5-31-2024

The Impact of Age on Second Language Vocabulary Learning

Steven J. Dessenberger

Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Psychology

Dissertation Committee:
Mitchell Sommers, Chair
Joe Barcroft,
Andrew Butler,
Henry Roediger,
Kristin Van Engen

The Impact of Age on Second Language Vocabulary Learning

by

Steven Jay Dessenberger III

A dissertation presented to
Washington University in St Louis
in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2024
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank my advisors, Mitchell Sommers and Henry Roediger, for supporting me throughout the graduate program and for helping me to achieve this degree. Their guidance and trust in me were critical to my growth as a researcher and I could not a have asked for better mentors. I would also like to thank the rest of my committee, Kristin Van Engen, Joe Barcroft, and Andrew Butler, for their feedback and time.

I would also like to thank everyone from the Speech and Hearing Research Lab as well as the Memory Lab for supporting me during my time in the lab as well as several of my fellow graduate students whom I collaborated with throughout my time at the University. Particularly, I would like to thank Taylor Levine for her constant support for the past five years and Abishek Dey for always being willing to lend a hand.

And of course, I must thank my family, especially my wife who has become my reason for everything. I would also like to thank our pets for their constant emotional support, and I would like to thank my son who gave me a reason to finish.

Steven Jay Dessenberger

Washington University in St. Louis

August 2024

Dedicated to Becca, my most interesting discovery throughout this process.

ABSTRACT OF THE DISSERTATION

The Impact of Age on Second Language Vocabulary Learning

by

Steven Jay Dessenberger III

Doctor of Philosophy in Psychological and Brain Sciences

Washington University in St. Louis, 2024

Professor Mitchell Sommers, Chair

Second language (L2) vocabulary learning is challenging for many would-be language learners, regardless of their cognitive abilities. The volume of to be learned vocabulary can be a roadblock to language fluency (Nation, 2006), especially for older adults who have diminished memory capabilities due to age-related memory decline, a degradation in our ability to remember information that naturally occurs as we age (Salthouse, 2009; Schaie, 1993). The process of L2 vocabulary acquisition can be separated into three components, the native language (L1) meaning of the new word (*cat*), the new L2 word form itself (*gato*) and the connection between the new word and its meaning (*gato means cat;* Barcroft, 2002; Kida & Barcroft, 2018). For young adults, there is little difference in their ability to encode and recall these three components, but prior research suggests older adults may struggle with forming a connection between the novel word and its meaning (Naveh-Benjamin, 2000).

This dissertation is comprised of two experiments that examine younger and older adult L2 vocabulary learning. There were 4 aims: (1) compare younger and older adults on L1 item memory, L2 item memory, and L1-L2 associative memory during L2 vocabulary learning, (2) identify the relationship between episodic memory and L2 vocabulary learning, (3) identify the unique contributions of working memory to L2 vocabulary learning in older and younger adults,

and (4) identify how receptive and productive retrieval practice affects long-term memory for L2 items and L1-L2 associative memory in older and younger adults.

Experiment 1 addresses aims 1-3, which seek to understand L2 vocabulary acquisition and how each age group encodes the three components. Participants were first given a working memory test, followed by a learning task for L1-L2 and L1-L1 word pairs and three recognition tests (one for L1 items, one for L2 items, and one for L1-L2 word pairs). Results indicated no age-related differences in the ability to recognize L1 items, L2 items and L1-L2 items, and both age groups were better able to recognize L2 items compared to L1 items. Results also indicated working memory was related only to the L1-L2 pair recognition test with higher working memory leading to improved performance. While older adults did have poor working memory scores compared to young adults, there was no statistical difference between the age groups on L1-L2 pair recognition performance. These findings suggest that age-related cognitive decline may not have a significant impact on L2 vocabulary learning.

Experiment 2 addressed aim 4 to examine how study method affected L2 vocabulary learning for both age groups. Participants were tasked with learning L1-L2 word pairs in one of three ways: repeated study, retrieval practice with a L2 target given an L1 cue, and retrieval practice with a L1 target given an L2 cue. After the learning session, participants were given the same three recognition tests as Experiment 1. Results indicated that retrieval practice and repeated study lead to equivalent performances for older adults, but young adults benefitted more from retrieval practice compared to repeated study. Additionally, for young adults retrieving the L1 word improved memory for the L1 item but retrieving the L2 item did not improve memory for the L2 item. These results suggest that retrieval practice may not be as an effective of a tool for older adults as it is for young adults for L2 vocabulary acquisition.

Chapter 1: Introduction

The population worldwide is getting older (Colby & Ortman, 2014), and with an aging population comes an increased need for research on older adult second-language acquisition (SLA). Older adult SLA is important because older adults may have to live in a region where they do not speak the local language and language barriers between older adults and their community can have far-reaching implications. For example, language barriers between older adults and medical providers have been linked to increases in severe medical complications such as misdiagnoses and increases in infections (Flores, 2006; Wasserman et al., 2014). Additionally, older adults are already considered at risk for feelings of depression and isolation and these conditions can be exacerbated when they live in a community that does not share their native language (Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults et al., 2020).

The current study seeks to expand our understanding of older adult SLA by investigating one of the most challenging areas within SLA: vocabulary learning. Second-language vocabulary learning necessitates encoding and recalling thousands of new words. For example, Webb and Rodgers (2009) found that to watch and understand tv (which requires no production of the language) requires knowing 3,000 words. Going beyond a passive observer to achieving language fluency which does involve production of the language, the number of words needed is closer to 6,000 (Nation, 2006). The volume of words is a challenge even for young adults with intensive language programs often taking years to complete (U.S. State Department, 2020). Older adults are likely to face even greater difficulty acquiring new second-language vocabulary words because older adults can have poorer memory abilities compared with young adults

(Commissaris et al., 1994; Rhodes et al., 2019). The proposed study seeks to expand our understanding of second-language vocabulary learning in younger and older adults by examining the learning process through a three-component model of second-language vocabulary learning. The goal is to highlight the key areas of decline in older adult SLA vocabulary learning such that future interventions can properly address the needs of these learners.

1.1 Three component model of second language vocabulary learning

Second-language (L2) vocabulary learning is a multi-stage cognitive process that necessitates encoding novel word forms and associating them with existing semantic concepts contained in the mental lexicon. To better understand how older and younger adults acquire new L2 vocabulary, we can divide the process into three distinct components or types of memory that need to be created and stored so that they can later be retrieved. As shown in Figure 1, these components consist of memory for the semantic meaning, the novel word form, and the association between the first two components. All three components are necessary for language fluency, which for the purposes of the proposed study, is defined as the ability to translate back and forth between the target language and the native language.

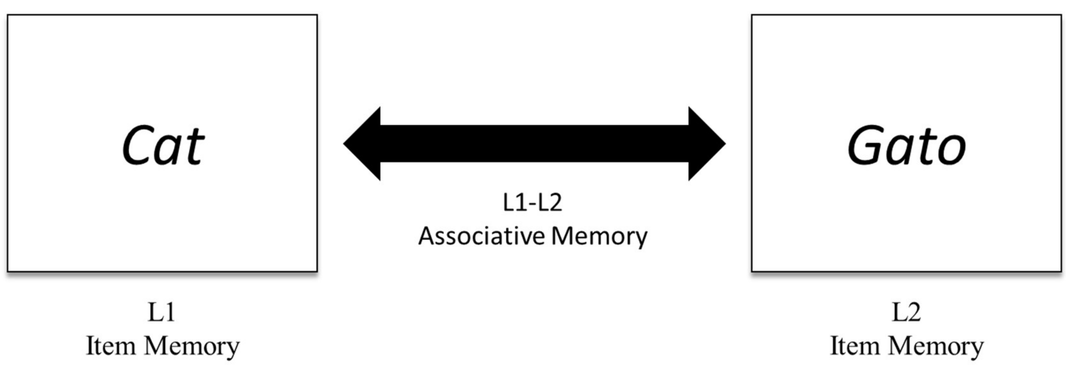

Figure 1. The three components of vocabulary learning.

1.1.1 Semantic Meaning

The first component in the model is the native language meaning or semantic concept for the vocabulary word and I will refer to this component as the L1 item memory. This could be something simple like the Mandarin Chinese word *gǒu* which means *dog* in English. For a native English speaker, *dog* is the L1 item to be remembered, the semantic concept represented by the novel word form *gǒu*. In other cases, the semantic concept can be something more abstract, such as the German word *schadenfreude* which means *the feeling of pleasure derived from another's misfortune*. Semantic concepts are stored within semantic memory, the memory system that deals with general knowledge (Tulving, 1972), but during L2 vocabulary learning, we create an episodic memory of the L1 item. Episodic memories are memories for experienced events or episodes. For instance, if someone encounters the word *truck* during study, they can access semantic memory to gain an understanding of what the word means but recalling that they saw the word *truck* during study is an episodic memory. It may seem like knowledge of the word truck within semantic memory is sufficient to support L2 learning, but the creation and recalling of the episodic memories of the L1 item is crucial to the L2 vocabulary learning process as the learner needs to recall that the L1 item was seen in relation to the new L2 vocabulary word. Once the learner can recall that the L1 item and L2 word are related by a singular episodic memory, they can form a connection between the two that will allow for translation back and forth between the L1 and the L2.

1.1.2 Novel Word Form

In addition to encoding an episodic memory for the L1 item, the learner has to encode an episodic memory for the second component or the novel word form, which I will refer to as an L2 item memory. Whether it is the written form or the spoken form, all words have some to-be-learned form. The L2 item memory is considered more difficult to recall compared to the L1 item memory. Evidence of this can be seen in research comparing memory for words to nonwords. Nonwords are similar to L2 words in that they are a sequence of phonemes that have no meaning for a novice learner unless paired with a semantic referent. Multhaup and colleagues (1996) compared memory span performance with familiar words to memory span performance for nonwords in both older and younger adults. Participants were tasked with tracking and maintaining a growing span of either words or nonwords. At test, both age groups demonstrated significantly greater ability to recall the words compared with the nonwords. Multhaup et al. argued that this ability is indicative of the role existing memory plays in new learning, such that familiar L1 words that already exist with the lexicon are more easily recalled after study compared to novel word forms that are not contained within the lexicon. Effectively, memory for both words and nonwords contained with the spans decayed but in theory participants could effectively reconstruct the decayed memories for the words using long-term memories. Memories for nonwords did not have an equivalent level of support from long-term memory and thus were more difficult to recall later in comparison to memories for words.

1.1.3 Associative Memory

The third component is the connection or association between the first two components. Referred to as *associative memory* or *mapping*, this third component is critical to the language learning process (in fact, it could be argued that this is the end goal of vocabulary learning). For example, a student who studied the *cat-gato* word pair has to know what a *cat* is and that *gato* is a Spanish word, and then they have to bind the two together, forming an association between *gato* and *cat.* A failure to bind the information together results in no understanding of the L2 word form, as it is now little more than a nonsense syllable from the student's perspective. According to the multi-process model of paired associate learning (McGuire, 1961), forming this association requires multiple stages of encoding, to include encoding that *cat* is associated with *gato* as well as a separate distinct encoding that *gato* is associated with *cat*. This bi-directionality in associations is critical for later recall as well as for language fluency.

It should be noted that when forming associative memories, it may not always be a one-to-one translation between the L1 and L2 when it comes to the semantic representation. For instance, the word "food" in one language may bear a similar literal meaning but due to cultural differences, the semantic representation in another language may be somewhat different, thus mapping the L2 to the L1 may leave something lost in translation. However, for the purposes of L2 vocabulary learning, we omit these subtle differences and focus on mapping the novel L2 word to the L1 even if it is not a perfect translation.

1.1.4 Episodic memory in L2 vocabulary learning

As suggested by the three-component model, episodic memory is crucial to the vocabulary learning process. When we study new L1-L2 word pairs, the novel word form and the corresponding association with the existing semantic representation have to be encoded and stored in long-term memory as episodic memories. When we have a failure of episodic memory, L2 vocabulary learning does not occur. For example, Zhang and colleagues (2021) had thirty-eight college-age, bilingual participants complete a nonverbal episodic memory task followed by a vocabulary learning exercise and found that performance on the nonverbal episodic memory task predicted vocabulary acquisition which is a verbal episodic memory task. This suggests the generalized episodic memory abilities (rather than a domain-specific episodic memory ability) predicts verbal acquisition abilities.

The influence of episodic memory on vocabulary learning is further reflected in multiple case studies with amnesiac patients who do not possess the ability to form and encode novel episodic memories (Hirst et al., 1988; Verfaellie et al., 1995). Unsurprisingly, the studies found that patients had extreme difficulty in acquiring novel L2 vocabulary words which indicates that individual differences in the ability to form, encode and recall episodic memories is crucial to the L2 learning process.

1.2 Age-related differences in episodic memory

General episodic memory is important to L2 vocabulary learning, but the three-component model highlights a difference between memory for items and memory for the association between the items. The ability to encode and recall item and associative memories both fall under the domain of episodic memory; however, the proposed three-component model makes the distinction that memory for the L1 and L2 items and the memory for the L1-L2 association are two different, albeit related, abilities. This distinction is not common in the young adult L2 vocabulary learning literature, likely due to two factors. The first being that item memory and associative memory do not differ greatly in young adults (Naveh-Benjamin, 2000; Vaughn & Rawson, 2011) and the second being that common forms of measurement in L2 vocabulary learning make it difficult to distinguish between associative and item memory abilities. However the distinction is present in the older adult memory literature, with many older adults demonstrating greater difficulty forming and recalling associative memories compared to their ability to form and recall item memories (Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008).

1.2.1 Young adult episodic memory abilities

Young adults do not often demonstrate significant differences in item memory and associative memory abilities (Naveh-Benjamin, 2000; Vaughn & Rawson, 2011). For example, Vaughn and Rawson investigated the influence of criterion training on L1-L2 word pairs in young adults. Participants studied and were tested repeatedly on L1-L2 word pairs. After successfully reaching the designated criterion of correct retrievals, participants were given a series of tests. The tests included a recognition test for the L2 items individually (L2 item memory) as well as a recognition test for the L1-L2 pair (associative memory). Results showed no difference between item memory and associative memory recognition for young adults.

The second potential factor for the lack of distinction between item memory and associative memory in the literature on L2 learning is that common measures of L2 vocabulary learning typically rely on both item memory and associative memory simultaneously. For example, two common measures of L2 vocabulary learning are receptive recall and productive recall. Both tests are cued recall tests, where a cue word is provided and a target must be retrieved from memory. A receptive recall test provides the L2 word form as the cue and the L1 meaning must be retrieved, while a productive recall test provides the L1 meaning and requests the L2 word form be retrieved (Davies, 1976; Webb, 2005). Cued recall tests like productive and receptive recall are ecologically valid measures of vocabulary learning as the end goal of learning a novel L2 word is to be able to translate back and forth from the target language. However, from the perspective of the three-component model, when there is a failure to recall during a cued recall test, it is often difficult if not impossible to determine whether it is a failing of item memory or associative memory. For example, if the learner was studying the English-Spanish word pair *cat-gato* and they failed to retrieve the word *gato* during a test it would be unclear whether it was due to the learner forgetting the word *gato* (item memory failure) or was it the case that they could remember the word *gato* but forgot that it means *cat* (associative memory failure). Essentially, cued recall tests are not designed to detect and measure differences between item and associative memories. This likely is not an issue for young adults given that their item and associative memory abilities do not differ greatly from one another (Vaughn & Rawson, 2011), but for older adults, this may be an important consideration.

1.2.2 Older adult general episodic memory decline

Older adults differ from young adults in their episodic memory abilities. According to a meta-analysis (Rhodes et al., 2019), recall and recognition of information are both negatively impacted by age, with an estimated standardized effect size for age-related decline in recalling episodic memories to be approximately .89 while the age-related decline for recognition of episodic memories was at a standardized effect size of .54. Older adults are aware of this decline as approximately half of older adults reporting fear of their increasing forgetfulness (Commissaris et al., 1994).

Despite their fear, increases in forgetfulness are not necessarily signs of dementia; rather as we age, the decrease in episodic memory is part of a normal aging process referred to as age-related cognitive decline (Craik, 1994; Craik & McDowd, 1987; Gordon & Clark, 1974; Park et al., 2002; Rhodes et al., 2019). Our cognitive capabilities fluctuate and change across the lifespan with many cognitive capabilities showing sharp rises in our youth followed by declines as we age past 25-30 years old with declines most evident beyond 65 years of age (Park et al., 2002; Salthouse, 2004, 2009; Schaie, 1993). Age-related decline is considered a natural part of the aging process and as shown by Figure 2, while our accumulated knowledge may grow across our lifespan, several processes decline as we age, including episodic memory.

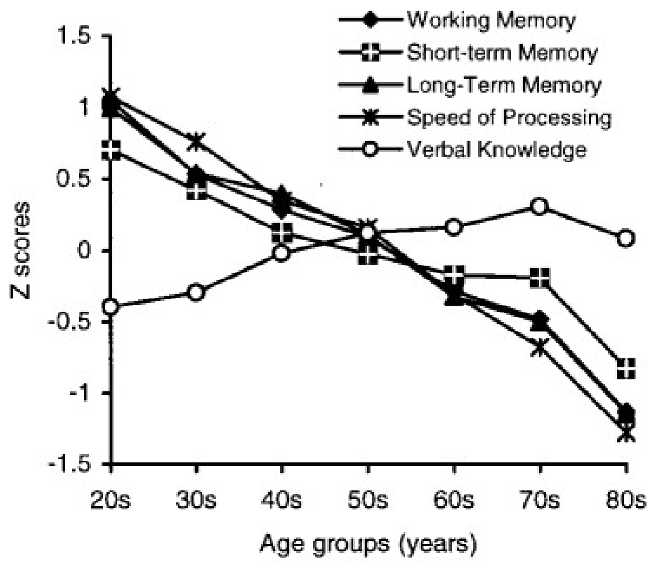

Figure 2. Age-related declines in memory; reprinted from Park et al., 2002

1.2.3 Older adult item memory and pair memory decline

Although there is a general decline in episodic memory as a result of aging, item memory and associative abilities decline at different rates (Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008; Rhodes et al., 2019). For example, Naveh-Benjamin (2000) measured differences in item memory and associative memory in a series of experiments examining cue-target word pair learning with both L1-L1 word pairs as well as L1 words paired with nonsense syllables (a suitable proxy for L1-L2 word pair learning). Older and younger adults studied word pairs and then completed two different kinds of tests: item tests and an associative test. As shown by Figure 3, the item tests presented single words, half of which had been seen during the study session (*old*) and half had not been seen (*new*). Participants were tasked with deciding which words were old and which were new. With the associative test, participants saw word pairs from the study session, but for half the pairs, the words were shuffled, resulting in cue words being matched with different targets such that new pairs were formed. Participants were tasked with deciding whether the pairs were the old or new relative to the pairs seen during the study session. When participants were instructed to study the word pairs, there was little difference between the young adults' abilities to recognize individual items and pairs, but older adults showed significantly greater declines in associative memory compared to item memory in relation to their young adult counterparts.

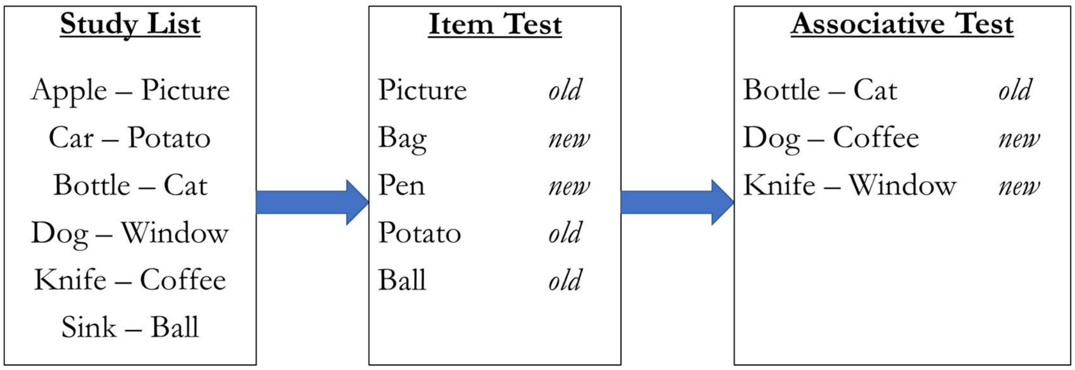

Figure 3. Example procedure for measuring item and associative memory using old/new recognition judgments

1.2.4 The Associative deficit hypothesis

Differences in the decline of associative memory in comparison to item memory as a result of age is referred to as the associative deficit hypothesis and has been demonstrated with a variety of stimuli including other word pairs (Naveh-Benjamin et al., 2007), name-face pairs (Naveh-Benjamin et al., 2009), and even object-location pairs (Siegel & Castel, 2018). According to a meta-analysis, associative memory differences between older adults and younger adults is greater (d = .85) than differences in item memory abilities between the two age groups (d = .65; Old & Naveh-Benjamin, 2008).

While the original study investigating associative memory deficits included an experiment using L1-nonwords as stimuli (Naveh-Benjamin, 2000), it should be noted that associative memory deficits have not been replicated with L1-L2 material. L2 words and nonwords are similar in that for a novice learner, neither has any inherent semantic meaning on its own. They differ though in that nonwords typically respect L1 phonetic structure and spelling while L2 words may not respect L1 phonology or may have additional requirements for pronunciation (e.g. Chinese Mandarin is a tonal language wherein various inflections change the meaning of a word while English words has no such distinction). Thus, while L2 words and nonwords are similar, research with L1-nonword pairs may not transfer to L1-L2 learning.

