

© 2014 Brennan R. Payne

THE EFFECTS OF VERBAL WORKING MEMORY TRAINING ON  
LANGUAGE COMPREHENSION IN OLDER ADULTHOOD

BY

BRENNAN R. PAYNE

DISSERTATION

Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in Educational Psychology  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2014

Urbana, Illinois

Doctoral Committee:

Professor Elizabeth A.L. Stine-Morrow, Chair  
Professor Kara D. Federmeier  
Associate Professor Kiel Christianson  
Associate Professor Duane G. Watson

## Abstract

Effective language understanding is crucial to maintaining cognitive abilities and learning new information through adulthood. However, age-related changes in cognitive abilities such as working memory (WM) have a profound influence on the products of language comprehension (e.g., problem solving, learning, following instructions). At the same time, the effects of age and working memory on the moment-to-moment processes underlying language comprehension are less well understood. The current project tests the causal role of working memory in language among older adults by examining the effects of a short-term working memory training program on changes in language comprehension. This dissertation describes the development of the *iTrain* program, a novel home-based computerized training program targeting complex verbal WM performance, and describes the results from a single 3-week randomized controlled training experiment testing the efficacy of *iTrain* on improving verbal working memory, language processing, and language comprehension outcomes in older adults. Results showed that individuals in the WM training group showed substantial improvements in the trained WM tasks and transfer to untrained verbal WM tasks. Additionally, results suggested that training led to selective improvements in aspects of language comprehension relative to an active control group, including improvements in sentence recall, verbal fluency, and comprehension of syntactically ambiguous sentences. Results from eye tracking did not reveal effects of training on on-line language processing. The results from this study provide some of the first causal evidence for the influence of WM on comprehension in aging.

*To my brothers, Kyle and Nathan*

*Dedicated to the memory of my mother, Rhonda D. Payne*

## Acknowledgements

It is my pleasure to thank those who have made this dissertation possible. First, I am very grateful to have support from my committee members: Dr. Elizabeth A.L. Stine-Morrow, Dr. Kara Federmeier, Dr. Kiel Christianson, and Dr. Duane Watson. It has been a pleasure to work with and learn from such great mentors during my time here as a graduate student. In particular, I would like to thank my advisor, Dr. Elizabeth A. L. Stine-Morrow. This work would not have been possible without her support, insight, and overall enthusiasm. I deeply appreciate the amount of time and effort Liz gave to working on this project (and many others) with me over the last several years. I would like to sincerely thank my friends and family both here and afar for their encouragement, feedback, and general greatness: Kyle Payne, Nathan Krieger, Mallory Stites, Matt Windsor, Pat Hill, Jennie Mae Fullington, Nick Oliver, Adam Williams, Joshua Morris, Erika Hussey, Alden Gross, and Jeanine Parisi. I want to thank my fiancée, Monika Lohani, who has provided me with endless support and love. She has sat up with me on endless nights listening to me ramble and think out loud as I worked through the many challenges of this project. Most of all, she has always remained patient, encouraging, and optimistic. I couldn't have done this without her. I wish to thank my lab mates and research assistants for their direct help on the development of the iTrain project, and testing: Sneha Gummulurri, Thomas Deegan, Andy Battles, Uttam Rajan, Verlisa Shanklin, Xiaomei Liu, and Laura Giffin. Finally, I would like to acknowledge the Campus Research Board, the language processing training grant, and the Beckman Institute pre-doctoral fellowship for providing research and training support during the duration of this project.

## Table of Contents

<b>Chapter I. Introduction and Literature Review .....</b>	<b>1</b>
<b>Chapter II. The iTrain Program for Training Verbal Working Memory .....</b>	<b>28</b>
<b>Chapter III. Randomized Controlled Experiment Methods and Design .....</b>	<b>34</b>
<b>Chapter IV. Responsiveness to Home-Based Working Memory Training.....</b>	<b>49</b>
<b>Chapter V. Transfer of Training to Working Memory and Language Comprehension .....</b>	<b>55</b>
<b>Chapter VI. Effects of Training on Syntactic Comprehension: Evidence from Eye Movement Control During Reading .....</b>	<b>61</b>
<b>Chapter VII. General Discussion .....</b>	<b>73</b>
<b>References .....</b>	<b>89</b>
<b>Tables and Figures .....</b>	<b>113</b>
<b>Appendix A: iTrain Screen Captions and Links to Video Demos.....</b>	<b>154</b>
<b>Appendix B: Expectation Survey and Results.....</b>	<b>159</b>

## **Chapter I.**

### **Introduction and Literature Review**

Effective language understanding is crucial to maintaining cognitive abilities (Many, Touradji, Tang & Stern, 2003; Stern, 2009) and learning new information through adulthood (Payne, Gao, Noh, Anderson, & Stine-Morrow, 2012; Stine-Morrow & Miller, 2009). However, normative age-related changes in cognitive ability have a profound effect on language understanding, especially for effortful processes related to comprehension and memory for language (Burke & Shafto, 2008; Wingfield & Stine-Morrow, 2000; Wlotko, Lee, & Federmeier, 2010). Working memory (WM) —the ability to temporarily store, maintain, and organize task-relevant information— is often implicated as a domain-general mechanism responsible for such age-related changes in language understanding (Kemper, 2012; van der Linden et al., 1999; Wingfield & Stine-Morrow, 2000). Although many models of sentence processing include some mechanism to account for memory constraints (see Caplan & Waters, 2013; Pickering & van Gompel, 2006 for reviews), the degree to which the WM system directly supports comprehension and the role of WM in immediate language interpretation are currently areas of much controversy.

The majority of research examining WM influences on language comprehension in healthy younger and older adults has relied on dual-task paradigms, which manipulate memory load (as a proxy for WM) (Fedorenko, Gibson, & Rohde, 2006; Kemper & Herman, 2006; Smiler, Gagne, & Stine-Morrow, 2008; Waters & Caplan, 1996), or individual difference paradigms, which correlate psychometric measures of WM with measures of language comprehension (Caplan & Waters, 1999; Caplan et al., 2011; DeDe et al., 2004; Just & Carpenter, 1992; King & Just, 1991; Payne et al., 2014; Stine, 1990; Stine-Morrow et al., 2008).

In contrast, the current study draws on a growing literature in working memory training (Morrison & Chein, 2010; Shipstead et al., 2013; Karbach & Verhaeghen, 2014) in order to adopt an experimental approach to examine the degree to which WM underlies language processing and performance in older adulthood. Specifically, this dissertation: (1) introduces the iTrain project, a novel home-based complex verbal WM training program, (2) assesses the effects of 3 weeks of training on changes in verbal WM performance and (3) assesses the degree to which short-term WM training leads to improvements in measures of language comprehension and on-line language processing, as assessed by eye tracking.

In the following sections, I introduce the major aims of this work by discussing (a) theories of the functional role of WM in language comprehension, (b) theories of how language comprehension is shaped by individual differences in working memory, (c) the cognitive aging of working memory and language comprehension, and (d) evidence of training-related WM plasticity in older adults.

### **Theories of the Functional Role of Working Memory in Language Comprehension**

Working memory resource limitations have historically been invoked in psycholinguistic models of language understanding to explain comprehension difficulties for linguistically complex material. One of the earliest examples comes from George Miller and Noam Chomsky (1963), who outlined clear limitations on human capacity of understanding certain syntactic constructions, such as multiple center-embeddings like (1) (see also Yngve, 1960):

(1) The rat the cat the dog chased ate died.

In order to process multiple center embeddings, each additional noun phrase (e.g., the rat, the cat, the dog) must be maintained in an immediate memory system that must be continually updated



in order to later connect each element to its respective predicate (e.g., chased, ate, died).

Although (1) is a short grammatical sentence with short high-frequency words, it is still difficult to understand, largely because the storage and maintenance of each incomplete dependency appears to overload the comprehension system. Indeed, a recent multi-language corpus assessment of several “standard average European” languages that permit such multiple center embeddings (e.g., English, Finnish, French, German) found that a maximum of only three center embeddings are ever found in such languages (Karlsson, 2007). These findings suggest that memory constraints are a real limiting factor of comprehension that may indeed shape the statistical properties of certain constructions in the language.

The introduction of Baddeley and Hitch’s (1974) multi-component working memory model laid the groundwork for a new era of investigation into individual differences in working memory in cognitive psychology. A series of clever experiments by Baddeley and colleagues (reviewed in Baddeley, 2003, 2012) clearly showed that simple short-term memory (STM) storage capacity is not predictive of higher-order cognition. This work was consistent with a growing literature showing that individual differences in STM capacity were uncorrelated with verbal ability and language comprehension (Daneman & Carpenter, 1980; Perfetti & Lesgold, 1977). Indeed, Baddeley and Hitch (1974) de-emphasized their focus on storage per se, and instead emphasized the functional properties of working memory— that is, the orchestration of storage and maintenance in WM along with the concurrent processing of incoming information. While there are many contemporary models of WM, each of which make slightly different predictions or have slightly different foci (e.g., Cowan, 2001; Engle, 2002, 2010; Kane & Engle, 2002), most models converge on a similar account that WM is “the ability to simultaneously

maintain information in an active and readily accessible state, while concurrently and selectively processing new information...” (Conway, Jarrold, Kane, Miyake, & Towse, 2007; p. 3).

Complex working memory measures such as the *reading span* task (Daneman & Carpenter, 1980) and the *operation span* task (Turner & Engle, 1989) were developed to measure an individual’s ability to coordinate the dual-task processing of short-term memory storage and the continuous manipulation of information in STM. These tasks share the requirement that participants must simultaneously hold a series of items in memory while performing some concurrent processing task (e.g., reading a sentence for comprehension or solving a mathematics problem). Studies examining performance in these tasks have shown that complex WM performance predicts substantial portions of variance in higher order cognitive abilities including inductive reasoning, episodic memory, and language comprehension (see Conway, Jarrold, Kane, Miyake, & Towse, 2007 for reviews). Given these findings, it is perhaps unsurprising that models of language processing call on such memory mechanisms to explain constraints in sentence processing and comprehension. I briefly review two prominent models that have invoked basic working memory mechanisms to account for processing difficulties associated with structural complexity.

*Syntactic Prediction Locality Theory*. Gibson’s (1998) syntactic prediction locality theory (SPLT) hypothesizes that processing difficulty is determined by two components: *storage costs* and *integration costs*, which draw on the same set of working memory resources. Storage costs occur when a primary element of a linguistic dependency has to be stored in short-term memory over some interval, while new information is being simultaneously processed and maintained, before that element can be integrated with some later dependent element. An integration cost occurs at the point at which the dependent element has been encountered and must be integrated

with the primary element in working memory, completing the dependency. The major claim of SPLT is that understanding a sentence requires some working memory system to maintain the partial products of language processing (i.e., incomplete dependencies), so that relations between distal parts of a sentence can be rapidly computed on-line.

Findings consistent with SPLT have been observed in studies manipulating long-distance dependencies (Chen, Gibson, & Wolf, 2005; Grodner & Gibson, 2005; Bartek, Lewis, Vasishth, & Smith, 2011; Balogh, Zurif, Prather, Swinney, & Finkel, 1998) and object-relative clause processing (Gibson, 1998; Grodner & Gibson, 2005; Wu & Gibson, 2008). For example, Grodner and Gibson (2005) showed that there are substantial costs in online processing time at an embedded verb when a long-distance dependency has been introduced between a head noun and the target verb (e.g., compare sentence (3) with sentence (2)).

(2) The boy who the girl *grabbed* lost his balance.

(3) The boy who the girl who fell down the stairs *grabbed* lost his balance.

These effects have been replicated and extended by Bartek, Lewis, Vasishth, and Smith (2011), who showed robust effects of long-distance dependency on both early (first fixation duration) and late-pass (regression path duration) eye-movements during reading, in both relative clause constructions (as in the above example sentences (2) and (3)), as well as in main clause constructions (e.g., The girl *grabbed* the boy... vs. The girl who fell down the stairs *grabbed* the boy...).

Indeed, a common finding in psycholinguistics is that object-relative (OR; sentence 5) constructions are more difficult to process than subject-relative (SR; sentence 4) constructions (see Gordon & Lowder, 2012; Pickering & van Gompel, 2006 for reviews), producing both increased reading times at the matrix verb, and increased errors in comprehension. SPLT

attributes this difficulty to increased memory demands while reading the OR clause, as in (4) and (5), because comprehension requires the retrieval of the matrix subject (*the reporter*) across the intervening noun phrase at the matrix verb (*admitted*).

(4) The reporter that attacked the senator *admitted* the error after the hearing.

(5) The reporter that the senator attacked *admitted* the error after the hearing.

Data from event-related potential studies of language comprehension have revealed reliable working memory maintenance effects in on-line sentence processing (reviewed in Kutas, Van Petten, & Kluender, 2006) consistent with SPLT. For example, a number of ERP studies have found a reliable slow anterior negative potential associated with processing long-distance dependencies (Fiebach et al., 2001; Kluender & Kutas, 1993; Phillips, Kazanina, & Abada, 2005) and object-relative clauses (King & Kutas, 1995; Mueller, King, & Kutas, 1997). These findings have been explained as WM costs associated with the continued maintenance of an element over the relative clause region, until its trace has been encountered (Fiebach et al., 2001; Munte et al., 1998; Phillips, Kazanina, & Abada, 2005), consistent with both the storage cost and integration costs mechanisms in SPLT.

*Retrieval-Based Parsing.* Another influential theory of memory mechanisms in sentence comprehension is the cue-based parsing framework, by Lewis, Vasishth and colleagues (Lewis & Vasishth, 2005; Lewis, Vasishth, & Van Dyke, 2006; Van Dyke & Lewis, 2003). The cue-based parsing framework focuses on interference in memory encoding and retrieval as a major source of parsing difficulties. This model is instantiated in a computational process model in the ACT-R architecture (Taatgen & Anderson, 2008), and successfully models on-line sentence processing with a small number of basic mechanisms: (1) a limited focus on attention in working memory (Cowan, 2001), (2) similarity-based interference in encoding and retrieval, and (3) fluctuating

activation in WM as a function of decay (Vasishth & Van Dyke, 2006). Like SPLT, this model can account for on-line syntactic processing costs, based on interference at encoding (similar to storage costs in SPLT) and interference and decay at retrieval (similar to integration costs in SPLT).

Evidence consistent with this model comes from studies showing that both semantic (Gordon et al., 2002, 2006; Fedorenko et al., 2006) and syntactic (Van Dyke & Lewis, 2003) sources of interference in working memory influence on-line sentence processing. Gordon et al., (2002) showed that when items in an external memory load matched in semantic category class (i.e., proper nouns like Joel, Andy Greg), with NPs in an object relative cleft construction (e.g., “It was [Sam/ the manager] that [Tony/ the clerk] liked before the argument began.”), reading times were greater than in a subject-relative cleft construction (see also Fedorenko et al., 2006). Similarly, syntactic similarity of lexical items has been argued to cause interference in parsing (Van Dyke & Lewis, 2003). In four experiments, Van Dyke and Lewis (2003) showed that introducing items that match in syntactic category (e.g., number of subject NPs) result in greater processing times at retrieval sites (e.g., verbs), holding length constant. Under the cue-based parsing account, when a retrieval site is encountered, readers activate syntactic features of the item to be retrieved in short-term memory, and when multiple items contain the same syntactic class information (e.g., subject NP), this interference slows retrieval (but see Caplan & Waters, 2013 for a critique).

### **Individual Differences in Working Memory and Language Comprehension**

Although the predictions from both the cue-based parsing framework and SPLT model explain data for groups of individuals well, knowledge about how individual differences in

working memory capacity fit within such theories in less clear. Indeed, a separate literature has developed attempting to account for individual differences in working memory and language comprehension in healthy adults and special populations. In the following, I briefly review three of these theories.

*Capacity Constrained Model.* Just, Carpenter, and colleagues (Just & Carpenter, 1980, 1992; Just, Carpenter, & Keller, 1996; Just & Varma, 2002, 2007) Capacity-Constrained (CC) theory of comprehension is arguably the most influential of these models, sparking much of the research and debate on individual differences in language comprehension. Indeed, the first article to describe the capacity-constrained model in detail (Just & Carpenter, 1992) has been cited over 3,000 times, indicating its widespread influence. The basic claims of the CC model were introduced by Just and Carpenter (1980), leading to the formalization and refinement of a computational architecture to model WM constraints in language comprehension (CC-READER, Thibadeau, Just, & Carpenter, 1982; 3CAPS, Just & Carpenter, 1992; 4CAPS, Just & Varma, 2007). The basic claims of the CC model, as it relates to language processing, are as follows (Just & Carpenter, 1992):

1. A general verbal WM system entails the “the set of processes and resources that perform language comprehension,” corresponding to “the part of the central executive...that deals with language comprehension” (p. 123) in Baddeley and Hitch’s (1974) model.

2. Maintenance and computational processing share resources in verbal WM, which is modeled as trade-offs in level of activation (cf. Anderson, 1983; Rogers & McClelland, 2004; 2008).

3. When task demands exceed available resources, both storage and computational functions are degraded.

4. The nature of an individual's language comprehension abilities is dependent upon individual differences in the capacity of the verbal WM system.

5. Tasks assessing performance on complex verbal WM span (e.g., the reading span task, listening span task) tap into the verbal WM system, and performance on these tasks will predict individual differences in language comprehension.

Indeed, in healthy college-aged adults, the meta-analytic correlations between verbal WM and offline measures of language comprehension performance (e.g., standardized reading comprehension, sentence and text recall, inference making, and ambiguity detection) are substantial (Daneman & Merikle, 1996), ranging in magnitude between .41 and .52. Thus, there does appear to be a robust relationship between complex working memory span performance and off-line language comprehension performance. At the same time, evidence for the influence of WM on on-line language processing, the immediate interpretation of language as it unfolds moment-to-moment, is less clear. Just and Carpenter (1992) review a series of behavioral, neuroimaging, and neuropsychological results that support the claim that individual differences in working memory immediately constrain language interpretation on-line. The most widely discussed of these findings (Caplan & Waters, 1999; MacDonald & Christiansen, 2002; Just & Varma, 2002; 2007; Wells et al., 2008) are those of King and Just (1991). In this study, younger adults were categorized into low- and high-span based on performance on the reading span task. Participants then read a series of subject-extracted relative clause sentences, such as (5), and object-extracted relative clause sentences, such as (4), in a self-paced reading paradigm.

King and Just (1991) found a reliable interaction between reading span performance and sentence type on reading times at the main verb (e.g., *admitted*), such that the low-span readers showed a object-relative processing cost of 197 ms, which was larger than the 87 ms cost among

high-span adults. These findings suggested that the increased memory load associated with object-relative clause processing was more costly among adults with lower verbal working memory resources available to allocate to on-line processing. A number of more recent studies have also presented data consistent with this claim. For example, as discussed above, increasing the similarity of items in an external memory load impacts both comprehension and on-line efficiency at the most demanding part of object-relative sentences (Fedorenko et al., 2006; Gordon et al., 2006). Traxler (2007, 2009; see also Felser et al., 2003; Swets et al., 2008), has also found that individual differences in WM impact “late pass” eye movement measures of syntactic ambiguity resolution, though these patterns are not always replicated (Traxler et al., 2005).

*Separate Language Interpretation Resource.* The CC model has been influential in motivating research primarily because of the directly targeted claims it makes about individual differences and comprehension. These claims have not gone without substantial debate. Caplan, Waters, and colleagues (Caplan & Waters, 1990; 1999; 2007; Waters & Caplan, 1996) strongly critiqued the CC model and introduced the Separate Language Interpretation Resource (SLIR) model as an alternative account of individual differences in language processing. SLIR is a fractionated working memory model, with a domain-specific resource for language interpretation that is independent from the conscious and controlled verbal working memory system (in the sense of Just & Carpenter, 1992) tapped by tasks like reading span.

The SLIR model makes a distinction between *interpretive* processes, which are “...an integrated set of largely unconscious, obligatory, on-line, first pass processes devoted to assignment of the literal, preferred, discourse-congruent meaning of utterances...” (Caplan & Waters, 1999; p. 128) and *post-interpretive* processes, which include conscious processes related



to remembering the semantic content of a sentence, using the meaning of a sentence to plan actions, and reasoning on the basis of sentence meaning. Evidence for SLIR is based largely on the findings from Caplan and colleagues that (1) on-line measures of the effects of syntactic processing difficulty are uncorrelated with individual differences in verbal working memory, (2) external memory loads do not always impact on-line language processing, and (3) neuropsychological data indicate that patients with central nervous system disorders that impair verbal working memory performance show no impairments in on-line language processing (Caplan & Waters, 1990; Caplan, Waters, & DeDe, 2007; Martin & Feher, 1990). Although the evidence for the SLIR model in healthy adults primarily relies on null findings (i.e., lack of relationship between syntactic processing and WM), Caplan and colleagues have replicated these null results with large and diverse samples, across a substantial number of studies (see Caplan, Waters, & DeDe, 2007 for a recent review), suggesting that these findings are not likely driven by power issues in detecting effect sizes.

One limitation of the SLIR model is that almost all of the experiments that have failed to find effects of working memory on language processing supported by SLIR have utilized a single behavioral paradigm, the auditory moving window (AMW) method (Ferreira et al., 1996). In this method, participants self-pace through segments of pre-recorded speech, and reaction times between the sectors are used as the “on-line” measure, uncorrected for presentation time for each segment. AMW has been critiqued as especially unnatural and less sensitive than other tasks, such as eye tracking (Kemper & Liu, 2007; Rayner, 1998) and ERPs (Kutas & King, 1999). This is particularly troublesome for SLIR, because insensitive measures are more likely to result in null findings, potentially confirming predictions based on methodological artifacts.

*Experience Constraint.* Lastly, MacDonald and Christiansen (2002) have strongly critiqued both the CC and SLIR theories, arguing that the working memory systems discussed by Just and Carpenter (1992) and Caplan and Waters (1996; 1999) are “theoretical soup stones” (cf. Navon, 1984) that do not offer any explanatory power in theories of language comprehension. They argue that the CC model does not distinguish between performance on working memory tasks and performance on tasks that index language ability. That is, “reading span, lexical decision, and reading are all just language processing tasks with slightly different task demands, and experiment participants marshal their comprehension in different ways to meet those demands” (MacDonald & Christiansen, 2002; p. 39). A major claim of their model is that individual differences in language performance largely reflect individual differences in language experience, which determines performance on both verbal WM tasks and language processing tasks.

At the same time, evidence for this theory is lacking. Primary evidence for this theory comes from computational simulations performed by MacDonald and Christiansen (2002), in which a series of simple recursive networks were provided with differing amounts of training experience on a simple grammar with various syntactic constructions, including subject-relative and object-relative sentences. Activation in response to OR sentences was dependent upon the degree of exposure to OR sentences in the training set, and MacDonald and Christiansen argued that these results mimicked the findings of King and Just (1991) without invoking working memory constraints per se. Rather, they argued, these findings suggested that the observed reading span effects are due to differences in language experience between high-span and low-span individuals. More recently Wells, Christiansen, Race, Acheson, and MacDonald (2008) have tested the claims in the experience-constraint model empirically, by exposing participants

to a relatively small set ( $n = 80$ ) of subject- and object-relative sentences over the course of 4 training sessions. Surprisingly, despite the small amount of training, there was evidence that the trained group showed facilitated processing of OR sentences (relative to SR sentences) compared to a control group that did not receive exposure to subject and object relative sentences.

It is important to note that, while the role of linguistic experience is significant in determining comprehension, this does not preclude the possibility that WM capacity is also a factor shaping comprehension (Engle, 2010; Just & Varma, 2002, 2007). This is especially true for older adults, who show both normative increases in verbal ability (Verhaeghen, 2003) and linguistic experience (Stanovich et al., 1995; Payne et al., 2012; Payne et al., 2014), as well as age-related declines in verbal working memory (Bopp & Verhaeghen, 2005). These findings suggest that, at least developmentally, verbal working memory and linguistic experience can be functionally dissociated and may have separable influences on on-line language processing (Payne et al., 2014).

### **Cognitive Aging of Working Memory and Language Processing**

Two divergent paths often characterize cognitive aging. In one route, aging is associated with monotonic declines in fluid cognitive abilities, which are based on the processing efficiency of the cognitive system (Park et al., 1996; Salthouse, 2008). However, abilities based on the accumulation of knowledge and experience, so-called crystallized abilities, are often stable or show selective growth into adulthood (Beier & Ackerman, 2005; Baltes, 1997; Schaie, 1994). Investigations into differential effects of age on cognitive ability have been studied since as early as the 1920's (Foster & Taylor, 1920). The distinction between these trajectories still remains a robust finding in contemporary research. Tracking crystallized cognition across the lifespan

illustrates that these abilities show relative age invariance, only declining in very late life. For example, older adults typically possess high levels of general world knowledge (Ackerman, 2008), and on average, also have a larger vocabulary from a lifetime of accumulated verbal knowledge (Verhaeghen, 2003). On the other hand, aging brings reductions in cognitive abilities including working memory capacity (Bopp & Verhaeghen, 2005), speed of processing (Salthouse, 1996), inhibition (Hasher & Zacks, 1988) and executive and attentional control processes (Kramer & Madden, 2008).

While some aspects of language use appear to be spared with advancing age, it is widely agreed upon that these age-related changes in cognitive ability influence how we process language in older adulthood (Burke & Shafto, 2008; Federmeier, 2007; Thorton & Light, 2006; Wingfield & Stine-Morrow, 2000; Stine-Morrow & Miller, 2009). In the following, I briefly review the literature suggesting that age-related changes in working memory are robust, and that age-related changes in language comprehension and language processing may, to some extent, be driven by declines in WM processes.

*Aging of Working Memory and Executive Control.* A number of studies have shown small age-related declines in STM, but large age-related declines in complex WM performance. A meta-analysis by Bopp and Verhaeghen (2005) summarized effect sizes of 124 studies of aging of STM and WM performance. Using Brinley function plots (i.e., older adults' performance as a function of younger adults' performance; Brinley, 1965), they showed that older adults' capacity in STM was 92% that of the young. At the same time, older adults' performance on complex WM capacity tasks, which involve dual-task costs of maintenance and processing, was only 74% that of their younger counterparts. Thus, complex WM performance shows robust normative age-related declines.

There has been a growing focus on explaining widespread age-related declines in a number of cognitive abilities, including working memory, in terms of localized declines in executive control functions (e.g., inhibitory control, task switching, goal maintenance, updating) (Balota et al., 2001; Lustig et al., 2007; West, 2001). However, there is reason to be critical of the hypothesis that executive control mechanisms such as inhibition control and task switching are responsible for age-related declines in higher order cognitive function (cf. Burke, 1997; Verhaeghen & Cerella, 2002). A series of meta-analyses by Verhaeghen (Verhaeghen, 2011, 2012; Verhaeghen & Cerella, 2002) have cast doubt on the contention that age-related declines in so-called “executive control” components of attention are reliable in the absence of working memory constraints. Using hierarchical linear models of Brinley functions, Verhaeghen tested whether measures of executive control showed differential age-related declines. Tasks of selective attention and inhibitory control (e.g., Flanker, Stroop), and local task switching costs showed no evidence for selective age-related deficits. However, tasks that required dual-task costs of storage and processing did show specific and selective age-related deficits over and above age-related slowing (see also Bopp & Verhaeghen, 2005; Verhaeghen & Salthouse, 1997). In line with theories of complex WM performance (Engle, 2010), maintaining a dual-task load of memory storage and concurrent processing results in substantial age-related deficits (Verhaeghen & Salthouse, 1997). Further results from meta-analytic structural equation models suggested that executive control tasks explained no additional age-related variance in reasoning or episodic memory over and above the substantial effects of psychomotor speed and verbal WM. These findings suggest that working memory is a unique predictor of age-related cognitive declines, and casts doubt on the view that executive control alone can account for age-related deficits in comprehension, in the absence of memory costs.

*Aging and Off-line Comprehension.* Age differences in off-line measures of language comprehension (Kemper, 1986, 1987; DeDe et al., 2004) and language memory (Johnson, 2003; Stine-Morrow et al., 2008) are robust. Indeed, estimates from a meta-analysis by Johnson (2003) have shown that, on average, older adults perform at about the 22<sup>nd</sup> percentile of the distribution of younger adults in memory for discourse, with effect sizes for age-group differences ranging between .60 *SD* and .92 *SD* across studies. Similar effect sizes have been found in a longitudinal study tracking changes in older adults memory for discourse over a 10- year period (Payne, Gross, Parisi, Sisco, Stine-Morrow, Marsiske, & Rebok, 2014).

Although memory for language is often treated as a component of episodic memory more generally (Hultsch, Hertzog, Dixon, & Small, 1998), the maintenance of sentences and connected discourse involves processes that are unique, including the continuous decoding and integration of phonological and lexical representations, parsing incoming strings into syntactic constituents, abstracting and retaining message-level semantics separate from the verbatim form, and integrating message-level propositional information across sentences (Frazier & Rayner, 1982; Kintsch & van Dijk, 1978; Kintsch, 1998). Not surprisingly, memory for language is supported by cognitive underpinnings that are distinct from those that underlie episodic memory, such as memory for word lists (Lewis & Zelinski, 2010). Maintaining a propositional representation from text is cognitively demanding and shows substantial declines in older adulthood (Johnson, 2003; Payne et al., 2014; Radvansky, 1999; Stine et al., 1995).

As discussed above, the association between complex WM span and language comprehension is substantial among younger adults (Daneman & Merikle, 1996). It may not be surprising then that verbal WM has been found to be a focal mediator of adult age differences in

both memory for text (Hertzog, Dixon, Hultsch, & MacDonald, 2003; Stine-Morrow, Miller, Gagne, & Hertzog, 2008; Van der Linden et al., 1999) and offline measures of language comprehension (DeDe et al., 2004). Age differences in sentence comprehension accuracy are also larger for sentences that are more complex, and these differences in performance have been shown to be dependent upon individual differences in WM (Christianson et al., 2006; Kemper, 1986, 1987, 1992; Stine & Hindman, 1994; Stine-Morrow et al., 2006; Stine-Morrow et al., 2000).

So-called “garden path” sentences such as (6) introduce a temporary syntactic ambiguity that must be resolved in order to comprehend the sentence, and have been used to examine the effects of working memory capacity on both on-line and off-line syntactic ambiguity resolution.

(6) The experienced soldiers warned about the dangers conducted the midnight raid.

Typically in this sentence, the first verb *warned* is initially (incorrectly) interpreted as the main verb of the sentence, rather than as the verb of the reduced relative clause (i.e., ...soldiers “[who were] warned about the dangers...”). Thus, individuals experience difficulty as they encounter the phrase “about the dangers,” and need to revise their initial main clause analysis in favor of the less common reduced relative parse in order to successfully understand the sentence (Bever, 1970; Christianson et al., 2006; Clifton, Traxler, Taha Mohamed, Williams, Morris, & Rayner, 2003; Ferreira & Clifton, 1986; Rayner, Carlson, & Frazier, 1983; Trueswell, Tanenhaus, & Garnsey, 1994). A number of studies have implicated working memory capacity as an important predictor of resolution processes in garden-path ambiguities in younger (Just & Carpenter, 1992; Just & Varma, 2002; MacDonald, Just, & Carpenter, 1992) and older adults (Christianson et al., 2006; Kemper et al., 2004; Kemtes & Kemper, 1997).

One argument is that individuals with greater working memory capacity are able to maintain multiple alternative syntactic representations of ambiguous phrases, which can be directly accessed at the point of disambiguation. However, low span readers are unable to maintain multiple syntactic representations, and therefore commit to one interpretation, so as to necessitate the allocation of more processing time at points of disambiguation in order to revise their incorrect interpretation (MacDonald, Just, & Carpenter, 1992; Kemper et al., 2004). Consistent with this account, across three experiments, Christianson and colleagues (2006) presented three experiments testing the degree to which aging and individual differences in WM influenced participants' off-line comprehension of garden path sentences. They showed evidence for a robust negative correlation between measures of reading span and the probability of incorrectly interpreting garden path sentences in an offline comprehension task among older adults, with correlations ranging between  $-.37$  and  $-.48$ . These findings suggest that older adults with low WM have particular difficulties in revising an initially incorrect interpretation (see also Payne et al., 2014 for similar evidence in the case of syntactic attachment ambiguities).

*Aging and On-line Comprehension.* As discussed above, while age differences in WM are robustly and reliably correlated with age differences in off-line measures of language performance, a relationship between age differences in WM and on-line language processing is less consistently found. In the syntactic processing literature, for example, the evidence that older adults' poorer accuracy for more syntactically complex sentences derive from on-line processing, during initial sentence interpretation, is highly contested. While several studies have found age differences in on-line syntactic processing (Kemper et al., 2004, 2007; Stine-Morrow et al., 2000) and WM (Fedorenko et al., 2006; King & Just, 1991), others have not. For example, across several studies, Caplan and colleagues have not found an association between age and online



processing of sentences that vary in syntactic complexity in the auditory moving window paradigm (Caplan et al., 2007 for a review). They have suggested that these findings are consistent with the SLIR model, in which normative declines in verbal working memory are independent of the memory system responsible for syntactic processing and other interpretive processes.

