point Value interaction was significant, $F(1, 104) = 5.03$, $p = .027$, $\eta_p^2 = .05$, which reflected faster response latency for high-value items than low-value items following the short retention interval ($M = -.02$ vs. -.14, respectively) and a reversal in the benefit following a long retention interval ($M = -.09$ vs. .06, respectively).

Figure 34. Mean nonconditional performance on the final cued recall test in Experiment 3 as a function of age, retention interval (RI), lag, and point value. Error bars represent ± 1 S.E.M.

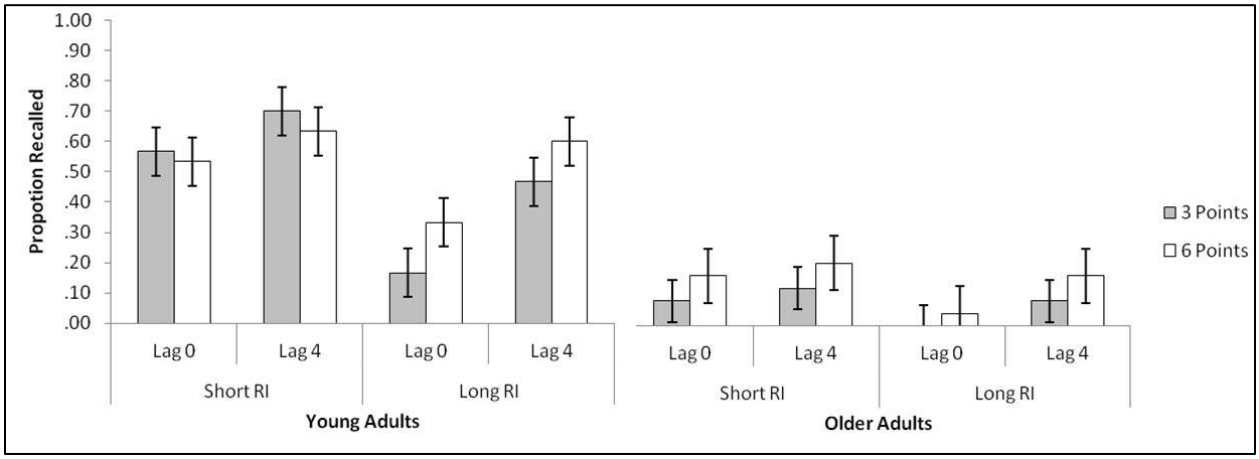
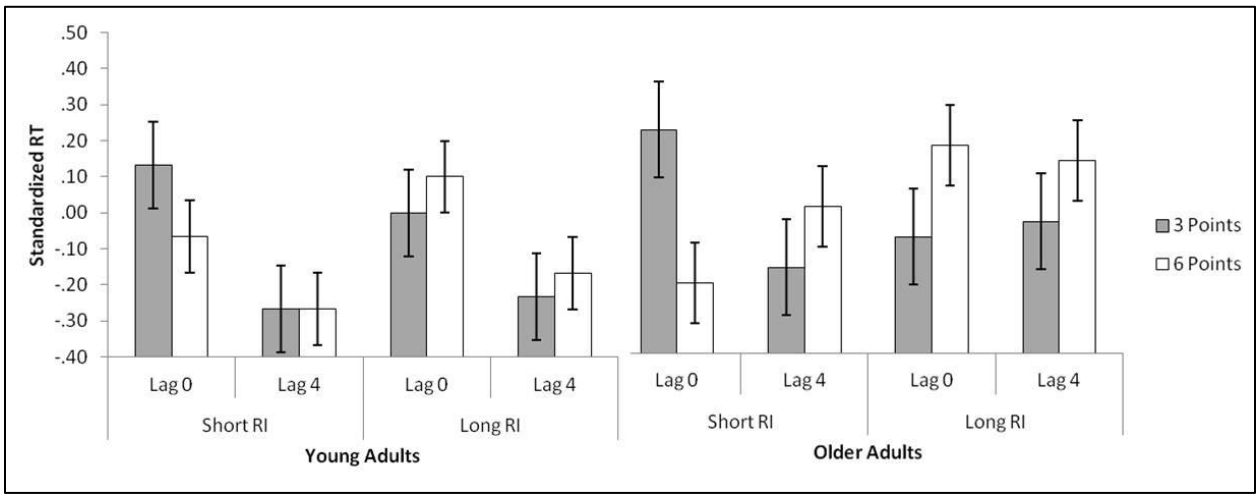


Figure 35. Mean nonconditional standardized response latency on the final cued recall test in Experiment 3 as a function of age, retention interval (RI), lag, and point value. Error bars represent ± 1 S.E.M.



Appendix F

Analysis of nonconditional final test accuracy yielded a significant WMC Group x Retention Interval x Lag interaction, $F(1, 101) = 6.42$, $p = .013$, $\eta_p^2 = .06$. This interaction reflected an increase in the lag effect across retention intervals for the low WMC group (M = 0% and 11%, for short and long retention intervals, respectively; $p = .008$) but no difference for the high WMC group (M = 8% and 6%, respectively; $p > .40$). Analysis of nonconditional final test response latency failed to yield a significant effect or any significant interactions involving WMC group (ps > .15).

The significant WMC Group x Lag x Retention Interval interaction in accuracy stands in contrast with analyses from the first two experiments that revealed significant WMC Group x Lag interactions in response latency but not accuracy. One possible reason for this discrepancy is the increase in attentional load associated with processing and maintaining point values. In the current experiment, it may be that that this increased attentional load is relatively more demanding for those with low WMC compared to those with high WMC which in turn results in a more effective spacing manipulation for this group, which is reflected at the longer retention interval. Indeed, Bui, Friedman, McDonough and Castel (in press) recently compared free recall of lists of words using Deese-Roediger-McDermott false memory paradigm (Roediger & McDermott, 1995). The critical comparison was between lists of items that were learned with a value-directed encoding procedure and lists of items that were learned without any associated point values. Results revealed significantly higher veridical recall and significantly lower false recall for the value-absent condition compared with the value-present condition. These results are consistent with an account in which processing point-value during encoding may have deleterious effects in terms of attentional processing of material