Additionally, one study that compared L1-L1 word pair learning to nonword-nonword pair learning found the expected age-related declines in item and associative memory for the L1-L1 word pairs such that associative memories declined further with age in comparison to item memories. However, there was no difference between age-related decline for item and associative memories with nonword-nonword pairs, which suggests no additional associative memory deficit for the nonword-nonword pairs (Badham & Maylor, 2011). While nonword-nonword pairs are not representative of L1-L2 word pairs in the same way that L1-nonword pairs are representative, the lack of research in this area leaves an unclear picture of whether older adults are likely to suffer from additional associative memory declines with L1-L2 material as has been seen with L1-L1 material.

1.2.5 Summary

L2 vocabulary learning necessitates forming two different types of episodic memory: item memories and associative memories. The distinction between the two kinds of memory is not common in the young adult literature, likely attributable to the relatively small difference in these memory abilities in the young adult population and a lack of sensitivity to item and associative memory differences in most measurements for L2 vocabulary learning. However, the distinction between item and associative memories may be more relevant to an older adult population who are known to have significantly greater age-related declines in associative memory compared to item memory.

1.3 The role of working memory in L2 vocabulary learning

Working memory is the short-term, limited-capacity system that allows for the simultaneous storage and manipulation of information as well as the orientation of attention to that information (Baddeley, 1983; Baddeley & Hitch, 1974). According to the original working memory model put forward by Baddeley and Hitch (1974), working memory is made up of multiple components: the phonological loop, the visuo-spatial sketchpad, and the central executive. A later model put forward by Baddeley (2000) added the episodic buffer resulting in a four-component model of working memory. The phonological loop, and the episodic buffer in particular are thought to play a crucial role in L2 vocabulary learning.

1.3.1 The role of the phonological loop during L2 vocabulary learning

The phonological loop allows for the temporary storage and maintenance of verbal information. In the case of L1 and L2 learning, it does not matter whether the L1 or L2 has an associated semantic meaning, in both cases the information is considered verbal information. When a word, such as a novel L2 word form, is heard or read, the information is stored and repeated on a loop (hence the name) within working memory. This potential storage and maintenance of the novel word form is critical for both initial encoding and later retrieval. For example, according to Papagno et al. (1991), suppressing the phonological loop negatively impacts L2 vocabulary word learning if the learner cannot rely on forming semantic associations between the L1 and L2. Across 7 experiments, participants were required to learn L1-L1 word pairs and L1-L2 pairs while either engaging in articulatory suppression which interferes with the phonological loop (saying the word "bla" once every second) or tapping the table every second (control condition). In 4 of the 7 experiments, results suggested that when the phonological loop was suppressed, memory for the L2 words was negatively impacted such that fewer L2 words were recalled compared to the memory for the L1 words. In the other 3 experiments, there was no significant difference between memory for L1 and L2 when the phonological loop was suppressed compared to the control condition.

Papagno and colleagues argued that, in the experiments without significant differences, the pre-existing associative strength between the L1 and L2 in those pairs was higher compared to the other experiments which may have allowed participants to engage in semantic processing, circumventing the need for the phonological loop. While cognates were avoided in word selection, many of the L2 words (Russian) selected for the experiment bore similarity to a word in the L1 (English) which gave the L2 words some inherent (albeit inaccurate) meaning that allowed participants to use their pre-existing semantic knowledge to create associative memories between the L2 and L1 words (the language was changed to Finnish for the other experiments to address this issue of similarity). This suggests that when semantic information about the L2 is available (e.g. prior knowledge of the language or at least a resemblance to a known language) the influence of the phonological loop is diminished. Put another way, working memory is likely a more powerful moderator for novice learners with no previous experience compared to advanced learners who are more likely able to form semantic associations with novel vocabulary, but there has yet to be a study that directly compares the influence of working memory on novice and advanced L2 learners. However, this does mirror our early L1 experience, as younger children are more reliant on phonological short-term memory to acquire novel vocabulary while later in development, L1 vocabulary learning is more strongly correlated with prior knowledge (scaffolding) rather than phonological memory (Gathercole et al., 1992).

1.3.2 The role of the episodic buffer during L2 vocabulary learning

The second component theorized to play a role in L2 vocabulary learning is the episodic buffer, which allows separate units of information in memory to be bound together into a single unit of information or episode. Typically, the episodic buffer is thought to play a pivotal role in binding information from different modalities (i.e. auditory and visual) to form singular episodic events in long-term memory. However, the episodic buffer has also been implicated in the role of binding novel word forms to their semantic referents (Wang et al., 2017). When we experience a novel word form and its associated meaning, the phonological loop maintains the novel verbal information while the episodic buffer is purportedly where the two pieces of information are associated or mapped together. In their study, Wang and colleagues had participants study L2 words paired with images, before completing an associative recognition task. The task mirrored the tests used by Naveh-Benajmin (2000) that shuffled some cue-target pairings while leaving others intact (see Figure 3). Additionally, participants completed a separate working memory task that measured their ability to bind visual information to auditory information. The working memory task consisted of the presentation of native language stimuli each paired with six-point shapes that were irregular and unnamable. After learning the pairs, the participants were tasked with two recognition tests (one for the words, one for the shapes), and then given a binding test (a multiple-choice test where participants were given the shape and had to recall the corresponding native language stimuli from a set of choices). The resulting correlation between associative recognition accuracy (images – L2 words) and binding ability (L1 word -- shapes) was significant (r =.35), which suggests that the episodic buffer plays a role in forming associative memories. It should be noted that the procedure used by Wang et al. relied on a cross-modal procedure, meaning that participants had to bind auditory and visual information, so it is not clear whether this binding ability extends to L1-L2 words presented in the same modality.

1.4 Age-related differences in working memory

A reliance on working memory for L2 vocabulary learning may indicate trouble for older adults, as working memory, like episodic memory, declines across the lifespan during healthy aging (see Figure 2; Park et al., 2002; Salthouse, 2009). Much like episodic memory, working memory abilities peak between ages 25-30 before slowly declining across the lifespan. Additionally, although episodic memory and working memory systems are distinct, there is considerable overlap between them, because episodic memories are typically created after the information is processed in working memory and transferred into long-term memory storage. Thus, both can potentially influence L2 vocabulary learning.

1.3.1 The influence of working memory on associative memory

Decreases in working memory may account for some of the variability in age-related declines in associative memory. According Hara and Naveh-Benjamin, (2015), when young adults have their working memory capacities challenged, they begin to mimic older adults such that their ability to encode and recognize novel associations is worse than their ability to encode and recognize individual items. In their study, Hara and Naveh-Benjamin had young adult participants study face-name pairs under divided attention or full attention. During the divided attention task, participants had to pair the appropriate name with its corresponding face while simultaneously encoding a series of letters followed by a probe which prompted the participants to decide whether the probe was contained in the letter series. A later test of item and associative memory indicated that face-name pairs learned under a divided attention condition resulted in decreased associative memory performance compared to performance under the full attention condition. Essentially, challenging the young adults' working memory resulted in associative memory performance that mimicked older adult associative memory scores, suggesting that age-related working memory declines may account for some of the associative memory deficits seen in older adults.

While the results of Hara and Naveh-Benjamin (2015) suggest that diminished working memory can account for age-related declines in associative memory, research with older adults is only partially in agreement. According to Bartsch et al. (2019), declines in working memory only account for some of the decline in associative memory; they argue that both working memory and episodic memory declines are symptomatic of a general cognitive decline due to aging. In their study, older adult and young adult participants were tasked with learning L1-L1 word pairs followed by a multiple-choice associative test wherein the cue word was matched with three different target words, one that was the correct pairing, one drawn from the studied list but not originally matched with the cue word, and one that was a new unstudied word. Participants had to assign each word to a category: *correct*, *old but not correct,* or *new*. This way, researchers could measure item memory and associative memory simultaneously. Participants were then given the same test after a delay to measure their episodic memory (the unrelated *old* and *new* words were changed). Bartsch et al. found that differences in working memory did account for some of the differences in episodic memory. However, when the experiment was repeated and the sample set size was varied (increased set size puts more strain on working memory, smaller set sizes decrease working memory strain), only the working memory tests were affected; outcomes on the delayed episodic memory test were not affected by set size regardless of age. These findings suggest working memory, while related to associative memory declines, accounts for only a portion of age-related associative memory deficits.

1.3.2 The influence of working memory on L2 learning in older adults

There is little research examining the interaction of working memory and L2 vocabulary acquisition specifically in older adults but the limited research does suggest that a positive relationship exists between the two (Mackey & Sachs, 2012; van der Hoeven & de Bot, 2012). In a study of 45 adults split into three age groups (young, middle-aged, and older adults), Van der Hoeven and de Bot found that with advanced learners (individuals who had taken language courses in the target L2), regardless of age, working memory was positively correlated with the acquisition of new vocabulary ($r = .54$), but it is unclear if the influence of working memory was different for the different age groups. Mackey and Sachs investigated the influence of working memory on L2 vocabulary learning with 9 older adult L2 learners who had all taken L2 language classes in the past and found that the only participants to show improvement in L2 listening span scores during the 5-week study were those with the highest working memory. Admittedly, the small sample size (N <= 15 per group) in both studies limits generalizability but both at least suggests that working memory continues to play a role for older adults.

1.3.3 Summary

Working memory may play a significant role in L2 vocabulary learning, particularly for forming associative memories (Hara & Naveh-Benjamin, 2015), although this view is under debate (Bartsch et al., 2019). Older adults, who suffer from age-related declines in working memory, were found to be reliant on working memory during L2 learning (Mackey & Sachs, 2012; van der Hoeven & de Bot, 2012). However, there has not been research into the influence of working memory on L1-L2 learning that both compares younger and older adults as well as considers the differences in L1 item, L2 item, and L1-L2 associative memories.

1.4 The TOPRA model: How study task impacts L2 vocabulary learning

Individual differences in abilities such as episodic memory and working memory play a critical role in L2 vocabulary learning(Badham & Maylor, 2011; Bartsch et al., 2019; Hara & Naveh-Benjamin, 2015; Mackey & Sachs, 2012; Naveh-Benjamin, 2000). However, prior research suggests that the nature of the task used to expose learners to L2 vocabulary also plays a significant role in L2 vocabulary learning. According to the type of processing resource allocation model or TOPRA model (Barcroft, 2002; Barcroft & Sommers, 2005; Kida & Barcroft, 2018), there are two types of processing that can occur during L2 vocabulary study, form processing and semantic processing. Form processing focuses resources on encoding the novel word form (L2 item memory) while semantic processing emphasizes the semantic properties of the target L2 item. According to the model, how we allocate our cognitive resources during study can impact memory for the word form and for the semantic meaning separately. Our available pool of cognitive resources is limited (see Kahneman, 1973); therefore, the allocation of cognitive resources to any one type of processing, either form processing or semantic processing, can improve retention for its respective type of memory, but it may come at a cost to the other type of memory.

For example, a method of study that encourages attention to semantic information instead of the L2 form can improve memory for the L1 at the cost of memory for the L2 word form (Dessenberger & Sommers, 2020; Harji et al., 2010; Ina, 2014). Dessenberger and Sommers had participants make inferences about the meaning of an L2 word form based on surrounding semantic context (i.e., L2 word forms embedded in native language sentences: "Cowboys often ride *farasi*"). Participants had to focus on the L1 meaning of the word and as a result their ability to recall the L1 word increased during a subsequent free recall test, but participants were less likely to recall the L2 word form during a free recall test compared to a control condition. According to the TOPRA model, these results are likely due to increased allocation of cognitive resources to the processing of the semantic meaning which left fewer resources dedicated to processing the novel word form. The lack of processing results in a decrease in memory retention for L2 word forms when compared to a control condition.

The benefits of type of processing also apply to the L2 word form as a procedure that emphasizes the processing of the L2 word form can improve memory for said form. For example Barcroft and Sommers (2005) had participants learn L1-L2 words using audio recordings where a speaker would say the L2 word aloud while a picture of the item was on the screen.

Participants encountered each word six times, but the speaker for each word could vary such that some words were only spoken by a single speaker six times, some words were spoken by three different speakers twice each, and the rest were spoken by six different speakers once each. While each speaker said the exact same word, because each voice had subtle but unique variations in their pronunciation, listeners could encounter a variety of different word forms for each word. This variety makes understanding and perceiving the word more difficult, but the increase in perception difficulty demanded additional attention and processing as the participants had to adapt to each voice as the number of speakers increased. The result was that as the number of speakers increased, the participants' ability to recall the L2 word forms also increased.

Additionally, later iterations of the TOPRA model also considered a mapping component and suggested that increased processing demands of form or semantic meaning can actually decrease learning for the association between L1 meanings and the L2 word forms (Kida & Barcroft, 2018). Kida and Barcroft had participants learning L1-L2 word pairs and either focus on the meaning (pleasantness ratings), focus on the form (count the number of letters in the word form) or on neither in hopes of promoting memory for the association (self-directed memorization). On the final cued recall tests (a test procedure that relies heavily on associative memory), words that were studied with an emphasis on meaning or form were recalled at a lower rate than when emphasis was not placed on either type of processing. These results lend further support to a three-component model of L2 learning such that any additional processing to one type of memory (L1 item, L2 item, or L1-L2 association) can come at the cost of the other types of memory.

The three-component model of L2 vocabulary learning and the TOPRA model both suggest a need for encoding semantic memory, form memory, and an association between the two for effective L2 vocabulary learning. The TOPRA model indicates that study method can have significant impact on learning outcomes for semantic and form memory, it may also be the case that if study method emphasizes the third type of memory (associative memory), such emphasis may promote improved memory of the association.

1.4.1 Retrieval practice and L2 vocabulary learning

One of the most well-researched and robust methods for improving memory for L2 vocabulary is through intentional learning, specifically retrieval practice (Candry et al., 2020; Kang et al., 2013; Rowland, 2014). The benefits of retrieval practice, commonly known as the *testing effect*, refers to the increase in memory retention after retrieving a memory during study in comparison to control conditions such as rereading information (Roediger & Karpicke, 2006; Rowland, 2014; Tse et al., 2010). Essentially testing your memory for vocabulary words through means such as flash cards or mobile language learning applications can improve the longevity of the memory compared to studying the same words by reading them multiple times (Akifumi, 2016; Fritz et al., 2007; Kang et al., 2013). For example, Kang and colleagues had participants study Hebrew-English word pairs by either retrieval practice or by repeating the words aloud during study. Retrieval practice in comparison to the repeating aloud condition resulted in a significant increase in both comprehension and the ability to produce the words.

The type of testing used during retrieval practice matters as some types of testing may emphasize different types of processing that, according the TOPRA model, may influence learning outcomes. For example, during language learning one of the most common forms of testing is cued recall testing which presents a cue word and requests the learner retrieve the corresponding target translation. There are two forms of cued recall testing, either through productive retrieval (L1 cue with L2 target) or receptive retrieval (L2 cue with L1 target) and the memory outcomes can depend on which type of cued recall is utilized. For example, Akifumi

(2016) had participants study 8 picture-L2 pairs a total of 12 times followed by either receptive retrieval practice which provided the L2 word as a cue and requested the participant recall the L1 word, productive retrieval practice which presented the L1 word as a cue and requested the participant recall the L2 word, or a restudy control condition which presented the L1-L2 word pairs such that it could be studied an additional time. Participants were then given both a productive and receptive final test over all 8 words in the training condition. Results indicated that productive cued recall retrieval practice improved scores on the final productive test compared to the other two conditions but both receptive and productive learning improved scores on the receptive final test compared to the control condition.

When Akifumi's (2016) findings are viewed from the perspective of the TOPRA model, it may be the case that productive retrieval practice training (L1 cue with L2 target) emphasized form processing thereby improving its encoding. However, we would then expect to see decreases in the receptive final test after productive retrieval practice, a finding that has been found in previous work (Webb, 2005), but there was no difference on the receptive test when training was conducted with receptive or productive recall. One explanation is that only L2 item memory is affected by retrieval direction (productive or receptive retrieval) and the L1-L2 associative memory is not impacted since both versions of the cued recall process necessitate retrieval of the associative memory which then bolsters its longevity through the testing effect. Effectively, during productive retrieval, both the associative memory and the L2 word form need to be recalled which then improves the retention of associative memory and the L2 item memory, respectively due to the benefits of retrieval practice. This suggests that during receptive recall, which requires the associative memory and the L1 memory be recalled, both the associative memory and the L1 item memory are improved which suggests that retrieval direction is less important if the goal is to improve associative memory for L1-L2 word pairs. Older adults are known to struggle with forming new associative memories (Naveh-Benjamin et al., 2009; Old & Naveh-Benjamin, 2008), but since both receptive and productive retrieval potentially improve memory for novel associations, either might be an effective tool for older adults who are seeking to improve memory for novel associations. In the current study, we will examine the influence of type of retrieval on memory for L2 vocabulary for both younger and older adults.

1.4.2 Age-related differences in retrieval practice benefits

Similar to younger adults, older adults also derive a memory benefit from retrieval practice compared to control conditions, at least for L1-L1 word pairs and prose material (Coane, 2013; Guran et al., 2020; Meyer & Logan, 2013; Rogalski et al., 2014; Tse et al., 2010). For example, Meyer and Logan had younger and older adults study native language passages followed by either a second study session (restudy) or a multiple-choice test (retrieval practice). Both age groups demonstrated significant testing effects on the subsequent final cued recall test, meaning that the retrieval practice group remembered more information compared to the restudy group. Additionally, after controlling for intelligence, the size of the testing effects was consistent between the age groups such that they experienced a nearly equivalent benefit of testing.

However, the literature on differences between older and younger adult L1-L2 word pair learning that relies on testing has been inconsistent and typically lacks control conditions for comparison (Gordon & Clark, 1974; Service & Craik, 1993; Whiting et al., 2011). For example, Service and Craik (1993) investigated younger and older adults’ ability to recall Finnish-English word pairs and found that older adults recalled the pairs at a significantly lower rate compared to their young adult counterparts. A total of 20 participants per age group were asked to study a list of 8 English-Finnish word pairs by listening to each word pair and repeating the words aloud.

Participants were then immediately given a cued recall task that tested their memory by giving the English cue word and asking participants to reproduce the Finnish target word. This process repeated four times and the results of the fourth test were analyzed, which means the first three tests could be considered retrieval practice while the fourth test can be considered the final test. Older adults performed significantly worse than their younger counterparts on the final test (the earlier tests were not analyzed).

Consistent with Service and Craik, Gordon and Clark (1974) investigated item memory differences using L1 and nonword material. In their study, 22 older and 22 younger adults studied lists consisting of a word and a nonsense syllable. Participants then completed two recognition tests with the first considered a retrieval practice and the second a final test. Test items consisted of two lists, one of single words and one of nonwords. Older adults performed significantly worse on both word and nonword recognition compared to young adults on both the first and second recognition tests. These findings, as well as those of Service and Craik (1993), are consistent with the idea that older adults have greater difficulty encoding and recognizing or retrieving both familiar and novel information, likely due to age-related cognitive decline.

Not all studies investigating age-related memory decline for nonwords indicates an advantage for younger adults compared to older adults when it comes to memory retention. Whiting and colleagues (2011) had older (N=17) and younger adults (N=11) learn new nonword names for familiar objects and found no differences between the age groups. Participants learned the nonword-object pairings during 5 learning sessions with exposure to the novel pairing once per session and two tests were given at the end of each session, one cued recall test and one associative recognition test with post tests given after one week and again after one month. Older and younger adults both improved in their ability to recall the correct name given the object with each subsequent session and both age groups improved in the associative recognition task with each session. Interestingly, across all sessions and during the post-test measures, there were no age differences in the recall test nor was there an age-difference in the associative recognition test. These results suggest that there are no reliable differences in associative memory between older and younger adults after retrieval practice.

1.4.3 Limitations of previous research

The limited research on older adult memory for L2 vocabulary after retrieval practice is inconsistent with respect to both results and methodology (Gordon & Clark, 1974; Service & Craik, 1993; Whiting et al., 2011). Additionally, the lack of control conditions makes it difficult if not impossible to determine whether retrieval practice has an impact on older adult L2 vocabulary learning that differs from young adults relative to a control condition (it could very well be that retrieval practice does not differ from more traditional forms of study in terms of words/nonwords retained for older adults). Finally, the sample sizes used in all three studies (i.e. Gordon & Clark, 1974; Service & Craik, 1993; Whiting et al., 2011) were relatively small (N = 20 per group) if the goal was to detect the effects of age on memory (Rhodes et al., 2019). This leaves an unclear picture of the impact of retrieval practice for L1-L2 word pair learning in older adults.

Study method can have a strong impact on L2 vocabulary learning, and over emphasis on one aspect of an L2 vocabulary word during study (e.g. emphasizing L2 item or L1-L2 associative memory) can strengthen that aspect but comes at the cost of learning other aspects of the word. Retrieval practice is one of the most robust methods for learning L2 vocabulary, but it is unclear how retrieval practice impacts older adult L2 vocabulary learning, with several studies highlighting differences in memory outcomes between older and young adults after retrieval practice (Gordon & Clark, 1974; Service & Craik, 1993) and another study (Whiting et al., 2011) suggesting no differences. In addition, the studies were relatively underpowered due to sample size limitations. It is also unclear how different forms of cued recall practice (productive vs receptive) affect younger and older adults' memory for L2 item memory compared to L1-L2 associative memory as no studies have compared these types of retrieval across age groups.