However, it is worth noting that two recent experiments from Caplan and colleagues (2011), adopting a self-paced reading paradigm, have found evidence that there are age- and WM- related deficits in online processing of syntactically complex sentences, calling into question to generalizability of findings from the auditory moving window studies. In a lifespan sample ranging between 19 and 90 years of age, both age and working memory were found to be correlated with comprehension for sentential complement (*The dealer indicated that the jewelry that was identified by the victim implicated one of this friends*) and doubly-embedded long-distance dependency sentences (*The dealer who the jewelry that was identified by the victim implicated was arrested by the police*). Additionally, age and WM were associated with self-paced reading times at the most demanding parts of the sentential complement and doubly embedded sentences.

Other evidence consistent with a domain-general view of WM and syntactic processing comes from a series of eye-tracking experiments by Kemper and colleagues examining age and WM differences in processing object relative clauses (Kemper & Liu, 2007) and garden-path sentences (Kemper et al., 2004). In both studies, older adults showed evidence of increased processing difficulty (e.g., inflated regression path durations, inflated total reading times, and a higher probability of regressing back) at the most demanding points in the complex syntactic constructions (object relative clauses and reduced relative garden path sentences), suggesting

increased processing difficulty among older adults. Effects of WM were sizable in explaining these effects, with standardized effects of WM ranging between .54 and .89 across studies. Importantly, these findings suggest that eye-movement measures provide an important source of information for examining age and WM differences in sentence processing. Thus, while the evidence for WM effects on on-line sentence processing is less consistent in the literature, this could be attributed in part to differences in methods of measuring moment-to-moment processing. It may be the case that by adopting more sensitive measures of language processing, such as eye tracking, individual variation in working memory may be more consistently found.

### **Plasticity of Working Memory: Evidence from Cognitive Training**

Despite WM declining with age (Bopp & Verhaeghen, 2005), recent training studies suggest that there exists the possibility of capacity for change in WM (Borella et al., 2010; Morrison & Chein, 2011; Olesen, Westerberg, & Klinberg, 2004; Jaeggi, Studer-Luethi, Buschkuhl, Jonides, & Perrig, 2010). Indeed, a quite controversial literature is emerging testing the effects of computerized WM training on cognitive outcomes in young adults, including effects on intelligence and attention deficit symptoms (Jaeggi et al., 2010; Shipstead et al., 2012; Melby-Verlag & Hulme, 2013). However, cognitive training has a rich history in aging research, dating back over thirty years (Bachman, 1989; Baltes & Willis, 1982; Ball et al., 2002; Rebok et al., 2008; Willis, Blieszner, & Baltes, 1981; Willis & Nesselroade, 1990). Older adults have shown targeted improvements in trained abilities, including episodic memory, inductive reasoning, task switching, psychomotor speed, and working memory capacity (see Stine-Morrow & Basak, 2011 for a review). Importantly, showing that extended training can have effects on targeted cognitive abilities is certainly not trivial among older adults, considering that age-related

declines in cognitive and brain plasticity are robust (see Mahncke, Bronstone, & Merzenich, 2006 for a review). However, the extant cognitive training findings among older adults indicate that, while the capacity to improve performance through repeated practice may become more limited in older adulthood, there still exists a capacity for long-term change in cognition through targeted practice (Stine-Morrow & Basak, 2011).

The benchmark example of such training-related improvements in older adulthood is the ACTIVE (Advanced Cognitive Training in Independent and Vital Elderly; Ball et al., 2002) trials, which was the largest randomized clinical trial of cognitive training among older adults in the United States, with a total sample size of 2,802. The results of their trial were clear. Training in psychomotor speed, episodic memory, and inductive reasoning resulted in large but targeted improvements in the trained abilities. However, there was no evidence of immediate transfer of training to other abilities, either at the mean level, or through examination of correlated changes in abilities (McArdle & Prindle, 2008). A recent 10-year longitudinal follow-up of the ACTIVE cohort (Rebok et al., 2014) showed that trained participants reported less difficulty with activities of daily living up and that groups trained in inductive reasoning and psychomotor speed showed maintained targeted training effects up to ten years after training. However, there was no evidence for transfer in training gains to non-trained cognitive abilities.

*Working Memory Training Interventions.* As opposed to interventions focusing on speed, episodic memory, and reasoning training, which show little evidence for transfer of training (Ball et al., 2002; Rebok, 2008), WM training has shown more promise for transfer. There is evidence of WM training leading to both near transfer (i.e., transfer to tasks that are proximal to the trained ability) and far transfer (i.e., transfer to tasks that are distal from the trained ability) to fluid abilities, cognitive control, episodic memory, and reasoning in various populations (Jaeggi

et al., 2008, 2011; Chein & Morrison, 2010; Klingberg et al., 2005), including older adults (Buschkuhl et al., 2008; Brehmer, Riekmann, Bellander, Westerberg, Fischer, & Backman, 2011; Borella et al., 2010; Li et al., 2008; Richmond et al., 2011). Although findings in the WM training literature have not all been positive, and have suffered from some methodological shortcomings (for critiques, see Shipstead et al., 2012; Melby-Verlag & Hulme, 2013), WM training has demonstrated greater success than previous interventions in showing broad influences of training.

Current issues clouding the literature, however, include the lack of adequate control groups (see Shipstead et al., 2012 for a discussion), very small sample sizes (e.g.,  $N = 13$ , Bushkuhl et al., 2008;  $N = 11$ , Dahlin et al., 2008), and heterogeneity in the tasks used to train working memory (Shipstead et al., 2012). For example, a recent meta-analysis by Melby-Verlag and Hulme (2013) collapsed across several different types of working memory training, including training programs that focused on multiple cognitive tasks simultaneously (Schmiedek et al., 2010; Mahnacke et al., 2006; Zinke et al., 2013), training that involved simple short-term memory capacity (Klingberg et al., 2005), and training on updating tasks, such as the  $n$ -back task (Jaeggi et al., 2008). Unsurprisingly, across these various training types, there was significant variability in effect sizes for change (see Morrison & Chein, 2011; Shipstead et al., 2012 for similar discussions). Indeed, researchers are strongly arguing for improved methodological and quantitative standards in cognitive training research (Shipstead et al., 2012; Walton, Mowszowski, Lewis, & Naismith, 2014).

*Complex Working Memory as a Target of Cognitive Training.* It is surprising that training on complex WM tasks has received less attention as a target for training, compared to training on tasks focusing on STM storage (Klingberg et al., 2005; Klingberg, 2010; McNab et al., 2009;

Olesen et al., 2004), *n*-back performance (Jaeggi et al., 2010, 2011; Li et al., 2008), or other tasks of STM updating (Dahlin et al., 2008). Indeed, only a small number of studies exist that explicitly train complex working memory capacity, defined by the simultaneous demands for memory storage and concurrent processing of new stimuli (Borella et al., 2010; Chein & Morrison, 2010; Richmond et al., 2011). While a larger literature exists examining the influence of other WM training paradigms, such as *n*-back training, there are several reasons to focus on training complex WM for examining effects of transfer to complex cognitive abilities in general, and language comprehension outcomes in particular.

First, the majority of the literature on individual differences and age deficits in WM and language comprehension has been conducted using complex verbal WM tasks (such as reading span and operation span). These studies have provided valuable data on the correlation between verbal WM, language processing, and language comprehension, as reviewed above (Caplan & Waters, 1999; Just & Carpenter, 1992; Just & Varma, 2007). Conversely, performance on the *n*-back task shares little or no correlation with language comprehension (Kwong See & Ryan, 1995; Roberts & Gibson, 2002). This is also true for performance on basic STM tasks, which relate poorly to language comprehension compared to tasks like the reading span (Daneman & Merikle, 1996; Friedman & Miyake, 2004). These findings suggest that the dual-task load of holding items in short-term memory while simultaneously processing new information is responsible for the predictive validity of complex span tasks (Engle et al., 2002; Lustig et al., 2001; Was, Rawson, Bailey, & Dunlosky, 2011).

Second, the measurement and psychometric properties of complex WM span performance are better understood compared to tasks such as *n*-back (Waters & Caplan, 2003; Stine-Morrow et al., 2001; Conway et al., 2005). Complex WM tasks are more reliable than tasks

such as *n*-back performance (Jaeggi et al., 2010). Moreover, complex span tasks are only weakly correlated with *n*-back performance (Kane et al., 2007), suggesting that the two tasks are tapping different underlying abilities, with the former more closely linked to language comprehension (Daneman & Merikle, 1996; Roberts & Gibson, 2002) unless task demands met by the language task overlap substantially with task demands inherent in *n*-back task (see Novick et al., 2012).

Lastly, the few recent studies to focus exclusively on complex WM training (Borella et al., 2010; Chein & Morrison, 2011; Richmond et al., 2011) have shown promising evidence for large effect sizes for training gains, as well as some early evidence for transfer of gains among older adults. Borella and colleagues (2010) trained older adults in complex verbal WM and found improvements on the order of two standard deviations as a function of training, as well as improvements in fluid ability, speed, and inhibitory control on the order of one standard deviation. Direct training effects and evidence of transfer have recently been replicated by Borella and colleagues in groups of older-old adults and in adults diagnosed with amnesic MCI (Borella et al., 2013; Carretti et al., 2013). Similarly, Chein and Morrison (2011), showed substantial improvements in complex verbal and visuospatial WM performance in younger adults, as well as transfer to measures of inhibitory control and reading comprehension (see below for more detail). Richmond et al. (2011) extended the training of Chein and Morrison (2010) to older adults, showing evidence of improvements in verbal WM and evidence of far transfer to a measure of episodic memory performance.

*Cognitive Training and Language Comprehension.* A clear causal link between WM and language comprehension would come from studies examining the effects of WM training on transfer to language performance outcomes. Surprisingly, there are very few studies that have examined language comprehension as an outcome of cognitive training interventions. In the

following, I briefly review the few existing studies examining language comprehension as targets of cognitive training.

As mentioned above, Chein and Morrison (2010) trained a group of college-aged adults in complex WM and included the Nelson-Denny reading comprehension test as part of their measurement battery. Despite limitations surrounding the broad assessment nature and sensitivity of this task as a measure of language comprehension (e.g., Coleman, Lindstrom, Nelson, Lindstrom, & Noel, 2010), individuals in the WM training condition did show substantial changes in Nelson-Denny performance, compared to an active control group (Cohen's  $d = .58$ ). These findings suggest that examining changes in language comprehension as a function of cognitive training is a fertile ground for future investigation with more diverse samples and with more precise, reliable, valid, and sensitive measures of language comprehension.

Shiran and Breznitz (2011) tested the effects of cognitive training on short-term memory and language performance in dyslexic and skilled readers. Both skilled readers ( $N = 35$ ) and dyslexics ( $N = 26$ ) were trained on CogniFit (Cognifit, 2003), a program that involves practicing serial short-term memory (forward and backward recall), as well as several other non-memory specific tasks (e.g., identifying whether the first or second of two sounds is longer or louder). Participants were tested on the Sternberg task both at baseline and at post-test. The Sternberg task is a delayed recognition task in which subjects view a string of letters and, after a delay, are probed with an item and must indicate whether the item was included in the previously viewed string of items or not. Scalp electrodes recorded concurrent brain activity during the Sternberg task in order to examine the effects of cognitive training on the neural mechanisms underlying

short-term recognition memory (cf. Patterson, Pratt, & Star, 1991). Participants also completed a simple self-paced reading and comprehension task at baseline and at post-test.

Following post-testing, both dyslexic and skilled readers showed improved performance on the comprehension items from the self-paced reading task, as well as faster reading times overall (relative to a control group not trained on CogniFit). Moreover, both dyslexic and trained individuals showed reduced latencies and amplitudes in the P300 component of the event-related potential (ERP) during Sternberg performance (relative to the control group), indicating that CogniFit training resulted in changes in efficiency of the neural system supporting STM performance at post-test. Dyslexic and skilled readers in an active control group showed no such effects on the P300.

Lastly, findings from Novick, Hussey, Teubner-Rhodes, Harbison, and Bunting (2013) are perhaps most relevant to the current study. In this study, 21 healthy younger adults were trained over 20 hours on an *n*-back task with lures (as part of a larger training battery including several cognitive tasks). Eye tracking was used to monitor on-line language processing during syntactic ambiguity resolution at pre-test and post-test. Participants read a series of reflexive absolute transitive garden path sentences (Christianson et al., 2001, 2006; Slattery et al., 2013) while their eye movements were recorded. Relative to a no-contact control group, participants in the training condition who showed significant improvements in *n*-back lure performance showed transfer to on-line processing as a function of ambiguity, such that those who responded to the training showed reduced disambiguation effects in late-pass measures (e.g., regression path duration) at post-test. The authors argued that these findings suggested that training on tasks of cognitive control (such as *n*-back with lures) resulted in transfer to ambiguity resolution, which



allowed individuals to more easily override early parsing decisions and recover from an initial misanalysis.

### **The Current Study**

Although these early findings described above are promising, several open questions still remain. First, can short-term home-based training in complex verbal WM span lead to improvements in trained tasks among older adults, who as a group, show reduced WM capacity and reduced cognitive plasticity? Second, do training gains in verbal WM lead to generalized improvements in verbal WM capacity; that is does training transfer to tasks that tap verbal WM but are not directly practiced during training? Third, does WM training lead to improvements in language comprehension in older adults? And lastly, are there dissociations between training related improvements in off-line vs. on-line measures of language comprehension, consistent with language-specific WM models (Caplan & Waters, 1999), or can complex span training transfer to on-line processing as would be predicted by shared-resource models (Gibson, 1998; Just & Carpenter, 1992). The current study aimed to address these questions by examining the influence of complex verbal WM training on multiple language processing and performance outcomes in healthy older adults. In the following sections, I introduce the iTrain program, a novel computerized method for training complex verbal WM performance at home, and discuss the results from a randomized controlled training experiment testing the efficacy of iTrain for improving verbal WM and language comprehension in older adults.

## **Chapter II.**

### **The iTrain Program for Training Verbal Working Memory**

As discussed above, the majority of the WM training literature consists of training studies focusing on improving performance by repeated practice on tasks such as the n-back task, or on tasks focusing on short-term memory processes. In this chapter, I introduce the *iTrain* protocol that was developed for use in the current study, which instead focuses on improving performance on complex dual-task verbal WM tasks. As described above, complex WM span was selected as the target of training, given the substantial literature indicating that individual differences in this ability is highly related to higher-order cognitive abilities, including language comprehension, and that such relationships with higher-order cognition do not exist with tasks more commonly adopted for training studies (see Shipstead et al., 2012 for a similar argument). Appendix A includes screenshots of the tasks as they appear on the iPad, along with hyperlinks to video demonstrations of trials in the iTrain program.

#### **Working Memory Training Protocol and Design**

The verbal working memory training program used in the current study was adapted and extended from a number of studies of lab-based complex verbal WM training (Chein & Morrison, 2010; Borella et al., 2010). The WM training protocol is called iTrain and was written in objective-C and implemented for use on iPad tablet computers via the Xcode environment (cf. Dufau et al., 2011). The program was designed for home-based training in order to allow participants to complete each training session without having to make daily visits to the lab. Prior studies suggest that home-based cognitive training shows gains on the same order of magnitude

as lab-based training, and also results in high retention rates in healthy older adults (Margrett & Willis, 2006; Payne et al., 2012; Wadley et al., 2006) in part because participants do not have to travel to the lab daily throughout the course of the intervention.

iTrain appears as an *app* on the home screen of the device. Three tasks are included in iTrain to practice verbal working memory, each resulting in a dual-task load of processing and storage. Each of the three tasks were presented in a random order in each session. The expectation was that by changing the response cues and surface level demands of the tasks while emphasizing the memory and dual-task demands, the training would be less likely to result in development of task-specific strategies, and thus, should be more likely to result in substantive changes in WM (Stine-Morrow & Basak, 2011; Lustig, Shah, Seidler, & Reuter-Lorenz, 2009). If training is indeed driven by improvements in core WM mechanisms, and not development of task specific strategies, then it would be expected that training related improvements over the 15 sessions would be strongly correlated across the three tasks. However, if training is driven by task-specific strategies, then we should expect to see divergence across tasks in training related improvements.

The training was designed to be individually adaptive (cf. Lustig et al., 2009). The set size (number of items to recall within a recall trial) fluctuates according to an individual's performance, so that each participant is always engaging in the task at a level that is matched to his or her current ability. Task difficulty follows a step function, such that when a perfect recall score is achieved on set size  $N$ , the set size for  $N+1$  is increased by one. If perfect recall is not achieved at a given set size, then the number of items in the following set is reduced by one. At the end of each set, feedback is presented to participants on both the accuracy of the judgment task (proportion correctly judged) and the proportion of items correctly recalled. Direct measures

of task performance (number of items correctly recalled) are derived from the iPad data during the training sessions in order to examine individual differences in training gains over the 15 sessions. The three tasks are described in detail below:

(1) In the *semantic category span* task, single words are presented one at a time, along with a particular category. Words are presented in sets of two or more items for recall. Participants are instructed to make a semantic category judgment for each word, judging if the word that is presented is a member of the category. A single category is presented for each set, and changes only between memory sets. For example, if the category were FURNITURE the word “sofa” would elicit a yes response, while the word “thunder” would elicit a no response. At the end of each set, participants are asked to recall each of the words that they categorized in the order in which they were presented.

Categories and exemplars were adapted from the Van Overschelde, Rawson, and Dunlosky (2004) category norms, which were updated from the Battig and Montague (1963) category norms. The final stimulus set included a total of 69 unique categories and over 1500 unique words. Words vary in length between 4 and 9 characters. Items are drawn randomly such that, within a set, each word has an equal probability of belonging to the presented category or not. Across the 15 sessions, items are rotated through such that all categories have to be selected at least once before a particular category could be repeated again. Participant’s reaction times and accuracy are recorded for each category judgment along with if each word was successfully recalled.

(2) In the *lexical decision span* task, letter strings constituting words (e.g., seek) or non-words (e.g., ceek) are presented on the screen one at a time. Words are presented in sets of two or more per set. Participants decide if each string of letters forms a word or not, by pressing the

button “word” if the string of letters forms a word and the button “not a word” if the string of letters does not form a word. Following the decision, a single letter is presented for the participant to hold in memory. At the end of each set, a probe appears and the participant must recall the string of letters in order. A total of 9,000 common and proper nouns and 10,000 phonologically regular and pronounceable non-words were generated from the English Lexicon Project database (Balota et al., 2007). Word/non-word strings ranged in length between 4 and 9 characters (for word stimuli: log word frequency range: 5-13.67). Items are drawn randomly such that, within a set, any given string had an equal probability of being a word or non-word. Across the 15 sessions, items are sampled such that all words or non-words have to be selected at least once before a repetition can occur. For the memory task, letters were chosen at random. Reaction times and accuracy are recorded for each lexical decision judgment along with if each letter was successfully recalled.

(3) In the *sentence reading span* task, participants read a series of sentences, presented in sets of two or more, and are asked to do two things. After they read the sentence, they must make an acceptability judgment on the sentence (cf. Waters & Caplan, 1996; Caplan & Waters, 2003). The acceptability judgment is made based on the message-level semantics of the sentence. An example of an acceptable sentence is:

*Development of the screenplay was done by a team of three authors.*

while an example of an unacceptable sentence is:

*As the ship gets better, your child needs to develop this oven.*

After reading each of the sentences within the set, the second task is to recall the last word of each sentence within each set in the order that they were presented. Sentences are presented in sets of two or more items for recall.

Acceptable sentences were drawn from two sources. The first set of sentences was adapted from the Nelson and Narens (1980) general information question norms. Declarative sentences were created by reforming questions and answers from the Nelson and Narens norms. For example, one sentence from these norms was:

*Frank Lloyd Wright was known professionally as an architect.*

This source yielded a total of 244 acceptable sentences. The second source of acceptable sentences was derived from the Manually Annotated Sub-Corpus (MASC) of the Open American National Corpus (Ide et al., 2013). A total of 301 sentences were collected from MASC, ranging widely in topic, length, and syntactic structure. An example sentence is:

*Prehistoric stone carvings show the continuity of totemic styles.*

A total of 346 unacceptable sentences were adapted from the “syntactic prose” conditions in earlier studies by Federmeier and colleagues (Stites, Federmeier, & Stine-Morrow, 2013; Lee & Federmeier, 2011). Unacceptable sentences have syntactically well-formed sentence frames, but contain no coherent message-level semantics. Sentences were created by replacing the content words in well-formed and semantically meaningful sentences with a randomly selected set of words from the same grammatical category as other well-formed sentences. Unacceptable sentences vary in length and syntactic structure. Thus, participants cannot make their decisions about the acceptability of the sentence without reading through the entire sentence. Sentences are presented randomly such that, within a set, each sentence trial has an equal probability of being acceptable or not. Across the 15 sessions, sentences are rotated through such that all acceptable or unacceptable sentences had to be selected at least once before a repetition could occur. Reading times and accuracy are recorded for each lexical decision

judgment along with if each letter was successfully recalled. All sentences ranged between 60 and 90 characters, and all sentence final words were between 4-9 characters.

## **Chapter III.**

### **Randomized Controlled Experiment Methods and Design**

The following chapter summarizes the methods used in the randomized controlled training experiment testing the efficacy of iTrain for improving verbal WM and language comprehension among a group of older adults. The section includes information on participant demographics in the training and control groups, the experimental design and overview of the procedure, handling of participant recruitment and participant retention throughout the training and testing, the active control group, and the neuropsychological battery administered at pre-test and post-test, along with the eye-tracking experiment administered at pre-test and post-test.

#### **Participants**

Volunteers were recruited from the Champaign-Urbana community through flyer advertisements, information booths at the farmer's market and related events, e-mail lists, and through phone recruitment from a database of older adult volunteers in the community who have previously participated in studies in the Adult Learning Lab. Participants were screened to exclude those who (a) had history of dementia or other neurological issues, (b) had health issues that would limit their ability to participate, (c) were non-native English speakers, (d) had functionally poor visual acuity, and (e) had recently (within the last three years) participated in an intervention study focused on cognitive training or physical exercise.

A CONSORT (CONsolidated Standards of Reporting Trials) diagram is presented in Figure 1 (Altman et al., 2001), which provides a graphical representation of the recruitment process and the flow of participants through the study, from eligibility to post-testing.



A total of  $N = 240$  individuals were contacted either by phone or e-mail from our recruitment database, or after expressing interest in the study. Of those,  $N = 134$  did not follow-up or reply to our invitation to participate in the study. A total of  $N = 106$  individuals were then assessed for eligibility. Of those,  $N = 64$  individuals were excluded for not meeting inclusion criteria, refusing to participate after learning more about the study, or for other various reasons. Thus, a total of  $N = 42$  individuals were pre-tested. One participant did not meet inclusion criteria at baseline, based on inability to complete pre-test cognitive tests. Thus  $N = 41$  individuals were randomly assigned to either a treatment ( $n = 22$ ) or control ( $n = 19$ ) group. Of those, 21 in the training group, and 17 in the control group, completed at least 80% of the training sessions. Table 1 presents demographics at baseline in the control and treatment groups. As can be seen, differences between the two groups in age, education, sex, MoCA score (a screening measure for mild cognitive impairment, see below), and vocabulary score were negligible.

### **Experimental Design and Overview of Procedure**

A pretest-posttest randomized controlled experimental design with an active control group was used to examine the effects of WM training. The duration of training was three weeks long. Participants were asked to complete a total of five 30-minute sessions in each week, for a total of 15 sessions over the three-week period (or 7.5 hours of total training). The interval between pretest and posttest sessions was held constant such that post testing occurred no more than 4 weeks from the pre-test date.

Figure 2 presents a graphical representation of the study procedure. At the onset of the study, all participants completed the neuropsychological battery and eye-tracking sessions (described in detail below) in a single laboratory session. Following the pre-testing battery, participants were given an iPad 2 tablet computer containing either the complex working memory training software (treatment group) or the active control training software, based on random assignment. Testers instructed participants on procedures for completing each of the tasks in the training program as described in Chapter II, and participants were given the opportunity to practice the tasks in the lab until they understood each task completely.

The testing was single blind, as testers were aware of the random assignment condition. This was necessary due to limited resources and pragmatic issues. Namely, testers had to prepare the iPads and, at the end of the pre-testing session, instruct participants on how to use the iPad and the training program software. However, testing sessions were designed to minimize the amount of contact with the participant, and testers were instructed to provide no information regarding the specifics of either training program or the study hypotheses. All data collection was conducted with the participant in a silent room without the presence of the tester, with the exception of the eye-tracking experiment, in which the tester sat silently on a separate machine that monitors and controls the eye tracking recording and presentation of stimuli. Testers were instructed to provide only minimum information about the training program before pre-testing completed.

### **Active Component Control Protocol**

A component-control design (Boot et al., 2013; Mohr et al., 2008; Brown, May, Nyman, & Palmer, 2012) was adopted in designing the active control group. In a component control

design, a multi-component intervention serves as the focal treatment, which is the iTrain verbal working memory training in the current study. An active control group is created by administering the same treatment absent a single component of the focal training. By matching the two groups on the surface level aspects of the tasks, along with presenting the same stimuli, this process results in reduced likelihood of placebo effects or differential expectancies for change (Boot et al., 2013).

Participants in the active control group complete the same three tasks as in the treatment group, without the recall component. Therefore, in the category task, participants train in making category judgments; in the lexical decision task, participants train in making lexical decision judgments; and in the sentence-reading task, participants train in making acceptability judgments. Importantly, both the treatment and control groups are matched in their exposure to stimuli (as the same items are used in both groups), the absolute magnitude of time allocated to training (15 30-minute sessions over three weeks), and the amount and type of linguistic exposure. Thus, findings comparing the treatment and active control groups cannot be driven by individual differences in exposure to linguistic stimuli (cf. MacDonald & Christiansen, 2002; Wells et al., 2008).

Because removing the memory load from the WM training makes the task less demanding and potentially less engaging, an individually adaptive speed threshold was added in order to (a) maintain continued interest in the task (b) de-confound memory load from task adaptivity (cf. Klinberg et al., 2005), and (c) reduce the potential for differences in expectancy for training benefits in the two groups (Boot et al., 2013). In the control training, participants are presented with stimuli in blocks of 15 items, and are told to make their judgments (lexical

decision, category, sentence acceptability) as quickly as possible. As participants improve in accuracy in the judgment decisions, presentation rates are increased following the function:

$$\text{presentationRate}_{\text{BlockN}} = (.95) \text{presentationRate}_{\text{BlockN-1}}$$

When accuracy falls below a criterion (80%), the presentation rate is increased, so that task adaptivity follows a similar step function as in the WM training. Participants are encouraged to practice speeded decisions in each of the linguistic tasks while maintaining high accuracy. A “speed level” score is presented, derived from change in presentation rate from the initial training block, after each block, so that participants can monitor their progress from the first block of the first session to the end of the training, as in the WM training protocol.

### **Neuropsychological Test Battery**

A cognitive battery consisting of computerized tasks and paper-and-pencil measures was administered. Following the measurement battery, an eye-tracking session was also administered. Both the measurement battery and eye tracking session were administered before training (pre-test) and after training (post-test). Total administration time at pre-testing and post-testing was between 3 and 3.5 hours. The measurement battery was chosen to target both complex working memory performance (targets of *near* transfer) as well as measures of off-line language performance and on-line language processing (targets of *far* transfer).

*Complex Verbal Working Memory.* Four complex working memory tasks were administered using the Psychophysics Toolbox in MATLAB (Brainard, 1997), adapted from the

CogToolbox (Fraundorf et al., 2014). Alternate forms of all four tasks were administered at pre-test and post-test.

First, the sentence reading span task (Daneman & Carpenter, 1980; Stine & Hindman, 1994) and sentence listening span task (Daneman & Carpenter, 1980) were administered. In the reading span task, participants read a set of sentences silently, and were asked to immediately make true/false judgments after each sentence. After reading all of the sentences in a group, participants recalled all of the target words (the last words of each sentence in that group) in order. The number of sentences per set increases with progress through the task (until eight sentences per set or when the participant can no longer recall each of the target words in a set successfully). If the participant fails at a particular set size, a second trial is administered. If the participant fails at the second trial within that set, the test terminates (Stine-Morrow et al., 2001; Waters & Caplan, 2003). A participant's final score is the number of target words recalled within the highest set with no errors, plus a fraction reflecting the proportion of correctly recalled words on the set with an error. The listening span uses the same administration and scoring, except that the sentences are presented in the auditory modality.

The operation span task (Turner & Engle, 1998; Conway et al., 2005) was administered. The operation span task follows a similar format to the reading and listening span tasks, in that participants hold items in memory while simultaneously performing a secondary task. Participants are presented with a set of simple three-term mathematic problems to solve (e.g., is  $[8/2] - 1 =$ ). They are then presented with a probe answer following the equation and are required to respond if the answer is true or false, given the prior problem (e.g., "3" would be True). Between problems, a letter is presented that the participant must hold in memory. At the end of a set of equations, participants are asked to recall the letters they saw in order. After a

brief practice session, 15 total sets are presented, with three sets at sizes (3-7). The total score is the total proportion of correct items in the correct position across all sets (Unsworth, Heitz, Schrock, & Engle, 2005).

The Minus-2 span task was also administered (Waters & Caplan, 2003). In this task, participants are presented with a series of digits one at a time, varying in length from trial to trial (between 3 and 8 digits per trial). Participants are asked to repeat the series of digits in the same order as presented after subtracting 2 from each digit. For example, if participants were presented with the string [8, 4, 3, 9], they would have to reply [6, 2, 1, 7]. The total score is the total proportion of correct items in the correct positions across all trials (Waters & Caplan, 2003).

*Sentence Memory.* Sentence memory was measured with an immediate sentence recall task (Stine-Morrow et al., 2006; Zelinksi & Lewis, 2003). A series of 8 18-word sentences were presented on the screen with presentation time self-paced. Following each sentence, a cue is presented on the screen for the participant to recall as much of the sentence from memory. Production was recorded and scored for sentence recall. Recall was scored as the proportion of individual words correctly recalled, as well as the proportion of propositions correctly recalled (Ferguson, Spencer, Craig, & Colyvas, 2013; Kintsch & Keenan, 1973; Kintsh & van Dijk, 1978; Snowdon et al., 1996). Propositional coding was conducted by a trained manual coder who was blind to condition. Sentence stimuli were from Stine-Morrow et al. (2001) and Stine-Morrow et al. (2008). Alternate sentences were presented at pre-test and post-test.

*Discourse memory.* Immediate discourse memory was measured with the Rivermead Behavioral Memory Task Paragraph recall subtest (Wilson, Cockburn, & Baddeley, 1985, 2003). A whole paragraph is presented on the screen, with presentation duration controlled by the participant. Following presentation of the paragraph, participants were cued to recall as much of

the paragraph out loud as they could remember. Production was coded and scored for number of words and propositions correctly recalled, using the same method as in the sentence recall task. Two paragraph-recall trials were presented separately at pre-test and post-test and alternate paragraphs were used at pre- and post-test. This administration has proven to result in high test-retest reliability (Sisco et al., 2012; Payne, Gross, et al., 2013).

*Reading Comprehension.* The Nelson-Denny Standardized Reading Comprehension subtest was administered at pre-test and post-test to assess general reading comprehension ability. The Nelson-Denny consists of eight prose passages and 36 multiple-choice questions. Participants were given 20 minutes to read the passages and answer the questions. Alternate forms were administered at pre-test and post-test.

*Verbal Fluency.* Verbal fluency was assessed with the FAS phonemic fluency task (Benton & Hamsher, 1978). In this task, participants are given a letter (at pre-test “F”, “A”, and “S”) and asked to produce as many words as they can think of that begin with that letter for 60 seconds. A total score is calculated as the sum of unique words correctly produced across the three trials. This task has been shown to be highly predictive of general cognitive status (Kemper & McDowd, 2008) as well as language comprehension (Federmeier, 2007) in older adults. An alternate form, the BDT, was used at post-test (Straus et al., 2006).

*Verbal Ability.* Vocabulary score was measured with the ETS extended range vocabulary task. This measure has been shown to influence sentence processing in older adults (Stine-Morrow et al., 2008; Payne & Stine-Morrow, 2014). Because this measure is based on knowledge and not expected to vary as a function of extended training, it was only assessed at baseline.

*MoCA*. The Montreal Cognitive Assessment (MoCA) was administered as a measure of general cognitive status at baseline. Cutoff scores have been published to screen for mild cognitive impairment (MCI) (Nasreddine et al., 2005), but recent studies on the MoCA have revealed that the specificity of cut scores vary substantially across samples and populations, such that they tend to be anticonservative (i.e., healthy adults are more likely to be categorized as “at risk” for MCI) (Rossetti et al., 2011; Waldron-Perrine & Axelrode, 2012; Larner 2011; Lee et al., 2008; Luis et al., 2009). Nevertheless, performance on such general cognitive status measures has been shown to be correlated with training-related cognitive plasticity (Stine-Morrow et al., submitted; Zinke et al., 2013), and language comprehension (Payne & Stine-Morrow, 2014). Therefore, a very conservative cut score (20) was used for inclusion in the study, and this measure was used instead to characterize our sample of older adults in terms of general cognitive status.