1.5 Rationale and specific aims

Prior research suggests older adults have greater difficulty learning new L2 vocabulary compared to young adults and this difficulty may stem from age-related declines in episodic and working memory capabilities (Mackey & Sachs, 2012; Naveh-Benjamin, 2000; Schaie, 1993). However, the research to date has been limited in nature, with few studies suggesting an age-related deficit in L2 vocabulary learning and, while it has been suggested that episodic and working memory declines play a role, it has yet to be confirmed. The present study sought to expand our understanding of L2 vocabulary learning in older and younger adults by applying the three-component model of vocabulary that stipulates the encoding and recall of L1 item memory, L2 item memory, and the L1-L2 associative memory. Additionally, the present study sought to understand the unique contributions of episodic memory and working memory to the vocabulary learning task with respect to the three-component model. Finally, the present study sought to resolve the inconsistencies in the previous research on L2 vocabulary learning involving retrieval practice by comparing the benefits of retrieval to a control condition with adequate power to detect age-related differences.

1.5.1 Aim 1: Determine whether younger and older adults differ in L1 item memory, L2 item memory, and associative memory abilities during L1-L2 pair learning.

Older adults are known to have poorer episodic memory capabilities compared to their young adult counterparts, with even greater declines seen in associative memories compared to item memory (Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008; Schaie, 1993). L2 vocabulary learning necessitates encoding L1 items, L2 items, and the association between them. According to Naveh-Benajmin (2000), older adults have poorer recognition for nonwords than for words (a suitable proxy for L2 items) and older adults also have decreased recognition for L1-nonword association compared to either L1 items or nonword items alone. However, this finding is contested by Badham and Maylor (2011) who found no additional deficit to associative memories when using nonword-nonword pairs, as well as Whiting et al. (2011) who found no age differences for associative memories when using nonword-object pairs as stimuli. Aim 1 sought to resolve the ambiguity in previous work and determine whether L1 item memory differs from L2 item memory in both age groups and how memory for the association differs from each type of item memory.

Summary of Methods. A fully detailed methodology can be found in the methods section of Experiment 1. To address Aim 1, Experiment 1 trained participants on L1-L2 word pairs twice followed by a 24h delay period. After the delay period, participants were asked to complete three recognition tests: an L1 item test, an L2 item test, and an associative memory test. Testing procedures were adapted from Naveh-Benjamin (2000) such that the item tests presented a single item from the appropriate list (either L1 or L2) and asked participants to specify if the item was *old* or *new*. Half of the words were taken from the studied list (*old*) and half were novel words (*new*). For the associative test, word pairs were taken from the studied list, but for half of the pairs, the L1 and L2 words were shuffled across pairs, producing new L1-L2 pairings. Participants were asked if the pairings were old (same as studied list) or new (different from studied list). Performance on all three tests was compared within and between age groups.

Hypotheses. Predicted outcomes constructed with hypothetical data can be seen in Figure 4. Similar to the findings of Naveh-Benjamin (2000), older adults were predicted to perform worse on recognition tests for L1 item memory, L2 item memory, and L1-L2 associative memory compared to younger adults, with significantly greater declines for associative memory compared to either item memory tests for older adults. Due to the increased cognitive demands of encoding unfamiliar word forms compared to familiar words, recognition rates for L2 item memory were predicted to be lower for both age groups compared to L1 item memory, but older adults were predicted to have significantly greater differences between L1 and L2 items.

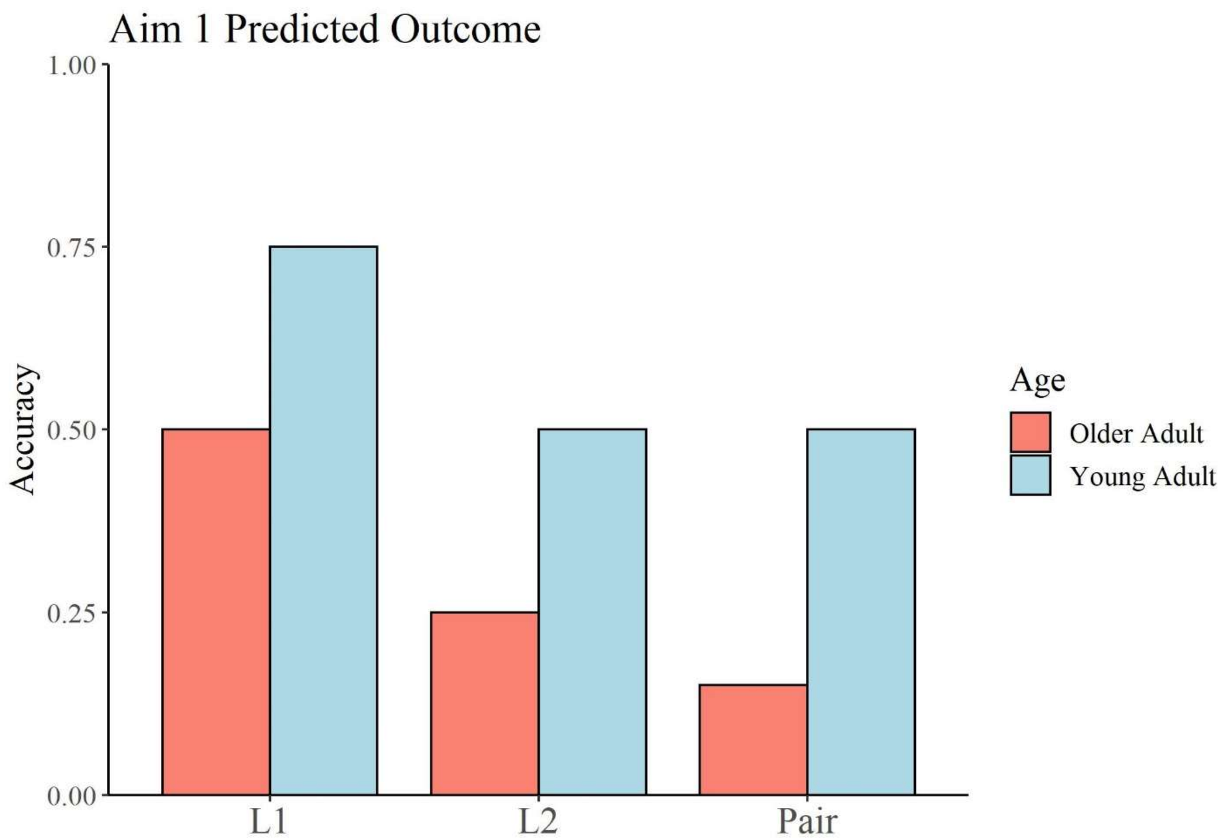

Figure 4. Predicted results for Aim1. Young adults were predicted to outperform older adults on all three recognition tests. I also predicted increased age-related deficits for L2 item memory compared to L1 item memory and an even greater deficits for associative memory compared to L2 item memory.

Statistical Analyses. A multi-level logistic regression model was used to analyze hits and false alarms with test type (L1 item memory vs L2 item memory vs L1-L2 associative memory) and age group as fixed effects and individual participant serving as the level-2 error term. Effectively, the model allowed us to investigate the influence of test type for each age group as well as compare across age groups while accounting for individual differences within each participant.

Summary. Older adults are known to have diminished episodic memory capabilities in comparison to young adults, with age differences significantly greater for associative memory compared to item memory for L1-L1 materials. However, relatively little is known regarding age-related decline in L1-L2 learning (Naveh-Benjamin, 2000; Rhodes et al., 2019). Additionally, the distinction between item memory and associative memory is less evident with young adults and L1-L1 material, but there have been few studies investigating young adult L1-L2 differences in associative and item memory (Naveh-Benjamin, 2000; Vaughn & Rawson, 2011). Aim 1 sought to expand our understanding of the L2 vocabulary learning process by exploring differences in L1 item, L2 item, and L1-L2 associative memory in young adults, as well as investigating the influence of age-related cognitive decline on all three types of memory. The primary goal was to identify and isolate the areas within the vocabulary learning process that presented the most difficulty for the language learner in both age groups so that future interventions could specifically target these areas during study.

1.5.2 Aim 2: Identify the relationship between episodic memory and L2 vocabulary learning.

Declines in episodic memory are a natural part of the aging process that potentially impact the ability to acquire new L2 vocabulary. However, the influence of language type (L1 vs L2) on various forms of episodic memory such as item memory and associative memory has yet to be explored. L1 words in an L1-L1 word pair have known semantic meanings which may facilitate pair memory as learners can form connections between the semantic meanings even if the two words are semantically unrelated (e.g. for the unrelated word pair *cat-truck*, a learner might envision a cat driving a truck or a truck in the shape of a cat, allowing them to recall the bizarre image and remember that *cat* and *truck* were a word pair). For L1-L2 learning, the novel L2 item has no intrinsic meaning to be associated with the L1, and a similar strategy cannot be adopted without extra steps involved. Therefore, it is unknown whether the influence of episodic memory on L1-L2 learning is similar to that of L1-L1. Aim 2 sought to investigate the influence of episodic memory on memory for L1-L2 word pairs by comparing memory performance on L1-L1 word pairs to memory performance for L1-L2 pairs to determine whether item or associative memory is impacted by the inclusion of the L2 word forms in the pairing. Additionally, Aim 2 sought to determine whether age-related declines in episodic memory for L1-L2 word pairs is greater than the impact of age-related decline on L1-L1 word pairs.

Summary of Methods. A fully detailed methodology can be found in the methods sections of Experiment 1. To investigate Aim 2, participants were asked to study L1-L1 word pairs in addition to L1-L2 word pairs. I classified recognition for the L1-L1 items and associations as the measure of a participant's episodic memory ability. This ability was then compared to recognition for L1 items, L2 items, and their association to determine the influence of language on each of these measures. Testing procedures were the same as those of Aim 1 for both sets of word-pair lists. For Aim 2, item memory comparisons were made based on the role of the word (either cue or target) with pair type (L1-L1 vs L1-L2) as a fixed effect. For example, in the word pairs *Truck-Dog* (L1-L1) and *Cat-Gato* (L1-L2), *Truck* and *Cat* are both cue words and would appear on the cue item memory test while *Dog* and *Gato* are both target words and would appear on the target memory test.

Hypotheses. Predicted outcomes constructed with hypothetical data can be seen in Figure 5. Similar to the findings presented by Multhuap et al. (1996) who examined memory for words and nonwords, age-related decline was predicted be evident for all recognition tests, but age-related memory declines were predicted to be greater for L1-L2 pairs compared to the L1-L1 word pairs. Both older and young adults were expected to show poorer memory for L1-L2 than for L1-L1 pairs, with differences between the two pair types greater for the older adults. Associative memory was expected to be similar to item memory in young adults for L1-L1 pairs, but age-related associative deficits were expected to lead to reduced recognition for associations compared with items in older adults. I predicted there would be greater declines as a result of pair-type for older adults compared to young adults, which suggests that age is more likely to impact the ability to recognize novel word forms as well as their associated meanings compared to familiar word forms and their associated meanings. Such results would be similar findings presented by Gordon and Clark (1974), who found greater age-related disparities between nonword recognition compared to word recognition as well as the case studies of patients with amnesia indicating that impaired episodic memory led to greater difficulty acquiring novel L2 vocabulary (Hirst et al., 1988; Verfaellie et al., 1995).

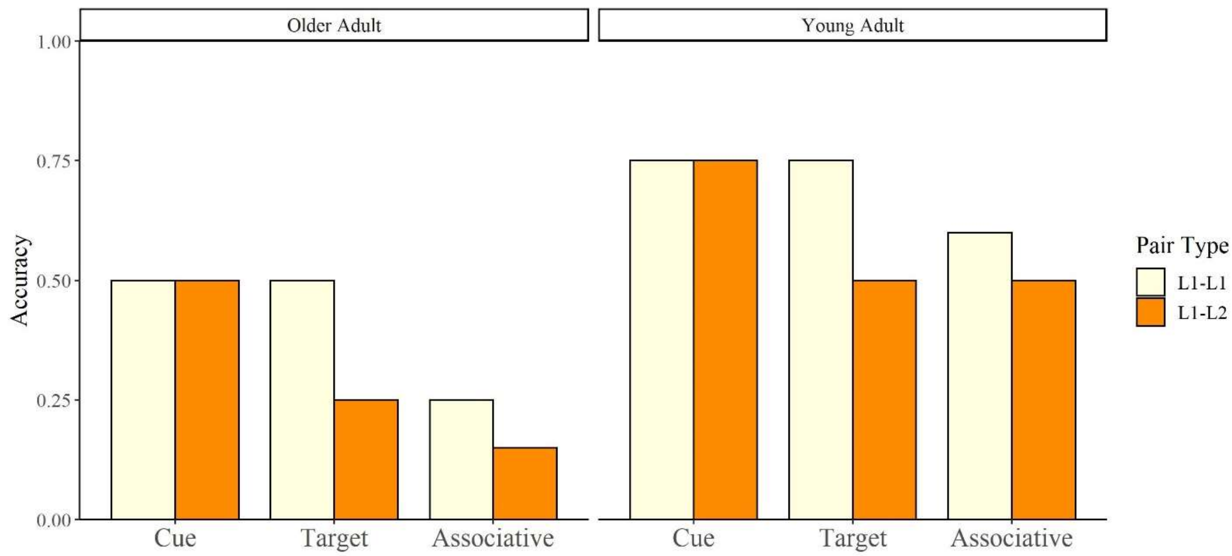

Figure 5. Predicted Results for Aim 2. Switching from L1-L1 to L1-L2 word pairs was predicted to have a greater impact on older adults compared to younger adults, specifically for target item memory. However, due to age-related associative memory declines, I predicted significant differences between L1-L1 and L1-L2 associative memory for both younger and older adults.

Statistical Analyses. A multi-level regression model with test type (cue memory vs target memory vs associative memory), pair type (L1-L2 vs L1-L1) and age group as fixed effects was used to analyze each of the variables' influence on recognition. Individual participant error terms were included in the model to account for individual differences between the test scores.

Summary. The influence of language (L1 vs L2) on memory is clear for young adults as they are more proficient at recalling native language material compared to foreign language material (Davies, 1976; Multhaup et al., 1996; Webb, 2008). For example, Webb (2008) compared receptive translation (the ability to translate a given L2 word to its L1 meaning) to productive translation (the ability to translate a given L1 meaning to its respective L2 word form) and found that the average language learner was much more proficient at receptive vocabulary compared to productive translation. However, it is unknown how older adults are impacted by recognition in

L1 compared to L2. The proposed study sought to compare memory for L1-L1 pairs to L1-L2 pairs, to determine what the influence of the presence of the L2 has on memory in older adults relative to both their recognition of L1-L1 pairs and of younger adults recognition of L1-2 pairs. Essentially, the present study used recognition for the L1-L1 stimuli as a proxy for an individual difference measure of episodic memory ability, to explore how changes in language may impact older adults in comparison to younger adults.

1.5.3 Aim 3: Identify the unique contributions of working memory to L2 vocabulary learning in older and younger adults.

As was the case for episodic memory, working memory plays a significant role in L2 vocabulary learning (Linck et al., 2014; Papagno et al., 1991) and suffers from age-related decline (Park et al., 2002; Salthouse, 2009). Working memory plays a stronger role in the case of novice learners than advanced learners because the latter are better able to rely on existing semantic information to encode novel L2 forms, a process known as scaffolding (Papagno et al., 1991). Older adults have a diminished working memory capacity which means they potentially are even more sensitive to the working memory demands of L2 learning placed on novice learners. Aim 3 sought to determine whether the relationship between working memory and L1-L2 word pair learning is consistent between age groups when L1 item, L2 item, and associative memory are measured separately.

Summary of Methods. A detailed methodology can be found in the methods sections of Experiment 1. In addition to the L1-L1 and L1-L2 word-pair lists and subsequent tests, participants will be asked to complete an automated online working memory battery (Dessenberger et al., in prep). The battery provides a latent measure of working memory that will be included as a covariate to determine the influence of working memory on L1 item memory, L2 item memory, and L1-L2 associative memory. Due to concerns regarding power and the demands of individual difference research, the episodic memory ability demonstrated in Aim 2 (the L1-L1 word list) will not be included in the full model. Additionally, the working memory variable will be converted into a 3-factor variable (low working memory, medium working memory, and high working memory) to preserve statistical power and allow for within group comparisons (low working memory older adults vs high working memory older adults) as well as cross group comparisons (low working memory young adults vs low working memory older adults).

Hypotheses. Predicted outcomes constructed with hypothetical data can be seen in Figure 6. Working memory was predicted to play a significant role in all three types of memory (L1 item, L2 item, and L1-L2 associative memory) such that greater working memory would lead to improved recognition. Working memory was predicted to play a greater role for L2 item memory and L1-L2 associative memory compared to L1 item memory, which would suggest that it plays a significant role in encoding novel word forms, likely due to the phonological loop and episodic buffer (Papagno et al., 1991; Wang et al., 2017). Working memory was predicted to have a significantly greater relationship for both L2 item and associative memory for older adults compared to its relationship with younger adult L2 item and associative memory, which would suggest that older adults are more reliant on working memory for encoding novel L2 word forms compared to young adults who may not be as sensitive to working memory demands, a finding supported by previous work in the literature (Mackey & Sachs, 2012; van der Hoeven & de Bot, 2012).

Statistical Analyses. A multi-level regression model measured the influence of test type (L1 item memory vs L2 item memory vs L1-L2 associative memory) and age-group as fixed effects with working memory as a covariate on recognition memory. Individual participant error terms will be included in the second level of the model to account for individual differences between the test scores.

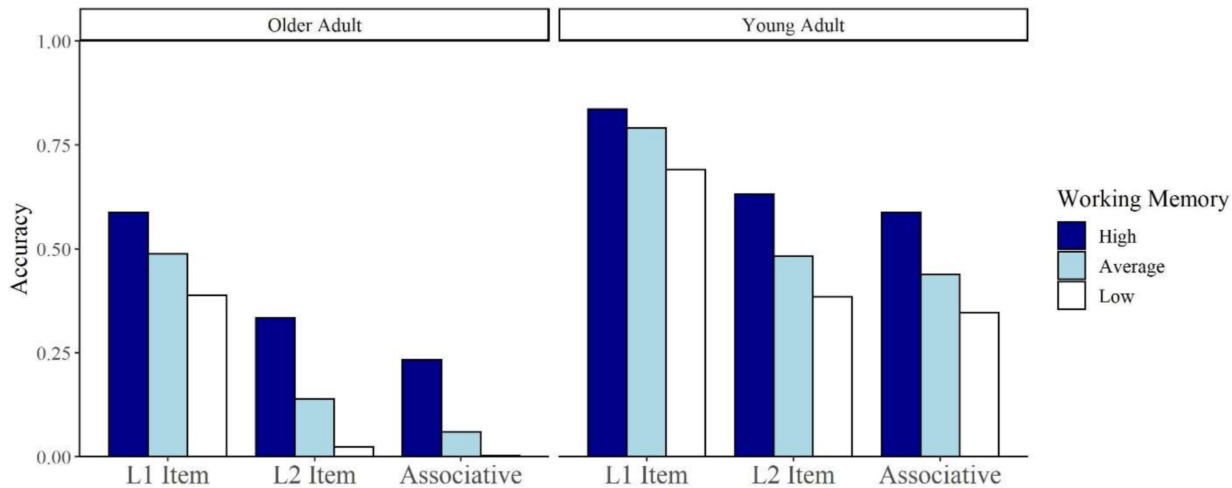

Figure 6. Predicted Results for Aim 3. Working memory was predicted to have a larger impact on older adults, such that decreasing working memory would result in steeper declines in all recognition memory test scores, with L2 item and associative memory suffering the most.

Summary. Working memory and episodic memory are distinct but related processes that are necessary for vocabulary learning but may be differentially impacted by language, age, and memory type (associative or item-memory). As indicated by Van der Hoeven & de Bot, (2012), working memory can play a pivotal role for older adult L2 learning. The present study sought to understand how working memory impacts learning for novice learners in the older adult population in comparison to the young adult population. Identification of the critical areas that suffer due to age-related decline would not only further our understanding of second language learning in general, but it will also allow for future interventions to be adapted to the learner's specifics needs to promote better learning.

1.5.4 Aim 4: Identify how receptive and productive retrieval practice affects long-term memory for L2 items and L1-L2 associative memory in older and younger adults.

Retrieval practice through cued recall is one of the most common methods for measuring L2 vocabulary learning and is coincidentally one of the most robust methods for improving memory of L2 vocabulary (Fritz et al., 2007; Kang et al., 2013). The benefits of retrieval practice are known to apply to both younger and older adults (Coane, 2013; Rogalski et al., 2014), but prior research on L1-L2 word pair learning after retrieval practice has obtained somewhat contradictory findings (Gordon & Clark, 1974; Service & Craik, 1993; Whiting et al., 2011). Additionally, the direction of the retrieval, whether it be receptive (L2 cue with L1 target) or productive (L1 cue with L2 target) may have a significant impact on L2 item memory and L1-L2 associative memory with the latter being a primary concern given that older adults are known to suffer from greater age-related associative memory deficits compared to item memory deficits. Aim 4 sought to investigate the influence of retrieval practice, specifically that of cued recall, on L1 item, L2 item, and L1-L2 associative memory in younger and older adults.

Summary of Methods. A detailed methodology can be found in the methods sections of Experiment 2. As was the case with Experiment 1 (Aims 1, 2 and 3), participants were asked to study L1-L2 word pair lists. However, the participants were asked to study the word lists in two different ways, (1) as either L1-L2 with L1 serving as the cue word on the left and the L2 word serving as the target and appearing on the right; or (2) the reverse, with an L2 cue word on the left and an L1 target word on the right. After studying the word pair list once, participants were then asked to either restudy or perform retrieval practice (restudy simply repeated the study procedure a second time). During retrieval practice, participants were given their respective cue word and asked to retrieve the target. Final test procedures were identical to those used in Aims 1, 2, and 3.

Hypotheses: Predicted outcomes constructed with hypothetical data can be seen in Figure 7. As predicted by the research on the benefits of testing (Kang et al., 2013; Roediger & Karpicke, 2006; Rowland, 2014), retrieval practice was predicted to improve memory retention compared to the restudy condition. As would be suggested by the TOPRA model (Barcroft, 2002; Barcroft & Sommers, 2005), receptive retrieval was predicted to improve L1 item memory more than productive learning because receptive retrieval (L2 cue with L1 target) emphasizes processing semantic information more than word form. Conversely, productive retrieval was predicted to improve memory for the L2 word form compared to receptive retrieval, because productive retrieval (L1 cue with L2 target) emphasizes processing of the word form rather than semantic information. I also predicted this pattern of findings will hold for both younger and older adults, as there is little evidence to suggest significant differences between the age groups and their respective benefits of retrieval practice (Rogalski et al., 2014).

Statistical Analyses: A multi-level regression model measured the influence of test type (L1 item memory vs L2 item memory vs L1-L2 associative memory), age-group, and study method (restudy vs receptive retrieval vs productive retrieval. Individual participant error terms were included in the second level of the model to account for individual differences between the test scores. Although the full model could have included two different restudy conditions (L1-L2 restudy and L2-L1 restudy), I sought to simplify the model to collapse across the restudy conditions as I predicted study direction would have no significant influence on recognition outcomes in the restudy condition.

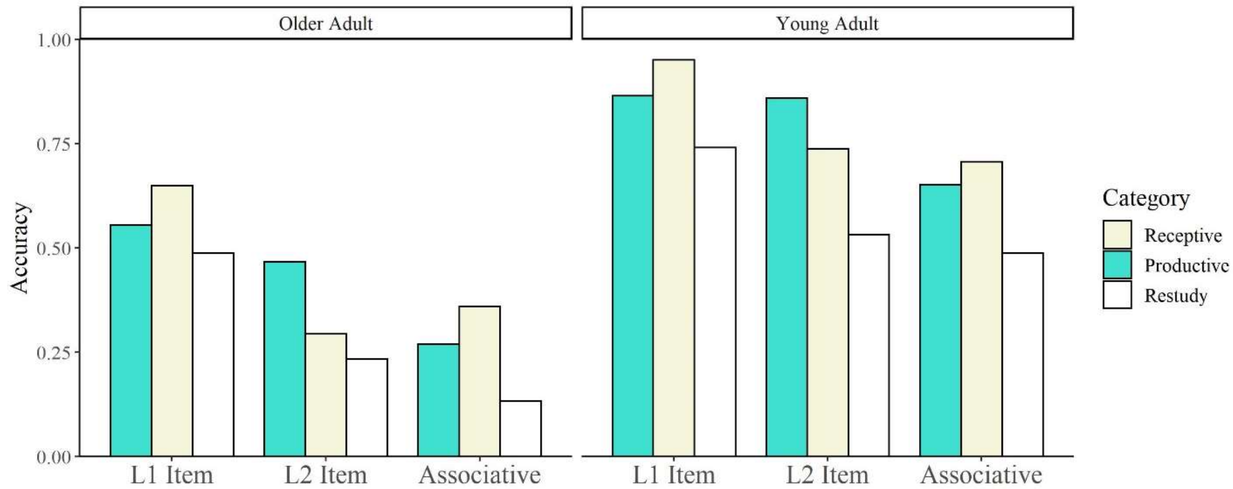

Figure 7. Predicted Results for Aim 4. Receptive recall (L2 cue with L1 target) was predicted to improve L1 item memory and associative memory in both age groups compared to productive recall (L1 cue and L2 target) and restudy. L2 productive recall was predicted to improve recognition memory compared to restudy in all three recognition tests and was predicted to improve L2 item recognition memory compared to receptive recall.

Summary. Retrieval practice is one of the most well-documented methods for improving L2 vocabulary learning, but there has been little investigation into exactly which part of the L1-L2 word pair is affected. According to the TOPRA model (Barcroft, 2002; Barcroft & Sommers, 2005), how we allocate our cognitive resources impacts memory for L2 vocabulary. If we emphasize word form during learning, we will improve L2 retention. If we focus instead on the semantic association, we will improve the associative memory retention more than memory for the word form. In comparing receptive and productive retrieval and measuring item memory and associative memory separately, I sought to highlight which method of study is best suited to addressing learners' needs. For example, if a learner has greater difficulty forming novel associations (such as older adults) they may need to rely on retrieval practice methods that focus on the association, which in this case would be receptive cued recall (the L2 is provided so less effort needs to be spent on recalling the novel word form). However, this benefit of receptive cued recall for associative memories and the benefit of productive cued recall for L2 word form learning has yet to be confirmed in both older adults and young adults.

Chapter 2: Experiment 1

Experiment 1 addressed Aim 1 by comparing item memory and associative memory for L1-L2 pairs in younger and older adults. Additionally, Experiment 1 addressed Aim 2 by including a separate study list of L1-L1 pairs to measure episodic memory ability. Finally, Experiment 1 addressed Aim 3 by including a working memory measure to investigate the impact of individual differences in working memory on item and associative recognition in younger and older adults.

2.1 Methodology

2.1.1 Participants

According to a boot-strapped power analysis, to reliably detect differences in item memory between age groups at an effect size of .65 (Old & Naveh-Benjamin, 2008), alpha level of .05, and power of .80 requires approximately 40 participants per group. The design of Experiment 1 necessitated two groups (older and younger), totaling in 80 participants. Prior research suggests age-related cognitive decline begins between 25-30 years old, so I recruited younger adults under the age of 25 and older adults who are 55 years or older. Typically, older adults are classified at 65 years or older, but age-related cognitive decline should be evident at 55 years old and allow for greater variability in working memory scores in the older population.

A total of 40 young adults (26 female, M_{age} = 22.25, SD_{age} = 2.10) and 40 older adults (27 female, M_{age} = 61.18, SD_{age} = 5.67) were recruited for the study. One younger adult participant and one older adult participant were removed from the study for failing to complete the final test. All participants were recruited from the Prolific web platform, a website dedicated to participant recruitment for research purposes. Participants were located primarily in Europe (51), with the rest located in Africa (17), North America (6), South America (2), and Australia

(2). Participants on average knew more than 1 language (M_{lang} = 1.96, SD_{lang} = 1.2). There was no difference between age groups for the number of languages known ($t(77) = 1.12$, $p = .24$), and all participants reported no familiarity with the target language (Swahili). Participants were compensated 10 dollars for their estimated 1 hour of participation.

2.1.2 Materials

An English word-pair list (L1-L1) and an English-Swahili vocabulary list (L1-L2) were included in the experiment (see Appendix I). Both lists contained 80 word-pairs with no words repeated between the lists. The Swahili-English language word pairs were chosen based on retrieval norms presented by Bangert and Heydarian (2017) while the English language pairs were drawn from the English Lexicon Project (Balota et al., 2007) such that they consisted of unrelated nouns that were similar in word frequency (M_{Freq_HAL} = 8.5) compared to the English words contained in the Swahili-English list.

A battery of working memory tests was used to assess individual differences in working memory. The battery includes an AlphaSpan task, the N-back task, the Operation Span task and the reverse digit span task. The AlphaSpan task was based on the task used by Craik et al. (2018). Two to seven words were presented in a random order and the maximum span length participants could retain and repeat in correct alphabetical order was recorded. The N-Back task was based on work by Kirchner (1958) and consisted of the presentation of a series of letters. Participants were tasked with identifying whether a presented letter matched the letter presented either 1-back, 2-back, or 3-back in the sequence with final scores being their average accuracy. The Operation Span task (based on Unsworth et al. (2005) presented a series of two-five math problems concurrently with words and asked participants to solve the math problems and retain the words. Final score was calculated as the maximum span length or the largest number of items that the participant correctly retrieved. Lastly the reverse digit span consisted of number sequences ranging from 3 to 11 digits and asked participants to repeat the number sequences in the reverse order. Final score was calculated as the maximum sequence length that participants could present in the reverse order. A latent measure of working memory was extracted from confirmatory factor analysis based on the procedure demonstrated by Dessenberger et al. (in prep).

2.1.3 Design

Participants were asked to study word pair lists followed by recognition tests after a 24hr delay. A 2x2x3 mixed design investigated the influence of age (younger vs older adults), pair type (L1-L1 vs L1-L2), and test type (cue recognition vs target recognition vs associative recognition) on memory for word pairs. Pair type was manipulated within-subject by drawing half of the studied word pairs from the L1-L1 pair list and the other half from the L1-L2 word list (presentation was blocked by pair type). Test type was manipulated within-subject as each participant was given a cue recognition test, target recognition test, and an association recognition test. Accuracy on these tests (hits-false alarms) served as the dependent variable.

Test procedures mimicked those used by Naveh-Benjamin (2000) such that each test contained two blocks, one for each language type with test order and block order counterbalanced. For the cue test for the L1-L1, participants were shown a subset of the cues from the word pair list along with new L1 items they had not seen before. Participants were asked to determine whether the item was *new* or *old*. The cue test for L1-L2 words followed an identical procedure as all cue words were still native language L1 words. For the L1-L1 word list, the target test followed the same procedure as the cue test, presenting studied words from the target position mixed with L1 items they had not seen before. The target test for the L1-L2 list was similar, but presented the L2 items the participants had studied along with L2 words the participants had not seen before. For the associative tests, both the L1-L1 list and L1-L2 list followed the same procedure: both tests contained pairs from the original study list, but for half of the pairings, the targets were shuffled across the pairs such that the cues each had new targets paired with them. Participants were then asked if the pairing was the *same* as they had seen before or *different*. For example, if originally the L1-L1 word pair was studied as *cat-mouse*, the word pair may appear again as *cat-mouse* or cat may be paired with a different word that had also been seen during the study phase. For the L1-L2 word pairs, the procedure was the same, with some L1 words appearing alongside the same L2 translations seen during the study phase while other L1 words appeared alongside different L2 translations. In all cases, during the associative test every word (both the L1 item and L2 item) had been seen before during the study phase so all were equally familiar and the test was able to isolate and measure participants' ability to recognize correct pairings. Additionally, all pairs were test-exclusive meaning if a word or its corresponding paired word appeared on one test, neither would appear on any other test. For example, for the word pair *cat-gato*, if *cat* appeared on the cue test, *gato* would not appear in the target test and neither *cat* nor *gato* would appear on the pair test.

2.1.4 Procedure

Participants were asked to complete the online working memory battery prior to signing up for the word-pair study task. The working memory task took approximately 30 minutes and was completed using a publicly available Psychopy program published on the Pavlovia web platform (Peirce et al., 2019). The working memory measure was completed up to 3 months prior to the word-pair study task.

The word-pair study task had participants view 80 word-pairs drawn evenly from each list (L1-L1 list and L1-L2 list) one at a time for 5s each (trial length is based on similar experiments, see Naveh-Benjamin, 2000). Cues appeared on the left side of the screen, targets appeared on the right, separated by a hyphen. Presentation order was blocked by list such that the first 40 pairs came from the same list while the remaining 40 were drawn from the other list (list order was counterbalanced). The 40-word blocks were further subdivided into four groups of 10 words each and each group of 10 was seen twice and within-list order randomized each time. For example, in the L1-L2 word pair block, participants studied 10 L1-L2 word pairs (one at a time, 5 seconds each) and then studied the same 10 word-pairs (one at a time, 5 seconds each), then moved onto the next set of 10 word-pairs.

Participants were then dismissed for 24h followed by three recognition tests. For the cue and target recognition tests, participants were shown a single word and were given up to 10s to determine whether the word was seen during the study session or whether it was a new word. While young adults were not expected to need the full 10s, prior research has shown that limiting response time for older adults can negatively impact memory performance (Mohanty & Naveh-Benjamin, 2018). As a result, shortened response times would exacerbate age-related differences in test scores attributable to processing speed rather than recognition abilities alone. The focus of this dissertation is age-related differences in memory and not processing speed, therefore I allowed for a longer response time. The associative memory test displayed word pairs for 10s and participants were asked whether the pairing was the same as the one they saw during study or if it was different. All tests contained two blocks (one for each list), test and block order were counterbalanced, and within-block order was randomized.

2.2 Results

Descriptive statistics for Experiment 1 can be found in Table 1. Recognition test performance was calculated as hits-false alarms and served as the dependent variable (hit rates and false alarm rates can be found in Appendix II). Three separate multi-level logistic regression models were constructed. The first model addressed Aim 1: assess age-related item and associative memory in L2 vocabulary learning. The model only examined recognition for words from the L1-L2 study list and included only test type (cue memory vs target memory vs associative memory) and age group (young vs old) as fixed effects. The second model addressed Aim 2: identify the influence of episodic memory ability in L1-L2 item and associative memory in L2 vocabulary learning. The model examined recognition for words from both lists and included test type, age, and language pair (L1-L1 vs L1-L2) as fixed effects (the L1-L1 words served as a proxy measure for episodic memory abilities). The third model addressed Aim 3: identify the influence of working memory on age-related item and associative memory in L2 vocabulary learning. The model only examined words from L1-L2 study list and included test type and age as fixed effects with working memory as a 3-factor fixed effect (high, medium, and low working memory).

Table 1: Descriptive Statistics for Experiment 1 Outcomes

Language	Age Group	Test Type	Mean	SD
L1-L1	Older Adult	Cue	.63	.48
		Pair	.65	.48
		Target	.68	.47
	Young Adult	Cue	.65	.48
		Pair	.67	.47
		Target	.63	.48
L1-L2	Older Adult	Cue	.64	.48
		Pair	.65	.48
		Target	.68	.46
	Young Adult	Cue	.61	.49
		Pair	.65	.48
		Target	.70	.46

For each aim, a multi-level logistic regression model was constructed in a step-wise fashion starting with a base model that contained no predictors, adding a single predictor or interaction term at each step. Each model was compared to the previous model using chi-square goodness of fit tests. The individual participant was set as the level-2 error term to account for differences across participants. The full model was determined for each aim a priori. Models were created using the R statistical analysis software with the *lme4* package (Bates et al., 2015; R Core Team, 2018), and post-hoc analyses were conducted using the *multcomp* package (Horthorn et al., 2008). Model formula, description, and code are all available in the supplemental materials. All post-hoc *p*-values were corrected using the Bonferroni method.

2.2.1 Aim 1 Results

To assess age-related item and associative memory in L2 vocabulary learning, a multilevel logistic regression model analyzed the influence of age (young adult vs older adult) and test-type (L1 item memory vs L2 item memory vs L1-L2 pair memory) on recognition memory. It should be noted that data for the L1-L1 pairs were not included in this analysis. The coefficients for the full model can be found in Table 2.

Step-wise Analysis. Age did not play a role in memory outcomes as adding the age predictor to the model did not significantly improve model fit compared to the base model ($X(1) = .003$, $p = .96$). The type of test did moderate test outcomes as adding the test-type significantly improved model fit ($X(2) = 14.27$, $p < .001$). The type of test was not significantly influenced by the age variable as adding the interaction term did not significantly improve fit ($X(2)$ 1.82, $p = .40$).

Table 2: Coefficients for Aim 1 Full Model
A multi-level logistic regression model fixed effects output for the final recognition test scores (Intercept is the Older Adult cue memory test). β and standard error are presented in logit units and effect size presented in Odds-Ratios (OR) with 95% Confidence Intervals (CI).

	Estimate	Std. Error	z value	*p* value	OR	95% CI
Intercept	.61	.11	5.51	<.001	1.84	[1.64, 2.05]
Pair memory	.04	.11	.33	.75	1.04	[.93, 1.15]
Target Memory	.19	.11	1.7	.09	1.2	[1.08, 1.34]
Young Adult	-.12	.15	-.76	.45	.89	[.76, 1.04]
Pair Memory x Young Adult	.14	.15	.90	.37	1.15	[.98, 1.33]
Target Memory x Young Adult	.20	.15	1.32	.19	1.23	[1.05, 1.43]

Post-hoc Analysis. As shown by Figure 8, post-hoc linear comparison of the full model indicated there was a significant main effect of test-type such that memory for the L2 words was greater compared to memory for the L1 ($z = 3.73$, $p < .001$). There was no significant difference between memory for L2 words and memory for the L1-L2 pairs ($z = 1.36$, $p = .52$) and there was no significant difference between memory for the L1 words and L1-L2 word pairs ($z = 2.37$, $p = .053$). These findings suggest there was no influence of age on recognition memory for any of the measured memory types. The only significant difference found was that recognition for the L1 item memory was lower than for L2 item memory and this was true for both age groups.

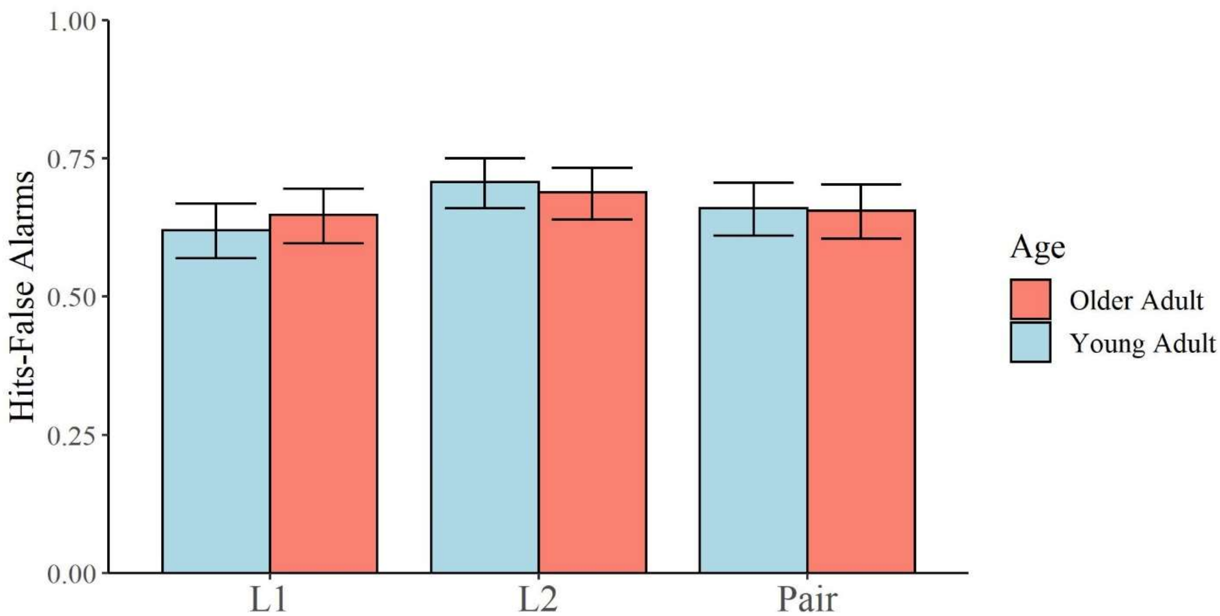

Figure 8. Aim 1 Results. Accuracy of recognition trials, separated by test type and age group. Accuracy is calculated as the proportional rate of accuracy or hits-false alarms. Error bars are 95% CI.

2.2.2 Discussion of Aim 1

Aim 1 was designed to investigate age-related differences for familiar (L1), unfamiliar (L2) items and memory for the associations between L1 and L2. As shown in Figure 8, no age-related differences for L1 item memory, L2 item memory, or L1-L2 item memory were found. This finding contradicts a large amount of literature suggesting that older adults have greater difficulty recognizing studied information compared to young adults (Chalfonte & Johnson, 1996; Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008; Rhodes et al., 2019). Moreover, another unexpected finding was that participants were actually better at recognizing L2 word forms compared with L1 items.

There are several potential explanations as to why older adults did not demonstrate memory decline relative to young adults. One explanation is that participants were presented with many more L1 than L2 words during the study and that may have diminished their abilities to distinguish L1 items at final test. Participants saw triple the number of L1 words (in addition to the L1-L2 pairs included in Aim 1 analysis, the participants also studied an equivalent number of L1-L1 words). This volume of L1 words may have produced greater difficulty in recognizing the L1 material for participants which would explain why both age groups were better able to correctly recognize the L2 word forms compared the L1 words. According to prior research, larger lists of stimuli can increased the difficulty of recognition tasks as the other stimuli can create confusion as to whether a target is new or old (Criss & Shiffrin, 2004; Underwood, 1978). However, if it was the case that participants were simply overwhelmed with the number of L1 words to maintain in their memory, we would expect older adults to suffer greater rates of memory decline in L1 recognition because older adults are known to have reduced memory abilities, but this was not the case as older adults did not differ from their young adult counterparts.