*Expectation Survey*. A survey was administered at post-test in order to assess individuals’ perceptions of improvement in performance on specific tasks as a function of training. Boot, Simons, Stothart, and Stutts (2013) have critiqued much of the cognitive training literature for not adequately assessing or controlling for differential expectation for improvements between treatment and control groups in psychological interventions. Based on the survey presented in Boot et al. (2013), a 14-item survey was created to assess individuals expectations that (1) they improved generally as a function of training (e.g., “I believe that iTrain helped improve my cognition”), and (2) they improved on specific tasks (e.g., “You completed a task called Listening Memory. In this task, you heard a series of sentences and you were asked to judge if the sentences were true or not. You were also asked to remember the last word of each of the sentences in that section in order. Do you believe that iTrain helped lead to better performance



on this task?”). The expectation survey and results are presented in full in Appendix B. Briefly, results showed no difference between the treatment and control groups in expectations for general improvements in cognition (see Figure C1) (*mean difference* = .13; 95% CI [-.15, .41]). There was a trend for the WM tasks to show greater self-reports of improvement in the treatment relative to the control, but this effect only reached marginal significance in one task (*mean difference* = .65; 95% CI [.004, 1.29]). The transfer tasks showed no evidence of differential expectation for improvement (see Figure B2).

### **Eye-Tracking Experiments**

Participants also completed an eye-tracking session to monitor on-line language processing at pre-test and post-test. Participants read sentences that differed in complexity (low demand vs. high demand) as their eye movements were monitored via a head-mounted eye tracker. Dependent variables from the eye-tracking data included fixation-based (e.g., gaze duration, regression path duration) and saccade-based (e.g., probability of regression) measures. Following each sentence, participants were probed for comprehension, and accuracy and reaction time to probe questions was recorded.

Different sentence sets that were approximately equivalent were presented at pre-test and post-test (i.e., alternate forms). Test sentences were presented on a 19 in. ViewSonic monitor (1024 x 768), while a head-mounted SR Research Eye-Link II (500 Hz) eye-tracking system monitored eye movements. Three different sets of sentences were presented in a random interspersed order. Each sentence set included a manipulation of syntactic complexity, described in more detail below. Sentences were counterbalanced across conditions at each testing occasion, resulting in each sentence having an equal opportunity of occurring in either a high demand or

low demand condition. At pretest and posttest, participants read 20 items from each sentence set (10 low complexity, 10 high complexity), resulting in a total of 60 sentence/ question pairs at each measurement occasion.

*Sentence Set 1. Subject/Object-Relative Clause Processing.* In the first sentence set, syntactic complexity was manipulated by varying whether a relative clause was subject-extracted (SR) or object-extracted (OR).

(4) SR (low demand): The farmer *that knew the barber* asked for a loan.

(5) OR (high demand): The farmer *that the barber knew* asked for a loan.

As reviewed above, theories of working memory and parsing attribute the difficulty of OR sentences to working memory costs, and working memory has been found to contribute to adult age-differences in processing OR clauses. Sentences were adapted from Staub (2010). The critical region is the relative clause region (between the relative pronoun *that* and the matrix verb (*knew*) or object noun phrase (*barber*)), which appears in italics in (4) and (5). All sentences included common name noun phrases for subject and object nouns (Fedorenko et al., 2006), had minimal pragmatic bias in noun-verb relationships (King & Just, 1991; King & Kutas, 1995), and were always animate (Gennari et al., 2012). These choices were made to generate the largest object-relative clause effect.

*Sentence Set 2. Syntactic Ambiguity Resolution.* The second set of sentences manipulated syntactic ambiguity to yield either simple unambiguous sentences (6) or late-closure garden-path ambiguous sentences (7).

(6) Unambiguous (low demand):

While the man hunted, the deer that was brown and graceful *ran into the woods*.

(7) Garden path (high demand):

While the man hunted the deer that was brown and graceful *ran into the woods*.

As reviewed above, such garden path effects are robust across individuals, but several studies have found that GP effects are larger among individuals with reduced working memory capacity, including older adults (Christianson et al., 2006; Kemper et al., 2004; but see Waters & Caplan, 1996). Sentences were adapted from Slattery, Sturt, Christianson, Yoshida, and Ferreira (2013) and use a simple comma manipulation to disambiguate the garden-path sentences. The critical disambiguation region appears in italics in (6) and (7).

*Sentence Set 3. Long-Distance Dependencies.* Finally, complexity was manipulated by introducing a long-distance dependency, increasing the number of discourse entities intervening between an embedded verb and its subject in relative clause sentences. Thus, the low-distance sentence was a single embedded object relative clause, while the high-distance sentence was a doubly embedded object relative clause.

(8) *Low distance* (low difficulty):

The administrator who the nurse supervised *scolded the medic* for being late.

(9) *High distance* (high difficulty):

The administrator who the nurse who was from the clinic supervised *scolded the medic* for being late.

Sentences were adapted from Grodner and Gibson (2005) and Bartek et al., (2011), who showed substantial disruptions in processing time and eye fixation durations for the doubly embedded relative clauses. Similarly, Caplan et al. (2011) showed that age and WM differences exist in the on-line processing and comprehension of doubly embedded relative clause sentences. Critical regions are shown in italics, which consist of an embedded verb and the spillover region.

## Analyses and Predictions

A series of linear mixed effects models were used to test for the effects of the intervention on each outcome measure from the cognitive battery, using an intent-to-treat approach (Hollis & Campbell, 1999). Effect sizes and 95% confidence intervals of the critical Training Group x Time interactions were estimated for each outcome measure in the cognitive battery. Analyses in the current study were focused on effect size estimation and quantification of the precision of these effects using confidence intervals (Cumming, 2013). Null hypothesis significance tests are referenced occasionally, but these are not used as the sole piece of information about the effectiveness of the intervention. Because sample sizes are relatively small, a robust bootstrapping approach, as described by Kirby and colleagues (2013), was used to estimate the Group x Time interactions.

For the eye-tracking data, fixation-based measures were modeled with a series of linear mixed-effects models, with crossed random effects for subjects and items, nested within measurement occasion, to test for difficulty by time interactions. Saccade-based measures (e.g., probability of regressing) and offline accuracy data are binary, and thus were analyzed with logit mixed models, with the same structure for the random effects as the linear mixed effects models (Jaeger, 2008). Eye-movement analyses were conducted on the critical regions for each sentence set, using a *slopes as outcomes model* (Singer & Willet, 2003) to test whether differences in fixation times and eye-movements between the low demand (e.g., SR) and high demand (e.g., OR) sentences changed over time as a function of group (i.e., a Training Group x Time x Sentence Complexity interaction). Tests were conducted separately for each of the three sentence sets, and effect sizes and 95% confidence intervals were estimated for the Training Group x Time x Sentence Complexity interactions for each eye movement measure.

Direct assessment of training efficacy was examined by analyzing the memory performance data over the 15 sessions of the home-based training. These data are used to assess (a) the overall effectiveness of the intervention and (b) individual differences in training gains (see Chein & Morrison, 2010; Jaeggi et al., 2011 for similar examples). The approach used to quantify individual differences in training gains is described in more detail in the following chapter.

If the training was successful, then older adults should show direct improvements in performance on the three WM tasks (category span, lexical decision span, and reading span) over the 15 sessions. Moreover, if the training results in improvements in complex working memory performance, abstracted from task specific practice, then participants in the training condition should show improved performance in measures of *near* transfer for complex working memory— that is the reading span, listening span, operation span, and minus-2 span tasks (i.e., Group x Time interactions)— relative to the control group.

If working memory training impacts general language comprehension in older adulthood, we would expect to see WM training gains transfer to language performance measures, including sentence and discourse memory, verbal fluency, comprehension accuracy (particularly for more complex sentences), and the Nelson-Denny reading comprehension task.

If the mechanisms of improved comprehension performance reside in moment-to-moment processes, then we would expect to see a change in the effects of complexity on eye movement behavior from pretest to posttest with working memory training. Specifically, at pretest, there should be longer gaze durations, regression path durations, and an increased probability of regressions associated with processing object relative clauses, garden-path sentences, and long-distance dependencies (relative to the respective low-demand sentences) for

both the training and control groups, along with poorer offline comprehension for these more complex sentences (relative to the low-demand sentences). To the extent that working memory training reduces the on-line processing costs for these more demanding sentences, we would expect to see the WM training group show reductions in online processing difficulty at post-test, effects that would be absent in the control group (i.e., Training Group  $\times$  Time  $\times$  Sentence Complexity interactions).

## **Chapter IV.**

### **Responsiveness to Home-Based Working Memory Training**

Before assessing the degree to which a training program results in broad cognitive change, it is key to establish that performance on the trained tasks was indeed improved over the course of the training program. The adaptive computerized cognitive training program used in this study made it possible to monitor session-to-session changes in performance throughout the course of multi-session interventions (see also Jaeggi, Bushkuehl, Jonides, & Shah, 2011; Morrison & Chein, 2010; Richmond, Morrison, Chein, & Olson, 2011). These data are important not only in establishing the learning curve of the dose-response relationship between training and improvements in trained tasks, but also in examining individual differences in responsiveness to training gains (Novick et al., 2012; Jaeggi et al., 2011; Payne et al., 2012). In this chapter, results are presented for session-to-session improvement in the three WM tasks in the iTrain program. Following that, a novel method for quantifying individual differences in non-linear changes in training gains is presented.

#### **Training Related Plasticity in WM Performance**

Trial-level performance was obtained from each iPad for the participants in the WM training group. From these data, the average span score was computed for each of the 15 sessions for the category span, lexical decision span, and sentence reading span tasks separately for each of the 21 subjects who completed at least 80% of the training ( $N = 21$ ; see Figure 1). Figure 3 plots the session-to-session effects of WM training on performance gains for each of the three verbal WM tasks. This plot shows that initial span scores were quite low. Performance on the lexical decision task was reliably larger at baseline. Note that this is likely a conservative

measure of baseline performance as well, since these scores are based on an average over 10 minutes of training. On average, training gains followed a non-linear trajectory, with comparably larger improvements in early sessions, relative to later sessions. Indeed, the largest improvements across the three tasks occurred from session 1 to session 2.

### **Quantifying Individual Differences in Working Memory Gains**

Raw scores on the span tasks were converted to a metric of percent change from baseline assessment in order to examine individual differences in training gains in each of the three tasks independently from task-specific differences in WM score and baseline individual differences in WM score (e.g., Morrison & Chein, 2010). Thus, for each task and each participant, baseline scores were normalized to 0, and span scores for the following sessions were computed separately for each subject and each task as below:

$$\text{percentChange}_{\text{SessionN}} = \frac{\text{score}_{\text{SessionN}} - \text{score}_{\text{Baseline}}}{\text{score}_{\text{Baseline}}}$$

Figure 4 plots the session-to-session training gains in each of the three WM tasks, expressed in percent of change in WM score from baseline. Participants showed an approximately 60% peak training improvement in WM score in the category and sentence span tasks over the 15 weeks. However, trainees showed an over 100% improvement from baseline score in the lexical decision span task, indicating that participants, on average, doubled their span score from their baseline performance.

A number of methods have been previously used to quantify individual differences in training gains. For example, Jaeggi et al. (2010) calculated a difference score between the



average of WM performance in the first two sessions and the average of WM performance in the last two sessions of a training study. However, such a method assumes that the training curve is linear, with increases in performance of the same magnitude across session and peak performance at the end of a training program. To the extent to which this is not the case, such simple difference scores will result in a loss of information at best, and a distortion of the data at worst. In order to account for the fact that peak training performance may not occur at the final training session across all individuals (i.e., there may be non-linear training effects), Bissig and Lustig (2007) scored individuals based on peak training gains, taking into account the session at which peak performance was reached (i.e., a higher score was given for individuals who reached peak performance earlier in the training). However, this Bissig and Lustig method relies on rank ordering subjects in terms of training benefits. Rank ordering results in a major loss of information about *magnitude* of improvement, however. Additionally, this method does not account for individual differences in overall performance as well, which is problematic given that there is often a relationship between initial level of performance and training gains (cf. Stine-Morrow & Basak, 2012).

For the current study, I used a novel method for quantifying individual differences in training gains that (1) accounts for individually varying non-linear growth trajectories, (2) preserves information about magnitude of improvement, (3) is de-confounded from initial level of performance, and (4) results in a single simple interpretable value for each individual, facilitating interpretation of results.

First, for each participant, a natural cubic smoothing spline was fit to the training data (using data expressed in the percent of change from baseline metric). Fitting separate smoothing splines for each individual allows for different non-linear trajectories for each subject. That is,

there is no parametric model to constrain the dose-response curve to be similar across individuals. Thus, the resulting individual-level splines fit the data very well and can vary substantially in the form of their trajectories. Following this, the area under the cubic spline interpolation was estimated over the interval  $[x = 1, x = 15]$ , separately for each participant. Numerical integration was used to calculate the area under the curve (auc) using adaptive quadrature methods. Integration was conducted using the *auc* function in the MESS package in R (Venables & Ripley, 2002). Figure 5 shows an illustration of the area under the curve calculation for two separate subjects who show very different levels of training gains, as well as very different dose-response curves. As can be seen, this method preserves the non-linear nature of the training curves for each subject, as well as the magnitude of the training gains.

Figure 6 plots the area under the curve (auc) for percent improvement from baseline in each of the three tasks separately for all 21 participants. The plot is rank ordered by average training gains (average auc across three tasks) and shows substantial heterogeneity in the overall improvement from baseline. This plot illustrates that individuals who showed substantial gains in one task, on average, tended to also show substantial gains in the other tasks, and likewise that individuals who showed low training gains in one task tended to also show low training gains in other tasks. This was confirmed by examining correlations between auc measures in each of the three training tasks. Figure 7 plots the bivariate scatterplot matrix among auc training estimates for the sentence span, category span, and lexical decision span tasks. The correlation between gain in the sentence span and category span task was  $r = .85$  (95% CI: [.66, .94]). The correlation between gain in the lexical decision span and category span task was  $r = .91$  (95% CI: [.79, .96]). Lastly, the correlation between gain in the lexical decision span and the sentence span task was  $r = .90$  (95% CI: [.76, .96]). Thus, training gains tended to cluster

together tightly, suggesting that training-related improvements occurred broadly across all tasks, and were not isolated to task-specific strategy development.

Lastly, Table 2 shows correlations between baseline measures and auc training gain estimates. In part because the sample size is relatively small, only correlations greater than  $r = .42$  are considered statistically significant at the  $p < .05$  level. This table shows that adults who were older, had greater verbal skill, and performed better on the Nelson-Denny showed larger training gains. Additionally there was a trend for older adults who scored better on the MoCA and had greater fluency scores to show larger training gains. Interestingly, correlations between baseline WM scores and training gains were variable and not reliable across tasks. Effect sizes were largest for age, vocabulary, and Nelson-Denny performance.

## **Summary**

The results presented in this chapter demonstrated that older adults who were assigned to the WM training group showed reliable and robust session-to-session improvements in all three verbal WM span tasks. Indeed, performance gains ranged between 60% to over 100% improvement in span score from baseline across tasks. Establishing that the WM training group showed session-to-session dose-response effects is important in establishing the logic of transfer effects. If far transfer effects are found without transfer to WM, this presents real interpretational challenges for such studies, which may be more easily explained by Hawthorne effects, or other expectancy or motivational factors (Shipstead et al., 2012).

In addition to showing large mean-level improvements in practiced tasks, the current study also established that there were robust individual differences in training gains across the three tasks. A novel method was adopted for assessing individual differences in training gains,

based on the observation that there was considerable heterogeneity in training curves across individuals (see Figure 5 for example). This process involved fitting restricted cubic splines to each individual's training data and using numerical methods to find the area under the training curve, which reflects the degree of training improvement over 15 weeks taking into account individual variation in individual specific training curves. Other methods, which rely on fitting parametric models to individuals' data (e.g., individual-level growth curve models), must constrain the dose-response curve to be similar in shape across individuals, which can lead to biased results.

This method proved to be fruitful, as it revealed that training-related improvements in each task are highly correlated across individuals, and that training gains varied as a function of age, verbal ability, and reading comprehension level. In sum, the results in this chapter showed that home-based training could lead to high retention rates and large practice-related improvements in WM span among older adults, a group who on average shows declines not only in WM span, but also in training-related cognitive plasticity. Thus, home-based computerized training may be particularly beneficial as a method for continued cognitive training among older adults.

## **Chapter V.**

### **Transfer of Training to Working Memory and Language Comprehension**

In chapter IV, data were presented to demonstrate that short-term training in verbal WM results in improved session-to-session performance on trained tasks among older adults. In this chapter, data are presented to test the degree to which WM training results in near transfer to untrained verbal WM tasks. If training results in improvements in the verbal WM system, then trained participants should show broad improvements in performance in the untrained complex verbal WM span tasks (reading span, listening span, operation span, minus-2 span) at post-test relative to the active control group. Following this, analyses are presented to test for far transfer effects of WM training to language measures in the neuropsychological test battery administered at pre-test and post-test. Specifically, control and treatment groups are compared in change in performance on the following tasks: Nelson-Denny reading comprehension, verbal fluency, sentence recall, and discourse recall.

#### **Transfer to Verbal Working Memory**

Table 3 presents pre-test and post-test mean scores and change scores for the WM and language tasks separately for the control group and treatment groups.

The top portion of Figure 8 presents summary effect sizes of the group differences in change in each of the verbal WM tasks (e.g., the Treatment x Time interaction) in Cohen's *d* units. Larger values indicate a difference in change from pre-test to post-test favoring the treatment group. Confidence intervals are non-parametric bootstrapped confidence intervals (Kirby et al., 2013). As can be seen in this figure, there was positive evidence for broad

training-related improvements in verbal WM across the tasks. Of the 4 WM tasks, all had effects sizes over .5, indicating an approximate half standard deviation difference between the treatment and control groups in change in WM. Reading span, however, had a negative lower-bound on the confidence interval, indicating that it did not reach traditional levels of statistical significance. The average effect size of training across the four tasks was  $d = .87$ . Each task is discussed in more detail below.

*Reading Span.* Figure 9 shows the pre-test and post-test reading span scores separately for WM training and control participants. A random intercept and slope linear change score model was fit to the data to test the Group x Time interaction. There was a relatively weak Group x Time interaction that did not reach statistical significance ( $b = .61$ , 95% CI [-.12, 1.33]). Both groups were approximately matched in span score at pre-test. However, at post-test the WM training group showed marginally larger test scores compared to the control group.

*Listening Span.* Figure 10 shows the pre-test and post-test listening span scores separately for WM training and control participants. A random intercept and slope linear change score model was fit to the data to test the Group x Time interaction. There was a reliable positive Group x Time interaction ( $b = 1.32$ , 95% CI [.53, 2.10]). As can be seen both groups were approximately matched in span score at pre-test. However, at post-test the WM training group showed larger test scores compared to the control group. The control group seemed to show a deficit in post-test performance, relative to pre-test performance. This could be due to either selective negative transfer of the speed training, regression to the mean, or form-effects, whereby post-test listening span stimuli were differentially more difficult than pre-test stimuli. However, without more than two waves of data, form effects cannot be dissociated from training effects (Gross et al., 2012; Payne, Gross, et al., 2014).

*Operation Span.* Figure 11 shows the pre-test and post-test operation span accuracy scores separately for the WM training and control participants. As can be seen, both groups were approximately matched in accuracy at pre-test. However, at post-test the WM training group showed larger accuracy scores compared to the control group. A random intercept and slope linear change score model was fit to the data to the Group x Time interaction. There was a reliable Group x Time interaction ( $b = .24$ , 95% CI [.08, .40]).

*Minus-2 Span.* Figure 12 shows the pre-test and post-test minus-2 span accuracy scores separately for WM training and speed training participants. A random intercept and slope linear change score model was fit to the data to test the Group x Time interaction. There was a reliable Group x Time interaction ( $b = .10$ , 95% CI [.02, .18]). As can be seen, both groups were approximately matched in accuracy at pre-test. However, at post-test the WM training group showed larger accuracy scores compared to the control group.

### **Transfer to Language Outcomes**

The test battery administered at pre-test and post-test included four tasks assessing different aspects of language use in older adulthood. (1) the Nelson-Denny task tapped general reading comprehension, (2) the FAS/BDT task tapped verbal fluency, (3) the sentence recall task probed memory for 18-word sentences, and (4) the Rivermead Memory task probed memory for longer multi-sentence discourse passages. The lower portion of Figure 8 presents summary effect sizes of the group differences in change in each of these language measures (e.g., the Treatment x Time interaction) in Cohen's  $d$  units, along with corresponding confidence intervals. To summarize the results, both the verbal fluency and sentence recall task showed evidence for positive transfer of WM training. In contrast, the two tasks focusing on

discourse understanding, the Rivermead and the Nelson-Denny tasks, showed no evidence for WM specific training-related improvements. Each task is discussed in more detail below.

*Reading Comprehension.* Figure 13 shows the pre-test and post-test Nelson-Denny accuracy scores separately for treatment and control participants. A random intercept and slope linear change score model showed no evidence for a Group x Time interaction  $b = .002$ , 95% CI [-.11, .12]. As can be seen, both groups were matched in accuracy at pre-test. At post-test, both the WM training group and the control group showed numerically lower scores than at baseline, with no group differences in performance.

*Verbal Fluency.* Figure 14 shows the pre-test and post-test verbal fluency scores for total number of words recalled separately for WM training and control participants. A random intercept and slope linear change score model was fit to the data to test the Group x Time interaction. There was a reliable Group x Time interaction ( $b = 6.57$ , 95% CI [1.32, 11.82]). As can be seen, both groups were matched at pre-test. However, at post-test the WM training group showed larger fluency scores compared to the control group.

*Sentence Recall.* Figure 15 shows the pre-test and post-test scores for percent of words recalled verbatim from the sentence recall task separately for WM training and control participants. A random intercept and slope linear change score model was fit to the data to test the Group x Time interaction. There was a reliable Group x Time interaction  $b = .08$ , 95% CI [.02, .14]. As can be seen, both groups were matched in recall at pre-test. However, at post-test the WM training group showed larger recall scores compared to the control group. A similar pattern was found when scoring the sentences as proportion of propositions correctly recalled (see Figure 16), although the Group x Time interaction was only marginal in this case ( $b = .05$ , 95% CI [-.01, .11]). The correlation between the two coding schemes (% words recalled



and %propositions recalled) was .91 at baseline, and both measures showed the same test-retest correlation ( $r = .66$ ).

*Discourse memory.* Figure 17 shows the pre-test and post-test Rivermead scores for percent of propositions correctly recalled, separately for control and treatment participants. At baseline, both groups were matched in accuracy. There was no group difference in change in discourse memory, with both groups showing similar scores at post-test. A random intercept and slope linear change score model showed no reliable Group x Time interaction ( $b = -2.85$ , 95% CI [-11.61, 5.91]).

## **Summary**

The results presented in this chapter support two general conclusions. First, the WM training group showed broad improvements in verbal working memory. Of the four verbal WM tasks employed in the test battery, there was evidence of robust training-specific improvements in three of the tasks, with the reading span task showing a trend of training-specific improvement. Interestingly, the sentence reading span task employed in the pre/post test battery was arguably the “nearest” near transfer WM task, as it matched closely with the surface-level demands of one of the training tasks in the iTrain task. Yet, it was this measure that showed the smallest effect size for training improvements, with the listening, operation, and minus-2 span tasks showing larger and statistically reliable training effects. Thus, this finding argues strongly against the notion that training-related improvements were task-specific. Indeed, the largest effect size for change occurred in the operation span task, which differed along many dimensions from the training tasks. These findings are consistent with the idea that targeted training of complex verbal WM can produce improvements in domain-general verbal WM

mechanisms that are abstracted from specific tasks and extend to untrained tasks. Secondly, the results in this chapter suggested that there is some positive evidence for far transfer of verbal WM to certain language tasks, including measures of sentence memory and fluency. However, measures tapping discourse comprehension (Rivermead and Nelson-Denny) did not show evidence of transfer of training gains. These findings are discussed in more detail in the general discussion.

## **Chapter VI.**

### **Effects of Training on Syntactic Comprehension:**

#### **Evidence from Eye-Movement Control During Reading**

In Chapter V, evidence was presented suggesting that the WM training group showed differential improvement not only in untrained verbal WM measures, but also selective improvements in certain aspects of language—specifically in sentence memory and verbal fluency. In this chapter, data are presented testing the degree to which WM training impacts the on-line processing and off-line comprehension of sentences that differ in syntactic complexity. As discussed previously, both treatment and control participants completed eye-tracking experiments at pre-test and post-test. This experiment probed participants' ability to resolve several kinds of syntactic difficulties that have previously been associated with WM costs: “garden-path” syntactic ambiguity resolution (Christianson et al., 2006), object-relative clause processing (Kemper & Liu, 2007), and long-distance syntactic dependency processing (Caplan et al., 2011). The goals of the analyses in this chapter are twofold. First, to test the degree to which WM training may affect comprehension of syntactically demanding sentences, and second, to test the degree to which training effects vary for off-line compared to on-line assessments of comprehension difficulty.

First, analyses of the baseline accuracy and fixation time data are presented for the full sample, in order to examine the locus of syntactic costs in accuracy and eye-fixation measures as they unfold across the sentence. Fixation and regression measures that are typically sensitive to syntactic effects in the eye-movement record are investigated: gaze duration (the sum of all first-pass fixations on a word before the eyes move to another word), regression path duration

(the sum of all fixations from when a reader first enters the target word, including the time spent refixating on earlier words, until he or she moves past the target word to the right), and the probability of regressing (the proportion of trials in which a regression was launched from the target word(s)) (Clifton et al., 2007). Analyses are then presented testing the degree to which WM training results in differential changes in off-line comprehension accuracy and on-line processing. For each sentence set, off-line comprehension accuracy and on-line eye-fixations and regressions in focal interest areas are compared between pre-test and post-test for the control and treatment groups.

### **Syntactic Comprehension in Older Adulthood**

*Sentence Set 1: Garden Path Ambiguity.* Analyses are first presented for accuracy, followed by eye-fixation measures.

A logit mixed-effects model was fit to the accuracy data to test the effects of syntactic ambiguity on comprehension accuracy for the full baseline sample. Comprehension was reliably poorer for ambiguous sentences ( $M = 67\%$ ) compared to unambiguous sentences ( $M = 82\%$ ) ( $b = 1.05$ ; 95% CI [.66, 1.44]). The first panel of Figure 18 presents the effect size of syntactic ambiguity on comprehension, expressed as an odds ratio.

Figures 19-21 show mean gaze duration, regression path duration, and probability of regressions for the ambiguous and unambiguous sentences across sentence region. The target region is the disambiguating region, which is the point in these sentences where an ambiguity is first detected (Slattery et al., 2013). In this region, there was no difference in gaze duration in between the ambiguous ( $M = 290$  ms) and unambiguous items ( $M = 297$  ms) ( $b = .94$ , 95% CI [-18, 20]) (see Figure 19). However, regression path duration revealed a reliable effect of

syntactic ambiguity ( $b = 270$ , 95% CI [107, 432]), such that ambiguous items showed longer regression path durations ( $M = 1373$  ms) than unambiguous items ( $M = 1116$  ms) (see Figure 20). Likewise, the probability of regression was reliably larger in the disambiguating region for garden path sentences ( $M = 52\%$ ) compared to sentences that were not ambiguous ( $M = 45\%$ ) ( $b = .28$ , 95% CI [.02, .53]) (see Figure 21).

An analysis of the baseline data was conducted as a function of random assignment, in order to test the degree to which there were pre-test differences between the randomly assigned control and treatment groups. This analysis revealed that there were no group differences at baseline in the effects of ambiguity on accuracy ( $b = .72$ , 95% CI [-.09, 1.52]) or on eye-movement measures showing ambiguity effects in the disambiguating region: regression path duration ( $b = -18$ , 95% CI [-347, 311]) or probability of regressions ( $b = .42$ , 95% CI [-.11, .94]).

*Sentence Set 2: Long Distance Dependency Processing.* Analyses are first presented for accuracy, followed by eye-fixation measures.

A logit mixed-effects model was fit to the accuracy data to test the effects of long distance dependency (i.e., presence of a doubly-embedded relative clause) on comprehension accuracy. Interestingly, comprehension was not reliably different between sentences with a single relative-clause embedding (low-distance) ( $M = 70\%$ ) and those with a double relative-clause embedding (high-distance) ( $M = 69\%$ ) ( $b = .11$ , 95% CI [-.26, .48]). The middle panel of Figure 18 presents the effect size of the difference between high and low distance embeddings on comprehension, expressed as an odds ratio.

However, the on-line eye-movement data revealed a different picture of comprehension difficulty. Figures 22-24 show mean gaze durations, regression path durations, and regression probabilities for the low-distance and high-distance sentences across interest areas. The focal

interest area in long distance dependencies is the embedded verb (V2) (see Bartek et al., 2011). In this region, there was no difference in gaze duration in between the single-embedded ( $M = 372$ ) and doubly embedded sentences ( $M = 391$ ) ( $b = 20$ , 95% CI [-21, 61]) (see Figure 22). However, regression path duration revealed a reliable effect of long distance dependency ( $b = 351$ , 95% CI [192, 509]), such that doubly-embedded sentences showed longer regression path durations ( $M = 911$ ) than single-embedded relative clause sentences ( $M = 536$ ) (see Figure 23). Likewise, the probability of regression was reliably larger at the embedded verb for long-distance sentences ( $M = 34\%$ ) compared to single-embedded relative-clause sentences ( $M = 18\%$ ) ( $b = .86$ , 95% CI [.35, 1.37]) (see Figure 24). Analyses of baseline data as a function of random assignment revealed that there were no group differences in the effects of ambiguity on accuracy ( $b = .25$ , 95% CI [-.51, 1.01]), regression path duration ( $b = -154$ , 95% CI [-473, 165]), or probability of regressing ( $b = -.67$ , 95% CI [-1.69, .35]) at baseline.

*Sentence Set 3: Subject vs. Object Relative Clauses.* Analyses are first presented for accuracy, followed by eye-fixation measures.

A logit mixed-effects model was fit to the accuracy data to test whether comprehension differed for subject and object-relative clauses. Comprehension was reliably worse for object relative sentences ( $M = 65\%$ ) compared to subject-relative sentences ( $M = 70\%$ ) ( $b = 1.13$ , 95% CI [.75, 1.50]). The last panel of Figure 18 presents the effect size of the difference in comprehension between subject-relative and object-relative clauses, expressed as an odds ratio.

However, eye-fixation data revealed no evidence of on-line processing difficulty. In gaze duration, regression path duration, and probability of regressing, there was no reliable difference between the subject-relative and object-relative clauses at the areas of interest (all  $t$ 's

> 1.17), suggesting that the costs seen in comprehension are not driven by increases in on-line processing difficulty in the older adults (cf. Stine-Morrow et al., 2000).

### **Effects of WM Training on Syntactic Comprehension and On-line Processing**

In this section, the effects of WM training on accuracy and on-line processing are tested for garden path, long-distance dependency, and subject vs. object relative sets, by examining the degree to which the baseline syntactic comprehension costs (i.e., reduced accuracy, longer regression path durations, and higher probability of regressing in more complex sentences) are moderated at post-test for the working memory training group, compared to the active control group. Analyses are presented first for accuracy, followed by the eye-tracking data. Because only regression path duration and probability of regressing appeared to be sensitive to syntactic costs at baseline, only these measures were assessed for training effects. Note that, in the subject-object relative sentence set, the training results are not presented for the eye-tracking data because there was no indication of baseline differences in regression-path durations or regressions between the conditions, despite large differences in accuracy.

*Garden Path Sentences.* A logit mixed-effects model with random effects for subjects and items was fit to the accuracy data with training group (control vs. treatment), testing occasion (pre vs. post), and syntactic ambiguity (ambiguous vs. unambiguous) treated as fixed-effect factors. Table 4 presents the parameter estimates and confidence intervals from this model. The results revealed a reliable Condition x Time x Treatment interaction ( $b = -.81$ ; 95% CI [-1.39, -.25]), suggesting that there were training group differences in change in accuracy for the ambiguous compared to the unambiguous items. Figure 25 plots mean accuracy for each sentence set at pre-test and post-test for the control and treatment groups. As can be seen in this

figure, in the control group, the ambiguity effect (difference between ambiguous and unambiguous conditions) is almost identical at pre-test and post-test. However, in the WM training group, the ambiguity effect at baseline is reduced at post-test and this is driven by increases in accuracy for the ambiguous condition from pre-test to post-test.

A linear mixed-effects model with random effects for subjects and items was fit to regression path durations, with training group, testing occasion, and syntactic ambiguity treated as fixed-effect factors. The results of this model are presented in Table 5. This model revealed no evidence for a Condition  $\times$  Time  $\times$  Treatment interaction. As can be seen in Figure 26, there was no evidence that the WM and control groups differed in the effects of syntactic ambiguity on regression path duration. A similar pattern was found for the probability of regressions. Table 6 reveals the results from a logit-mixed model testing group differences in change in probability of regressing for ambiguous and unambiguous sentences. There was no reliable 3-way interaction, and no reliable lower-order interactions, as can be seen in Figure 27.

Thus, while WM training appeared to improve comprehension accuracy for garden-path sentences relative to the control group, there was no evidence that this improvement in accuracy derived from changes in processes reflected in eye-movement behavior during sentence reading.

*Long Distance Dependency Comprehension.* A logit mixed model was fit to the accuracy data to test for group differences in change in accuracy for the long-distance dependency sets. Results from this model are presented in Table 7. The Condition  $\times$  Time  $\times$  Treatment interaction was weak and not statistically significant ( $b = -.83$ , 95% CI [-1.81, .15]). Figure 28 plots mean accuracy for each sentence condition at pre-test and post-test for the control and treatment groups. The figure shows no difference in accuracy at pre-test between the treatment and control groups, and a slight decline in accuracy in the control group for the high-



distance dependency sentences only. However, this effect did not reach traditional levels of statistical significance.