An additional possibility is that when older adults are given sufficient time during testing there are minimal age differences in recognition memory, a result reported by Mohanty and

Naveh-Benjamin (2018). In their study, participants were tasked with learning paired associates and were given recognition tests covering both item memory and associative memory. Some test items had a response deadline of 1 second while other items allowed for up to 5 seconds of response time. There was a significant interaction of age and response deadline, such that the expected age-related decline in recognition test scores was attenuated when participants had a longer time to respond. Participants in the current study were provided with up to 10s to respond. This additional time could potentially have diminished any-age related declines. Effectively, this would mean that given sufficient time to process the presented information during recognition testing, older adults do not differ from their young adult counterparts.

Although the results of Aim 1 may seem unusual given the large number of studies (Gordon & Clark, 1974; Old & Naveh-Benjamin, 2008; Rhodes et al., 2019; Salthouse, 2004) demonstrating age differences in recognition memory, the results are not unique as both Badham and Maylor (2011) as well as Whiting and colleagues (2011) a reported no differences between age groups in recognition of L1-L2 words. It may be the case that there is less of a difference between age groups for recognizing unfamiliar material (L2) because each of the words is salient without competing with memory traces of existing knowledge as is the case with familiar material (L1). The reduced influence of age-related memory decline on unfamiliar material when combined with the influence of lengthier response times during testing may have contributed to findings of no difference between the age groups.

2.2.3 Aim 2 Results

To assess the influence of episodic memory ability for L1 and L2 item memory as well as for associative memory, a multilevel logistic regression model analyzed the influence of language pair type (L1-L1 vs L1-L2), age (young adult vs older adult) and test-type (cue item memory vs target item memory vs associative pair memory) on recognition memory. The coefficients for the full model can be found in Table 3.

Table 3: Coefficients for Aim 2 Full Model
A multi-level logistic regression model fixed effects output for the final recognition test scores (Intercept is the Older Adult cue memory test for the L1-L1 word list). β and standard error are presented in logit units and effect size presented in Odds-Ratios (OR) with 95% Confidence Intervals (CI).

	Estimate	Std. Error	*z* value	*p* value	OR	95% CI
Intercept	.56	.11	5.24	<.001	1.76	[1.58, 1.96]
L2 Language	.05	.11	.48	.63	1.05	[.95, 1.17]
Young Adult	.10	.15	.69	.49	1.11	[.95, 1.29]
Pair Memory	.07	.11	.64	.52	1.07	[.96, 1.19]
Target Memory	.20	.11	1.84	.07	1.22	[1.10, 1.36]
L2 language x Young Adult	-.23	.15	-1.51	.13	.80	[.68, .93]
L2Language x Pair Memory	-.03	.15	-.23	.82	.97	[.83, 1.12]
L2 Language x Target Memory	-.02	.15	-.10	.92	.98	[.84, 1.15]
Young Adult x Pair Memory	.01	.15	.09	.93	1.01	[.87, 1.18]
Young Adult x Target Memory	-.30	.15	-1.99	.05	.74	[.63, .86]
L2 Language x Young Adult x Pair Memory	.12	.22	.57	.57	1.13	[.91, 1.4]
L2 Language x Young Adult x Target Memory	.51	.22	2.34	.02	1.66	[1.34, 2.06]

Step-wise Analysis. Neither the type of language pair nor the age variable impacted memory outcomes as adding the language-pair type (L1-L1 vs L1-L2) did not significantly improve model fit compared to the base model ($X(1) = .30$, $p = .58$) and adding the age variable also did not significantly improve model fit ($X(1) < .001$, $p = .98$). Adding the interaction term between language-pair type and age condition also did not significantly improve model fit compared to the base model ($X(1) = .06$, $p = .81$). Similar to Aim 1, adding the test-type variable (cue test vs target test vs pair test) to the model did improve model fit ($X(2) = 9.37$, $p = .01$), but adding the interaction between age and test type did not significantly improve model fit ($X(2) = 1.41$, $p = .50$). Adding the interaction term between language-pair type and test type did not significantly improve model fit compared ($X(2) = 5.83$, $p = .054$) nor did adding the three-way interaction between age, language, and test-type significantly improve model fit compared to the base model ($X(1) = 5.91$, $p = .052$).

As show by Figure 9, post-hoc linear comparison of the full model indicated there was a significant main effect of test-type such that memory for the target words was greater than memory for the cue words ($z = 3.08$, $p = .01$) There was no significant difference between memory for target words and memory for the word pairs ($z = 1.42$, $p = .46$) and there was no significant difference between memory for the cue words and memory for word pairs ($z = 1.66$, $p = .29$). As was the case with Aim 1, the only significant difference observed in the analyses addressing Aim 2 was that the difference between memory for the target item was significantly higher than for the cue item. This ability did not differentiate across language pair types which would suggest that there is an inherent difference in how memory for cue words and item words are encoded during study and this difference was consistent across age groups.

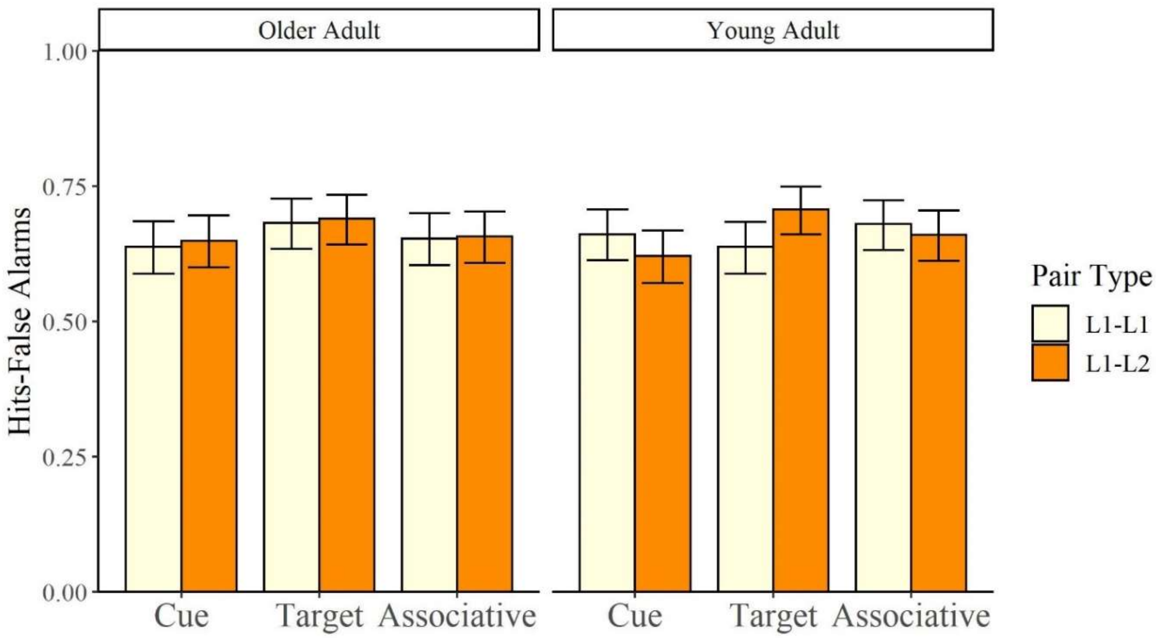

Figure 9. Aim 2 Results. Accuracy of recognition trials separated by test type, age group and language-pair type. Recognition accuracy separated by age, test type, and relative working memory ability. Accuracy is calculated as the proportional rate of accuracy or hits-false alarms. Error bars are 95% CI.

2.2.4 Discussion of Aim 2

Aim 2 sought to examine how episodic memory abilities as demonstrated by memory for L1-L1 pairs contributed to memory abilities for L1-L2 material. According to the analysis, there was no difference between the ability to recognize L1 material and the ability to recognize L2 material. Put another way, similar mechanisms likely mediate memory for both types of materials, such that if a participant demonstrates high capability during study of L1 material, they are likely to also demonstrate high capabilities when studying L2 material. Additional analyses investigating the relationship between L1 recognition ability and L2 recognition ability as well as corresponding graphics can be found in Appendix II.

The same pattern demonstrated in Aim 1 was again seen in Aim 2 which included analysis of both the L1-L1 word pairs as well as the L1-L2 word pairs. No effect of age was found in the recognition data and no effect of language pair type (L1-L1 vs L1-L2). Again, this unexpected finding may be explained by providing a lengthy response window, but something unusual was found when looking at L1-L1 words pairs in isolation. As was the case with Aim 1, there was an increase in the rates of recognition for the target word relative to the cue word, keeping in mind that for the L1-L1 words pairs, the target was simply another L1 item. This would suggest that regardless of language-pair type, there is an emphasis in focusing on the target words, or the words that appeared on screen in the second position if reading from left to right. During the study procedure, participants were informed their task was to commit the pair of words to memory as they would be tested on them later. Though the wording does not indicate any additional emphasis should be placed on the word in the target position, participants may have assumed that the target word alone rather than both cue and target would be tested.

Viewed from the perspective of the TOPRA model, increases in memory for L2 words come from an increased focus on L2 word form processing. A similar phenomenon could have occurred during the study phase for L1-L1 word pairs such that there was an increase in processing of the target word simply because it appears in the target position. As such, it is likely that participants were focusing their efforts towards encoding the target word rather than the cue word during the study phase. During study of the L1-L1 word pairs, participants were unaware that the L1 cue word held any unique significance relative to the L1 target word with the exception that the L1 cue word appeared on the left while the L1 target word appeared on the right (participants were not informed of the recognition test procedures). That being the case it would seem participants may orient their attention to the word on the right throughout the experiment. It may be due to some participants viewing the L1-L2 word pairs first and thus learning that the second word is of greater difficulty and applied a compensatory strategy that they developed during the L1-L2 study phase to the L1-L1 study phase. Future research should investigate how presentation order of L1-L2 word pairs (either presented as L1-L2 or as L2-L1) may impact learning.

While we would normally expect to see improved memory for L1 targets over L2 targets, it may be the case that memory for L1 targets suffered because participants had to three times the number of L1 items to learn. One explanation for the results is that, consistent with previous research (Gordon & Clark, 1974; Service & Craik, 1993), memory for L1 items is better than memory for L2 items, but the large volume of L1 items in Experiment 1 by comparison reduced recognition rates to the point they were equivalent to the recognition rates of L2 items. There was even a sizeable numerical advantage for L2 target memories, but it was most prominent for young adults and not older adults who are theorized to have greater difficulty with remembering target items (Old & Naveh-Benjamin, 2008).

2.2.5 Aim 3 Results

To assess age-related differences in the influence of working memory on item and associative memory deficits in L2 vocabulary learning, a multilevel logistic regression model analyzed the influence of age (young adult vs older adult), test-type (L1 item memory vs L2 item memory vs L1-L2 pair memory) and working memory (low vs medium vs high) on recognition memory.

A simple t-test compared working memory performance in older and younger adults and found that the young adults did indeed have higher levels of working memory compared to the older adults ($t(67) = 2.70$ p = .009, $d = .65$) which is consistent with previous research suggesting that working memory declines across the lifespan (Park et al., 2002). To account for this difference, scores were categorized within age group such that participants were divided based on their working memory performance within their age group. Higher performers were classified as participants with a WM score equal to the mean performance of their age group plus the standard deviation, low performers classified as those with WM score equal to the mean minus the standard deviation and everyone else scoring as a medium performer. This division based on age group allows for a relative look at how varying levels of WM affect each age group. It should be noted that as with Aim 1, data for the L1-L1 pairs were not included in this analysis. The coefficients for the full model can be found in Table 4.

Table 4: Coefficients for Aim 3 Full Model

A multi-level logistic regression model fixed effects output for the final recognition test scores (Intercept is the Older Adult cue memory test for the for older adults with High Working Memory (WM). β and standard error are presented in logit units and effect size presented in Odds-Ratios (OR) with 95% Confidence Intervals (CI).

	Estimate	Std. Error	z value	*p* value	OR	95% CI
Intercept	.94	.31	2.99	<.001	2.56	[1.87, 3.51]
Pair Memory	.27	.33	.82	.41	1.31	[.94, 1.81]
Target Memory	.10	.32	.32	.75	1.11	[.80, 1.53]
Young Adult	-.37	.44	-.85	.39	.69	[.44, 1.07]
Low WM	-.17	.44	-.40	.69	.84	[.54, 1.30]
Medium WM	-.36	.35	-1.04	.30	.70	[.49, .99]
Pair Memory x Young Adult	.22	.45	.48	.63	1.24	[.79, 1.96]
Target Memory x Young Adult	.44	.45	.97	.33	1.55	[.99, 2.42]
Pair Memory x Low WM	-.71	.44	-1.6	.11	.49	[.32, .77]
Target Memory x Low WM	-.10	.44	-.23	.82	.90	[.58, 1.40]
Pair Memory x Medium WM	-.26	.36	-.72	.47	.77	[.54, 1.10]
Target Memory x Medium WM	.13	.35	.37	.71	1.14	[.80, 1.62]
Young Adult x Low WM	.04	.6	.07	.95	1.04	[.57, 1.90]
Young Adult x Medium WM	.28	.48	.59	.55	1.33	[.82, 2.15]
Pair Memory x Young Adult x Low WM	-.21	.60	-.35	.72	.81	[.44, 1.48]
Target Memory x Young Adult x Low WM	-.29	.61	-.47	.64	.75	[.41, 1.38]
Pair Memory x Young Adult x Medium WM	.06	.49	.13	.90	1.06	[.65, 1.74]
Target Memory x Young Adult x Medium WM	-.24	.49	-.49	.63	.79	[.48, 1.29]

The full model from Aim 1 served as the base model for Aim 3 as their structure and statistical outputs are identical up until the inclusion of the working memory fixed effect. Adding the working memory categorizations did not significantly improve model fit compared to the base model ($X(2) = 4.21$, $p = .12$). Adding the interaction between working memory and test type did significantly improve model fit ($X(4) = 9.66$, $p = .047$). Adding the interaction term between working memory and age condition did not significantly improve model fit compared to the base model ($X(2) = 1.04$, $p = .59$). In addition, the three-way interaction between working memory, test type, and age did not significantly improve model fit ($X(4) = .79$, $p = .94$).

As shown by Figure 10, post-hoc linear comparison of the full model indicated there was a significant interaction between test-type and working memory such that there were significantly more L2 words correctly recognized in comparison to L1 item recognition by individuals with a medium level of WM ($z = 3.40$, $p = .01$). There was also a significant increase in the number of correctly identified word pairs for individuals with high WM relative to those with low WM ($z = 3.17$, $p = .03$). The remaining comparisons all resulted in p. > 05 and can be found in the supplemental materials section. These results suggest while working memory may play a vital role in forming associative memories it does not influence memory for individual items regardless of language.

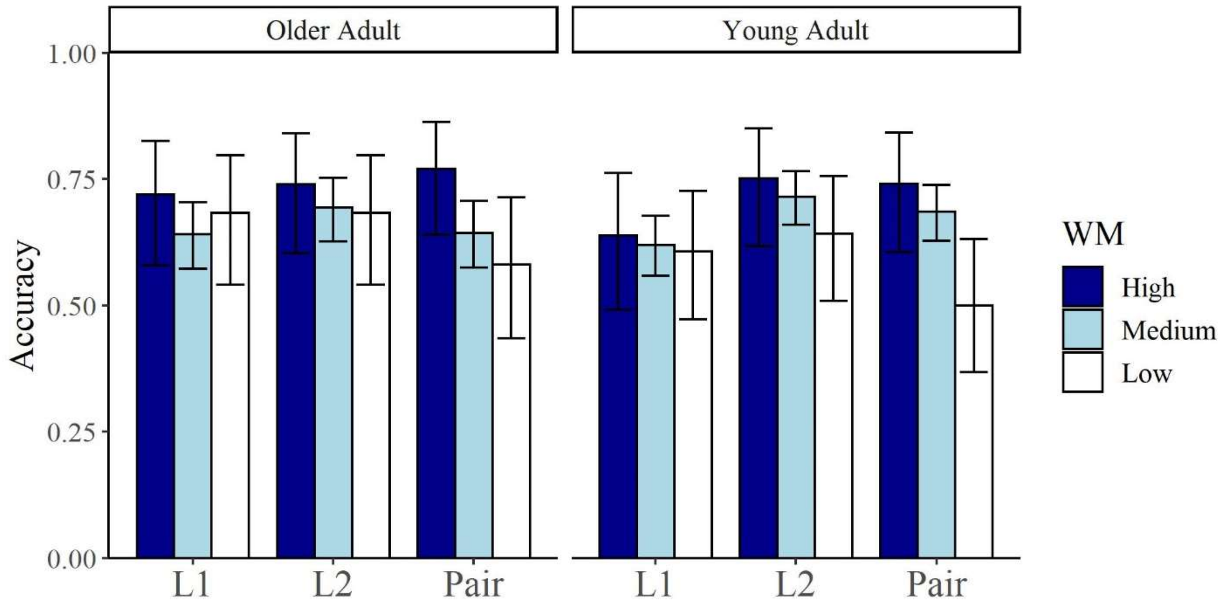

Figure 10. Aim 3 Results. Recognition accuracy separated by age, test type, and relative working memory ability. Accuracy is calculated as the proportional rate of accuracy or hits-false alarms. Error bars are 95% CI.

2.2.6 Aim 3 Discussion

Aim 3 sought to assess differences in the influence of working memory on age-related item and associative memory deficits in L2 vocabulary learning and found that working memory does indeed contribute to associative memory deficits relative to item memory performance. It should be noted that working memory scores were standardized within age-group and that older adults on average had a lower working memory score compared to their young adult counterparts. Since older adults tend to have lower working memory and I found that working memory plays a more prominent role in enabling recognition of associative memories relative to its role in item memory recognition, I conclude that age-related associative memory deficits can in part be accounted by age-related working memory decline.

According to the working memory model put forward by Baddeley (2000), the episodic buffer, one of the four components of working memory, likely plays an important role in forming connections between novel L2 word forms and their respective meanings. Wang et al. (2017) found evidence supporting this, identifying working memory as key in binding novel L2 word forms to their semantic meanings although they used images instead of written L1 meanings. The use of images to demonstrate the relationship between working memory and memory for L1-L2 associations raised concerns that potentially working memory and more specifically the episodic buffer would only play a role if the L1 was presented as an image rather than written (images are governed by the visuospatial sketchpad while written words are typically associated with the phonological loop). However, our results suggest that working memory plays are role even if the L1 and L2 items are presented in the same written modality.

The influence of working memory on the rates of recognition for associative memories would suggest that the associative deficit hypothesis put forward by Naveh-Benjamin (2000), which states our associative memory abilities decline more quickly than our item memory abilities as we age, could in part be due to the gradual decline of working memory across the lifespan (Park et al., 2002). While an age-related associative deficit was not found in Experiment 1, I would argue the results do indirectly support the associative deficit hypothesis because working memory is expected to decline across the lifespan and according to the results of Aim 3, working memory significantly related to the ability to form associative memories.

It should be noted that the participants included in Experiment 1 were all novice learners of the target L2 and it may be that the results, specifically those investigating the influence of working memory on L2 learning, do not generalize to advanced learners. Previous investigations into the influence of working memory and L2 learning suggest that as we gain more experience in the target L2 language, the less we rely on working memory to facilitate associative memory creation (Papagno et al., 1991). This is likely because advanced learners can rely more so on scaffolding or building on their existing semantic memory (Gathercole et al. 1992) to encode novel word forms and form associations much like they would with their native language. For example, verb conjugation may be a significant challenge for novice language learners that emphasizes working memory, but advanced learners need not rely on working memory to process and encode new verbs because the conjugation patterns have already been stored in long-term memory. The same could be said for nouns in latin-root based languages or eastern languages written with components or stroke patterns that inform their meaning. As we advance in our understanding of a language, the ability to process and encode becomes more reliant on existing knowledge rather than the raw processing power of working memory. As such, the results found while investigating Aim 3 may not generalize to an advanced learner population.

2.3 General Discussion of Experiment 1

The purpose of Experiment 1 was to investigate the influence of age, episodic memory ability, and working memory on L2 vocabulary learning with respect to the three-component model (see Figure 1). While each of the three aims address separate questions, each analysis resulted in a general finding of no age-related memory decline in any of the three components, including memory for L1 items, memory for L2 items and memory for L1-L2 associations regardless of episodic and working memory abilities (although the latter provides some indirect support).

2.3.1 Age-related decline and the associative deficit hypothesis

Experiment 1 did not provide evidence for an age-related decline in memory for item (L1 or L2) or associative memory. The only evidence in support of an age-related memory decline is that older adults have lower levels of working memory on average and working memory plays a more prominent role in encoding novel associative memories in comparison to its role in encoding novel item memories. Based on the results of Experiment 1, the current study does not support age-related declines in the ability to encode and recognize novel L1-L2 word pairs. The lack of any age-related differences in Experiment 1 suggests that age-related memory decline may be less prominent than previously suggested at least when applied to L2 vocabulary learning. There are a variety of potential explanations that account for such a finding. One may be that the present series of tests were simply not sensitive enough to detect differences between age groups. Alternatively, it may be that there is relatively little difference in older adult and young adult L1-L2 learning abilities which is a similar finding to both Badham and Maylor (2011) who found no evidence of age-related associative memory deficits and Whiting et al. (2011) who found no difference between older adults and young adults in their L1-nonword learning task. Additionally any differences in L1-L2 ability may have been attenuated during Experiment 1 due to the increased response time allowed for all age groups during recognition testing, a factor that has previously been shown to diminish age-related differences (Mohanty & Naveh-Benjamin, 2018).

2.3.2 The Three component model revisited

The three-component model proposes that there are three types of memory necessary to learn a novel foreign language word: memory for the L1 item, L2 item and the L1-L2 association. In Experiment 1, memory for the target item was better than that of memory for the cue item regardless of language (both L1-L1 and L1-L2 material). This finding suggests that in accordance with the three-component model, when learning novel word pairs, each word is encoded separately rather than as only a part of a whole unit (and this was true regardless of language).

Interestingly, memory for the association did not statistically differ from memory for either L1 or L2 items despite the predictions of the associative deficit hypothesis. Experiment 1 suggests there was no difference between the types of memory as far as the ability to recognize them and these results do not necessarily support a three-component model of L2 vocabulary learning. To support the model with the data available, there would need to be evidence of differences in the rates of recognition, such as deficits to associative memory or differences between cue and target memory in the L1-L2 word pairs, but only the latter was found. While there is a large volume of previous work that highlights an age-related associative memory deficit (for a meta-analyses see Old & Naveh-Benjamin, 2008) which suggests there is some validity to a model that makes a distinction between memories for L1 items, L2 items, and their association, the only evidence of differences between item memories and associative memories in Experiment 1 was that working memory had a unique influence on associative memories that was not present when it came to item memories. To further explore the differences between item memories and associative memories, Experiment 2 repeated a similar study paradigm as Experiment 1, but manipulated how participants studied the word pairs with study methods requiring the cognitive processing of the association through cued recall practice in comparison to a restudy condition that does not necessarily require the processing of associations during learning.

Chapter 3: Experiment 2

Experiment 2 addresses Aim 4: how does retrieval practice affect long-term memory for L2 items and L1-L2 associative memory in older and younger adults. Experiment 2 compared memory retention after retrieval practice versus restudy for older and younger adults. Additionally, type of retrieval (whether participants were asked to retrieve L1 items or retrieve L2 items) was manipulated between groups to explore the influence of retrieval on item and associative memory for older and younger adults.

3.1 Methodology

3.1.1 Participants

As was the case with Experiment 1, boot-strapped power analysis indicated that to reliably detect differences in item memory between age groups at an effect size of .65 (Old & Naveh-Benjamin, 2008), alpha level of .05, and power of .80 requires approximately 40 participants per group. In addition to the two age groups, the design of Experiment 2 included two different retrieval groups (one retrieving the L1, the other retrieving the L2), necessitating a total of 4 groups or 160 participants in total.

A total of 94 young adults (52 female, M_{age} = 19.31, SD_{age} = 1.16) and 72 older adults (54 female, M_{age} = 68.50, SD_{age} = 7.08) were recruited for the study. The young adult participants were recruited from a research university participant pool and the older adults were recruited from the local community through a research university volunteer pool. Older adult recruitment did not reach the target goal of 80 participants due to time constraints. Participants were compensated 5 dollars for their estimated half hour of participation.

3.1.2 Materials

A Swahili-English vocabulary list (L1-L2) comprised of 80 word-pairs was used in the experiment (the list can be found in Appendix I). As was the case with Experiment 1, all words were nouns and had similar lexical frequency ($M_{Freq\_HAL} = 8.5$).

3.1.3 Design

Participants were asked to study L1-L2 word pair lists followed by recognition tests after a 24hr delay. A 2x3x3 mixed design investigated the influence of age (younger vs older adults), study condition (receptive retrieval practice vs productive retrieval practice vs restudy), and final test type (cue recognition vs target recognition vs associative recognition) on memory for word pairs. Age was manipulated between subjects such that half the participants were young adults (between the ages of 18 and 24) and half were older adults (ages 55+). Study condition was manipulated such that all participants restudied half of the word pairs and for the other half of the word pairs they were given a cued recall test of the word pair as a form of retrieval practice.

For half of the participants, the cued recall test consisted of the L1 item serving as the cue word participants were asked to retrieve the L2 item as the target, a process we will refer to as productive retrieval (they had to produce the L2 item from memory). For the other half of participants, they received the L2 item as the cue and were asked to produce the L1 item, a process we will refer to as receptive retrieval (they received the L2 item instead of having to produce it). Retrieval condition assignments were counterbalanced with the age groups such that half of each age group was assigned to the receptive retrieval condition and the other half were assigned to the productive retrieval condition. Additionally, to maintain consistency, the restudy condition was different for the receptive and productive retrieval condition group. While both groups restudied half the word pairs, the order the word pairs were presented in differed such that the word appearing on the left side of the screen was always the cue word and the word appearing on the right side of the screen was always the target word. For example, in the productive retrieval condition, the L1 item was the cue and the L2 item was the target, so during both study and restudy, the L1 item appeared on the left side of the screen and the L2 item appeared on the right side (e.g. heart -*moyo*). For the receptive retrieval condition, the L2 item was the cue and the L1 item was the target so during study and restudy, the order of the words were reversed such that the L2 item appeared on the left and the L1 item appeared on the right (e.g. *moyo*-heart). The goal was to maintain consistency in presentation between the receptive and productive recall tests such that the cue was always the item on the left and the target was the item on the right during study and restudy. This change in study direction was not predicted to affect learning outcomes for the restudy condition (thus we only include one restudy condition in the final model instead of two). Each participant received three recognition tests, one for the cue items, one for the target items and one for the association between the word pairs (recognition test order was counterbalanced between participants). Like Experiment 1, each word pair was assigned to one of the three recognition tests such that if a cue word from a pair appeared on the cue test, the target word for that pair would not appear on the target recognition test nor would the word pair itself appear on the association recognition test (test assignments were counterbalanced between participants). Accuracy on these tests (hits-false alarms) served as the dependent variable.

Final test procedures were identical to Experiment 1.

3.1.4 Procedure

The word-pair study task had participants view 40 word-pairs randomly selected from the 80-word list, one at a time for 5s each. Cues appeared on the left side of the screen, targets appeared on the right, separated by a hyphen. For half the participants, the L1 served as the cue and the L2 the target with the other half of participants having the reverse, L2 cue with L1 target.

Participants completed two blocks of training with 40 word-pairs per block. Within each block, participants viewed word pairs in groups of ten (presented one at a time.) Once a full group of ten was studied, if they were in the restudy block, participants restudied the list of ten words and then moved onto the next set of ten words. For the retrieval practice block, after studying the list of ten words participants completed a cued recall task which presented the cue and asked participants to type out the target for those ten words (presented one at a time). Once an answer was submitted, participants were shown the correct answer as feedback for 5s.

After completing both blocks of study, participants were then dismissed for 24h. Once they returned the following day, I administered the three recognition tests. For the cue and target recognition tests, participants were shown a single word and given 10s to determine whether the word had been seen during the study session or whether it is a new word. The association test displayed word pairs for 10s and participants were asked whether the pairing was the same as the one they saw during study or if it was different.

3.2 Results

The same R software and packages used in Experiment 1 was used in Experiment 2. Recognition test performance was calculated as hits-false alarms and served as the dependent variable. Descriptive statistics for Experiment 2 can be found in Table 5.

Table 5: Descriptive Statistics for Experiment 2 Outcomes

Study Condition	Age Group	Test Type	Mean	SD
Productive Retrieval	Older Adult	L1 Item	.78	.42
		L2 Item	.73	.45
		Pair	.59	.49
	Young Adult	L1 Item	.80	.40
		L2 Item	.83	.38
		Pair	.64	.48
Receptive Retrieval	Older Adult	L1 Item	.78	.41
		L2 Item	.66	.47
		Pair	.64	.48
	Young Adult	L1 Item	.81	.39
		L2 Item	.76	.43
		Pair	.62	.49
Restudy	Older Adult	L1 Item	.76	.43
		L2 Item	.72	.45
		Pair	.60	.49
	Young Adult	L1 Item	.68	.47
		L2 Item	.71	.45
		Pair	.60	.49

3.2.1 Initial retrieval practice performance

Initial retrieval practice performance was analyzed to determine whether there was a difference between older and younger adults during the cued recall training as accuracy during retrieval practice is predictive of the size of the memory benefits (Dessenberger & Sommers, 2020; Rowland, 2014). Initial retrieval practice results were scored in two ways. The first method was a strict score where spelling of the target word whether it was English or Swahili had to be exact and any spelling mistakes would be counted as failure to recall. The second method was a Levenshtein Distance (LD) metric which calculates how many changes must be made to a submitted answer to match the corresponding target (e.g. from *moldy* to *molly* is an LD of one since only one letter has to be changed while *latter* to *talk* is an LD of 5 since five letters need be changed or removed). Most previous retrieval practice research has relied on strict scoring methodology (Akifumi, 2016; Kang et al., 2013); however, the LD scoring method was included as it can identify whether weak memory traces for the target word exist. For example, if a participant were tasked with recalling the word *manzana,* but spelled the word *mansana*, the strict scoring methodology would indicate the participant failed to recall the item, but the LD scoring method would indicate the participant had recalled all but one letter correctly. We include both scoring methods to allow for a direct comparison with previous research with the strict method as well as further identifying weaker memory traces using the LD method (descriptives statistics in Table 5 relied on the strict method).

The model used in both methods was a multi-level regression model constructed in a step-wise fashion starting with a base model that contained no predictors, adding a single predictor or interaction term at each step. Each model was compared to the previous model using chi-square goodness of fit tests. The individual participant was set as the level-2 error term to account for differences across participants.

For the strict scoring method, adding the age predictor to the model significantly improved model fit compared to the base model ($X(1) = 10.64$, $p = .001$). Adding the retrieval type (productive vs receptive) also significantly improved model fit ($X(1) = 4.54$, $p = .03$) but adding the interaction did not ($X(1) = 0.12$, $p = .73$). These findings suggest that younger adults on average were better able to recall the target words during retrieval practice ($M = .25$, $SD = .43$) compared to older adults ($M = .15$, $SD = .34$), and participants were better able to recall the target words during receptive recall training ($M = .17$, $SD = .38$) compared to productive recall training ($M = .25$, $SD = .43$).

For the LD scoring method, adding the age predictor to the model significantly improved model fit compared to the base model ($X(1) = 4.35$, $p = .04$). Adding the retrieval type (productive vs receptive) did not significantly improve model fit ($X(1) = .07$, $p = .80$) nor did adding the interaction (*X*(1) = 0.08, $p = .78$). These findings suggest that young adults (*M* = .36, *SD* = .43) were more likely to recall the target words during retrieval practice compared to their older adult counterparts (*M* = .29, *SD* = .39) even when accounting for differences in spelling abilities.

3.2.2 Study Direction

Study direction (whether participants studied *heart-moyo* or *moyo-heart*) was included in the design of the experiment to maintain consistency between the restudy and retrieval practice conditions. However, I did predict that study direction would have no impact on the restudy condition as there should be little reason for emphasis on a word in a given word pair based on whether it appeared on the left or right side of the screen. It was determined apriori that if study direction did not influence the restudy condition that it should be removed from the full model to increase statistical power. This would result in the 2x2x2x3 full model that included predictors for age, study direction, study type, final test type being converted into a 2x3x3 full model including predictors for age, study method (one restudy condition, a productive retrieval condition and a receptive retrieval condition), and final test type. The model used to evaluate this hypothesis was a multi-level regression model constructed in a step-wise fashion starting with a base model that contained no predictors, adding a single predictor or interaction term at each step. Each model was compared to the previous model using chi-square goodness of fit tests. The individual participant was set as the level-2 error term to account for differences across participants. This model only included data from the restudy condition to maximize statistical power for detecting differences attributable to study direction.

Adding the age predictor to the model significantly improved model fit compared to the base model (*X*(1) = 4.16, $p = .04$). Adding the study direction (L1-L2 vs L2-L1) did not significantly improve model fit (*X*(1) = 2.47, $p = .12$) nor did adding the interaction (*X*(1) = 0.07, *p* = .78). Adding the test type variable did improve model fit (*X*(2) = 23.12, $p < .001$), but adding the test type by age interaction did not improve model fit (*X*(2) = 3.37, *p* = .18), nor did adding the test type by study direction interaction (*X*(1) = 0.51, *p* = .78). Finally, adding the three way interaction between age, study direction, and test type did not significantly improve model fit (*X*(2) = 2.12, *p* = .35) These findings suggest that younger adults on average were better able to recognize the target words after restudy (*M* = .25, *SD* = .43) compared to older adults (*M* = .17, *SD* = .37) but study direction was not a significant factor and was removed from the primary analysis of Aim 4.

3.2.3 Aim 4 Results

To assess the impact of receptive and productive retrieval practice on long-term memory for L2 items and L1-L2 associative memory in older and younger adults, a multi-level regression model was constructed which included test type (cue memory vs target memory vs associative memory), age (young adult vs older adult), and study condition (productive retrieval vs receptive retrieval vs restudy) as fixed effects. Recognition test accuracy served as the dependent variable. The coefficients for the full model can be found in Table 6 (hit rate and false alarm rates can be found in Appendix III)

Table 6: Coefficients for Aim 4 Full Model
A multi-level logistic regression model fixed effects output for the final recognition test scores (Intercept is the Older Adult cue memory test for the Productive Training Condition). β and standard error are presented in logit units and effect size presented in Odds-Ratios (OR) with 95% Confidence Intervals (CI).

	Estimate	Std. Error	*z* value	*p* value	OR	95% CI
Intercept	1.22	.15	8.34	<.001	3.38	[2.92, 3.91]
Young Adult	.19	.21	.91	.37	1.21	[.98, 1.48]
Receptive Training	.16	.22	.74	.46	1.18	[.94, 1.47]
Restudy Training	-.05	.18	-.30	.77	.95	[.79, 1.13]
Target Memory	-.27	.17	-1.61	.11	.76	[.64, .90]
Pair Memory	-.84	.19	-4.33	<.001	.43	[.35, .52]
Young Adult x Receptive Training	-.04	.3	-.12	.90	.96	[.72, 1.30]
Young Adult x Restudy Training	-.58	.25	-2.35	.02	.56	[.44, .72]
Young Adult x Target Memory	.46	.25	1.84	.07	1.58	[1.23, 2.03]
Young Adult x Pair Memory	-.02	.27	-.07	.94	.98	[.75, 1.28]
Receptive Training x Target Memory	-.34	.26	-1.29	.20	.71	[.55, .93]
Restudy Training x Target Memory	.11	.21	.52	.60	1.11	[.91, 1.37]
Receptive Training x Pair Memory	.06	.28	.21	.84	1.06	[.80, 1.40]
Restudy Training x Pair Memory	.11	.27	.42	.68	1.12	[.85, 1.47]
Young Adult x Receptive Training x Target Memory	-.18	.36	-.50	.62	.84	[.58, 1.20]
Young Adult x Restudy Training x Target Memory	-.15	.29	-.53	.60	.86	[.64, 1.15]
Young Adult x Receptive Training x Pair Memory	-.16	.37	-.42	.68	.86	[.59, 1.24]
Young Adult x Restudy Training x Pair Memory	.39	.36	1.06	.29	1.47	[1.02, 2.12]

Step-wise Analysis. Adding the age predictor to the model did not significantly improve model fit compared to the base model ($X(1) = .065$, $p = .80$). Adding the study condition significantly improved model fit ($X(2) = 11.70$, $p < .001$). Adding the interaction term between age and study condition significantly improved model fit ($X(2)$ 14.65, $p < .001$). Adding the test type to the model significantly improved model fit ($X(2)$ 185.48, $p < .001$). Adding the two-way interaction between age and test type did significantly improved model fit ($X(2)$ 7.92, $p = .02$) and adding the two-way interaction between study condition and test type significantly improved model fit ($X(4)$ 14.10, $p = .007$). Adding the three-way interaction between age, test type, and study condition did not significantly improve model fit ($X(4)$ 3.10, $p = .44$).

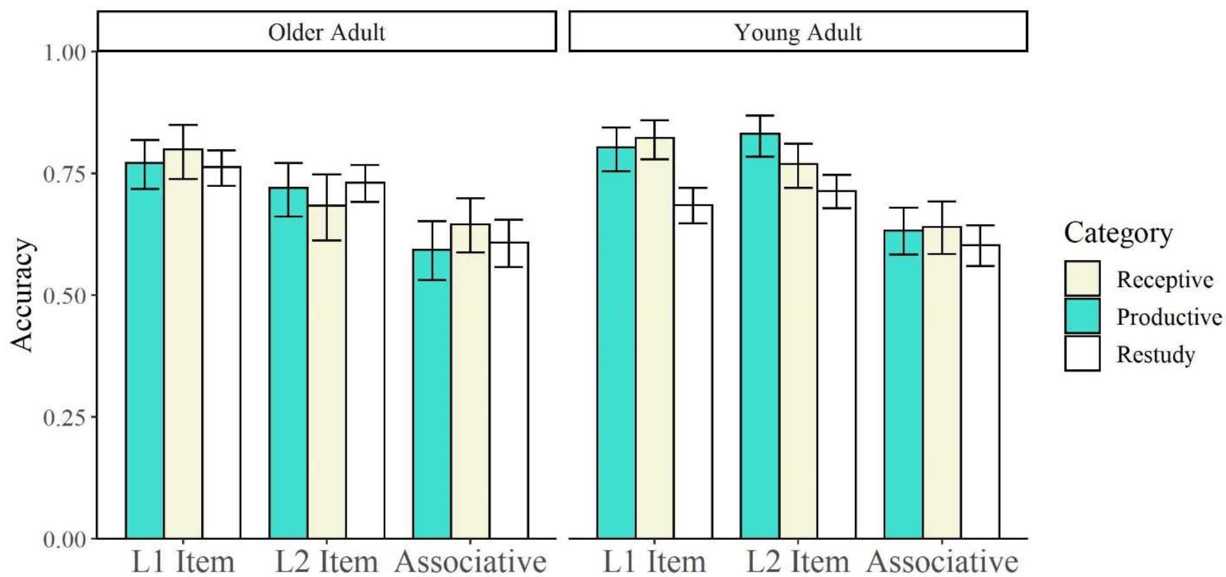

Figure 11. Aim 4 Results. Experiment 2 recognition accuracy separated by age, test type, and study condition. Accuracy is calculated as the proportional rate of accuracy or hits-false alarms. Error bars are 95% CI.

Post-hoc Analysis. Post-hoc analyses were completed simultaneously for each interaction such that Bonferroni corrections were made considering all post-hoc comparisons to account for potential family-wise error. A total of 36 linear comparisons were constructed to evaluate both of the two-way interactions.

For the two-way interaction between age and study condition, linear comparison of the full model indicated there was a significant interaction of age and study condition such that young adults were better able to recognize the words and pairs after productive recall relative to the restudy condition ($z = 5.36$, $p < .001$). Additionally, young adults were better able to recognize the word and pairs after receptive recall relative to the restudy condition ($z = 4.70$, $p < .001$). No other significant differences were found. These findings suggest a benefit of retrieval regardless of retrieval type (receptive vs productive) over the restudy condition, but the benefits are only present for young adults while older adult performance did not differ between the productive recall, receptive recall, and restudy conditions.

For the two-way interaction between age and test type, linear comparison of the full model indicated there was a significant interaction of age and test type such that young adults were better able to recognize the L2 items compared to older adults ($z = 3.02$, $p = .02$). Both older and younger adults were more likely to recognize L1 items relative to word pairs (older adults: $z = 8.24$, $p < .001$; young adults: $z = 8.65$, $p < .001$) and both age groups were more likely to recognize L2 items relative to word pairs (older adults: $z = 4.78$, $p < .001$; young adults: $z = 8.63$, $p < .001$). As for recognition of L1 items vs. L2 items, only older adults demonstrated a difference such that they were more likely to recognize L1 items in comparison to L2 items ($z = 3.67$, $p = .002$). These findings suggest the only difference between older and younger adults is the ability to recognize L2 items and this difference is likely the result of older adults having greater difficulty with L2 item memory in comparison to their L1 item memory and associative memory.

For the two-way interaction between study condition and test type, linear comparison of the full model indicated there were significantly more L1 cue words remembered compared to L1-L2 word-pairs correctly recognized after all three study conditions, including restudy ($z = 5.71$, $p < .001$), productive retrieval ($z = 6.43$, $p < .001$), and receptive retrieval ($z = 6.16$, p

< .001). Additionally, there was an increase in correctly recognized L1 cue words relative to correctly recognized L2 target words after receptive retrieval ($z = 3.65$, $p = .004$). There was also an increase in correctly recognized L2 target words in comparison to correctly recognized L1-L2 word pairs after restudy ($z = 5.58$, $p < .001$) as well as after productive retrieval ($z = 6.10$, $p < .001$). The difference between L2 target word recognition and L1-L2 word pair recognition did not reach significance after receptive retrieval practice ($z = 2.97$, $p = .05$). Lastly, there was a benefit to memory for L1 cue words after receptive recall relative to the restudy condition ($z = 3.74$, $p = .003$). No other significant differences were found. These results suggest L1 and L2 items were remembered better individually than the associations between L1 and L2 items with the exception of receptive recall training which benefitted L2 item memories and L1-L2 associative memories equally. Additionally, receptive recall improved memory for L1 items relative to the restudy condition.

3.3 Aim 4 discussion

The goal of Experiment 2 was to assess the influence of retrieval practice, both receptive and productive recall, on memory for L1 and L2 items as well as L1-L2 word pairs. The results of Experiment 2 suggest that only young adults were likely to benefit from retrieval practice and that receptive retrieval practice is best for promoting memory for the L1 meaning of the L1-L2 word pair, but counter to my predictions, productive retrieval does not promote memory for the novel L2 word form to a greater degree compared to receptive retrieval.