Figure 29 shows mean regression path durations, and Figure 30 shows probability of regressing for the high-distance and low-distance items at the embedded verb region, separately as a function of training group and testing occasion. A linear mixed-effects model with random effects for subjects and items was fit to regression path durations with training group, testing occasion, and syntactic complexity treated as fixed-effect factors. The results of this model are presented in Table 8. This model revealed no evidence for a Condition  $\times$  Time  $\times$  Treatment interaction. As can be seen in Figure 29, there was no evidence that the WM and control groups differed in the effects of syntactic complexity on regression path duration. However, both groups showed a reduction in long-distance dependency costs over time, as revealed by a reliable Condition  $\times$  Time interaction ( $b = -285$ , 95% CI [-535, -35]). Because both groups showed a reduction in regression path duration, it is not clear if this effect is due to re-test effects, syntactic learning (i.e., increased facilitation after increased exposure), or perhaps due to the possibility that both training methods lead to improved on-line syntactic comprehension.

A similar pattern was found for the probability of regressions. Table 9 reveals the results from a logit-mixed model testing group differences in change in probability of regressing for long-distance and low-distance sentences. There was no evidence for a 3-way interaction, as can be seen in Figure 30. However, both groups showed a trend for a reduction in probability of regressing at pre-test relative to post-test for the more difficult long distance dependencies, though this did not reach statistical significance.

*Subject and Object Relative Sentences.* Table 10 presents the parameter estimates and confidence intervals from a logit mixed model testing group differences in change in accuracy

for the subject-object relative sentence set. The Condition x Time x Treatment interaction was not reliable, but showed evidence of a trend ( $b = -.95$ , 95% CI [-2.00, .11]). Figure 31 plots mean accuracy for each sentence condition at pre-test and post-test for the control and treatment groups. The figure shows that, in the control group, there were no differences in accuracy between the subject-relative and object-relative sentences at pre-test or post-test. However, in the working memory training group, there was a difference between SR and OR sentences at pre-test and disappeared at post-test. However, while, accuracy for object-relative sentences improved from pre-test to post-test, accuracy for subject-relative sentences declined, generating the trend toward an interaction.

## **Summary**

The goals of the eye-tracking and comprehension analyses presented in this chapter were threefold (1) to investigate the degree to which various forms of syntactic complexity that putatively strain working memory resources impact on-line and off-line language comprehension in older adults, (2) to investigate the degree to which WM training can reduce such syntactic costs in off-line comprehension, and (3) to investigate the degree to which WM training transfers to on-line comprehension.

Towards the first goal, the results of the baseline analyses were clear. Although these three sentence sets are argued to index similar working-memory dependent syntactic constraints, the different manipulations of syntactic complexity resulted in different patterns of costs in the eye-movement record and accuracy data for each sentence set. Garden path sentences showed increased regression path durations and probability of regressions at the disambiguating region, along with correspondingly worse accuracy for syntactically ambiguous items. Long-distance

dependency sentences showed evidence for on-line comprehension difficulty at the embedded verb in regression path duration and probability of regression, although there was no evidence for off-line comprehension costs. And lastly, subject-and object-relative clause items showed evidence for off-line differences in accuracy, but no differences in processing were found in the eye-movement record.

Although a domain-general verbal working memory resource constraint has been theoretically invoked to explain the source of syntactic complexity in each of these sentences across different studies, it is clear that there are qualitatively different patterns in how, when, and whether this complexity is resolved. In particular, it is surprising that there was no evidence for off-line comprehension differences in the long-distance dependency set at baseline, given the relatively large on-line effect in regression path duration and probability of regression. It is possible that older adults were able to use this continued allocation of effort in order to resolve the long-distance dependency and maintain higher comprehension (Payne, James, Stine-Morrow, & Watson, 2014; Schotter, Tran, & Rayner, 2014). However, the subject and object-relative clauses showed the exact opposite pattern. Consistent with a number of studies (Kemper et al., 1997; Stine-Morrow et al., 2000; Caplan et al., 2011), we found that older adults showed worse accuracy for OR sentences relative to SR sentences. However, this accuracy difference was not manifested on-line in terms of increases in processing at the main verb (Traxler et al., 2001) or relative-clause region (Staub et al., 2010), as has been previously reported. One possibility is that older adults were not allocating enough time to resolve the dispreferred OR structure, and this resulted in reduced accuracy, which is consistent with results from a self-paced reading study by Stine-Morrow and colleagues (2000), in which reduced allocation of effort to OR processing was related to worse accuracy in understanding these sentences among older adults.

Alternatively, it could be that the overall increased proportion of relative clauses in this study could have resulted in exposure or learning-related facilitation across the testing session at pre-test (Fine, Jaeger, Farmer & Qian, 2013), that impacted on-line processing, but not comprehension for these items.

The second goal was to test the degree to which syntactic comprehension accuracy was improved in the WM training group relative to the control group. Results of this analysis were mixed. In the garden path items, there was positive evidence for WM-specific-improvements in comprehension of syntactically ambiguous items. Both the treatment and control groups showed the canonical garden-path ambiguity effect in comprehension (Christianson et al., 2001; 2006). However, at post-test, only the WM training group showed evidence for reduced ambiguity effects on comprehension. This effect is driven by a selective increase in comprehension for the ambiguous items only. While the effect size of this increase in comprehension is small overall, this effect is theoretically interesting as it suggests that the WM system is engaged in older adults' ability to suppress the initial infelicitous interpretation of the ambiguous sentences. This is consistent with findings from Christianson et al. (2006), who showed that older adults' comprehension of similar RAT verb ambiguous sentences was highly correlated with complex verbal WM span performance. Thus, the training data presented here corroborate prior correlational results and extend these by suggesting that the WM system subserving off-line ambiguity resolution is plastic and responsive to training.

Somewhat similar results were found for the subject- and object-relative clause items. At baseline, both training and control groups showed evidence for worse accuracy for object-relative sentences compared to subject-relative sentences. At post-test, the control group showed the same SR-OR difference. However, the training group showed a different pattern, with a

marginally lower difference between the SR and OR sentences at post-test. This pattern is muddled by the fact that the reduction in cost at post-test between the training and control groups was driven partially by increased comprehension for OR sentences but also driven partially by decreases in comprehension for SR sentences. Because there were form-differences between pre-test and post-test sentence sets, it is possible that the SR items overall were more difficult at post-test.

The LDD sentence set showed no clear evidence for a training-related improvement in accuracy. There was evidence for a marginal three-way interaction, however this was complicated by a lack of pre-test difference in accuracy to begin with, along with the interaction being driven by a decrease in accuracy at post-test only for the control group.

Alternatively, it is possible that WM training resulted in an increased expectancy for object-relative clauses, based on the local language statistics in the eye-tracking session. That is, it is possible that WM training lead to an increased sensitivity to the presence of OR sentences, so that there was greater expectancy for such constructions, given that object-relative clauses were more frequent in this study. Evidence consistent with this explanation is given by Farmer, Christensen, and Kemtes (2005), who showed that high-span participants show greater statistical learning than low-span participants. Given the increased probability of being exposed to an OR sentence within the eye-tracking session, it is possible that the high-span adults at post-test (those who were trained on complex WM) had more working memory resources available to predict and develop biases for the non-canonical OR syntactic construction, leading to both an *increase* in comprehension for OR items, but a concomitant *decrease* in comprehension for the more common SR construction.

Towards the last goal, on-line processing (regression path duration and probability of regressing) was analyzed as a function of training group, syntactic complexity, and measurement occasion for the two sentence sets showing on-line syntactic costs, the garden-path sentence set, and the long-distance dependency set. The results from our eye-tracking analyses consistently suggested no evidence for training-specific facilitation in on-line syntactic processing for either garden-path ambiguity resolution or long-distance dependency processing. This was true despite robust on-line processing costs at pre-test, which have been argued to be due to on-line WM resource constraints. For both regression path duration and probability of regressing, the garden path sentences showed no clear differences at pre-test and post-test. For the LDD set, long-distance dependency processing was reduced at post-test relative to pre-test. However, this effect was equivalent across both the training and active control groups, suggesting that there was no differential improvement in the training group. The difference in training effects found between the on-line and off-line measures of sentence comprehension are discussed in more detail in the following chapter.

## Chapter VII.

### General Discussion

Maintaining effective language understanding and communication into old age is crucial not only to learning new information in adulthood, but also to continued cognitive resilience (Stern, 2009; Stine-Morrow & Payne, 2014). At the same time, age-related changes in component cognitive abilities such as working memory have a profound effect on language comprehension. This is especially troubling because educational opportunities are front-loaded to early in life, so that reading is one of the major ways through which older adults learn new information and seek mental stimulation. Importantly however, while the effects of age and working memory are robust for certain aspects of language understanding, the majority of the evidence implicating the efficiency of the WM system in language comprehension comes from correlational studies. Thus, definitive conclusions about underlying mechanisms are difficult to make. In order to tackle this issue of causal ambiguity, the current study tested the degree to which cognitive training in verbal WM could transfer to aspects of language comprehension in older adulthood, using a novel home-based training program called *iTrain*.

More specifically, the goals of this dissertation were to develop a home-based verbal working memory training program, to examine its effectiveness in improving working memory among older adults, and to test a causal account of the role of WM resources in language comprehension and on-line language processing. I argued in the introduction that such training data would provide key evidence to adjudicate between current models of memory and language in the psycholinguistics and cognitive aging literature. In the remainder of the chapter, I aim to address the goals discussed above and identify the conclusions that can be drawn from the current study and what implications these conclusions have for theories of aging, cognitive

training, working memory, and language comprehension. Specifically, in this final chapter, I will address the following questions:

*Is home-based working memory training effective in older adults?*

*Does working memory training improve language use in older adulthood?*

*Do gains from WM training transfer to language processing?*

*What implications do these results have for memory and language models?*

In the final section, I address the limitations and directions for future research.

### **Home-Based Verbal Working Memory Training: The Promise of Self-Initiated Training**

Shipstead, Redick, and Engle (2012) noted that a majority of recent WM training studies have produced findings that are difficult to interpret, in part because (a) the methods have relied on training in a large number of ill-defined tasks or (b) the tasks have largely focused on short-term memory or other abilities unrelated to WM. The issue with the former case is that, with such “kitchen-sink” interventions, it is impossible to discern what aspects of the training tasks were responsible for observed improvements. That is, it is equally possible that the presence of a single aspect of multi-modal training is responsible for observed benefits (cf. Novick et al., 2013), or that training gains are dependent upon some synergistic combination of practice within multiple domains, consistent with evidence from variable priority training (Kramer et al., 1999; Kramer & Willis, 2002). The latter issue is driven by the selection of training tasks that are not well motivated by theory. Baddeley and Hitch (1974) showed that simple short-term memory storage does not contribute to higher-order cognitive abilities, such as reasoning and language comprehension (see also Turner & Engle, 1989). In contrast, complex span tasks, which require that information be maintained in short-term memory in the face of ongoing



processing and distraction, have been shown to reliably predict recall and to decline substantially with advancing age (Bopp & Verhaeghen, 2005).

These concerns lead to the development of the iTrain program used in the current study, which focused on improving complex verbal WM span through repeated practice in the core abilities tapped by such tasks; that is, these tasks focused on promoting the continued maintenance of information in memory while simultaneously processing novel incoming information. The training program adopted for the current study was similar to lab-based complex span training tasks used in Morrison and Chein (2010), Borella et al. (2010), and Richmond et al. (2011). Each of these studies shared in common that training involved participants' repeatedly practicing tasks that required making some speeded judgment on a stimulus while holding in memory either information about that stimulus item (e.g., a content embedded span task) or novel information unrelated to the item (see Was, Rawson, Bailey, & Dunlosky, 2011). Although iTrain was similar to these complex span training programs in some ways, it also differed along a number of dimensions.

First, because the focus of the study was on investigating transfer to language ability, the training focused on verbal WM only. Second, and perhaps most importantly, the training protocol in the current study was home-based. This study is, to our knowledge, the first to employ a home-based variant of complex WM span training. In order to establish the efficacy of this method as a valid training paradigm, it is important to establish that home-based training can result in improved performance. Home-based training differs from lab-based training in several ways. The benefits of home-based training include convenience for the participant and a reduction of resources devoted to weekly testing sessions in the lab.

Additionally, offering home-based training as an option is likely to lead to a more diverse sample of training participants. Lab-based training likely leads to biased sampling of individuals who are highly mobile, healthy, and able to allocate substantial amounts of time each week to participating in laboratory sessions. Offering a home-based training is likely to lead to more heterogeneous sampling at both ends of the age distribution (i.e., younger-old adults who are still in the workforce do not have time for daily lab visits and older low-mobility adults who cannot make daily lab visits), as it reduces the burden on the participant to visit the lab over several sessions.

However, a major component of home-based training is that it requires the trainees to self-administer and self-monitor their training progress throughout the course of the intervention. In two experiments, Wadley and colleagues (2006) directly compared training gains in a useful-field-of-view training program among healthy older adults in laboratory and home settings. Both home and lab groups underwent eight to ten 1-hour cognitive training sessions. Both groups showed significant improvements in processing speed relative to a control group that underwent no training. Gains in the home-based group were 74% that of those in a lab-based training condition. These data suggest that self-administration of cognitive training is indeed feasible, though effect sizes may be smaller and more heterogeneous (see Payne et al., 2012 for similar evidence in a home-based reasoning training intervention). Such home-based training is likely to be more sensitive to individual differences in motivational factors, which may directly influence the amount of effort allocated to the training (cf. Payne et al., 2012). The data from the current study indicate that self-administration of the WM training is feasible, though important individual differences in the magnitude of effect sizes did exist.

The individual difference analyses showed that there was considerable variability across participants in the magnitude of improvement. Moreover, individual differences in training gains were highly correlated across the three tasks, indicating broad improvements in span, rather than the development of task-specific strategies (Schmiedek, Lovden, & Lindenberger, 2010). Specifically, training participants who were older, higher in verbal ability, and had better baseline reading comprehension showed the most improvement overall.

Indeed, as a group, older adults are typically less responsive to cognitive training programs than the young (e.g., Richmond et al., 2011), likely due to age-related declines in cognitive plasticity. For example, Baltes and Kliegl (1992) adopted a “testing-the-limits” paradigm to test for age-differences in the effects of repeated practice in the method-of-loci mnemonic on serial recall. Younger adults showed greater training-related improvements over 35 sessions compared to the old, so that age differences in performance were actually magnified at post-testing. Richmond and colleagues (2011) showed similar effects in a program of lab-based complex verbal and visuospatial WM training. While both younger and older adults showed training-related improvements, the improvements were greater among the young. However, it is important to note that, in these studies, older adults still showed reliable practice-related improvements from their baseline performance. Negative correlations between age and training gains are consistent with the so-called *Matthew effect* (Brehmer, Westerber, & Backman, 2012; McDougal & House, 2012), whereby individuals with better performance at baseline also show increased improvement through training. An analysis of the correlations of training gains with verbal ability and reading comprehension were more consistent with a Matthew effect however, as higher ability individuals showed greater improvements. A larger and more diverse sample of older adult training participants would likely reveal the extent to

which the age-related deficits in cognitive ability and improvements in verbal skill differentially impact responsiveness to verbal WM training.

In addition to showing that training resulted in improvements in the practiced WM tasks, a key test of the effectiveness of the training program was the assessment of the degree to which training effects transferred to untrained complex verbal WM span tasks. There was positive evidence for transfer across the complex span tasks measured in the current study, with all four tasks showing at least a half standard deviation improvement in WM for the training group relative to the control, with three of the four tasks reaching conventional standards of statistical significance. The average effect size across all four tasks was  $d = .87$ , indicating that training resulted in a near transfer improvement in verbal WM span of slightly less than a standard deviation. Thus, the evidence from the current study suggest that home-based training of WM can be effective in improving both trained and untrained complex verbal WM span tasks in the short-term. The question whether the training results in broader transfer to non-memory specific tasks is discussed in the following sections.

### **Working Memory Training and Language Comprehension: Implications for Models of Memory and Language**

The primary aim of the current study was to test the degree to which training-related improvements in WM led to improvements in language comprehension. To date, several WM training studies have shown promising results for training gains and transfer to so-called “far transfer” measures in older adults, suggesting that there exists some age-related maintenance of plasticity in the WM system (Buschkuhl et al., 2008; Borella et al., 2010; Zinke et al., 2013),

including evidence for training effects in the oldest-old (over 75 years; Borella et al., 2012) and among individuals with mild cognitive impairment (Caretta et al., 2013). As reviewed above, the findings from the iTrain project suggested that training does result in near transfer to untrained complex span tasks. However, the findings indicating whether these training gains impacted language comprehension in older adulthood were less consistent.

Positive evidence for transfer of training was found for several language measures. Older adults in the WM training group showed differentially larger improvements in both sentence recall and verbal fluency relative to the active control group. It is perhaps unsurprising that short-term sentence recall showed transfer, as sentence recall performance is highly related to WM (Payne et al., 2012; Stine-Morrow et al., 2008; Zelinski & Lewis, 2011), and, at least for the reading span task, overlaps substantially in task demands (Roberts & Gibson, 2002; MacDonald & Christiansen, 2002). However, the demonstration of training-related transfer to sentence recall is not trivial for at least two reasons. First, although verbal WM and sentence recall share a substantial amount of variance in older adults, this does not necessarily imply that training should result in transfer. Indeed, individual differences in WM and fluid intelligence share upwards of 50% of the same variance (Engle, 2010), and yet evidence for transfer of WM training to fluid ability is inconsistent at best (see Melby-Verlag & Hulme, 2013; Shipstead et al., 2012 for reviews). Second, sentence memory shows some of the largest effect sizes for age-related declines among measures of language comprehension (Stine-Morrow et al., 2008; Johnson, 2003).

The demonstration that WM training transferred to verbal fluency indicates that training can lead to far transfer, as the fluency tasks shared very little overlap with the tasks involved in the training. However, interpretation of training effects on verbal fluency are clouded by the fact

that tasks such as the *FAS* are used in both research and clinical settings to index a range of theoretically different cognitive operations such as executive control functioning (Mayr & Kliegl, 2000), semantic processing efficiency (Troyer et al., 1997), predictive language production mechanisms (Federmeier, 2007), and lexical knowledge (Nagels et al., 2012). For example, studies investigating verbal fluency in Alzheimer's disease suggest that the source of observed difficulties is driven by disease-related deficits in semantic memory (Laatu et al., 2003). Moreover, Federmeier and colleagues (Federmeier, Kutas, & Schul, 2010; Federmeier, McLennan, De Ochoa, & Kutas, 2002) have presented evidence that age-related declines in verbal fluency are implicated in age-related deficits in predictive processing in sentence comprehension.

Recently, McDowd and colleagues (2011) reported results from an individual difference study examining verbal fluency in healthy aging, Alzheimer's disease, and Parkinson's disease. Their data suggested that individual differences in verbal fluency could be largely accounted for by variation in processing speed, inhibitory control, and working memory, but that verbal ability played only a minor role in fluency performance. These data suggest that measures of verbal fluency may be tapping aspects of executive control functioning to a greater extent than they are tapping linguistic knowledge. More work is necessary in order to understand the role that WM training may play in improving executive control components compared to aspects of language production and semantic processing tapped by such fluency measures.

Two tasks tapping discourse comprehension showed no evidence of transfer of training gains, the Nelson-Denny reading comprehension task, and the Rivermead behavioral memory task, a measure of discourse memory (see Payne et al., 2014). While this may be surprising given that both reading comprehension and discourse recall are correlated with WM, one

explanation is that age-related declines in discourse understanding are rare, with some arguing that discourse comprehension is age-invariant (Radvansky, 1999; Radvansky & Dijkstra, 2007; Stine & Wingfield, 1990). Radvansky and colleagues have argued that discourse comprehension relies on the establishment of a situation model in memory, and this level of understanding is “relatively durable in the face of deficits at other, more abstract levels of processing... This durability may arise from the use of more fundamental representational processes that would be less likely to suffer under the relatively mild neurological disruption that accompanies normal aging” (Radvansky & Dijkstra, 2007, *p.* 1039). Under this account, WM training would not be expected to impact measures of discourse comprehension, as older adults can rely on situational representations as a compensatory mechanism in order to maintain comprehension despite reduced WM resources. However, for the context-free sentences in the sentence recall task, where it is less likely that a situational representation can be established, WM effects are larger, and effects of training are found.

In a recent study, Caretti et al. (2012) trained older adults in verbal working memory and showed transfer to measures of discourse comprehension. However, several methodological weaknesses in this study cloud the interpretation of these findings. Most importantly, the training involved what was called “WM updating during reading.” In this training task, participants read short stories and were asked to recall specific actions or thoughts of a protagonist in the story. This focus on training in reading comprehension in addition to verbal WM confounds an examination of “far transfer” to measures of language comprehension (such as the Nelson-Denny) as the training task itself included overlapping features of the language measures.

While we found some positive evidence for transfer to off-line measures of comprehension, most importantly we did not find evidence that training resulted in transfer to measures of on-line language processing during syntactic comprehension. This occurred in spite of some evidence that WM training lead to increases in accuracy for the same set of sentences for which on-line processing was assessed. In brief, the eye-tracking experiments demonstrated some evidence that syntactic comprehension accuracy appeared to be sensitive to WM training, though these effects were selective to sentence sets focusing on syntactic ambiguity resolution and object-relative clause processing. At the same time, there was no evidence that this increase in accuracy was accompanied by a change in on-line language processing.

Models of WM and language comprehension that assume a domain-general verbal memory system is brought on-line in real time to assist in encoding and retrieval operations might predict that complex span training would result in reduced on-line comprehension costs for more complex sentences (Just & Carpenter, 1992 Gibson, 1998). Models such as Gibson's SPLT (1998), or the cue-based parsing framework (Van Dyke & Lewis, 2003), predict that syntactic comprehension involves the on-line maintenance of incomplete syntactic relationships, and that the encoding, binding, and retrieval of these dependencies must be dependent upon an attentionally constrained WM system that actively updates information. However, these models do not make explicit claims about how individual differences in the efficiency of such a system might influence on-line comprehension. Just and colleagues' 4-CAPS model however makes very explicit claims that individuals with increased efficiency of WM will show facilitation in processing specifically for linguistic items that are most taxing to WM. Thus, resource-dependent models like Just and colleagues' 4-CAPS model would predict that improving WM through repeated training should not only affect accuracy in comprehending more complex



syntactic constructions, but that this effect should be detected in the initial moment-to-moment encoding of the language, during on-line processing.

However, as reviewed in the introduction, other models of individual differences in memory and parsing make very different predictions. First, Caplan and colleagues' SLIR model hypothesizes that there is a fractionated WM system with a dedicated sub-system for on-line language processing. Under this model, age-related reductions in WM result in poorer representations of the language in memory so that so-called "post-interpretive" operations of language comprehension are diminished with age. However, on-line language processes are not tapped by this general verbal WM resource that declines with age (Caplan et al., 2011), leading to age-related preservation in on-line syntactic comprehension. The SLIR model is supported most clearly by results from DeDe and colleagues (2004), who used structural equation models to show that individual differences in WM mediated age-related changes in off-line measures of syntactic comprehension, but that this mediational relationship was not found for measures of on-line processing derived from self-paced listening tasks. Thus the SLIR model predicts that complex verbal WM span training should result in improvement in off-line measures of language comprehension, but that measures of on-line processing would not be sensitive to WM training.

Lastly, MacDonald and colleagues (MacDonald & Christiansen, 2002) argue that the WM system subserving language comprehension is an epiphenomenon of the influence of linguistic knowledge and exposure on language processing. Individuals with greater language experience perform better on both measures of complex verbal WM (based on the increased efficiency in processing linguistic information) and on measures of language comprehension. Although there is reason to believe that verbal abilities and verbal WM share some overlap in

task demands, developmental evidence shows that language experience and verbal WM resources show divergent trajectories over the lifespan (Payne et al., 2014; Stanovich et al., 1995), and have dissociable influences on language comprehension (Payne et al., 2014; Payne & Stine-Morrow, 2014). Nevertheless, there was some evidence that individuals with greater verbal ability showed greater WM training gains over 15 sessions, suggesting that there may be more complex interactions between WM and language skill than has previously been assumed (see also Caretti et al., 2009; Payne et al., 2012). MacDonald's experience constraint model makes clear predictions in the current study. Because the training and control groups were matched on the exact linguistic items they were exposed to, and differed only in their goals to recall information versus make speeded judgments about those items, both groups should perform the same at post-test on all measures of language comprehension, given the equivalence in exposure to the language through the training tasks.

Overall, the evidence for transfer to language comprehension was isolated to measures of off-line comprehension, memory, and production. The eye-movement data revealed no evidence for improvements in processing as a function of training. Thus, the training results appear to be consistent with Caplan and colleagues SLIR model. Complex span training can improve the efficiency of domain-general verbal WM systems in older adulthood, and this may transfer to tasks of language comprehension that are dependent upon such a system. However, immediate moment-to-moment processing is not sensitive to improvements in complex WM span, at least as assessed by eye-movements during reading.

Eye-tracking data is inherently noisier than data derived from psychometric tests, and hardly any data exist testing the psychometric properties (reliability, stability over measurement occasions, convergent validity) of eye-fixation data during reading. Given that reduced

reliability can lead to underpowered tests of effects, more work needs to be conducted to understand the relative value of eye-movement data for testing longitudinal predictions, an issue currently being explored in other measures of on-line language processing such as fMRI (cf., Uttal, 2013), and EEG/ERP (Cassidy, Robertson, & O'Connell, 2012).

### **Limitations and Future Directions**

In this last section, I address important caveats and assumptions of the current study, as well as limitations and areas of future direction for WM training research more generally. The primary limitation in the current study is that the training study was certainly underpowered to detect small-to-moderate effect sizes. The issue of small sample sizes is rampant in the WM training literature. Indeed, the sample in the current study of  $N = 41$  makes this one of the largest WM training studies with older adult samples (see Melby-Verlag & Hulme, 2013). This issue is largely driven by resource constraints associated with conducting adequately powered cognitive training studies due to issues with recruitment, retention, maintaining intention-to-treat, and the overall costs of conducting longitudinal research. Nevertheless, cognitive training research needs to overcome these methodological shortcomings, and current calls in the literature exist to create standardized training protocols that meet the standards of medical trials in terms of methodological rigor (Walton, Mowszowski, Lewis, & Naismith, 2014). Several advances were made in the current study to meet the criteria of a randomized controlled trial, as laid out in the CONSORT statement.

Great care was taken to evaluate the effects of iTrain against the appropriate control group in the context of a literature in which inadequate control groups plague many cognitive training studies. A number of cognitive training studies use no-contact control groups, which

only control for retest effects. Indeed, because the control groups are not matched on their expectancies to improve, differential change can be attributed to *Hawthorne* effects, in which task-related expectancy to improve drives motivational factors to improve performance at post-testing. Even in studies with so-called “active” control groups, different groups may vary substantially in their expectations for improvement generally as well as on specific tasks (Boot et al., 2013). In this study, we adopted a “component control” design to keep control and treatment groups as well matched as possible. Indeed, post-testing surveys revealed that individuals in both groups had similar endorsement of perceived training improvements. That only moderate perceived change was found in the presence of observable improvement suggests that these effects are not likely attributable to so-called “Hawthorne” effects (Boot et al., 2013). In addition, an intention-to-treat approach was used, in order to downwardly bias effect sizes with differential drop from the training. However, because the home-based training resulted in such high retention, the issue of differential drop-out causing the observed training benefits is not likely.

Future work in training of complex WM should focus on larger and more diverse samples. This may be accomplished more easily and inexpensively with home-based training and assessments, as these approaches require less lab resources to be allocated to each individual subject. One goal of this work is to show that home-based training is a valid option in future studies and may be able to help move towards studies with optimally powered sample sizes to detect more nuanced effects of training, as the overall cost-per-subject is lower in home-based relative to lab-based training studies.

This is the first study to my knowledge to demonstrate successful far transfer of WM training to language outcomes in older adults. However, despite the relative breadth of

measurement, the battery was sparse in assessments of language production. Indeed the two tasks in the neuropsychological battery to show the strongest evidence of far transfer each involved verbal production (FAS and sentence recall), and it is possible that production measures may be more sensitive to WM training than measures of language reception (Acheson & MacDonald, 2009; Acheson et al., 2011).

This leads to a larger question for future work regarding the potentially relatively narrow nature of cognitive training on outcomes. In clinical and applied fields, it is common for training programs to adopt a very broad cognitive battery and expect broad-based changes. In contrast, it is possible that training programs will lead to very specific improvements only to transfer tasks that are subserved by the core mechanisms being continually taxed in the training tasks. A goal of future work is to establish theoretically sound training and transfer tasks and reintroduce the training paradigm as a method to target specific mechanisms and test mechanistic accounts of theoretical models (Baltes et al., 1994; Hussey & Novick, 2012).

## **Conclusion**

The data presented in the current study yield important insights into the nature of the verbal WM system in older adulthood, as well as the degree to which language comprehension is plastic and dependent upon WM resources in adulthood. Specifically, the results suggested that verbal WM is capable of short-term change in adults through less than 10 hours of home-based training over the course of 3 weeks, and that this training transfers to untrained verbal memory measures, as well as measures of language fluency, memory, and comprehension. However, the system underlying on-line language interpretation was not modulated by improvements in complex verbal WM. Caplan and colleagues' SLIR model appears to be most

consistent with the current data, though future work will need to establish the reliability of eye-tracking as an assessment measure in training studies in order to conclusively rule out the possibility of transfer of WM training to eye-movement control during reading. In summary, the findings from the current suggest the presence of WM plasticity in aging and are among the first to indicate that selective aspects of language performance can be modified through targeted practice in working memory.

## References

- Anderson, J. R. (1983). *The Architecture of Cognition*. Cambridge, MA: Harvard University Press.
- Baayen, R.H., Davidson, D.J. and Bates, D.M. (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390-412.
- Bachman, L. (1989). Varieties of memory compensation by older adults in episodic remembering. In L. Poon, D. Rubin, & B. Wilson (Eds.). *Everyday cognition in adulthood and late life*. (pp. 509-544). New York: Cambridge Press.
- Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, 63, 1-29.
- Baddeley, A., & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.) *The psychology of learning and motivation: Advances in research and theory*, 8, 47-89. New York: Academic Press.
- Ball, K., Berch, D.B., Helmers, K.F., Jobe, J.B., Leveck, M.D., Marsiske, M., Morris, J.N., Rebok, G.W., Smith, D.M., Tennstedt, S.L., Unverzagt, F.W., & Willis, S.L. (2002). Effects of cognitive training interventions with older adults: A randomized controlled trial. *Journal of the American Medical Association (JAMA)*, 288, 2271-2281.
- Balota, D.A., Yap, M.J., Cortese, M.J., Hutchison, K.I., Kessler, B., Loftis, B., Neely, J.H., Nelson, D.L., Simpson, G.B., Treiman, R. (2007). The English lexicon project. *Behavior Research Methods*. 39, 445-459.
- Balogh, J., Zurif, E., Prather, P., Swinney, D., & Finkel, L. (1998). Gap-filling and end of sentence effects in real-time language processing: Implications for modeling sentence comprehension in aphasia. *Brain and Language*, 61, 169–182.

- Baltes, P. B., & Kliegl, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology*, 28(1), 121.
- Baltes, P.B. & Willis, S.L. (1982). Enhancement (plasticity) of intellectual functioning in old age: Penn State's Adult Development and Enrichment Project (ADEPT). In F.I.M Craik & S.E. Trehub (Eds.) *Aging and Cognitive Processes* (.353-389). New York: Plenum Press.
- Bartek, B., Lewis, R. L., Vasishth, S., and Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(5):1178-1198.
- Beier, M.E. & Ackerman, P.L. (2005). Age, ability and the role of prior knowledge on the acquisition of new domain knowledge. *Psychology & Aging*, 20, 341-355.
- Benton, A. L., & Hamsher, K.S. (1978). *Multilingual Aphasia Examination* (Rev. ed.). Iowa City: Department of Neurology, University of Iowa Hospitals.
- Bever, T.G (1970). The cognitive basis for linguistic structures. In: J.R. Hayes, Editor, *Cognition and the Development of Language*, Wiley, New York (1970), pp. 279–362.
- Bissig, D., & Lustig, C. (2007). Who benefits from memory training? *Psychological Science*, 18(8), 720-726.
- Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The pervasive problem with placebos in psychology: Why active control groups are not sufficient to rule out placebo effects. *Perspectives on Psychological Science*, 8, 445-454.
- Bopp, K. L., & Verhaeghen, P. (2005). Aging and verbal memory span: A meta-analysis. *Journal of Gerontology: Psychological Sciences*, 60, 223–233.