3.3.1 Age related memory decline and retrieval practice

Retrieval practice was predicted to improve associative memory, one of the primary areas of age-related memory decline, because the cued recall test format requires participants to focus on not only the target word but also its association to the cue word. The potential to improve associative memory meant retrieval practice could directly address one of the key purported issues in older adult vocabulary learning by aiding in associative learning.

Results indicated that retrieval practice was only effective for young adults with both receptive and productive recall leading to higher rates of recognition overall compared to the restudy condition. No three-way interaction was found which suggests that the influence of retrieval practice is consistent for item memory (both L1 and L2) as well as for the associative memory (L1-L2 word pair). The benefit of retrieval over restudy was not present for older adults which suggests they do not benefit from retrieval practice in the same way.

There are several explanations for why older adults did not benefit from retrieval practice. First, older adults performed worse on the cued recall task compared to the young adults. Prior research on the benefits of retrieval practice indicate that the benefits of retrieval are predicated on accuracy during retrieval practice (Butler et al., 2007; Kornell et al., 2011; Rowland, 2014). For example, according to a meta-analysis of the benefits of retrieval practice by Rowland (2014), the benefits of retrieval relative to a restudy control condition can be severely lessened when participants are unable to accurately recall the target word even when feedback is provided. It may be the case that due to low accuracy on the practice trials, older adults did not benefit from retrieval practice.

An alternate explanation for the absence of retrieval practice benefits for older adults in the current study is that retrieval practice does not improve memory for older adults to the same magnitude that it does for younger adults. The reason retrieval practice promotes memory retention is a subject of frequent debate, but whatever the cognitive reason, retrieval practice may simply not affect older adults in the same way it affects young adults. While most studies indicate both younger and older adults benefit from retrieval practice (Coane, 2013; Guran et al., 2020; Meyer & Logan, 2013; Rogalski et al., 2014; Tse et al., 2010), Guran and colleagues found that older adults have a smaller magnitude of benefit compared to their young adult counterparts. Participants were tasked with learning scene stimuli and either restudying the stimuli or undergoing a new/old recognition test (retrieval practice) followed by a final test of the stimuli. Both older and young adults were better able to recognize the stimuli during the final test after retrieval practice relative to the control condition, but older adults had a smaller magnitude of benefit (d = 1.96) relative to their young adult counterparts (d = 3.30). It may be the case that older adults do not gain as much of a benefit from retrieval practice as their young adult peers and the present series of tests were not sensitive enough to detect any differences.

Most likely the full explanation is a combination of reduced accuracy during retrieval practice and a reduced magnitude of benefit from retrieval practice. This raises questions as to whether retrieval practice is a suitable exercise for older adult L2 vocabulary learning when a simpler approach that relied on restudying information is as effective.

3.3.2 The three component model and retrieval practice

Viewing the three-component model under the lens of the TOPRA model, if one component is emphasized during study, there should be a corresponding increase in memory for that component. The TOPRA model indicates that when a task emphasizes L1 item, memory for the L1 item increases and if a task emphasizes L2 item, memory for the L2 item should increase. Cued recall practice is a process that emphasizes two components simultaneously. It stands to reason that a task that emphasizes processing the association should improve memory for the association. Participants are provided with the cue word and have to not only recall the target, but also have to recall the association between the cue and the target to successfully perform the retrieval. As this was the case, I had predicted that retrieval practice should promote memory relative to control conditions for the retrieved item and associative memory for the pair.

The results of Experiment 2 suggest retrieval practice primarily promotes memory for the L1 words when retrieval practice consisted of receptive recall (L2 cue with L1 target) which required retrieval of the L1 item. According to the analysis, productive recall which necessitates retrieval of the L2 item did not promote memory of L2 items. However, as can be seen in Figure 11, this can likely be reflective of the lack of retrieval practice benefits to memory for older adults for either productive or receptive recall.

However, in both age groups it is clear the retrieval practice provided no benefit to associative memories. The lack of retrieval practice benefit for associative memories (memory for the L1-L2 word pairs) was particularly concerning given that retrieval practice is a task that on its surface would appear to be heavily reliant on processing the association between L1 and L2 items. In fact, cued recall tests such as those used during the study phase in Experiment 2 are frequently used in classroom textbooks as a measure of language translation ability. According to these results, memory for associations between L1 and L2 items is not bolstered by retrieval practice relative to the restudy condition.

One explanation for the lack of an associative memory benefit is participants were not focusing on the association during retrieval practice. Rather than identifying the cue word and then identifying the association in their memory followed by the target word, it may be the case that participants identified the cue word and then scanned an existing word bank contained within their memories of all the possible answers they had studied thus far. Once they consider a target word that sparks recognition, they have successfully "retrieved" the association when in reality minimal effort was given towards processing the association. According to the three-component model, the cue word is being processed, the target word is being processed, but the association is not being interacted to any significant degree resulting in a poor associative memory. If it is the case that cued recall as performed in Experiment 2 does not appropriately emphasize the third component (the association between the L1 and L2 words), retrieval practice on its own may be a poor method for addressing any age-related associative memory declines.

Chapter 4 General Discussion

The current study was designed to investigate L2 vocabulary learning in older and younger adults. In addition, the study examined the relationship between episodic memory, working memory, and L2 vocabulary learning. Finally, Experiment 2 investigated how study method (restudy vs retrieval practice) affected L2 vocabulary learning and whether these effects were similar for older and younger adults.

The literature to date has often relied on cued recall as a measure of L2 vocabulary learning, which may be ecologically valid but has provided a limited picture of the underlying processes that occur during L2 vocabulary learning. For example, when a failure to retrieve occurs during cued recall when an L1 cue is provided, it is unclear whether the failure can be attributed to weak item memory for the L2 word form or a weak associative memory the L1-L2 pair. The proposed study relied on multiple recognition tests instead of a single cued recall test to identify which components of the three-component model (memory for L1, memory for L2, associative memory) were affected by the manipulations and whether this differed for older and younger adults.

4.1 Age-related differences in L2 vocabulary learning

The results of the studies addressing Aims 1,2,3, and 4 suggest older and younger adults are nearly equally effective at learning novel vocabulary. My initial hypotheses was that older adults would have greater difficulty than young adults recognizing L1-L2 word pairs due to age-related cognitive decline (Craik, 1994; Craik & McDowd, 1987; Gordon & Clark, 1974; Park et al., 2002; Rhodes et al., 2019). Furthermore, consistent with the associative memory deficit hypothesis (Old & Naveh-Benjamin, 2008), it was predicted that associative memories would suffer higher age-related declines compared to item memories, but this hypothesis was not supported by the data. In the following sections, I review the main findings and provide possible explanations for the pattern of results.

4.1.1 Age-related memory decline

The present series of experiments did not find a consistent age-related decline in recognition memory for L1-L2 word pairs. Young and older adults performed at comparable levels in tests measuring recognition accuracy for L1 and L2 items as well as for the associations between them. There were only two cases in which significant differences between older and younger adults were observed: older adults had poorer working memory than young adults and they had greater difficulty recognizing L2 items after retrieval practice relative to the young adults (there was no difference between age groups in the restudy control condition).

The lower levels of working memory for older than younger adults is consistent with previous research (Park et al., 2002). Critically, however, these differences in working memory had no significant effect on the ability to learn novel L2 vocabulary. For both older and younger adults, levels of working memory relative to their own peer group were predictive of test performance: low working memory resulted in poor test performance on the associative memory test while high working resulted in higher levels of performance. This influence of working memory was consistent across age groups. However, differences in working memory did not influence memory for individual items (neither L1 nor L2 items), only influencing memory for associations. Working memory did influence memory associations such that increases in working memory corresponded to increases in associative memory. It would seem that working memory primarily plays a role in forming connections between items and plays less of a role in encoding the individual items themselves, regardless of familiarity with those items (L1 vs L2).

As for method of study, young adults and older adults differed on their learning outcomes after retrieval practice. Young adults derived a significant benefit from retrieval over restudy on the item memory tests but not the associative test. Older adult test performance was not affected by study condition such that restudy was found to be as effective as retrieval practice for older adults on all three recognition measures. While previous literature (Coane, 2013; Guran et al., 2020; Meyer & Logan, 2013; Rogalski et al., 2014; Tse et al., 2010) suggests older adults do benefit from retrieval practice, the findings of this dissertations suggests that careful selection of study method is more important for young adults while older adults may be able to employ a wider variety of study methods.

4.1.2 Assessing the associative deficit hypothesis during L2 vocabulary learning

The associative deficit hypothesis states that as we age, our ability to encode and recall memories for associations declines at a faster rate compared to our ability to encode and recall memory for individual items (Naveh-Benjamin, 2000). Associative memory is of particular importance during L1-L2 word pair learning as the ability to form connections between novel L2 word forms and their meanings is a key component of language learning. Any age-related declines in associative memory would suggest that older adults should have more trouble learning new vocabulary words compared to younger adults.

In Experiment 1 and Experiment 2, minimal evidence was found in support of the associative deficit hypothesis during L1-L2 learning. There was evidence that working memory plays a more significant role in encoding novel associations compared to encoding L1 and L2 items, and there was evidence that older adults had lower levels of working memory compared to young adults. However, the results of Experiment 1 suggest that any differences between age groups in their associative memory was not significant. In Experiment 2, there was also evidence that both age groups had greater difficulty encoding novel associations compared to individual items but there was again no difference between the age groups. There is a wealth of research suggesting an age-related difference in associative memory (see Old & Naveh-Benjamin, 2008), but the present study found no associative memory deficit when it comes to L1-L2 material. The present findings are consistent with those of Badham and Maylor (2011), who found a similar result when investigating memory for word-nonword pairs. From these findings I would argue that there is insufficient evidence to suggest our ability to form novel associations with L1-L2 words diminishes across the lifespan.

However, one potential explanation for a lack of associative memory deficit is that the sample size in Experiment 1 was insufficient to detect differences in the influence of working memory on item memory abilities due to the influence of working memory on item memories being markedly smaller compared to its influence on associative memories. As shown in Figure 10, there was a numerical advantage in recognition memory for L1 items and L2 items for individuals with high working memory relative to medium and low working memory, but this "outcome" was not statistically significance. If it was the case that the influence of working memory on item memories is significantly smaller than associative memories, it would mimic the associative deficit hypothesis exactly (with the understanding that working memory is playing a key role), with associative memories declining at a faster rate than individual item memories.

4.1.3 Limitations and future research into age-related differences in L2 vocabulary learning

Despite previous research which suggests that older adults perform worse on memory tests relative to young adults and that it is more difficult to learn L2 material compared to L1 material, the present study failed to demonstrate significant difference between age groups and language types with limited exceptions. While this is a favorable outcome from a pedagogy standpoint (research relying on young adults may be applicable to older adults), it does raise concerns as to why the older adults who were recruited from varied sources did not demonstrate the expected levels of cognitive decline seen in previous research (with the exception of declines in working memory). It could very well be a cohort effect such that the influence of age-related decline has diminished as health and living standards have increased as evidenced by increasing life expectancy rates (Dwyer-Lindgren et al., 2022; Woolf & Schoomaker, 2019) . The present study did not include a health survey so it is difficult to determine whether the individuals in the present study are simply healthier than their counterparts in previous research which may delay cognitive aging (Murman, 2015; Pardon & Bondi, 2012).

Future research should address the health of the participants in the study in addition to their age to determine whether the physical health of the individual impacts the influence of age-related cognitive decline on L1-L2 word pair learning. The implications if there are differences would be that methodology provided to the language learner may be more dependent on their current health state rather than simply their age alone.

4.2 The three-component model of L2 vocabulary learning

The three-component model proposed that three types of memory were necessary for L2 vocabulary learning: L1 item memory, L2 item memory, and the memory for the association between them. Creation of the model was informed by two lines of research, the first being the TOPRA model, which suggests that any emphasis on one type of memory (be it form processing of the L2 item or semantic processing of the L1 item) could affect learning outcomes (Barcroft & Sommers, 2005; Kida & Barcroft, 2018). The second line of research was the literature on the associative deficit hypothesis which indicates that as we age, our memories for associations are worse than our memories for items (Naveh-Benjamin, 2000; Old & Naveh-Benjamin, 2008). Combining the ideas behind the TOPRA model and the associative deficit hypothesis, I proposed the three-component model with the assumption that when studying L1-L2 word pairs, three unique types of memory were created: L1 items, L2 items, and L1-L2 associations.

4.2.1 Assessing the three-component model

The results of both Experiment 1 and Experiment 2 provided some support for the three-component model. In particular, Experiment 1 demonstrated that working memory played a significant role in recognition of associative memories, but not in recognition of item memories. Additionally, Experiment 2 also demonstrated an increase in item memory relative to associative memories, further dividing the two types of memory. There were no differences between L1 and L2 items in Experiment 1 or Experiment 2, only that older adults demonstrated a better memory for L1 items than L2 items in Experiment 2. Taken together, it would seem that as we age, differences between L1 items and L2 items as well as differences between item memory and associative memory becomes more apparent. It may be the case that L2 items and associative memories are more cognitively demanding, which has a limited effect on young adults as their cognition is at its peak in the lifespan. Older adults in turn may have greater difficulty encoding L2 items and the L1-L2 association because they have reduced cognitive capabilities relative to younger adults. The present study provides limited support for the three-component model such that three components are distinguishable only under certain conditions (e.g. when cognitive resources are diminished).

4.2.2 Future investigations of the three-component model

Due to the limited support for the three-component model by the present study (particularly with young adults), I suggest additional research test the validity of the model. Unfortunately, it is difficult to make a backwards comparison to previous literature because few studies rely on testing methodology that highlights memory abilities for each of the three components separately without confounding one component with another (as is the case with cued recall tests). As such, future research should measure each component separately to determine whether the three-component model effectively highlights how memories are created and stored during L2 learning.

In addition, the use of the recognition testing procedures may provide insight into each component, but the procedure itself is limited as the ability to recognize information and the ability to recall are inherently different. For effective language-use, learners need to be able to recall L2 word forms and their associated meanings, not just recognize them when their encounter them. Future research should seek new methods of measurement that include recall rather than just recognition to better understand the relationship between L1 item memories, L2 item memories, associative memories, and the ability to translate to and from the target language. In addition to being a more ecologically method of measurement (cued recall better reflects translation in comparison to recognition testing), it is also a more difficult method of testing that likely demands more cognitive resources. If young adults were to have an increased cognitive demand placed on them during learning, the difference in the cognitive demands of L2 item learning and L1-L2 associative learning in comparison to L1item learning may become apparent.

4.3 Conclusion

The population as a whole is getting older, which increases the need to understand older adult L2 language learning. One of the most significant hurdles to L2 learning, vocabulary acquisition, has received little attention in older adult L2 learning research. The proposed study sought to not only address this gap in the literature by investigating individual differences in older adult L2 vocabulary learning, but also to expand our understanding of L2 vocabulary learning in young adults through the application of a three-component model of vocabulary learning.

The present study found no difference between younger and older adults in their ability to learn L1-L2 word pairs with the exception that young adults derive a greater benefit from retrieval practice in comparison to their older adult counterparts. The present findings suggest that older adults are able to encode individual L1 item memories, L2 item memories and L1-L2 associations at nearly an equivalent rate as their young adult counterparts. These findings are fortunate as the dearth of literature on L2 learning in older adults may not be indicative of a problem, as much of the literature that focuses on young adults is likely to be applicable to older adults as well. The present study only scratched the surface however and raises several questions regarding differences (or lack thereof) between young adults and older adults.

References

Akifumi, Y. (2016). *The Effects of Receptive and Productive Word Retrieval Practice on Second Language Vocabulary Learning*. 15.

Baddeley, A. (1983). Working Memory. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *302*(1110), 311–324. JSTOR. http://www.jstor.org/stable/2395996

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

Baddeley, A., & Hitch, G. (1974). Working Memory. In *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Elsevier. https://doi.org/10.1016/S0079-7421(08)60452-1

Badham, S. P., & Maylor, E. A. (2011). Age-related associative deficits are absent with nonwords. *Psychology and Aging*, *26*(3), 689–694. https://doi.org/10.1037/a0022205

Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. A., Kessler, B., Loftis, Neely, J. H., Nelson, D. L., Simpson, G. B., & Treiman, R. (2007). The English Lexicon Project. *Behavior Research Methods*, *39*, 445–459. https://elexicon.wustl.edu/about.html

Bangert, A. S., & Heydarian, N. M. (2017). Recall and response time norms for English–Swahili word pairs and facts about Kenya. *Behavior Research Methods*, *49*(1), 124–171. https://doi.org/10.3758/s13428-015-0701-1

Barcroft, J. (2002). Semantic and structural elaboration in L2 lexical acquisition. *Language Learning*, *52*(2), 323–363. https://doi.org/10.1111/0023-8333.00186

Barcroft, J., & Sommers, M. S. (2005). Effects of acoustic variability on second language vocabulary learning. *Studies in Second Language Acquisition*, *27*(03). https://doi.org/10.1017/S0272263105050175

Bartsch, L. M., Loaiza, V. M., & Oberauer, K. (2019). Does limited working memory capacity underlie age differences in associative long-term memory? *Psychology and Aging*, *34*(2), 268–281. https://doi.org/10.1037/pag0000317

Butler, A. C., Karpicke, J. D., & Roediger, H. L. (2007). The effect of type and timing of feedback on learning from multiple-choice tests. *Journal of Experimental Psychology: Applied*, *13*(4), 273–281. https://doi.org/10.1037/1076-898X.13.4.273

Candry, S., Decloedt, J., & Eyckmans, J. (2020). Comparing the merits of word writing and retrieval practice for L2 vocabulary learning. *System*, *89*, 102206. https://doi.org/10.1016/j.system.2020.102206

Chalfonte, B. l., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition*, *24*(4), 403–416. https://doi.org/10.3758/BF03200930

Coane, J. H. (2013). Retrieval practice and elaborative encoding benefit memory in younger and older adults. *Journal of Applied Research in Memory and Cognition*, *2*(2), 95–100. https://doi.org/10.1016/j.jarmac.2013.04.001

Colby, S. L., & Ortman, J. M. (2014). *Projections of the Size and Composition of the U.S. Population: 2014 to 2060* (pp. 25–1143). U.S. Census Bureau.

Commissaris, C. J. A. M., Verhey, F. R. J., Ponds, R. W. H. M., Jolles, J., & Kok, G. J. (1994). Public education about normal forgetfulness and dementia: Importance and effects. *Patient Education and Counseling*, *24*(2), 109–115. https://doi.org/10.1016/0738-3991(94)90004-3

Committee on the Health and Medical Dimensions of Social Isolation and Loneliness in Older Adults, Board on Health Sciences Policy, Board on Behavioral, Cognitive, and Sensory Sciences, Health and Medicine Division, Division of Behavioral and Social Sciences and Education, & National Academies of Sciences, Engineering, and Medicine. (2020). *Social Isolation and Loneliness in Older Adults: Opportunities for the Health Care System* (p. 25663). National Academies Press. https://doi.org/10.17226/25663

Craik, F. I. M. (1994). Memory Changes in Normal Aging. *Current Directions in Psychological Science*, *3*(5), 155–158. https://doi.org/10.1111/1467-8721.ep10770653

Craik, F. I. M., Bialystok, E., Gillingham, S., & Stuss, D. T. (2018). Alpha span: A measure of working memory. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *72*(3), 141–152. https://doi.org/10.1037/cep0000143

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 474–479. https://doi.org/10.1037/0278-7393.13.3.474

Criss, A. H., & Shiffrin, R. M. (2004). Context Noise and Item Noise Jointly Determine Recognition Memory: A Comment on Dennis and Humphreys (2001). *Psychological Review*, *111*(3), 800–807. https://doi.org/10.1037/0033-295X.111.3.800

Davies, N. F. (1976). Receptive Versus Productive Skills in Foreign Language Learning. *The Modern Language Journal*, *60*(8), 440–443. https://doi.org/10.1111/j.1540-4781.1976.tb03667.x

Dessenberger, S. J., Levine, T. F., & Sommers, M. S. (In Prep). *A Latent Variable Approach to Working Memory Measured with an Automated Online Battery*.

Dessenberger, S. J., & Sommers, M. S. (2020, November 19). *Lexical inferencing as a generation effect* [Poster Presentation]. Psychonomic Society- Virtual Psychonomics 2020 Annual Meeting, Virtual.

Dwyer-Lindgren, L., Kendrick, P., Kelly, Y. O., Sylte, D. O., Schmidt, C., Blacker, B. F., Daoud, F., Abdi, A. A., Baumann, M., Mouhanna, F., Kahn, E., Hay, S. I., Mensah, G. A., Nápoles, A. M., Pérez-Stable, E. J., Shiels, M., Freedman, N., Arias, E., George, S. A., … Mokdad, A. H. (2022). Life expectancy by county, race, and ethnicity in the USA, 2000–19: A systematic analysis of health disparities. *The Lancet*, *400*(10345), 25–38. https://doi.org/10.1016/S0140-6736(22)00876-5

Flores, G. (2006). Language Barriers to Health Care in the United States. *New England Journal of Medicine*, *355*(3), 229–231. https://doi.org/10.1056/NEJMp058316

Fritz, C. O., Morris, P. E., Acton, M., Voelkel, A. R., & Etkind, R. (2007). Comparing and combining retrieval practice and the keyword mnemonic for foreign vocabulary learning. *Applied Cognitive Psychology*, *21*(4), 499–526. https://doi.org/10.1002/acp.1287

Gathercole, S. E., Willis, C. S., Emslie, H., & Baddeley, A. D. (1992). Phonological Memory and Vocabulary Development During the Early School \ears: A Longitudinal Study. *Developmental Psychology*, *28*(5), 887–898.