- Borella, E., Carretti, B., Riboldi, F., De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging, 25*, 767-778.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433-436.
- Brehmer, Y., Riekmann, A., Bellander, M., Westerberg, H., Fischer, H., & Backman, L. (2011). Neural correlates of training-related working memory gains in old age. *NeuroImage, 58*, 1110–1120. doi:10.1016/j.neuroimage.2011.06.079
- Brinley, J. F. (1965). Cognitive sets, speed, and accuracy of performance in the elderly. In A. T. Welford & J. E. Birren (Eds.), *Behavior, Aging, and the Nervous System* (pp. 114–149). Springfield, IL: Charles C. Thomas.
- Brown, C, Snodgrass, T, Kemper, S, Herman, R, & Covington, M,A., (2008). Automatic measurement of propositional idea density from part-of-speech tagging. *Behavior Research Methods, 40*, 540-545.
- Brown, C., May, R., Nyman, J., & Palmer, E. (2012). Effects of friend vs. foe discrimination training in action video games. *Paper presentated at the Human Factors and Ergonomic Society Annual Meeting*. Boston, Massachusetts.
- Burke, D. M., & Shafto, M. A. (2008). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (3rd ed.) (pp. 373–443). New York:Psychology Press.
- Burke, D. M. (1997). Language, aging and inhibitory deficits: Evaluation of a theory. *Journal of Gerontology: Psychological Sciences, 52B*, P254-P264.
- Buschkuehl, M., Jaeggi, S. M., Hutchison, S., Perrig-Chiello, P., Dapp, C., Muller, M.,... Perrig, W. J. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology and Aging, 23*, 743–753. doi:10.1037/a0014342.

- Caplan, D. & Waters, G.S. (2013). Memory mechanisms supporting syntactic comprehension. *Psychonomic Bulletin and Review*, 20, 243-268.
- Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, 22, 77-94.
- Caplan, D., DeDe, G., Waters, G., Michaud, J., & Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychology and Aging*, 26, 439-450.
- Caplan, D., Waters, G., & DeDe, G. (2007). Specialized verbal working memory for language comprehension. In A.R. Conway, C. Jarrold, M.J. Kane, A. Miyake, & N. Towse, *Variation in Working Memory*, Oxford.
- Carretti, B., Borella, E., Zavagnin, M., & Beni, R. (2013). Gains in language comprehension relating to working memory training in healthy older adults. *International Journal of Geriatric Psychiatry*, 28(5), 539-546.
- Carretti, B., Borella, E., Fostinelli, S., & Zavagnin, M. (2013). Benefits of training working memory in amnesic mild cognitive impairment: Specific and transfer effects. *International Psychogeriatrics*, 25, 617-626.
- Cassidy, S. M., Robertson, I. H., & O'Connell, R. G. (2012). Retest reliability of event-related potentials: Evidence from a variety of paradigms. *Psychophysiology*, 49, 659-664.
- Chein, J. & Morrison, A. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17, 193-199.
- Chen, E., Gibson, E. & Wolf, F. (2005). Online syntactic storage costs in sentence comprehension. *Journal of Memory and Language*, 52, 144-169.

- Clifton, C., Traxler, M.J., Mohamed, M.T., Williams, R.S., Morris, R.K., & Rayner, K. (2003). The use of thematic role information in parsing: Syntactic processing autonomy revisited. *Journal of Memory and Language, 49*, 317-334.
- Clifton, C., Staub, A., & Rayner, K. (2007). Eye movements in reading words and sentences. *Eye Movements: A window on Mind and Brain*, 341-372.
- Coleman, C., Lindstrom, J., Nelson, J., Lindstrom, W., & Noel, G.K. (2010). Passageless comprehension on the Nelson-Denny reading test: Well above chance for university students. *Journal of Learning Disabilities, 43*, 244-249.
- Conway, A.R., Kane, M.J., Bunting, M.F., Hambrick, D.Z., Wilhelm, O., & Engle, R.W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12*, 769-786.
- Conway, A.R., Jarrold, C., Kane, M.J., Miyake, A., & Towse, N. (2007). *Variation in Working Memory*. Oxford: Oxford University Press.
- Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87-185.
- Dahlin, E., Neely, A.S., Larsson, A., Backman, L. & Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science, 320*, 1510-1512.
- Daneman, M. & Merikle, P.M. (1996). Working memory and comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3*, 422-433.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*, 450-466.
- DeDe, G., Caplan, D., Kemtes, K., & Waters, G. (2004). The relationship between age, verbal working memory and language comprehension. *Psychology and Aging, 19*, 601-616.

- Dufau, S., Duñabeitia, J. A., Moret-Tatay, C., McGonigal, A., Peeters, D., Alario, F. X., ... & Grainger, J. (2011). Smart phone, smart science: how the use of smartphones can revolutionize research in cognitive science. *PloS One*, *6*, e24974.
- Engle, R.W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*.
- Engle, R.W. (2010). Role of working-memory capacity in cognitive control. *Current Anthropology*, *51*, 17-26.
- Farmer, T. A., Christiansen, M. H., & Kemtes, K. A. (2005). Sentence processing in context: The impact of experience on individual differences. In *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 642-647).
- Federmeier, K.D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology*, *44*, 491–505.
- Federmeier, K. D., Kutas, M., & Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain and Language*, *115*(3), 149-161.
- Federmeier, K. D., McLennan, D. B., Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, *39*(2), 133-146.
- Fedorenko, E., Gibson, E., & Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources. *Journal of Memory and Language*, *54*, 541-553.
- Felser, C., Marinis, T., & Clahsen, H. (2003). Children's processing of ambiguous sentences: A study of relative clause attachment. *Language Acquisition: A Journal of Developmental Linguistics*, *11*, 127–163.

- Ferguson, A., Spencer, E., Craig, H., & Colyvas, K. (2013). Propositional idea density in women's written language over the lifespan: Computerized analysis. *Cortex*.
- Ferreira, F., & Clifton, C. (1986). The independence of syntactic processing. *Journal of Memory and Language*, 25, 348-368.
- Ferreira, F., Henderson, J. M., Anes, M. D., Weeks, P. A., Jr. & McFarlane, D. K. (1996) Effects of lexical frequency and syntactic complexity in spoken language comprehension: Evidence from the auditory moving window technique. *Journal of Experimental Psychology: Learning, Memory and Cognition* 22:324–35.
- Fiebach CJ, Schlesewsky M and Friederici AD (2002). Separating syntactic memory costs and syntactic integration costs during parsing: The processing of German WH-questions. *Journal of Memory and Language*, 47, 250-272.
- Fiebach, C. J., Schlesewsky, M., & Friederici, A. D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, 30, 321-338.
- Fine, A. B., Jaeger, T. F., Farmer, T. A., & Qian, T. (2013). Rapid expectation adaptation during syntactic comprehension. *PloS one*, 8(10), e77661.
- Foster, J. C., & Taylor, G. A. (1920). The applicability of mental tests to persons over fifty years of age. *Journal of Applied Psychology*, 4, 39.
- Fraundorf, S. H., Diaz, M. I., Finley, J. R., Lewis, M. L., Tooley, K. M., Isaacs, A. M., Lam, T. Q., Trude, A. M., Brown-Schmidt, S., & Brehm, L. (2014). CogToolbox for MATLAB [computer software]. Available from <http://www.scottfraundorf.com/cogtoolbox.html>
- Friedman, N.P. & Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory and Language*, 51, 136-158.

- Gennari, S. P., Mirkovic, J., & MacDonald, M. C. (2012). Animacy and competition in relative clause production: A cross-linguistic investigation. *Cognitive Psychology*, *65*, 141-176.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1-76.
- Gibson, E.A. (2000) The dependency locality theory: a distance-based theory of linguistic complexity. In *Image, Language, Brain* (Miyashita, Y. et al., eds), MIT Press.
- Goh, J. O. S., An, Y., Resnick, S. M. (2012). Differential trajectories of age-related changes in components of executive and memory processes. *Psychology and Aging*, *27*, 707-719.
- Gordon, P.C., & Lowder, M.W. (2012). Complex sentence processing: A review of theoretical perspectives on the comprehension of relative clauses. *Language and Linguistics Compass*, *6/7*, 403-415.
- Gordon, P.C., Hendrick, R., & Levine, W.H. (2002). Memory-load interference in syntactic processing. *Psychological Science*, *13*, 425-430.
- Gordon, P.C., Hendrick, R., Johnson, M., & Lee, Y. (2006). Similarity-based interference during language comprehension: Evidence from eye tracking during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *32*, 1304-1321.
- Grodner, D., & Gibson, E. (2005). Consequences of the serial nature of linguistic input. *Cognitive Science*, *29*, 261-291.
- Haberlandt, K., & Graesser, A. C. (1989). Processing new arguments at clause boundaries. *Memory & Cognition*, *17*, 186-193.
- Henrich, J., Heine, S.J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 1-75.

- Hertzog, C., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. S. (2003). Latent change models of adult cognition: Are changes in processing speed and working memory associated with changes in episodic memory? *Psychology and Aging, 18*, 755-769.
- Hollis, S., & Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ, 319*(7211), 670-674.
- Ide, N., et al. (2013). *Manually Annotated Sub-Corpus Third Release*. Linguistic Data Consortium, Philadelphia.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the USA, 108*, 10081–10086. doi:10.1073/pnas.1103228108
- Jaeggi, S. M., Studer-Luethi, B., Buschkuhl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning: Implications for training and transfer. *Intelligence, 38*, 625–635. doi:10.1016/j.intell.2010.09.001
- Jaeggi, S.M., Buschkuhl, M., Jonides, J., & Perrig, W.J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*, 6829-68f.
- Jaeggi, S.M., Buschkuhl, Perrig, W.J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory, 18*, 394-412.
- Just, M. A., & Carpenter, P. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329–354.
- Just, M. A., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*, 122-149.

- Just, M. A., & Varma, S. (2002). A hybrid architecture for working memory: Reply to MacDonald and Christiansen (2002). *Psychological Review*, *109*, 55-65.
- Just, M., Carpenter, P. A, Keller, T.A, Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, *274*, 114-6.
- Just, M.A. & Varma, S. (2007). The organization of thinking: What functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cognitive, Affective, and Behavioral Neuroscience*, *7*, 153-191.
- Kaan, E., Harris, T., Gibson, E., & Holcomb, P. J. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159- 201.
- Kane, M. J., Conway, A. R. A., Hambrick, D. Z., & Engle, R. W. (2007). Variation in working memory capacity as variation in executive attention and control. In A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, J. Towse (Eds.), *Variation in Working Memory*. Oxford University Press.
- Karbach, J. & Verhaeghen, P. (2014). Making working memory work: A meta analysis of executive control and working memory training in older adults. *Poster presented at the 2014 Cognitive Aging Conference*, Atlanta, GA.
- Karlsson, F. (2007). Constraints on multiple center-embedding of clauses. *Journal of Linguistics* *43* (2): 365-392.
- Kemper, S. (1986). Imitation of complex syntactic constructions by elderly adults. *Applied Psycholinguistics*, *7*, 277-288.
- Kemper, S. (1987). Syntactic complexity and elderly adults' prose recall. *Experimental Aging Research*, *13*, 47-52.



- Kemper, S. (1992). Language and aging. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 213–270). Hillsdale, NJ: Erlbaum.
- Kemper, S. (2012). The interaction of linguistic constraints, working memory, and aging on language production and comprehension. In M.N. Benjamin & N. Ohata., *Memory and Aging: Current Issues and Further Directions*.
- Kemper, S., & Herman, R. E. (2006). Age differences in memory-load interference effects in syntactic processing. *Journal of Gerontology: Psychological Sciences*, *61*, P327-P332.
- Kemper, S., & Liu, C. (2007). Eye movements of young and older adults during reading. *Psychology and Aging*, *22*, 84-93.
- Kemper, S., & McDowd, J. (2008). Dimensions of Cognitive Aging: Executive Function and Verbal Fluency. In S. M. Hofer and D. F. Alwin (Eds.), *Handbook of Cognitive Aging: Interdisciplinary perspectives*. Thousand Oaks, CA: Sage.
- Kemper, S., Crow, A., & Kemtes, K. (2004). Eye-fixation patterns of high- and low-span young and older adults: Down the garden path and back again. *Psychology and Aging*, *19*, 157-170.
- Kemtes, K. A. & Kemper, S. (1997). Younger and older adults' on-line processing of syntactically ambiguous sentences. *Psychology and Aging*, *12*, 362-371.
- King, J. W., & Just, M. A. (1991). Individual differences in syntactic parsing: The role of working memory. *Journal of Memory and Language*, *30*, 580–602. doi:10.1016/0749-596X(91)90027- H.
- King, J. W., & Kutas, M. (1995). Who did what and when? Using word- and clause related ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*, 378-397.

- Kintsch, W. and Keenan, J. (1973). Reading rate and retention as a function of the number of propositions in the base structure of sentences. *Cognitive Psychology*, 5, 257-274.
- Kintsch, W. and Van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363-394.
- Kirby, K. N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, 45, 905-927.
- Klingberg, T., Fernell, E., Olesen, P., Johnson, M., Gustafsson, P., Dahlström, K...Westerberg, H. (2005). Computerized training of working memory in children with ADHD – a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 177-186.
- Kluender, R., & Kutas, M. (1993a). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, 5, 196-214.
- Kramer, A. F., & Willis, S. L. (2002). Enhancing the cognitive vitality of older adults. *Current Directions in Psychological Science*, 11, 173-177.
- Krushke, J. (2011). *Doing Bayesian Data Analysis: A Tutorial in R and BUGS*.
- Kutas, M. Van Petten, C.K., & Kluender, R. (2006). *Psycholinguistics Electrified II: 1994-2005*. In Handbook of Psycholinguistics, ed. M. Traxler & M.A., Gernsbacher, 659–724. New York: Elsevier.
- Kutas, M. & King, J. (1999). In-line measures of syntactic processing using event-related brain potentials. *Behavioral and Brain Sciences*, 22, 104–105.
- Kwong See, S. T., & Ryan, E. B. (1995). Cognitive mediation of adult age differences in language performance. *Psychology and Aging*, 10, 458– 468.

- Larner A.J. (2011). Screening utility of the Montreal Cognitive Assessment (MoCA): In place of—or as well as—the MMSE? *International Psychogeriatrics* 3, 1–6.
- Lee J.Y., Lee D.W., Cho S.J., Na D.L., et al. (2008). Brief screening for mild cognitive impairment in elderly outpatient clinic: Validation of the Korean version of the Montreal Cognitive Assessment. *Journal of Geriatric Psychiatry and Neurology*, 21, 104–110.
- Lee, C, and Federmeier, K. D. (2011). Differential age effects on lexical ambiguity resolution mechanisms. *Psychophysiology*, 48, 960-972.
- Li, S.-C., Schmiedek, F., Huxhold, O., Roßcke, C., Smith, J., & Lindenberger, U. (2008). Working memory plasticity in old age: Practice gain, transfer, and maintenance. *Psychology and Aging*, 23, 731-742.
- Luis C.A., Keegan A.P., & Mullan M. (2009). Cross validation of the Montreal Cognitive Assessment in community dwelling older adults residing in the Southeastern US. *International Journal of Geriatric Psychiatry*, 24, 197–201.
- Lustig, C., Shah, P., Seidler, R. D., & Reuter-Lorenz, P. (2009). Aging, training, and the brain: A review and future directions. *Neuropsychology Review, special issue on Plasticity in the Aging Brain*, 4, 504-522
- MacDonald, M. C., & Christiansen, M. C. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychological Review*, 109, 35– 54.
- MacDonald, M., Just, M., & Carpenter, P. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognitive Psychology*, 24, 56-98.

- Mahncke, H. W., Bronstone, A., & Merzenich, M. M. (2006). Brain plasticity and functional losses in the aged: Scientific bases for a novel intervention. *Progress in Brain Research*, *157*, 81-109.
- Mahncke, H., Connor, B., Appelman, J., Ahsanuddin, O., Hardy, J., Wood, R., . . . Merzenich, M. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. *Proceedings of the National Academy of Sciences, USA of the United States of America*, *103*, 12523–12528.
- Manly, J.J., Touradji, P., Tang, M-X., & Stern, Y. (2003). Literacy and memory decline among ethnically diverse elders. *Journal of Clinical and Experimental Neuropsychology*, *25*, 680- 690.
- Margrett, J. A., & Willis, S. L. (2006). In-home cognitive training with older married couples: Individual versus collaborative learning. *Aging, Neuropsychology and Cognition*, *13*, 173–195.
- Martin, R. & Feher, E. (1990) The consequences of reduced memory span for comprehension of semantic versus syntactic information. *Brain and Language*, *38*,1–20.
- Mayr, U., & Kliegl, R. (2000). Complex semantic processing in old age: Does it stay or does it go?. *Psychology and Aging*, *15*(1), 29.
- McDowd, J., Hoffman, L., Rozek, E., Lyons, K. E., Pahwa, R., Burns, J., & Kemper, S. (2011). Understanding verbal fluency in healthy aging, Alzheimer's disease, and Parkinson's disease. *Neuropsychology*, *25*, 210.
- McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, *11*, 103–107. doi:10.1038/nn2024

- Melby-Verlag, M & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*, 270-291.
- Miller, G.A. and Chomsky, N. (1963) Finitary models of language users. In *Handbook of Mathematical Psychology* (Vol.II) (Luce, D.R. et al., eds), John Wiley
- Morrison, A.B. & Chein, J.M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*, 46-60.
- Muller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials to relative clause processing in spoken sentences. *Cognitive Brain Research*, *5*, 193-203.
- Munte, T. F., Heinze, H. -J., Matzke, M., Wieringa, B. M., & Johannes, S. (1998). Brain potentials and syntactic violations revisited: No evidence for specificity of the syntactic positive shift. *Neuropsychologia*, *36*, 217-226.
- Nagels, A., Kircher, T., Dietsche, B., Backes, H., Marquetand, J., & Krug, A. (2012). Neural processing of overt word generation in healthy individuals: The effect of age and word knowledge. *NeuroImage*, *61*, 832-840.
- Nasreddine, Z.S., Phillips, N.A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., Cummings, J.L., & Chertkow, H. (2005). The Montreal cognitive assessment (MoCA): A brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*, 695-699.
- Navon, D. (1984) Resources – A theoretical soup stone. *Psychological Review* *91*, 216-234.
- Novick, J., Hussey, E.K., Teubner-Rhodes, S., Harbison, J.I., & Bunting, M. (2013). Clearing the garden path: Improving sentence processing through executive control training. *Language and Cognitive Processes*. doi: 10.1080/01690965.2012.758297

- Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*, 75–79. doi:10.1038/nn1165.
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N., Smith, A. D., & Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult lifespan. *Psychology and Aging*, *17*, 299–320.
- Patterson, J. V., Pratt, H., & Starr, A. (1991). Event-related potential correlates of the serial position effect in short-term memory. *Electroencephalography and Clinical Neurophysiology*, *78*, 424–437.
- Payne, B.R. & Stine-Morrow, E.A.L (2012). Aging, parafoveal preview, and semantic integration in sentence processing: Testing the cognitive workload of wrap-up. *Psychology and Aging*, *27*, 638-649. doi: 10.1037/a0026540.
- Payne, B.R. & Stine-Morrow, E.A.L. (in press). Risk for mild cognitive impairment is associated with semantic integration deficits in sentence processing. *Journal of Gerontology: Psychological Sciences*.
- Payne, B.R., Gao, X., Noh, S.R., Anderson, C.J., & Stine-Morrow, E.A.L. (2012). The effects of print exposure on sentence processing and memory among older adults: Evidence for efficiency and reserve. *Aging, Neuropsychology, and Cognition*, *19*, 122-149. doi: 10.1080/13825585.2011.628376.
- Payne, B.R., Grison, S., Gao, X., Christianson, K., Morrow, D.G., & Stine-Morrow, E.A.L. (2014). Aging and individual differences in binding during sentence comprehension: Evidence from temporary and global syntactic attachment ambiguities. *Cognition*.

- Payne, B.R., Gross, A., Parisi, J., Marsiske, M., Rebok, G., & Stine-Morrow, E.A.L. (2014; in press). Modeling longitudinal changes in older adults' memory for spoken discourse: Findings from the ACTIVE cohort. *Memory*.
- Payne, B.R., Jackson, J.J, Hill, P.L., Gao, X., Roberts, B.W., & Stine-Morrow, E.A.L. (2012). Memory self-efficacy predicts responsiveness to inductive reasoning training in older adults. *Journal of Gerontology: Psychological Sciences*, 67, 27-35. doi: 10.1093/geronb/gbr073.
- Perfetti, C.A., & Lesgold, A.M (1977). Discourse comprehension and sources of individual differences. In M. A. Just & P. A. Carpenter (Eds.), *Cognitive processes in comprehension* (pp. 141–183). Hillsdale, NJ: Erlbaum.
- Phillips C, Kazanina N, Abada SH (2005). ERP effects of the processing of syntactic long-distance dependencies. *Cognitive Brain Research*, 22, 407- 428.
- Pickering, M.J. & van Gompel, R.P.G. (2006). Syntactic parsing. In M.J Traxler & M. Gernsbacher. (Eds.), *Handbook of Psycholinguistics*, 2<sup>nd</sup> Ed. (pp. 455-503), Amsterdam: Academic Press.
- Radvansky, G. A. (1999). Memory retrieval and suppression: The inhibition of situation models. *Journal of Experimental Psychology: General*, 128, 563.
- Radvansky, G. A., & Dijkstra, K. (2007). Aging and situation model processing. *Psychonomic Bulletin & Review*, 14, 1027-1042.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372 – 422.

- Rayner, K., Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing: Eye movements in the analysis of semantically biased sentences. *Journal of Verbal Learning and Verbal Behavior*, 22, 358-374.
- Rebok, G.W. (2008). *Cognitive training: Influence on neuropsychological and brain function in later life*. State-of-Science Review: SR:E22. UK Government Foresight Mental Capital and Mental Wellbeing Project. Government Office for Science.
- Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H. Y., King, J. W., ... & Willis, S. L. (2014). Ten-year effects of the Advanced Cognitive Training for Independent and Vital Elderly cognitive training trial on cognition and everyday functioning in older adults. *Journal of the American Geriatrics Society*, 62, 16-24.
- Richmond, L. L., Morrison, A., Chein, J., Olson, I. R. (2011). Working memory training and transfer in older adults, *Psychology and Aging*. Online Advance Access.
- Roberts, R., & Gibson, E. (2002). Individual differences in sentence memory. *Journal of Psycholinguistic Research*, 31, 573–598.
- Rogers, T. T. & McClelland, J. L. (2004). *Semantic Cognition: A Parallel Distributed Processing Approach*. Cambridge, MA: MIT Press.
- Rogers, T. T. & McClelland, J. L. (2004). Precis of semantic cognition: A parallel distributed processing approach. *Behavioral and Brain Sciences*, 31, 689-749.
- Rossetti H.C., Lacritz L.H., Cullum C.M., Weiner M.F. (2011). Normative data for the Montreal Cognitive Assessment (MoCA) in a population-based sample. *Neurology*, 29, 261-291.
- Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review*, 103, 403–428.



- Schmiedek, F., Lovden, M., & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: Findings from the COGITO study. *Frontiers in Aging Neuroscience*, 2, 1–10. doi:10.3389/fnagi.2010.00027
- Schotter, E. R., Tran, R., Rayner, K., & Tran, R. (2014). Don't believe what you read (only once): Comprehension is supported by regressions during reading.
- Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*. Advance online Access.
- Shiran, A., & Breznitz, Z. (2011). The effect of cognitive training on recall range and speed of information processing in the working memory of dyslexic and skilled readers. *Journal of Neurolinguistics*, 24, 524–537.
- Singer, J. D. and Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Methods for Studying Change and Event Occurrence*. (New York: Oxford University Press)
- Slattery, T. J., Sturt, P., Christianson, K., Yoshida, M., Ferreira, F. (2013). Lingering misinterpretations of garden path sentences arise from competing syntactic representations. *Journal of Memory and Language*, 69, 104–120.
- Snowdon, D., Kemper, S., Mortimer, J., Greiner, L. H., Wekstein, D. R., and Markesbery, W. R. (1996). Linguistic ability in early life and cognitive function and Alzheimer's disease in late life. *Journal of the American Medical Association (JAMA)*, 275: 528–532.
- Stanovich, K. E., West, R. L., & Harrison, M. R. (1995). Knowledge growth and maintenance across the life span: The role of print exposure. *Developmental Psychology*, 31, 811–826.
- Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, 116, 71–86

- Stern, Y. (2009). Cognitive reserve. *Neuropsychologia*, *47*, 2015-2028. doi: 10.1016/j.neuropsychologia.2009.03.004.
- Stine, E. A., & Wingfield, A. (1990). How much do working memory deficits contribute to age differences in discourse memory? *European Journal of Cognitive Psychology*, *2*, 289-304.
- Stine-Morrow, E. A. L., Miller, L. M. S., Gagne, D. D., & Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychology and Aging*, *23*, 131-153.
- Stine-Morrow, E.A.L., & Basak, C. (2011). *Cognitive interventions*. In K.W. Schaie & S.L. Willis (Eds.) *Handbook of the Psychology of Aging* (7th ed.). New York: Elsevier.
- Stine-Morrow, E.A.L., & Miller, L.M.S. (2009). Aging, self-regulation, and learning from text. *Psychology of Learning and Motivation*, *51*, 255-285.
- Stine-Morrow, E.A.L., Payne, B.R., Hill, P.L., Jackson, J.J., Roberts, B.W., Kramer, A.F., Morrow, D.G., Payne, L., Janke, M., Noh, S.R., & Parisi, J.M. (in preparation). Training versus engagement as paths to cognitive optimization with aging.
- Stine, E.A.L. (1990). On-line processing of written text by younger and older adults. *Psychology and Aging*, *5*, 68-78.
- Stites, M. C., Federmeier, K. D., and Stine-Morrow, E. A. L. (in press). Cross-age comparisons reveal multiple strategies for lexical ambiguity resolution during natural reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.
- Taatgen, N.A. & Anderson, J.R. (2008). ACT-R. In R. Sun (ed.), *Constraints in Cognitive Architectures*. Cambridge University Press, pp 170-185.
- Thibadeau, R., Just, M. A., & Carpenter, P. A. (1982). A model of the time course and content of reading. *Cognitive Science*, *6*, 157-203.

- Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*, 204–224.
- Traxler, M.J. (2009). A hierarchical linear modeling analysis of working memory and implicit prosody in the resolution of adjunct attachment ambiguity. *Journal of Psycholinguistic Research*, *38*, 491-509.
- Traxler, M. J. (2007). Working memory contributions to relative clause attachment processing: A hierarchical linear modeling analysis. *Memory & Cognition*, *35*, 1107–1121.
- Trueswell, J.C., Tanenhaus, M.K., & Garnsey, S.M. (1994). Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution. *Journal of Memory and Language*, *33*, 285-318.
- Troyer, A. K. (2000). Normative data for clustering and switching on verbal fluency tasks. *Journal of Clinical and Experimental Neuropsychology*, *22*(3), 370-378.
- Tse, C-S., Balota, D.A., Yap, M.J., Duchek, J.M., & McCabe, D.P. (2010). Effects of healthy aging and early stage dementia of the Alzheimer's type on components of response time distributions in three attention tasks. *Neuropsychology*, *24*, 300-315.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, *28*, 127-154.
- Turner, A. & Greene, E. (1977). *The Construction and Use of a Propositional Text Base*. Boulder, CO: University of Colorado Psychology Department, The Institute of Cognitive Science, University of Colorado Boulder, (<http://www.colorado.edu/ics/sites/default/files/attached-files/>), 77-63.
- Uttal, W. R. (2013). *Reliability in Cognitive Neuroscience: A Meta-meta-analysis*. MIT Press.

- Van der Linden, M., Hupet, M., Feyereisen, P., Schelstraete, M., Bestgen, Y, Bruyer, M.R. et al., (1999). Cognitive mediators of age-related differences in language comprehension and verbal memory performance. *Aging, Neuropsychology, and Cognition*, 6, 11-24.
- Van Dyke, J.A. & Lewis, R.L. (2003). Distinguishing effects of structure and decay on attachment and repair: A retrieval interference theory of recovery from misanalyzed ambiguities. *Journal of Memory and Language*, 49(3), 285-316.
- Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, 50, 289-335.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer.
- Verhaeghen, P. (2011). Aging and executive control: Reports of a demise greatly exaggerated. *Current Directions in Psychological Science*, 20, 174-180.
- Verhaeghen, P. (2012). Working memory still working: Age-related differences in working memory functioning and cognitive control. In M. Naveh Benjamin and N. Ohta (Eds.) *Memory and aging: Current issues and future directions* (pp. 1-30). New York, NY: Psychology Press
- Verhaeghen, P., & Cerella, J. (2002). Aging, executive control, and attention: A review of meta-analyses. *Neuroscience and Biobehavioral Reviews*, 26, 849-857 .
- Verhaeghen, P., & Salthouse, T. A. (1997). Meta-analyses of age-cognition relations in adulthood: Estimates of linear and non-linear age effects and structural models. *Psychological Bulletin*, 122, 231-249.

- Wadley, V.G., Benz, R.L., Ball, K.K., Roenker, D.L., Edwards, J.D., & Vance, D.E. (2006). Development and evaluation of home based speed of processing training for older adults. *Archive of Physical and Medical Rehabilitation, 87*, 757-763.
- Waldron-Perrine, B., & Axelrod, B.N. (2012). Determining an appropriate cut score for indication of impairment on the Montreal Cognitive Assessment. *International Journal of Geriatric Psychiatry, 27*, 1189-1194.
- Walton, C. C., Mowszowski, L., Lewis, S. J., & Naismith, S. L. (2014). Stuck in the mud: time for change in the implementation of cognitive training research in ageing? *Frontiers in Aging Neuroscience, 6*.
- Was, C. A., Rawson, K. A., Bailey, H., & Dunlosky, J. (2011). Content-embedded tasks beat complex span for predicting comprehension. *Behavior Research Methods*.
- Wasylyshyn, C., Verhaeghen, P., & Sliwinski, M. J. (2011). Aging and task switching: A meta-analysis. *Psychology and Aging, 26*, 15–20.
- Waters, G.S. & Caplan, D. (1996). The capacity theory of sentence comprehension: Critique of Just and Carpenter (1992). *Psychological Review, 103*, 761-772.
- Waters, G.S. & Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behavior Research Methods, Instruments, and Computers, 35*, 550-564.
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., & MacDonald, M. C. (2009). Experience and sentence comprehension: Statistical learning and relative clause comprehension. *Cognitive Psychology, 58*, 250-271.
- Willis, S. L., Blieszner, R., & Baltes, P.B. (1981). Intellectual training research in aging: Modification of performance on the fluid ability of figural relations. *Journal of Educational Psychology, 73*, 41-50.

- Willis, S. L., & Nesselroade, C. S. (1990). Long-term effects of fluid ability training in old-old age. *Developmental Psychology, 26*, 905-910.
- Wilson, B.A., Cockburn, J., & Baddeley, A.D. (1985). *The Rivermead Behavioural Memory Test*. London. Pearson Assessment.
- Wilson, B.A., Cockburn, J., & Baddeley, A.D. (2003). *The Rivermead Behavioural Memory Test – Second Edition*. London. Pearson Assessment.
- Wingfield, A., & Stine-Morrow, E.A.L. (2000). Language and Speech. In F.I.M. Craik & T.A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (2nd ed., pp. 359-416). Mahwah, NJ; Erlbaum.
- Wlotko, E.W., Lee, C-L., & Federmeier, K.D. (2010). Language of the aging brain: Event-related potential studies of comprehension in older adults. *Language and Linguistics Compass, 4*, 623-638.
- Yngve, V. H. (1960). A model and a hypothesis for language structure. *Proceedings of the American Philosophical Society, 104*, 444-466.
- Zelinski, E. M. & Lewis, K.L. (2003). Adult age differences in multiple cognitive functions: Differentiation, dedifferentiation, or process-specific change? *Psychology and Aging, 18*, 727-745.
- Zelinski, E. M. & Stewart, S. T. (1998). Individual differences in 16-year memory changes. *Psychology and Aging, 13*, 622-630.
- Zinke, K., Zeintl, M., Rose, N.S., Putzmann, J., Pydde, A. & Kliegel, M. (2013). Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology*. Advance online publication.

## Tables and Figures

*Table 1. Baseline Demographics in Control and Treatment Groups*

	Control ( <i>N</i> = 19)		Treatment ( <i>N</i> = 22)		diff	95% CI
	<i>M</i> or Count	<i>SD</i> or %	<i>M</i> or Count	<i>SD</i> or %		
1. Age	68.11	6.24	67.68	2.77	-.42	[-2.55, 3.41]
2. Years of Education	17.47	2.38	17.09	2.20	-.38	[-1.83, 1.07]
3. MoCA	27.21	2.39	27.77	1.93	.56	[-.81, 1.92]
4. Vocabulary	.63	.23	.68	.18	.04	[-.09, .17]
5. % Female	14	74%	16	73%	.01	[-.26, .26]

*Table 2. Correlations between baseline measures and training gains (AUC)*

	Sentence AUC	Category AUC	LD AUC	Average AUC
1. Age	.62	.58	.68	.66
2. Education	.05	-.05	-.02	-.01
3. Vocabulary	.39	.62	.58	.55
4. MoCA	.38	.20	.33	.32
5. Reading Span	.08	-.01	.12	.08
6. Listening Span	-.19	-.21	-.29	-.25
7. Operation Span	.02	-.19	-.17	-.12
8. Subtract 2 Span	.29	.02	-.01	.09
9. FAS	.38	.27	.36	.35
10. Nelson Denny	.54	.57	.61	.60



*Table 3. Change in Working Memory and Language Measures in Control and Treatment Groups*

	Control (N = 19)						Treatment (N = 22)					
	Pretest		Posttest		Δ		Pretest		Posttest		Δ	
	M	SE	M	SE	M	95% CI	M	SE	M	SE	M	95% CI
<i>Verbal Working Memory</i>												
1. Reading Span	2.74	.25	2.98	.34	.22	[-.20, .65]	2.85	.20	3.70	.22	.84	[.28, 1.40]
2. Listening Span	3.50	.37	2.88	.31	-.64	[-1.31, .03]	3.68	.16	4.40	.19	.72	[.24, 1.18]
3. Operation Span	.40	.07	.44	.06	.02	[-.04, .09]	.39	.05	.63	.05	.30	[.19, .41]
4. Minus 2 Span	.58	.05	.63	.06	.01	[-.03, .05]	.63	.04	.76	.02	.12	[.03, .21]
<i>Language Understanding</i>												
5. Nelson Denny	.87	.02	.82	.04	-.06	[-.13, .01]	.86	.02	.81	.05	-.05	[-.11, .02]
6. Verbal Fluency	41.00	2.47	42.71	2.07	.71	[-3.15, 4.56]	43.82	2.22	51.36	3.00	7.55	[3.96, 11.13]
7. Sentence Memory	.66	.04	.71	.03	.04	[-.01, .07]	.65	.02	.77	.02	.12	[.08, .16]
8. Discourse Memory	.46	.04	.44	.04	-.02	[-.86, .43]	.52	.03	.47	.03	-.05	[-.11, .01]

Table 4. Model Results for Effects of Training on Garden Path Ambiguity Accuracy

	Estimate	95% CI
Intercept	1.02	[.40, 1.65]
Amb	.58	[.02, 1.15]
Time	.09	[-.47, 0.65]
Treatment	-.19	[-.76, 0.38]
Amb x Time	.20	[-.64, 1.04]
Amb x Treatment	.68	[-.09, 1.45]
Time x Treatment	.45	[-.29, 1.20]
Amb x Time x Treatment	-.82	[-1.39, -.25]

*Table 5.* Model Results for Effects of Training on Garden Path Ambiguity in Regression Path Durations

Parameter	Estimate	95% CI
Intercept	1438	[1262, 1615]
Amb	465	[326, 604]
Treatment	-203	[-532, 127]
Time	50	[-97, 196]
Amb x Treatment	-248	[-526, 29]
Amb x Time	-6	[-283, 271]
Treatment x Time	-185	[-473, 103]
Amb x Treatment x Time	100	[-452, 653]

*Table 6.* Model Results for Effects of Training on Garden Path Ambiguity in Probability of Regressing

	Estimate	95% CI
Intercept	.02	[-.20, .24]
Amb	.28	[.10, .46]
Treatment	-.11	[-.42, .20]
Time	.11	[-.09, .31]
Amb x Treatment	-.25	[-.62, .12]
Amb x Time	-.09	[-.46, .28]
Time x Treatment	-.14	[-.51, .23]
Amb x Time x Treatment	.33	[-.41, 1.07]

*Table 7.* Model Results for Effects of Training on Long-Distance Dependency Processing in Accuracy.

	Estimate	95% CI
Intercept	.86	[.34, 1.38]
Distance	-.03	[-.54, .48]
Time	-.34	[-.86, .18]
Treatment	.06	[-.47, .59]
Distance x Time	.49	[-.23, 1.23]
Distance x Treatment	.17	[-.50, .85]
Time x Treatment	.38	[-.31, 1.07]
Distance x Time x Treatment	-.83	[-1.80, .14]

*Table 8. Model Results for Effects of Training on Long-Distance Dependency Processing in Regression Path Duration.*

	Estimate	95% CI
Intercept	718.58	[598, 838]
Distance	247.08	[122, 371]
Treatment	95.31	[-120, 311]
Time	-40.40	[-169, 88]
Distance x Treatment	69.92	[-179, 319]
Distance x Time	-285.51	[-535, -35]
Time x Treatment	-133.32	[-387, 121]
Distance x Time x Treatment	-18.44	[-516, 479]

*Table 9.* Model results for Effects of Training on Long-Distance Dependency Processing in Probability of Regressing

	Estimate	95% CI
Intercept	-1.19	[-.19, .23]
Distance	.52	[.10, .45]
Treatment	.15	[-.42, .20]
Time	.10	[-.08, .30]
Distance x Treatment	.47	[-.62, .12]
Distance x Time	-.69	[-.46, .28]
Time x Treatment	-.60	[-.51, .23]
Distance x Time x Treatment	-.31	[-.41, 1.07]

*Table 10.* Model Results for Effects of Training on Object-Relative Clause Processing in Accuracy.

	Estimate	95% CI
Intercept	.82	[.31, 1.33]
SR	.79	[.22, 1.35]
Time	.10	[-.43, .64]
Treatment	-.19	[-.82, .44]
SR x Time	-.10	[-.90, .70]
SR x Treatment	.51	[-.24, 1.26]
Time x Treatment	.33	[-.37, 1.03]
SR x Time x Treatment	-.97	[-1.95, .01]



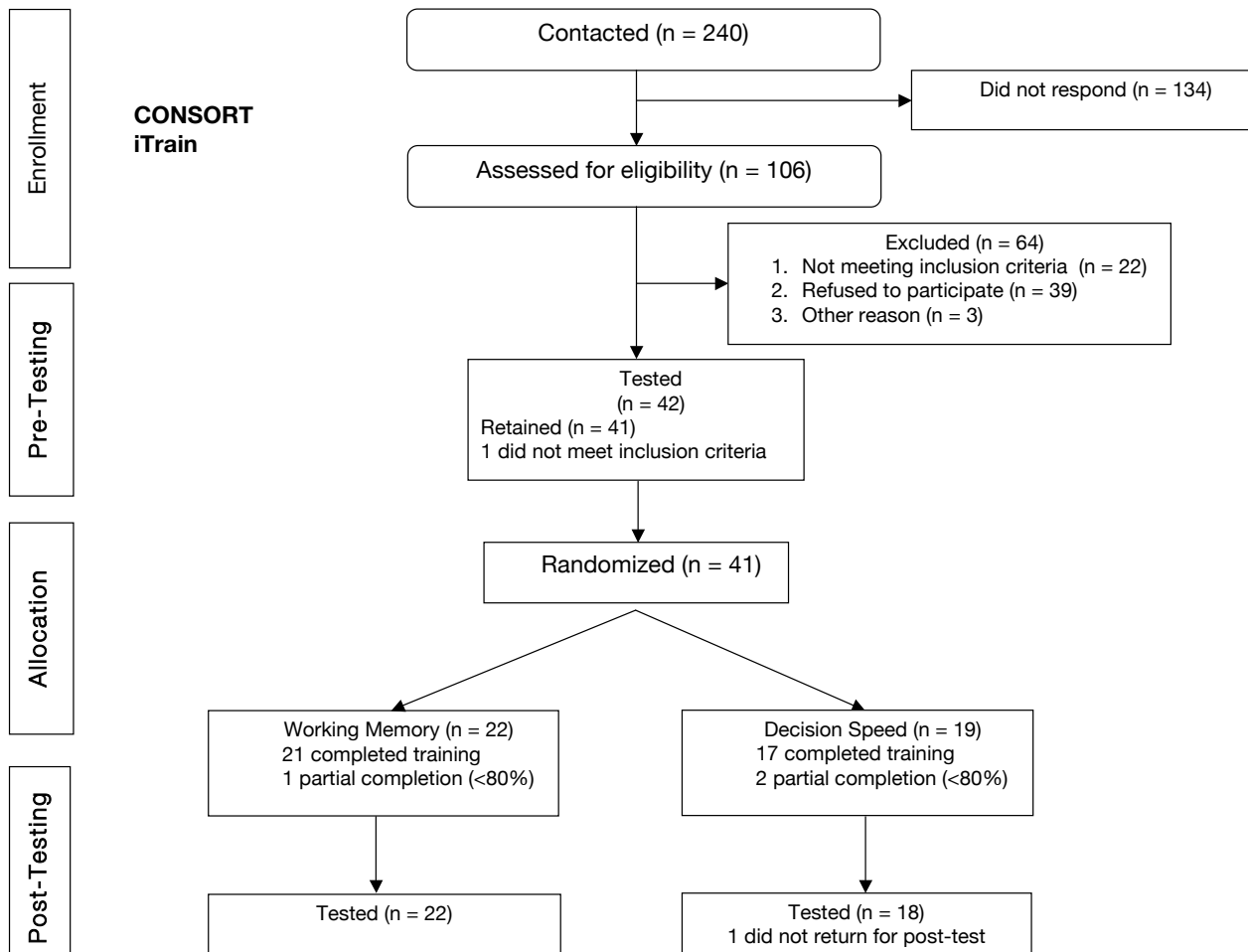


Figure 1. CONSORT Diagram for iTrain Study

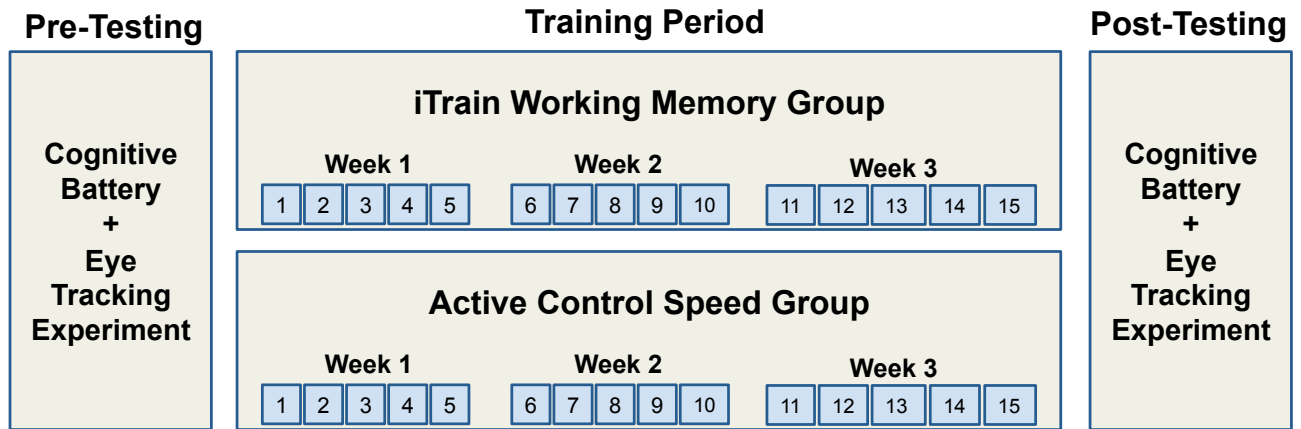


Figure 2. Diagram of Study Procedure

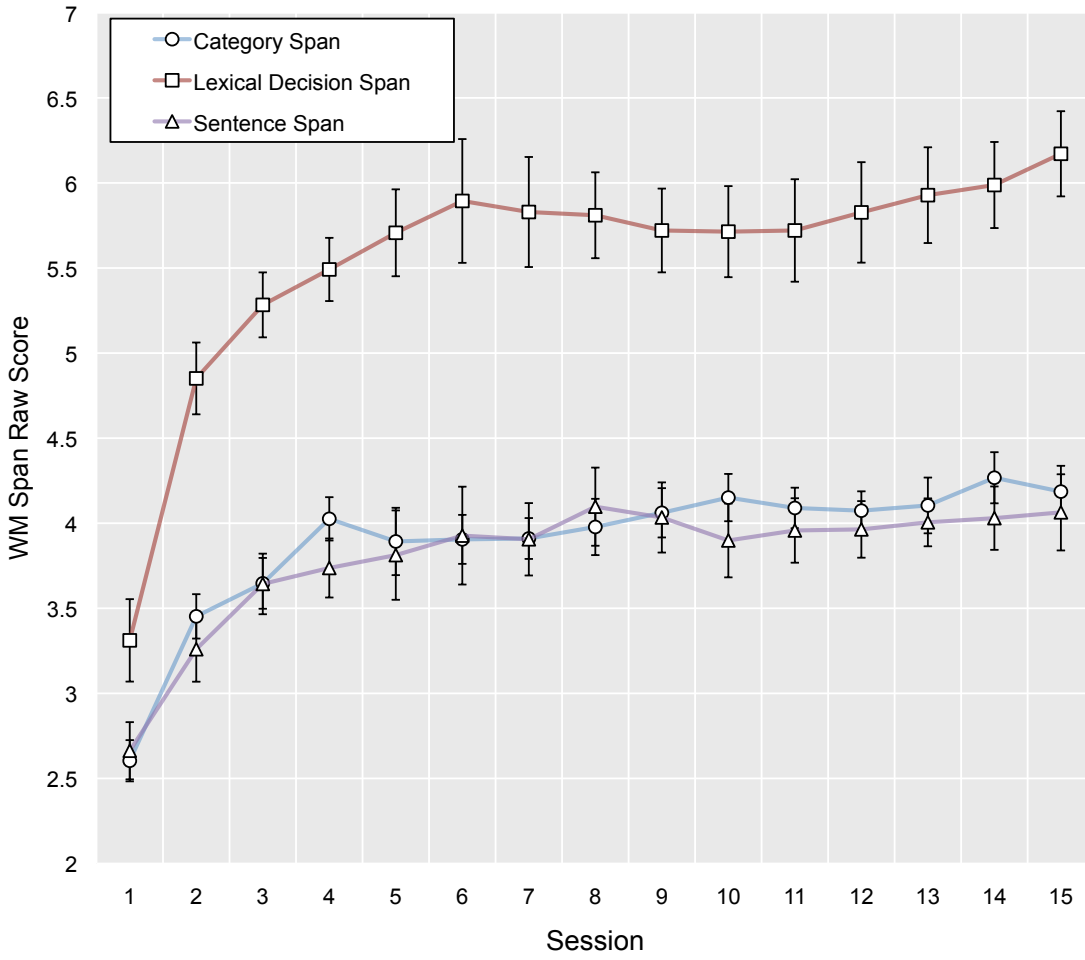


Figure 3. WM Span Raw Score Over Fifteen Weeks in WM Training Tasks

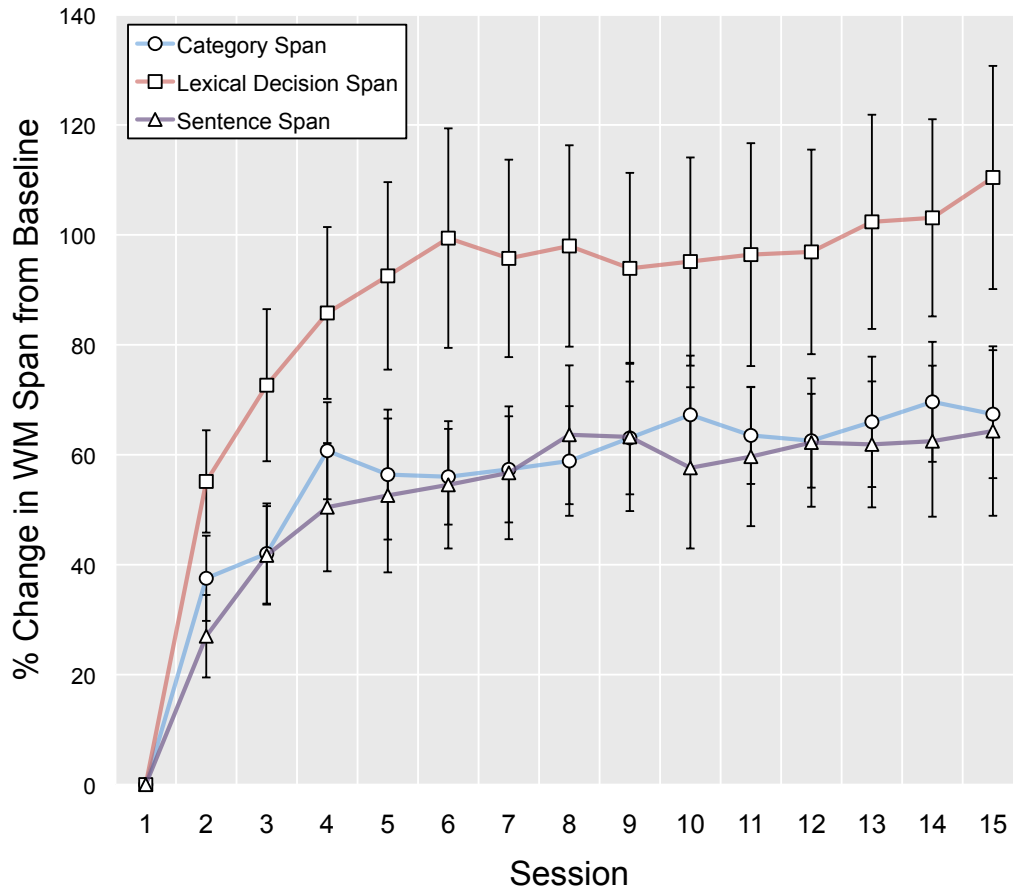


Figure 4. Percent Change in WM Span from Baseline Over Fifteen Weeks in WM Training Tasks

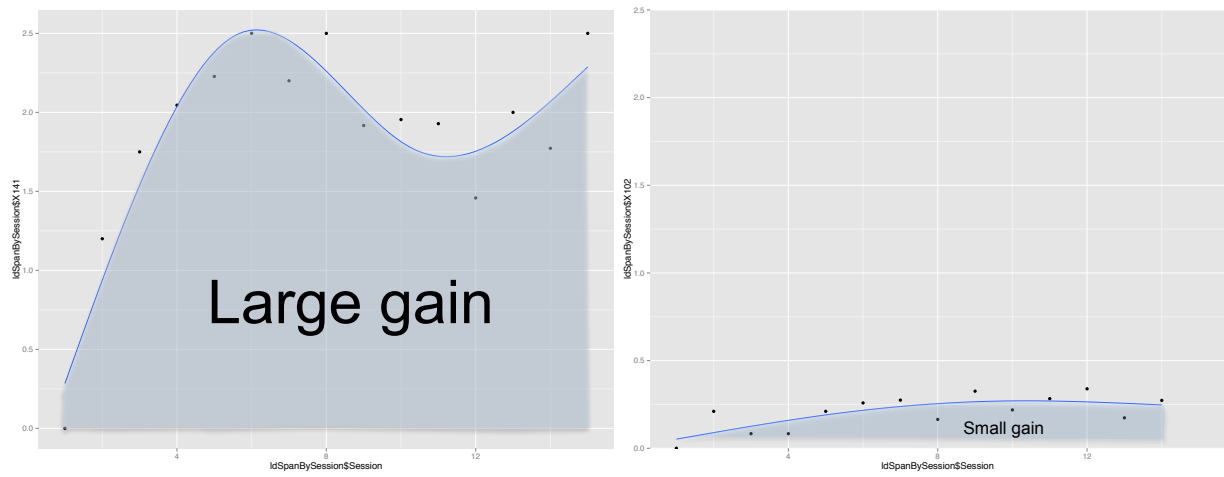


Figure 5. Demonstration of *auc* Method for a Participant with Large Training Gains and a Participant with Small Training Gains

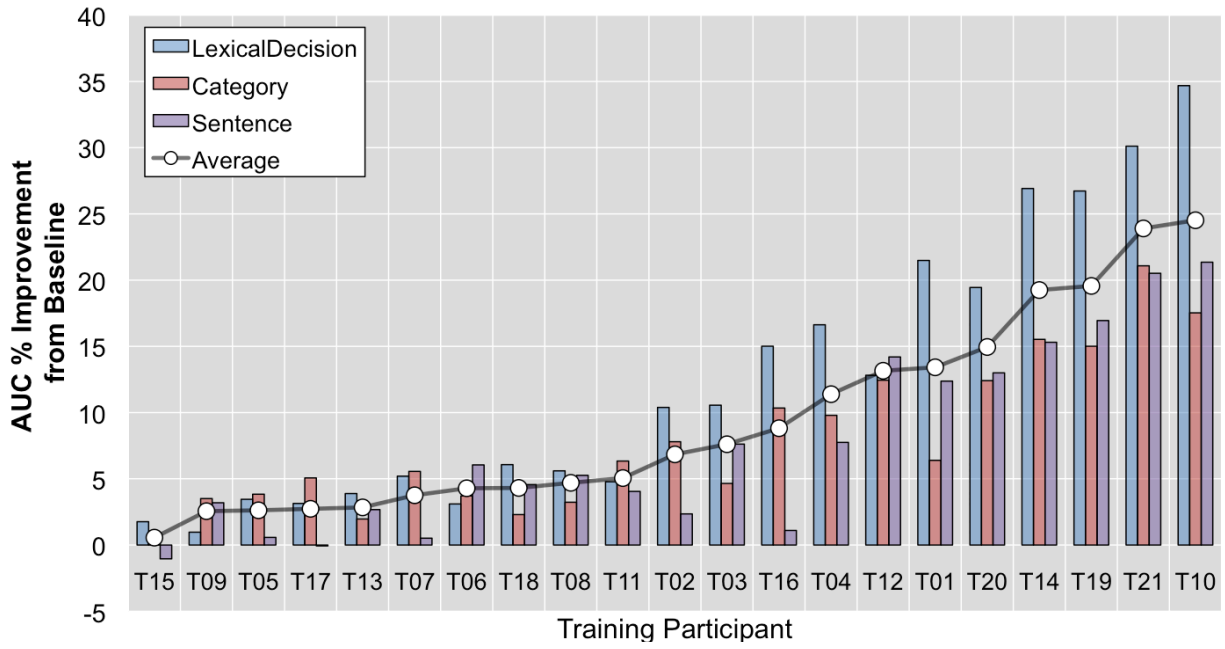


Figure 6. Rank Order Improvements in AUC Measure of Training for all Participants in the WM Training Program

### Gain Correlations

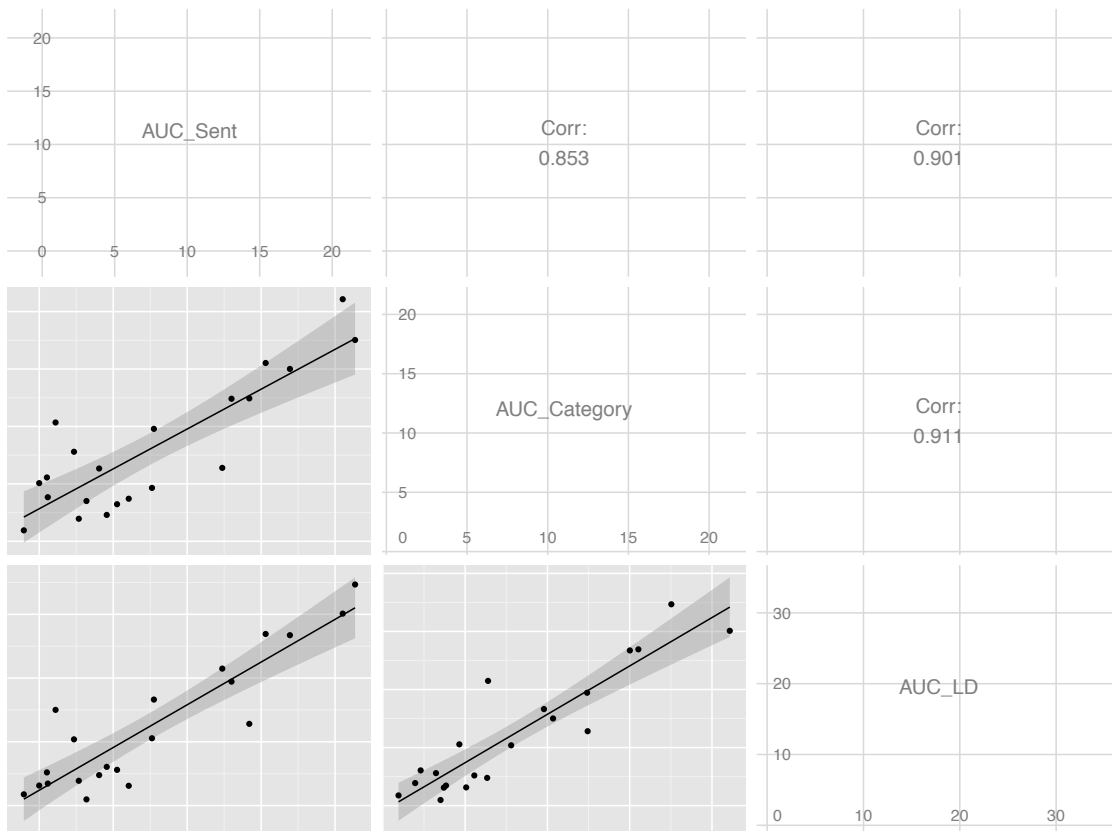
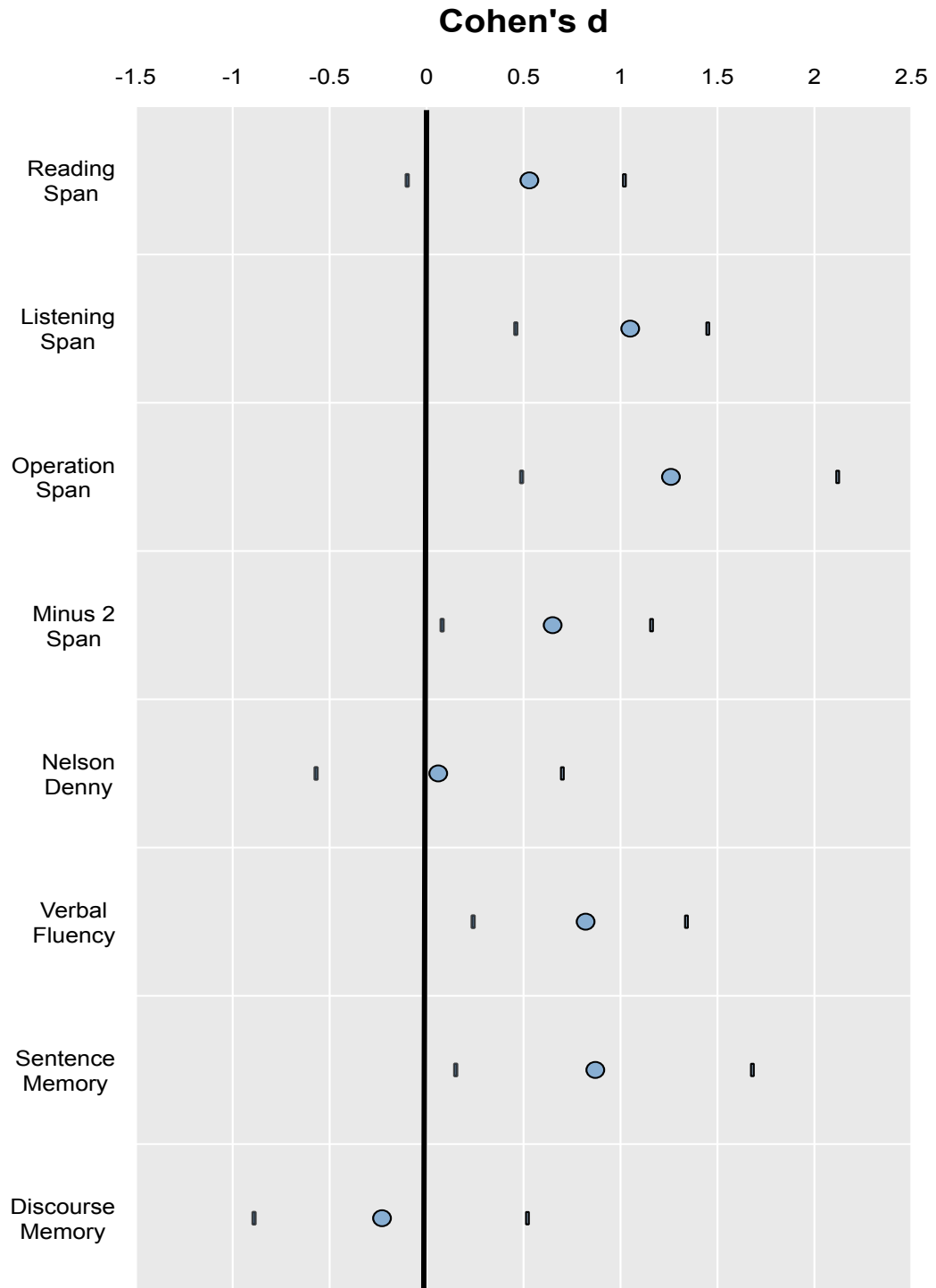


Figure 7. Across Task Correlations in WM Training Gains



*Figure 8.* Effect Sizes and 95% Bootstrapped Confidence Intervals of the Training Group x Time Interaction for Measures from Neuropsychological Battery