Gordon, S. K., & Clark, W. C. (1974). Adult Age Differences in Word and Nonsense Syllable Recognition Memory and Response Criterion. *Journal of Gerontology*, *29*(6), 659–665. https://doi.org/10.1093/geronj/29.6.659

Guran, C.-N. A., Lehmann-Grube, J., & Bunzeck, N. (2020). Retrieval Practice Improves Recollection-Based Memory Over a Seven-Day Period in Younger and Older Adults. *Frontiers in Psychology*, *10*, 2997. https://doi.org/10.3389/fpsyg.2019.02997

Hara, Y., & Naveh-Benjamin, M. (2015). The role of reduced working memory storage and processing resources in the associative memory deficit of older adults: Simulation studies with younger adults. *Aging, Neuropsychology, and Cognition*, *22*(2), 129–154. https://doi.org/10.1080/13825585.2014.889650

Harji, M. B., Woods, P. C., & Alavi, Z. K. (2010). The effect of viewing subtitled videos on vocabulary learning. *Journal of College Teaching & Learning (TLC)*, *7*(9). https://doi.org/10.19030/tlc.v7i9.146

Hirst, W., Phelps, E. A., Johnson, M. K., & Volpe, B. T. (1988). Amnesia and second language learning. *Brain and Cognition*, *8*(1), 105–116. https://doi.org/10.1016/0278-2626(88)90042-5

Ina, L. (2014). Incidental foreign-language acquisition by children watching subtitled television programs. *The Turkish Online Journal of Educational Technology*, *13*(4), 81–87.

Kahneman, D. (1973). *Attention and effort*. Prentice-Hall.

Kang, S. H. K., Gollan, T. H., & Pashler, H. (2013). Don't just repeat after me: Retrieval practice is better than imitation for foreign vocabulary learning. *Psychonomic Bulletin & Review*, *20*(6), 1259–1265. https://doi.org/10.3758/s13423-013-0450-z

Kida, S., & Barcroft, J. (2018). Semantic and structural tasks for the mapping component of L2 vocabulary learning: Testing the TOPRA model from a new angle. *Studies in Second Language Acquisition*, *40*(3), 477–502. https://doi.org/10.1017/S0272263117000146

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358. https://doi.org/10.1037/h0043688

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*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861–883. https://doi.org/10.3758/s13423-013-0565-2

Mackey, A., & Sachs, R. (2012). Older Learners in SLA research: A First look at working memory, feedback, and L2 development: Older learners in SLA research. *Language Learning*, *62*(3), 704–740. https://doi.org/10.1111/j.1467-9922.2011.00649.x

McGuire, W. J. (1961). A multiprocess model for paired-associate learning. *Journal of Experimental Psychology*, *62*(4), 335–347. https://doi.org/10.1037/h0042676

Meyer, A. N. D., & Logan, J. M. (2013). Taking the testing effect beyond the college freshman: Benefits for lifelong learning. *Psychology and Aging*, *28*(1), 142–147. https://doi.org/10.1037/a0030890

Mohanty, P. (Pam), & Naveh-Benjamin, M. (2018). Mitigating the adverse effects of response deadline on recognition memory: Differential effects of semantic memory support on item and associative memory. *Journal of Memory and Language*, *102*, 182–194. https://doi.org/10.1016/j.jml.2018.05.010

Multhaup, K. S., Balota, D. A., & Cowan, N. (1996). Implications of aging, lexicality, and item length for the mechanisms underlying memory span. *Psychonomic Bulletin & Review*, *3*(1), 112–120. https://doi.org/10.3758/BF03210750

Murman, D. (2015). The Impact of Age on Cognition. *Seminars in Hearing*, *36*(03), 111–121. https://doi.org/10.1055/s-0035-1555115

Nation, I. S. P. (2006). *Learning vocabulary in another language* (8. print). Cambridge Univ. Press.

Naveh-Benjamin, M. (2000). Adult age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), 1170–1187. https://doi.org/10.1037/0278-7393.26.5.1170

Naveh-Benjamin, M., Brav, T. K., & Levy, O. (2007). The associative memory deficit of older adults: The role of strategy utilization. *Psychology and Aging*, *22*(1), 202–208. https://doi.org/10.1037/0882-7974.22.1.202

Naveh-Benjamin, M., Shing, Y. L., Kilb, A., Werkle-Bergner, M., Lindenberger, U., & Li, S.-C. (2009). Adult age differences in memory for name–face associations: The effects of intentional and incidental learning. *Memory*, *17*(2), 220–232. https://doi.org/10.1080/09658210802222183

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on item and associative measures of memory: A meta-analysis. *Psychology and Aging*, *23*(1), 104–118. https://doi.org/10.1037/0882-7974.23.1.104

Papagno, C., Valentine, T., & Baddeley, A. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language*, *30*(1), 331–347.

Pardon, M.-C., & Bondi, M. W. (Eds.). (2012). *Behavioral Neurobiology of Aging* (Vol. 10). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23875-8

Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, *17*(2), 299–320. https://doi.org/10.1037/0882-7974.17.2.299

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Rhodes, S., Greene, N. R., & Naveh-Benjamin, M. (2019). Age-related differences in recall and recognition: A meta-analysis. *Psychon Bull Rev*, 19.

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Rogalski, Y., Altmann, L. J. P., & Rosenbek, J. C. (2014). Retrieval practice and testing improve memory in older adults. *Aphasiology*, *28*(4), 381–400. https://doi.org/10.1080/02687038.2013.870965

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

Salthouse, T. A. (2004). What and When of Cognitive Aging. *Current Directions in Psychological Science*, *13*(4), 140–144. https://doi.org/10.1111/j.0963-7214.2004.00293.x

Salthouse, T. A. (2009). When does age-related cognitive decline begin? *Neurobiology of Aging*, *30*(4), 507–514. https://doi.org/10.1016/j.neurobiolaging.2008.09.023

Schaie, K. W. (1993). The Seattle Longitudinal Studies of Adult Intelligence. *Current Directions in Psychological Science*, *2*(6), 171–175. https://doi.org/10.1111/1467-8721.ep10769721

Service, E., & Craik, F. I. M. (1993). Differences between young and older adults in learning a foreign vocabulary. *Journal of Memory and Language*, *32*(5), 608–623. https://doi.org/10.1006/jmla.1993.1031

Siegel, A. L. M., & Castel, A. D. (2018). Memory for important item-location associations in younger and older adults. *Psychology and Aging*, *33*(1), 30–45. https://doi.org/10.1037/pag0000209

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*(4), 833–845. https://doi.org/10.1037/a0019933

Tulving, E. (1972). Episodic and semantic memory. In W. Donaldson (Ed.), *Organization of Memory* (p. 423). Academic Press.

Underwood, B. J. (1978). Recognition memory as a function of length of study list. *Bulletin of the Psychonomic Society*, *12*(2), 89–91. https://doi.org/10.3758/BF03329636

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

U.S. State Department. (2020). *Foreign Language Training*. Foreign Language Training. https://www.state.gov/foreign-language-training/

van der Hoeven, N., & de Bot, K. (2012). Relearning in the Elderly: Age-Related Effects on the Size of Savings: Relearning in the Elderly. *Language Learning*, *62*(1), 42–67. https://doi.org/10.1111/j.1467-9922.2011.00689.x

Vaughn, K. E., & Rawson, K. A. (2011). Diagnosing criterion-level effects on memory: What aspects of memory are enhanced by repeated retrieval? *Psychological Science*, *22*(9), 1127–1131. https://doi.org/10.1177/0956797611417724

Verfaellie, M., Croce, P., & Milberg, W. P. (1995). The role of episodic memory in semantic learning: An examination of vocabulary acquisition in a patient with amnesia due to encephalitis. *Neurocase*, *1*(4), 291–304. https://doi.org/10.1080/13554799508402374

Wang, S., Allen, R. J., Fang, S.-Y., & Li, P. (2017). Cross-modal working memory binding and L1-L2 word learning. *Memory & Cognition*, *45*(8), 1371–1383. https://doi.org/10.3758/s13421-017-0731-2

Wasserman, M., Renfrew, M. R., Green, A. R., Lopez, L., Tan-McGrory, A., Brach, C., & Betancourt, J. R. (2014). Identifying and Preventing Medical Errors in Patients With Limited English Proficiency: Key Findings and Tools for the Field. *Journal for Healthcare Quality*, *36*(3), 5–16. https://doi.org/10.1111/jhq.12065

Webb, S. (2005). Receptive and productive vocabulary learning: The Effects of reading and writing on word knowledge. *Studies in Second Language Acquisition*, *27*(01). https://doi.org/10.1017/S0272263105050023

Webb, S. (2008). Receptive and productive vocabulary sizes of L2 learners. *Studies in Second Language Acquisition*, *30*(01). https://doi.org/10.1017/S0272263108080042

Webb, S., & Rodgers, M. P. H. (2009). Vocabulary Demands of Television Programs. *Language Learning*, *59*(2), 335–366. https://doi.org/10.1111/j.1467-9922.2009.00509.x

Whiting, E., Chenery, H. J., & Copland, D. A. (2011). Effect of aging on learning new names and descriptions for objects. *Aging, Neuropsychology, and Cognition*, *18*(5), 594–619. https://doi.org/10.1080/13825585.2011.598912

Woolf, S. H., & Schoomaker, H. (2019). Life Expectancy and Mortality Rates in the United States, 1959-2017. *JAMA*, *322*(20), 1996. https://doi.org/10.1001/jama.2019.16932

Zhang, Y., Ridchenko, M., Hayashi, A., & Hamrick, P. (2021). Episodic memory contributions to second language lexical development persist at higher proficiencies. *Applied Cognitive Psychology*, *35*(5), 1356–1361. https://doi.org/10.1002/acp.3865

Appendix I: Experiment 1 Stimuli

Table 7. List of L1-L1 word pairs used in Experiment 1

L1	L1	L1	L1	L1	L1
blend	beak	lime	cow	skirt	draw
box	bomb	lion	burn	smash	elephant
call	blouse	mix	camel	snap	bear
celery	brush	mouth	basement	sock	destroy
chat	boots	murder	feather	sofa	ascend
cherry	acquire	mushroom	comb	spinach	exchange
cough	fawn	neck	advise	spoon	flame
crowbar	bean	nose	ascent	spray	giraffe
cry	broccoli	oink	descent	squint	cliff
dagger	cigar	peach	drool	swallow	axe
drive	bark	pencil	belt	swan	donation
dustpan	arrive	pig	airplane	table	bracelet
eye	baseball	plum	burp	teeth	clang
flash	command	potato	barn	tiger	drill
fork	follow	pull	crown	touch	camisole
fox	cobra	push	flea	train	cat
frown	cactus	rabbit	door	truck	cook
get	bird	raisin	flicker	tweezers	find
goat	camp	razor	chirp	van	fish
grin	duck	receive	carrot	wade	drop
growl	floor	repair	congress	watch	bicycle
gun	bounce	return	clash	window	bridge
hammer	coconut	roast	elbow	wolf	crash
hand	catapult	see	arm	wrist	clatter
hatchet	blink	shield	escape	yawn	bureau
leave	couch	shoulder	buy	yell	cabbage
leg	chime	shovel	cushion		

Table 8. List of L1-L2 word pairs used in Experiment 1 and Experiment 2

L1	L2	L1	L2	L1	L2
agent	wakili	forehead	paji	plate	sahani
anchor	nanga	forgery	ubini	poem	utenzi
barrel	pipa	friend	rafiki	poison	sumu
beer	pombe	frog	chura	prayer	sala
boat	mashua	gate	lango	prophet	nabii
bribe	rushwa	glue	ambo	quarry	chimbo
bucket	ndoo	grapes	zabibu	queen	malkia
bull	fahali	harbor	bandari	rumor	fununu
carpet	zulia	honor	adhama	sailor	baharia
cheese	jibini	horse	farasi	scarf	leso
cinnamon	dalasini	invoice	ankra	science	elimu
cloud	wingu	knee	goti	silk	hariri
corn	nafaka	lake	ziwa	sleep	usingizi
cotton	pamba	leaf	jani	snow	theluji
curtain	pazia	leech	ruba	soul	roho
custom	desturi	leisure	wasaa	spite	inda
divorce	talaka	lung	pafu	story	hadithi
doctor	tabibu	maggot	buu	tailor	mshoni
dog	mbwa	manure	samadi	tomato	nyanya
donkey	punda	mattress	godoro	treasure	dafina
dust	vumbi	merchant	tajiri	trench	handaki
economy	iktisadi	monkey	tumbili	trouble	adha
egg	yai	oath	yamini	wheel	duara
enemy	adui	olives	zeituni	wound	jeraha
envelope	bahasha	orphan	yatima	yeast	hamira
flavor	ladha	oyster	chaza	yoke	nira
flood	gharika	pearl	lulu		

Appendix II: Additional Analyses for Experiment 1

Table 9. Hit rate for Experiment 1

Language	Age Group	Test Type	Mean	SD
L1-L1	Older Adult	Cue	.62	.49
		Pair	.64	.48
		Target	.62	.49
	Young Adult	Cue	.64	.48
		Pair	.62	.49
		Target	.55	.50
L1-L2	Older Adult	Cue	.61	.49
		Pair	.70	.46
		Target	.69	.46
	Young Adult	Cue	.55	.50
		Pair	.68	.47
		Target	.67	.47

Table 10. False Alarm rate for Experiment 1

Language	Age Group	Test Type	Mean	SD
L1-L1	Older Adult	Cue	.35	.48
		Pair	.34	.47
		Target	.26	.44
	Young Adult	Cue	.33	.47
		Pair	.27	.45
		Target	.29	.45
L1-L2	Older Adult	Cue	.32	.47
		Pair	.39	.49
		Target	.32	.47
	Young Adult	Cue	.32	.47
		Pair	.37	.48
		Target	.26	.44

Table 11. Correlation Matrix for Experiment 1 Recognition Outcomes

		L1-L1 Cue	L1-L1 Target	L1-L1 Pair	L1-L2 Cue	L1-L2 Target	L1-L2 Pair
Older Adults	L1-L1 Cue	-					
	L1-L1 Target	.48	-				
	L1-L1 Pair	.13	.30	-			
	L1-L2 Cue	.51	.30	.03	-		
	L1-L2 Target	.40	.40	.24	.62	-	
	L1-L2 Pair	.17	.33	.37	.13	.31	-
Younger Adults	L1-L1 Cue	-					
	L1-L1 Target	.70	-				
	L1-L1 Pair	.57	.52	-			
	L1-L2 Cue	.40	.34	.32	-		
	L1-L2 Target	.62	.51	.42	.41	-	
	L1-L2 Pair	.36	.36	.45	.58	.43	-

Table 12: Coefficients for Alternate Aim 2 model

A multi-level linear regression model fixed effects output for the final recognition test scores (Intercept is the Older Adult cue memory test for the L1-L2 word list). β and standard error are presented in percentage accurate units. Only significant outcome indicated is the main effect of L1-L1 score's on all L1-L2 scores, consistent with analysis presented in Aim 2.

	B	Std. Error	df	*t*	*p*
(Intercept)	.37	.10	209.12	3.69	<.001
Young Adult	.04	.13	215.37	.26	.79
Pair Memory	.08	.14	189.34	.56	.58
Target Memory	.08	.14	169.43	.55	.58
L1-L1 Score	.42	.16	205.29	2.71	.01
Young Adult : Pair Memory	-.10	.18	179.72	-.57	.57
Young Adult : Target Memory	-.04	.18	164.92	-.23	.82
Young Adult: L1-L1 Score	-.11	.20	212.19	-.52	.60
Pair Memory: L1-L1 Score	-.11	.21	191.20	-.51	.61
Target Memory: L1-L1 Score	-.09	.22	169.94	-.41	.69
Young Adult : Pair Memory: L1-L1 Score	.19	.27	181.93	.72	.47
Young Adult : Target Memory: L1-L1 Score	.17	.27	165.44	.63	.53

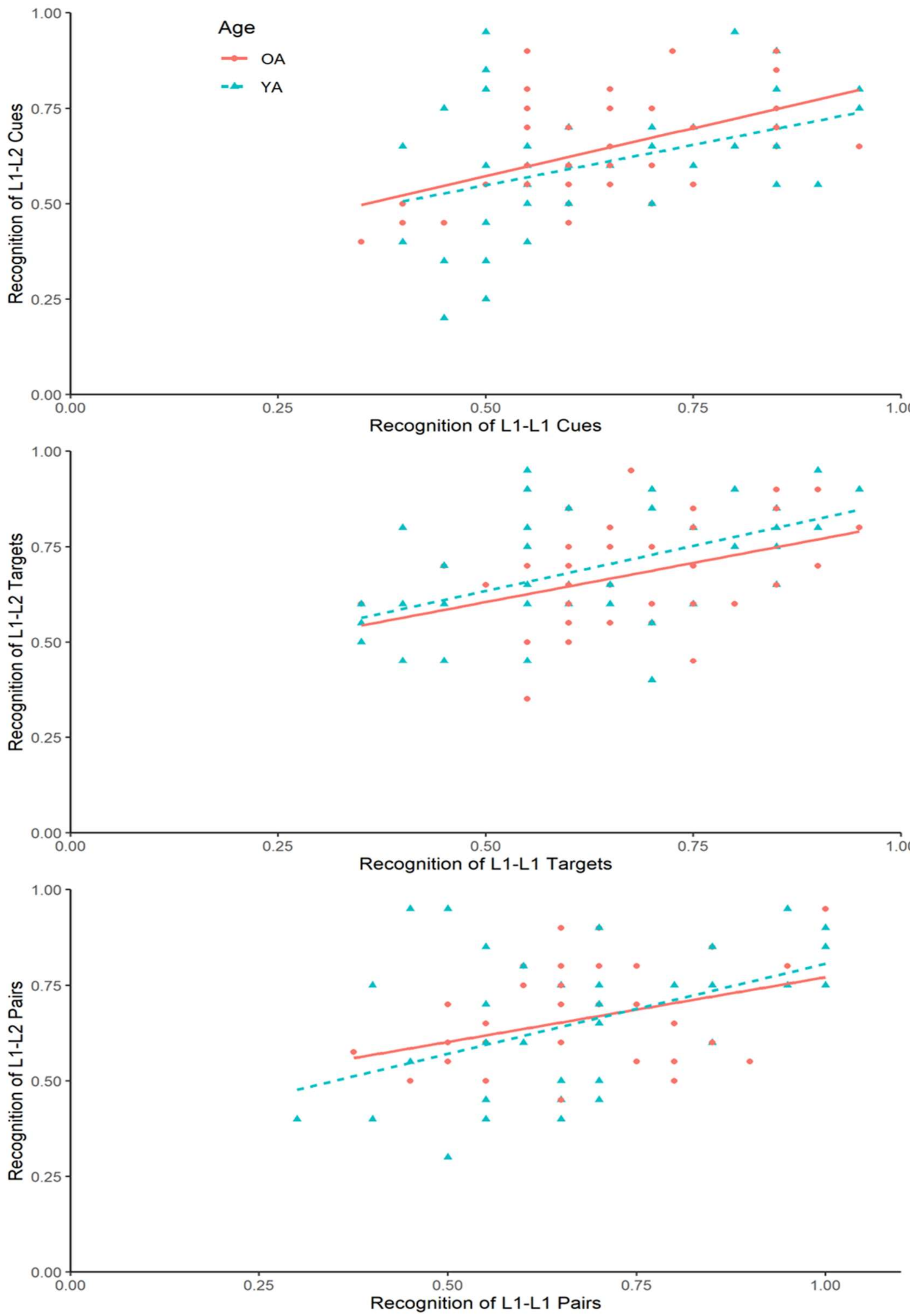

Figure 12. Experiment 1 Scatterplot. Rates of recognition separated by test type (Cue vs Target vs Pair) and Age Group.

Appendix III: Additional Analyses for Experiment 2

Table 13. Hit rate for Experiment 2

Training Type	Age Group	Test Type	Mean	SD
Restudy	Older Adult	Cue	.70	.46
		Target	.71	.45
		Pair	.75	.44
	Young Adult	Cue	.62	.48
		Target	.69	.46
		Pair	.77	.42
Productive	Older Adult	Cue	.76	.43
		Target	.69	.46
		Pair	.76	.43
	Young Adult	Cue	.81	.39
		Target	.84	.36
		Pair	.80	.40
Receptive	Older Adult	Cue	.80	.40
		Target	.60	.49
		Pair	.81	.39
	Young Adult	Cue	.83	.38
		Target	.75	.43
		Pair	.80	.40

Table 14. False Alarm rate for Experiment 2

Training Type	Age Group	Test Type	Mean	SD
Restudy	Older Adult	Cue	.19	.39
		Target	.26	.44
		Pair	.54	.50
	Young Adult	Cue	.26	.44
		Target	.27	.44
		Pair	.57	.50
Productive	Older Adult	Cue	.21	.41
		Target	.24	.43
		Pair	.58	.49
	Young Adult	Cue	.22	.41
		Target	.19	.39
		Pair	.52	.50
Receptive	Older Adult	Cue	.24	.43
		Target	.28	.45
		Pair	.52	.50
	Young Adult	Cue	.20	.40
		Target	.23	.42
		Pair	.55	.50