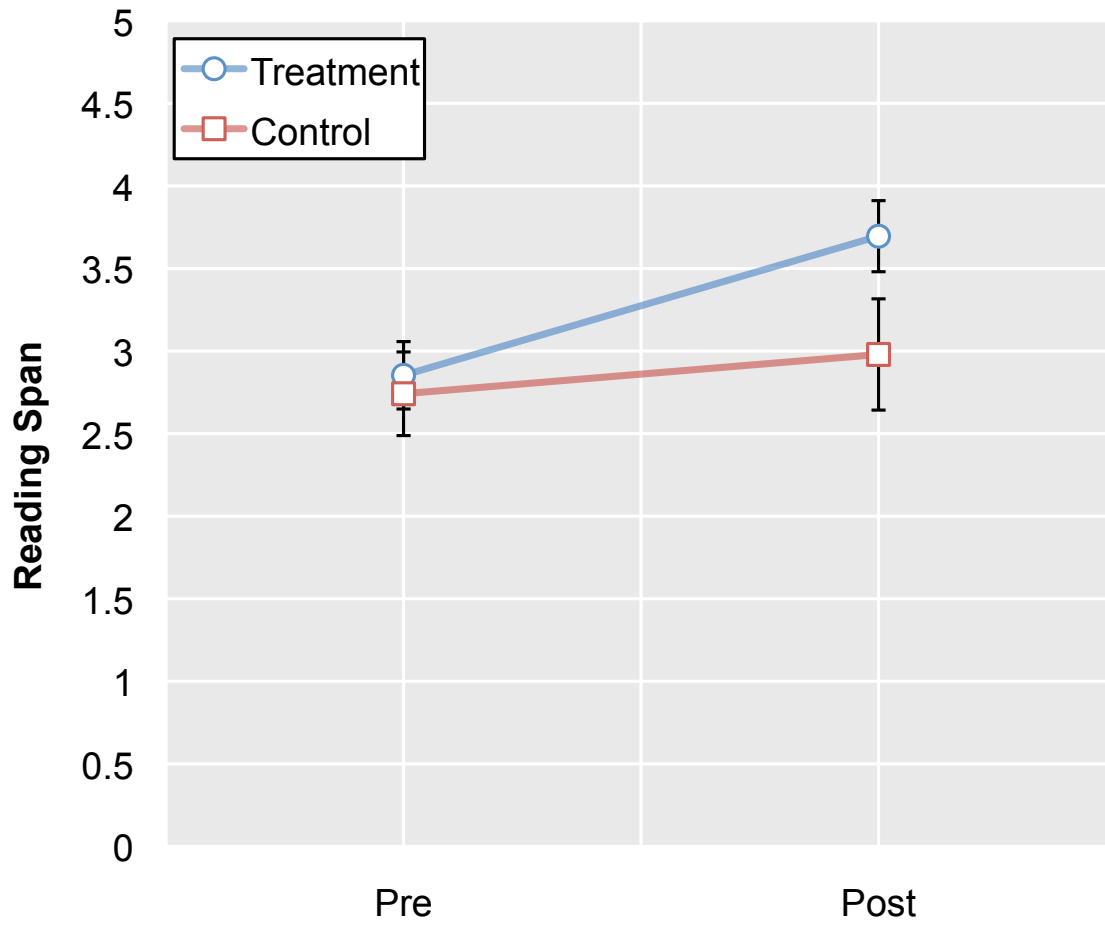


Figure 9. Reading Span Training Effects

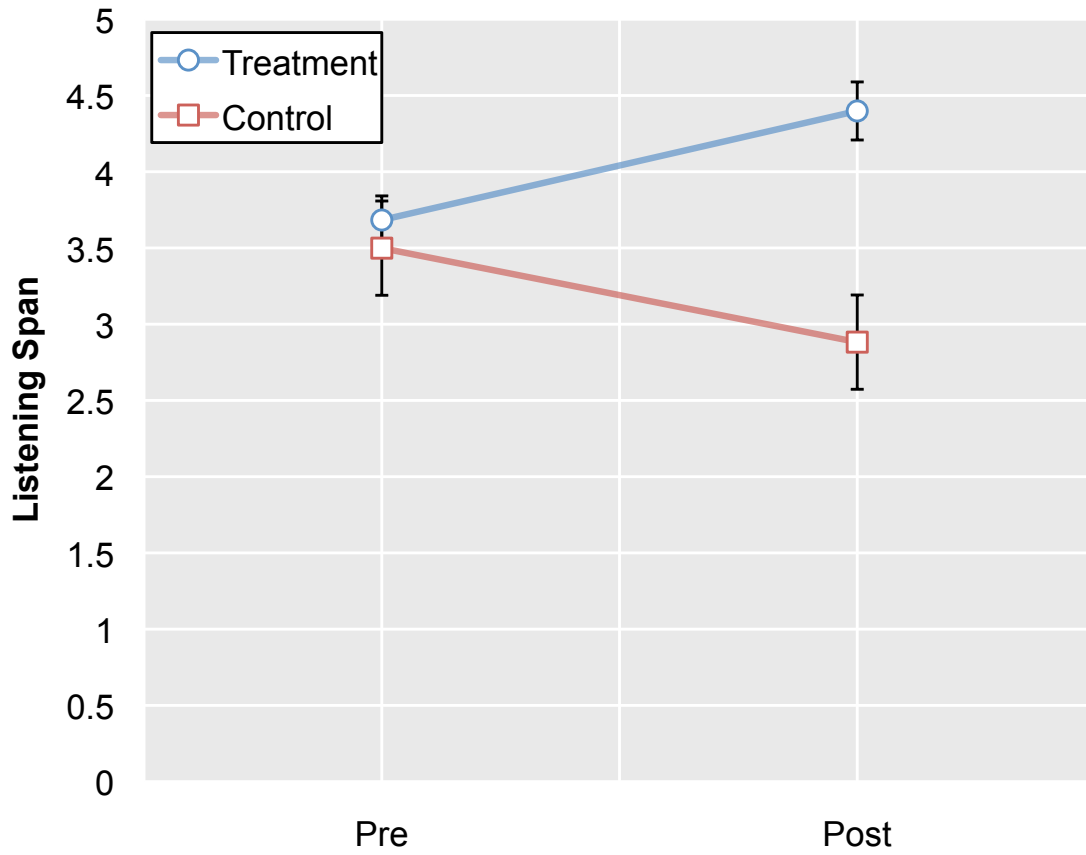


Figure 10. Listening Span Training Effects

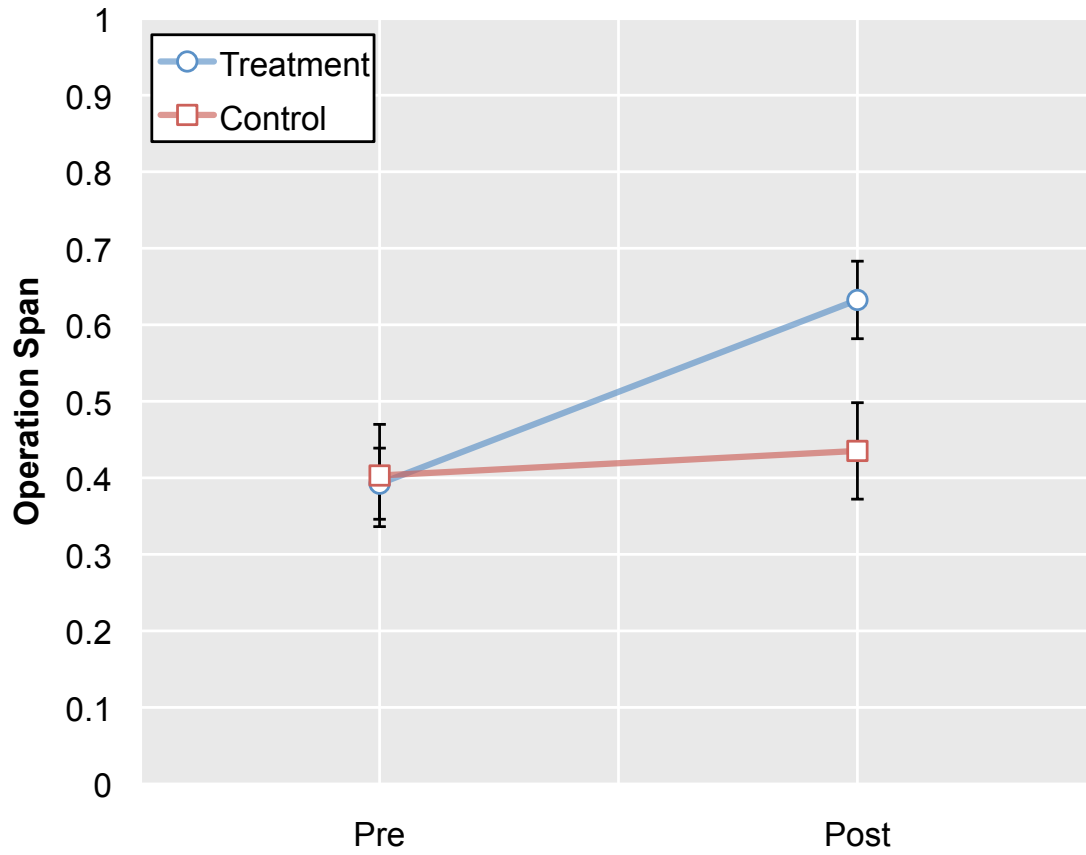


Figure 11. Operation Span Training Effects

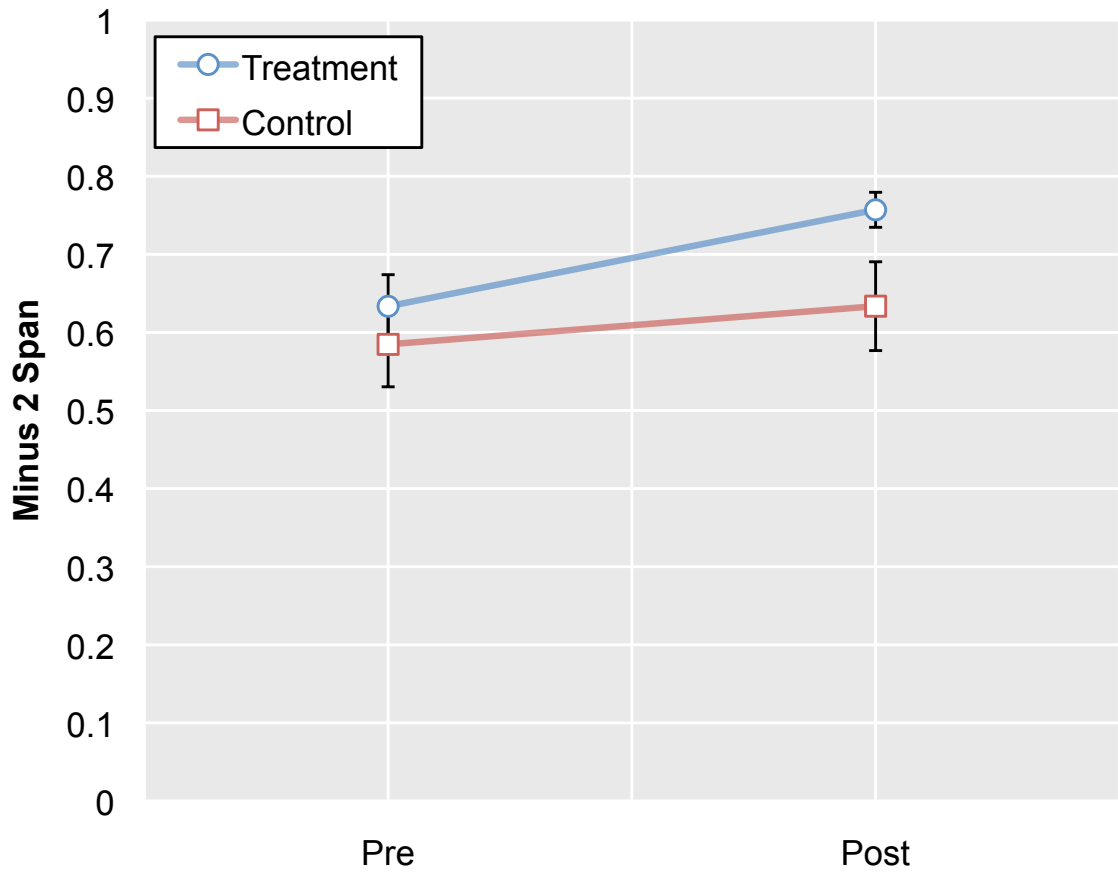


Figure 12. Minus-2 Span Training Effects

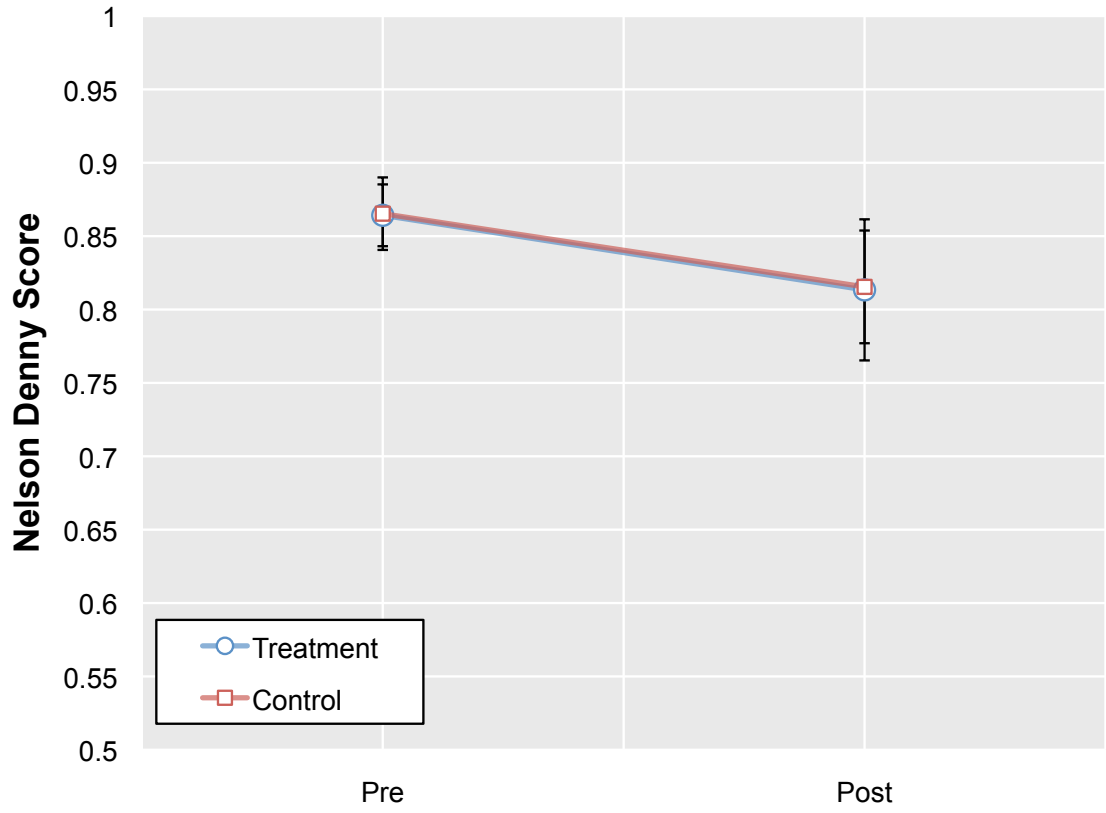


Figure 13. Nelson-Denny Training Effects

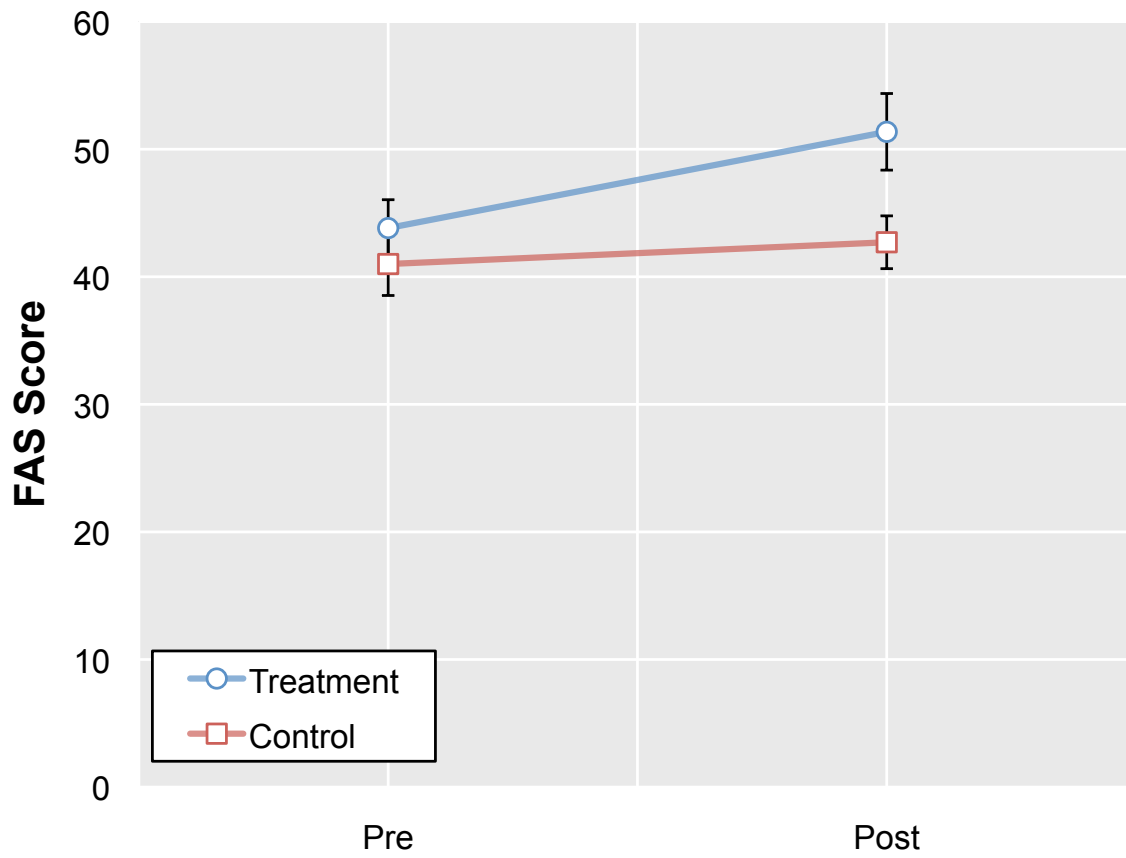


Figure 14. FAS Score Training Effects

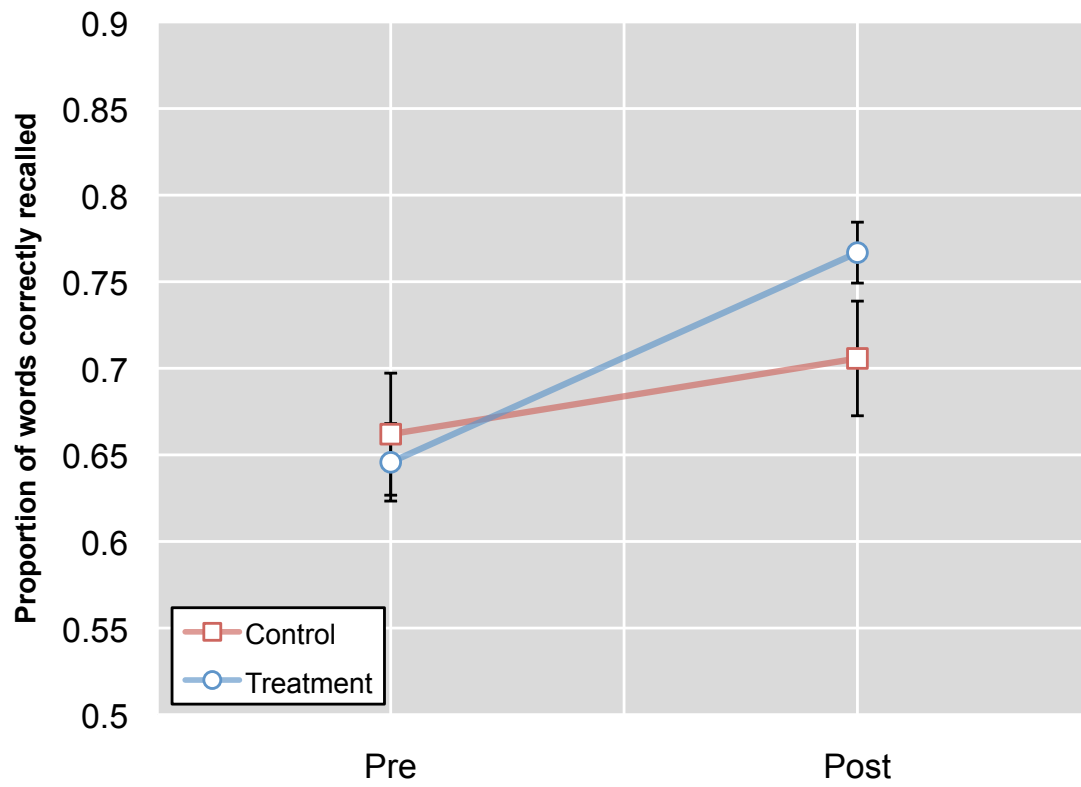


Figure 15. Sentence Recall (word scoring) Training Effects

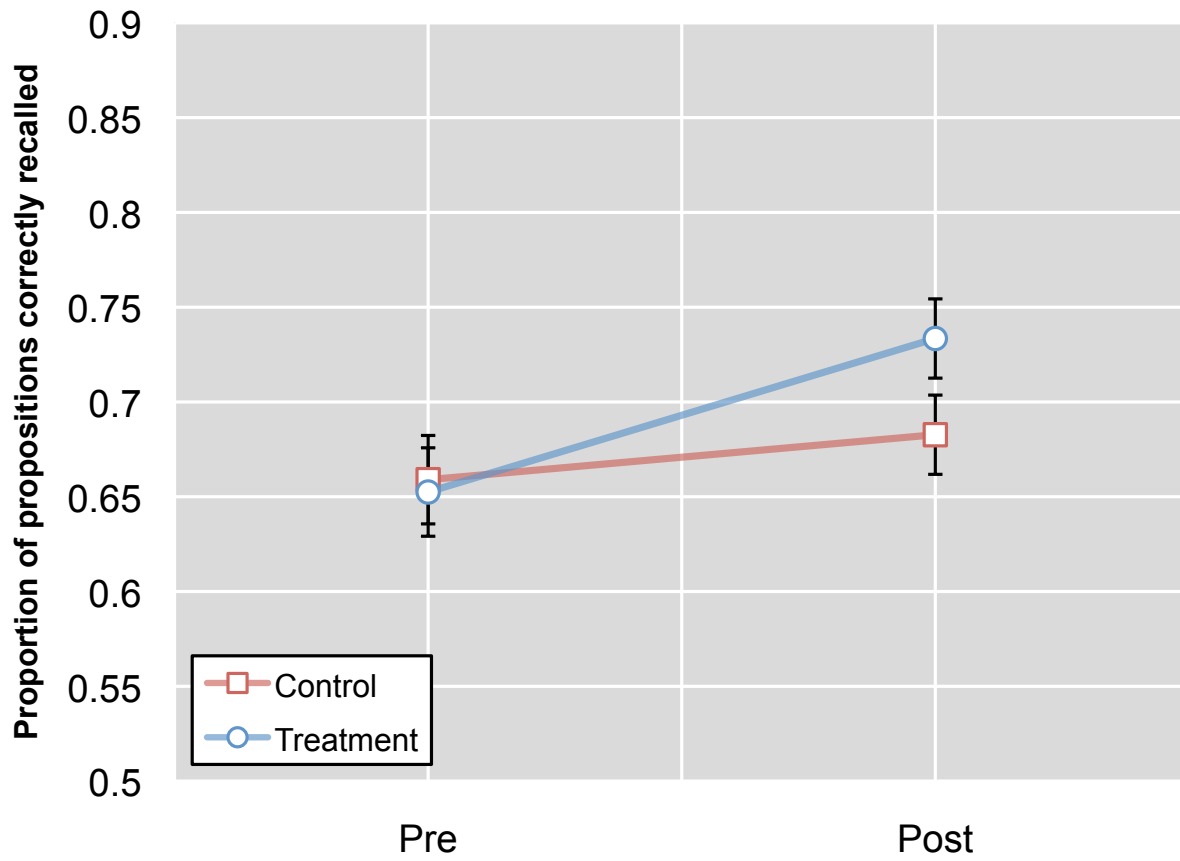


Figure 16. Sentence Recall (propositional scoring) Training Effects



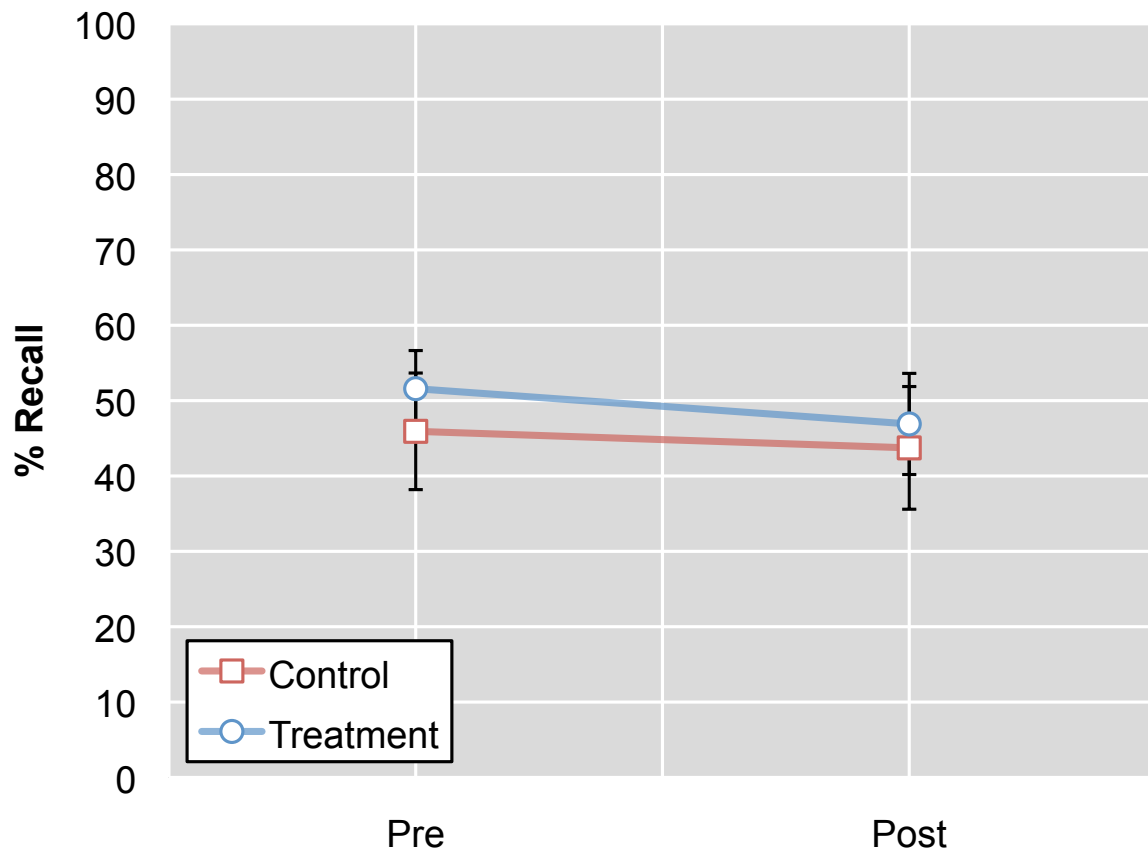
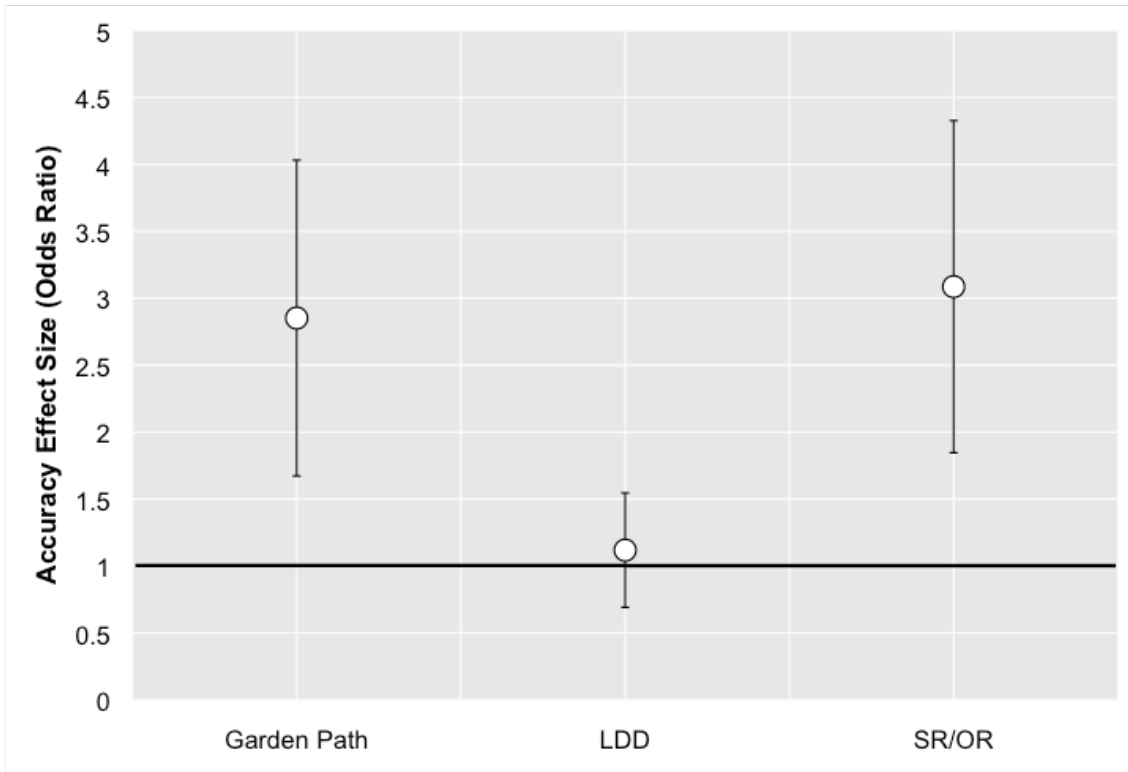


Figure 17. Rivermead Discourse Recall Training Effects



*Figure 18.* Effect Size of Syntactic Complexity Effects on Accuracy at Baseline. *Note.* Larger value means worse accuracy in syntactically complex condition. LDD = long distance dependency.

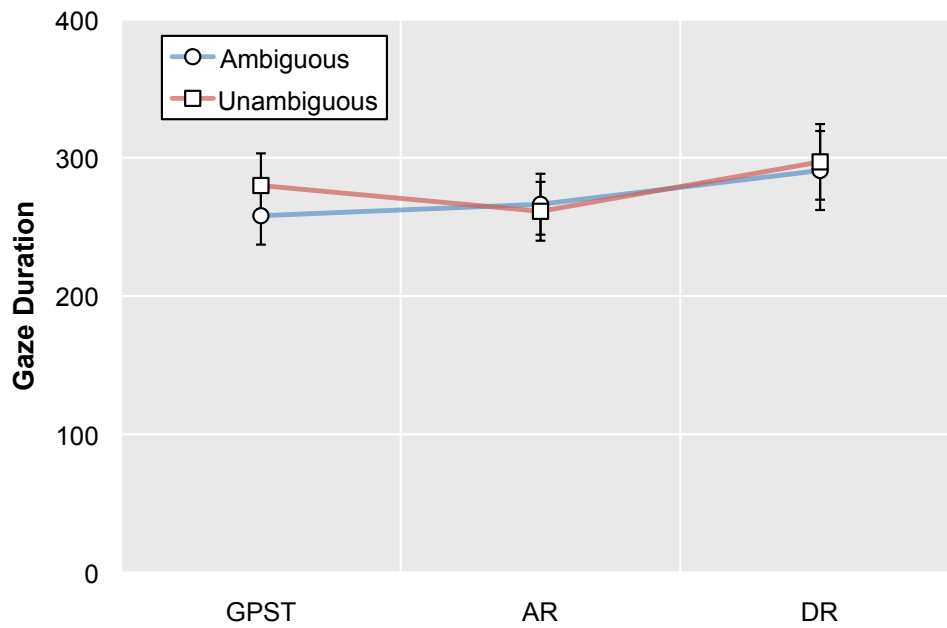


Figure 19. Baseline Syntactic Ambiguity Resolution (Gaze Duration). Note. GPST = Beginning of Garden Path Sentence; AR = Ambiguous Region; DR = Disambiguating Region

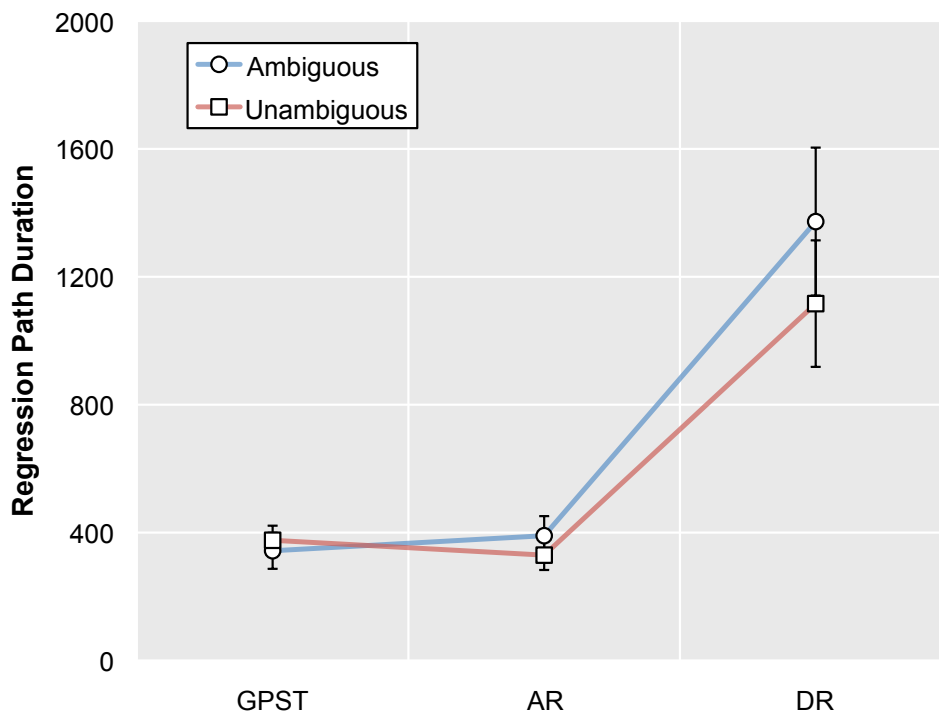
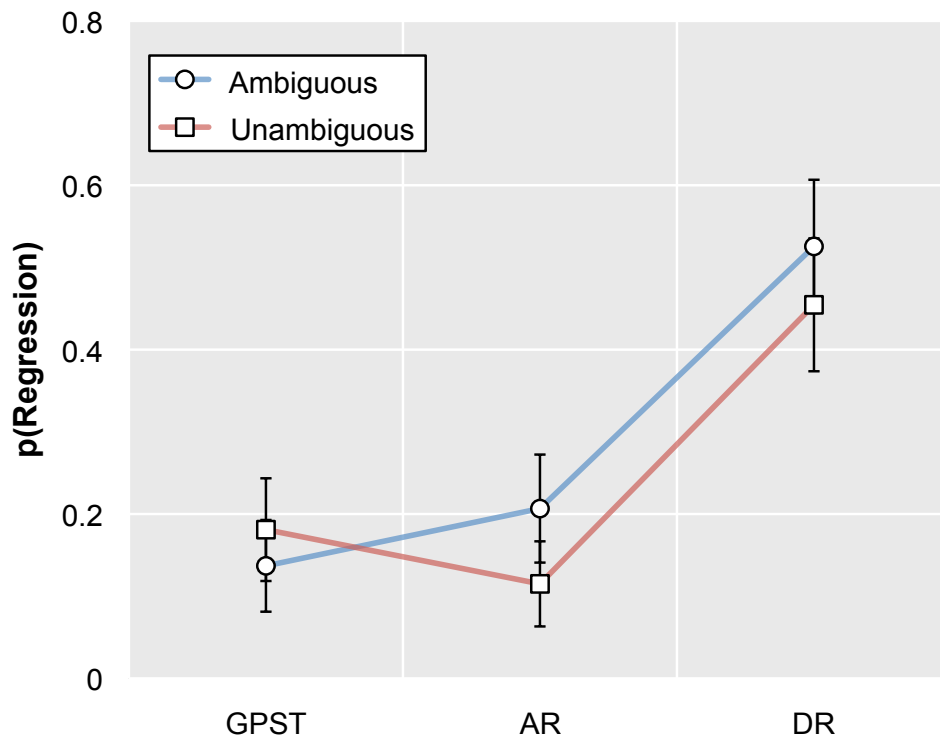


Figure 20. Baseline Syntactic Ambiguity Resolution (Regression Path Duration). *Note.* GPST = Beginning of Garden Path Sentence; AR = Ambiguous Region; DR = Disambiguating Region



*Figure 21.* Baseline Syntactic Ambiguity Resolution (Regression Probability). *Note.* GPST = Beginning of Garden Path Sentence; AR = Ambiguous Region; DR = Disambiguating Region

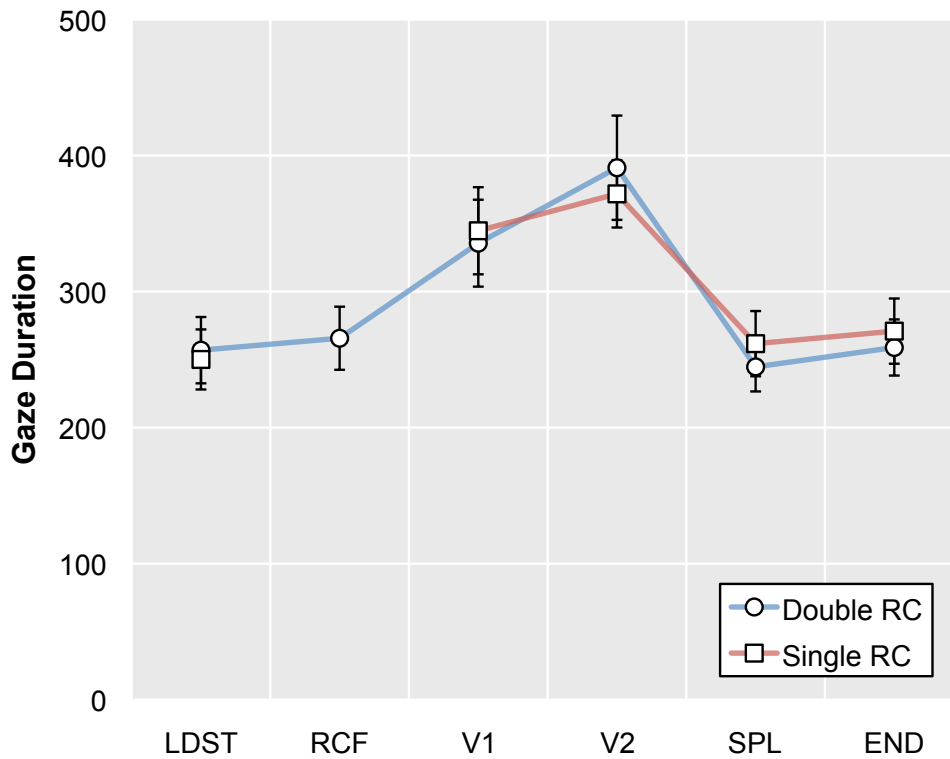


Figure 22. Baseline Long Distance Dependency Processing (Gaze Duration). *Note.* LDST = Beginning of Long Distance Sentence; RCF = Relative Clause Region (present only in double RC); V1 = First Verb; V2 = Second Verb; SPL = Spillover Region; END = End of Sentence.

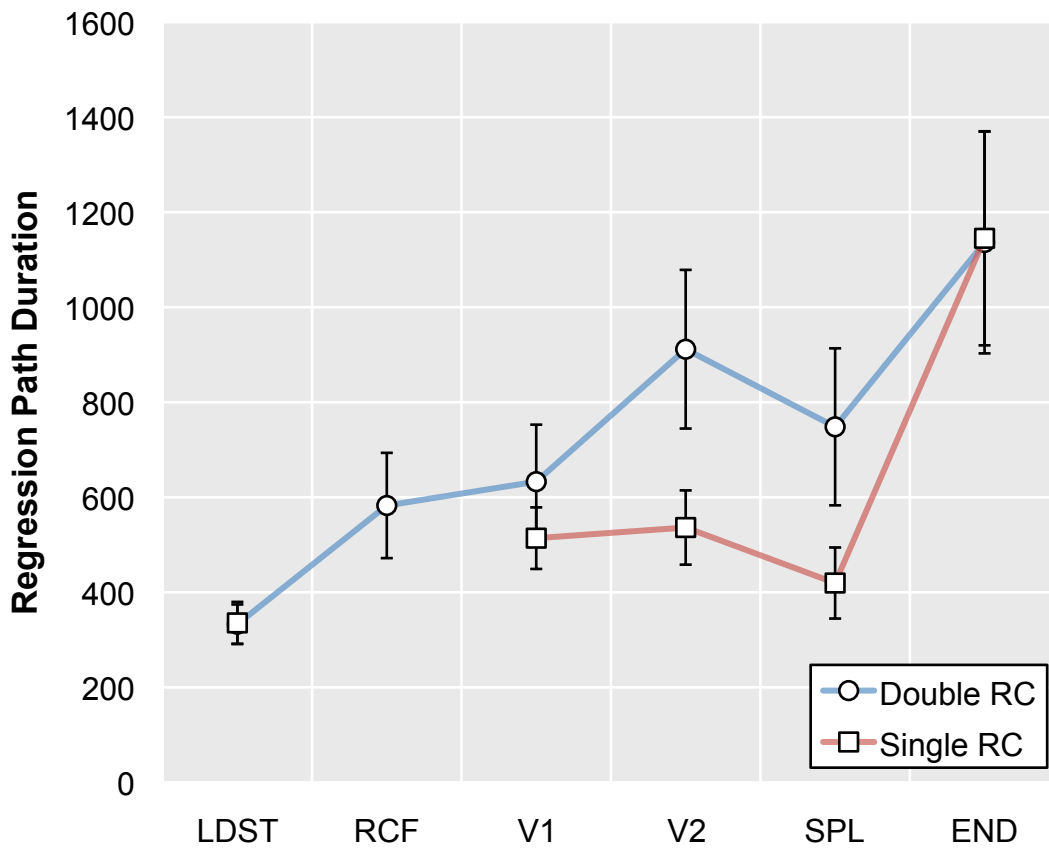


Figure 23. Baseline Long Distance Dependency Processing (Regression Path Duration). *Note.* LDST = Beginning of Long Distance Sentence; RCF = Relative Clause Region (present only in double RC); V1 = First Verb; V2 = Second Verb; SPL = Spillover Region; END = End of Sentence.

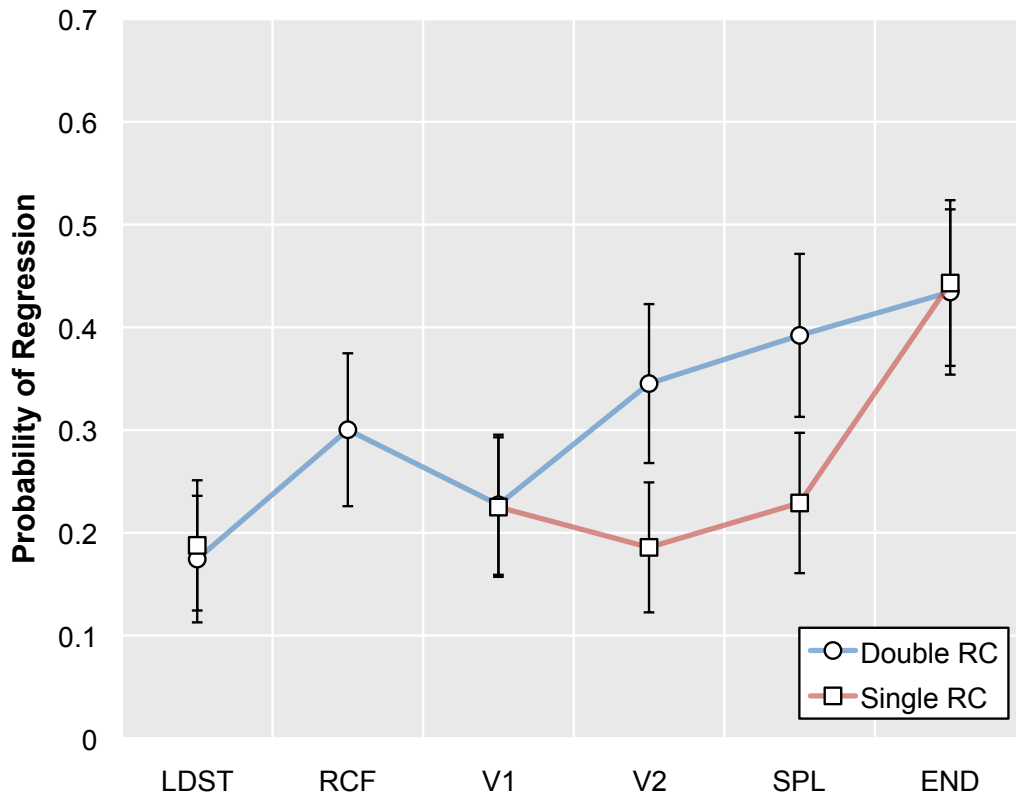


Figure 24. Baseline Long Distance Dependency Processing (Regression Probability). *Note.* LDST = Beginning of Long Distance Sentence; RCF = Relative Clause Region (present only in double RC); V1 = First Verb; V2 = Second Verb; SPL = Spillover Region; END = End of Sentence.



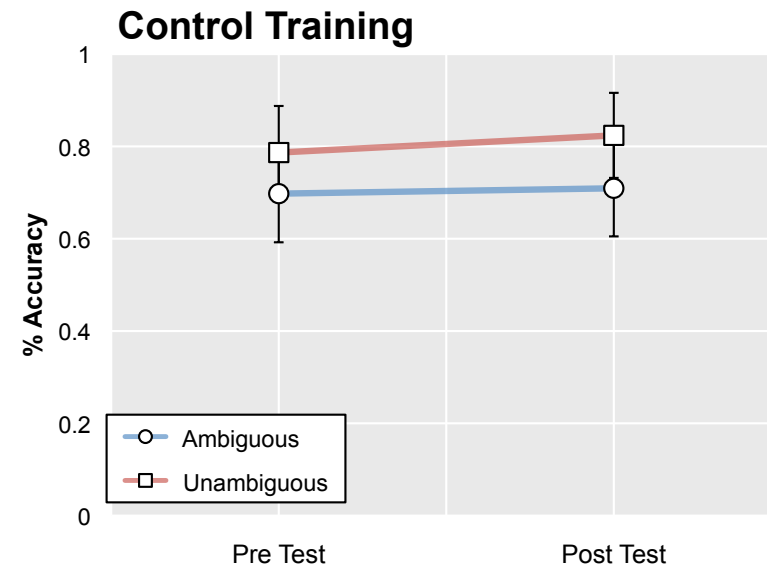
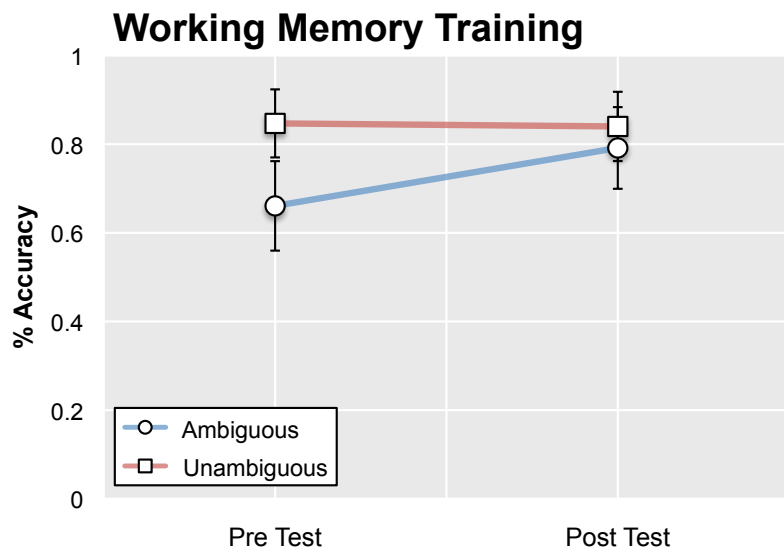


Figure 25. Effects of Training on Syntactic Ambiguity Resolution in Accuracy.

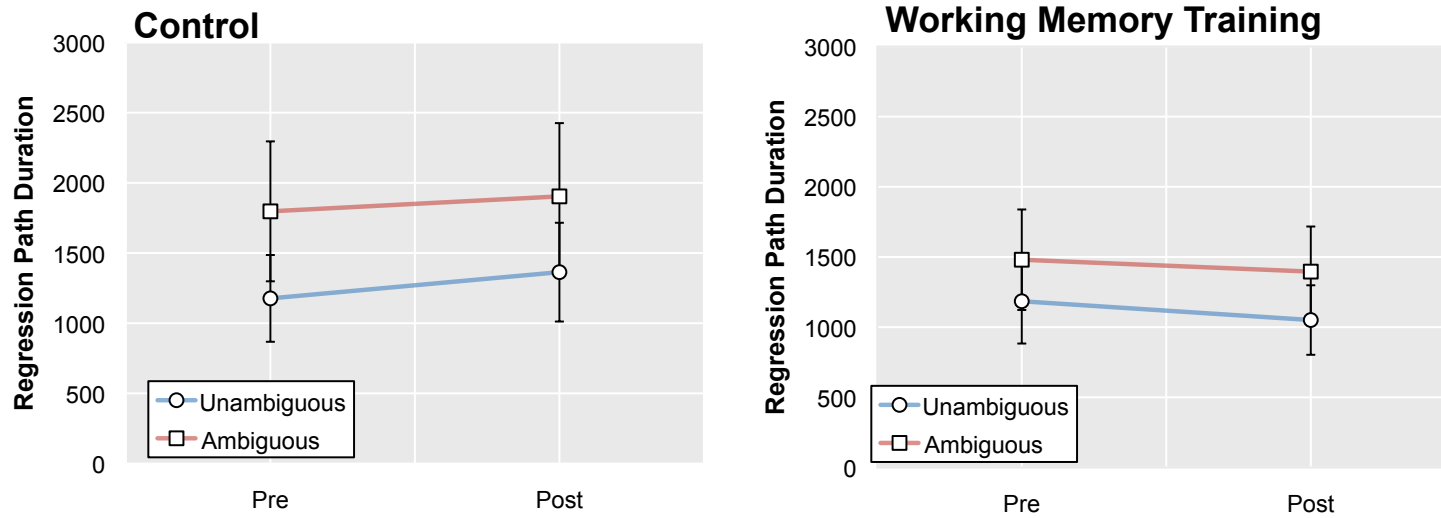


Figure 26. Effects of Training on Syntactic Ambiguity Resolution in Regression Path Duration.

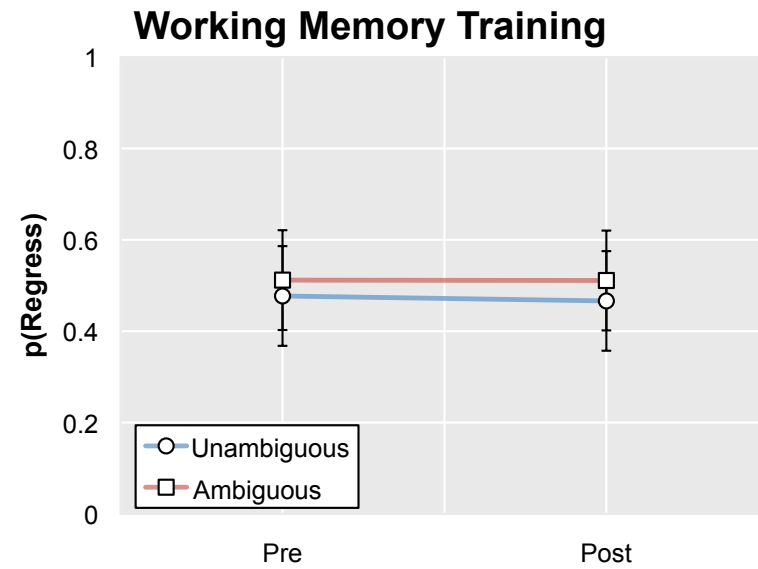
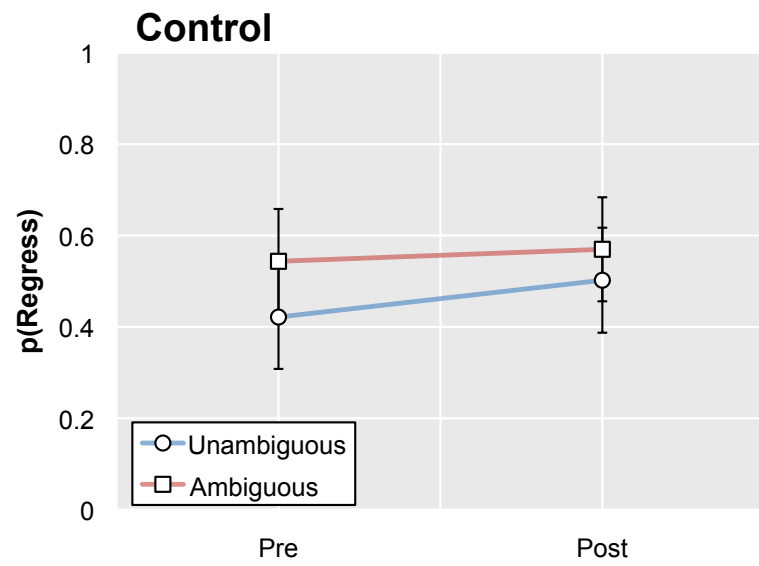


Figure 27. Effects of Training on Syntactic Ambiguity Resolution in Regression Probability.

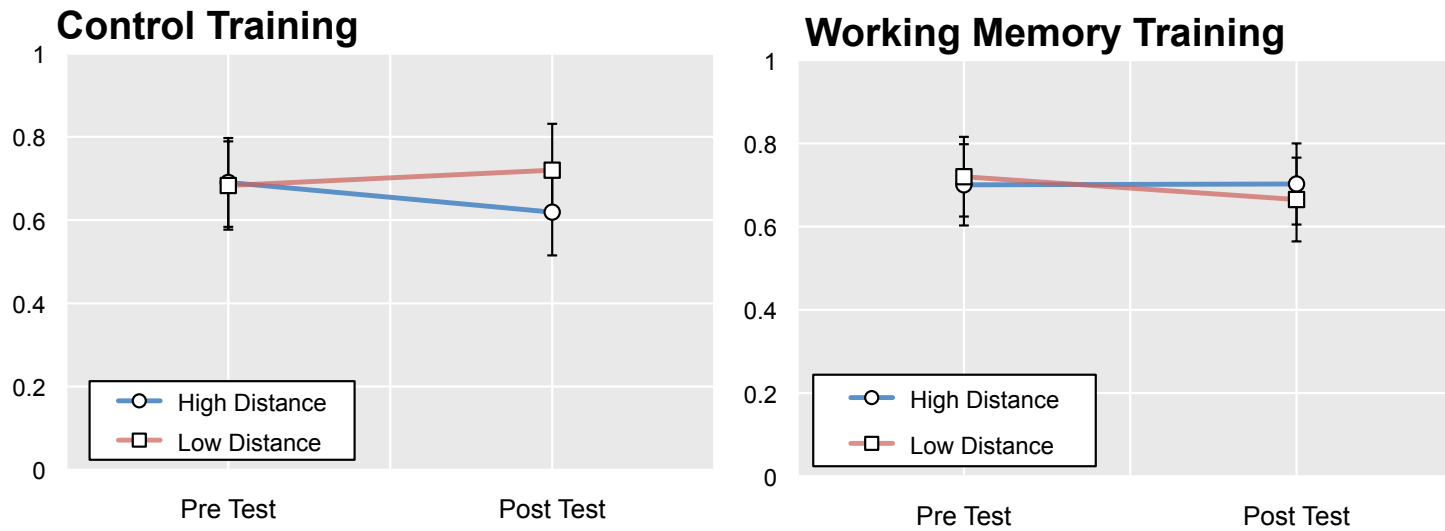


Figure 28. Effects of Training on Long Distance Dependency Processing in Accuracy.

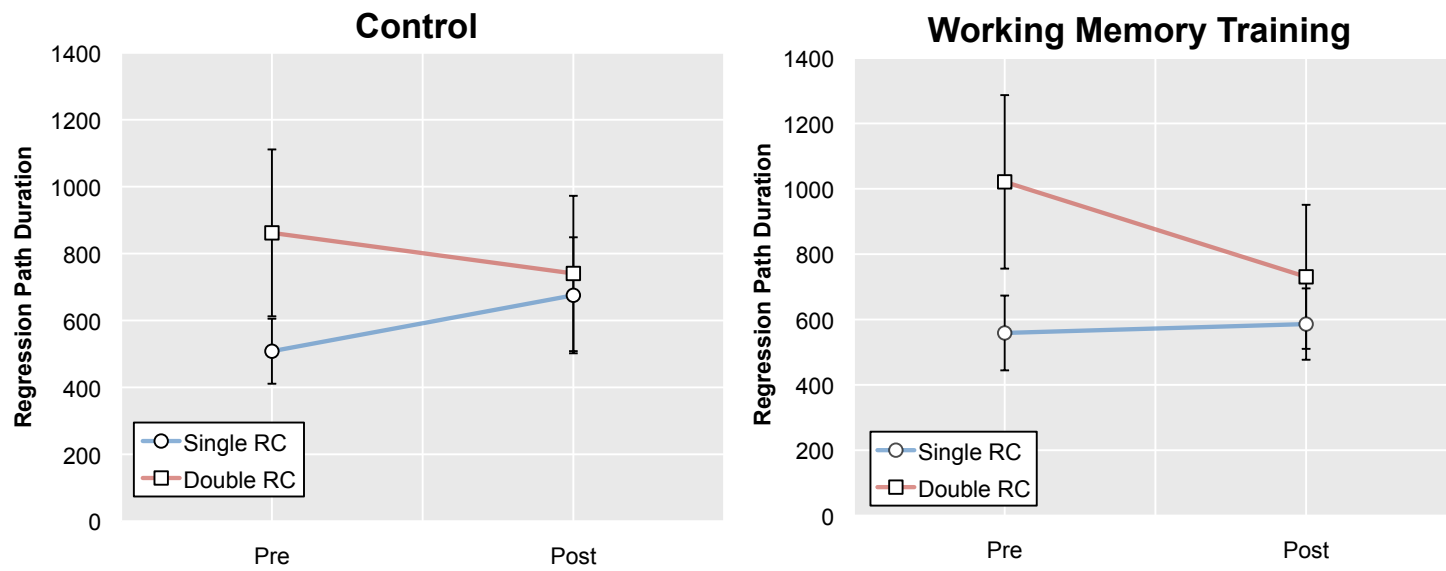


Figure 29. Effects of Training on Long Distance Dependency Processing in Regression Path Duration.

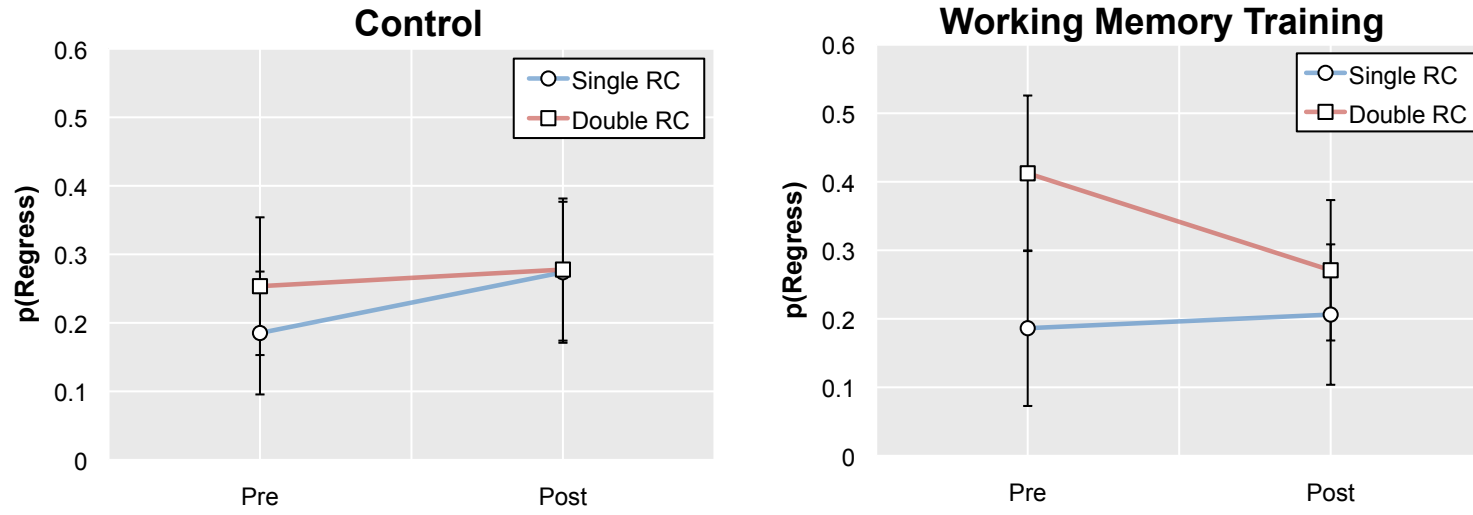


Figure 30. Effects of Training on Long Distance Dependency Processing in Regression Probability.

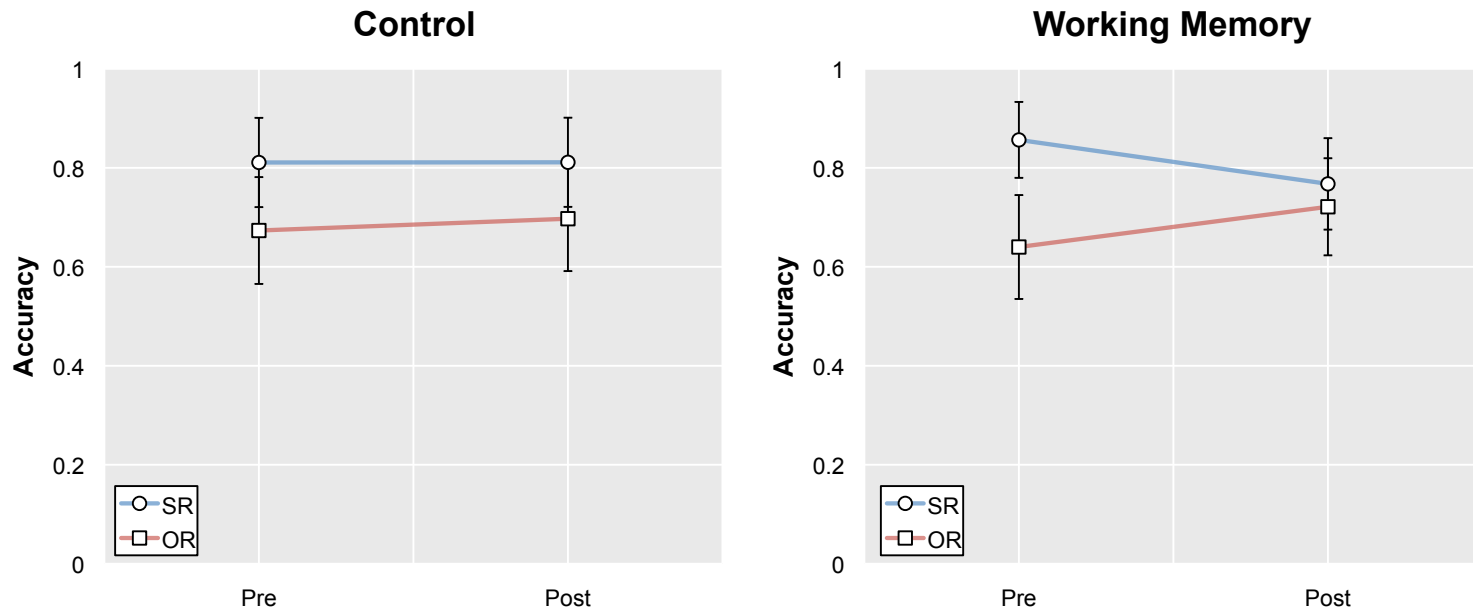
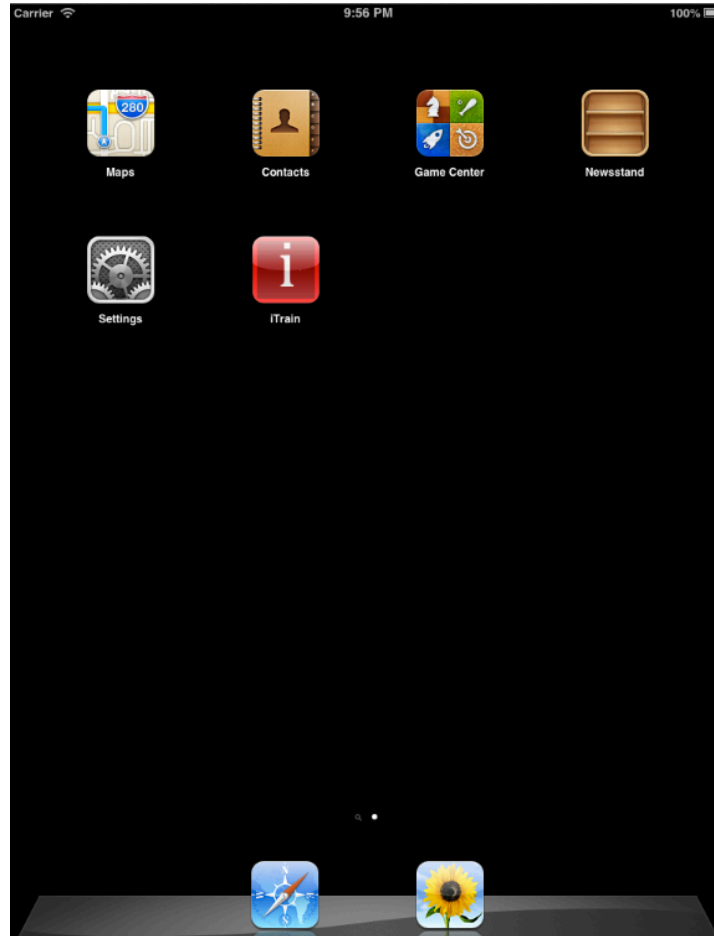


Figure 31. Effects of Training on Object-Relative Processing in Accuracy.

## Appendix A: iTrain Screen Captions and Links to Video Demos



*Figure A1.* Screen Caption of the iPad “home” window containing the iTrain software as an *app*.

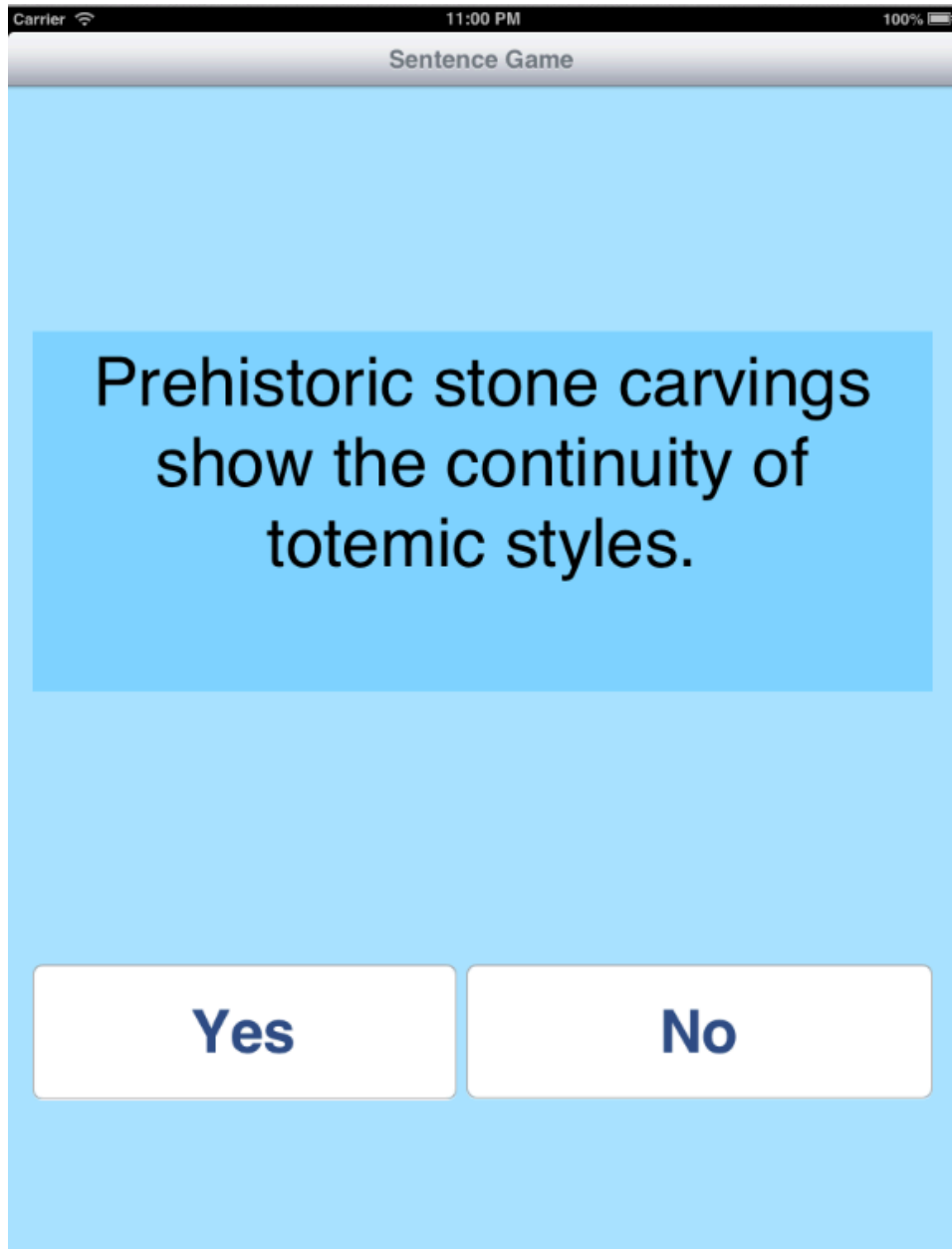




*Figure A2.* Category Span Training in the iTrain program



*Figure A3.* Lexical Decision Span Training in the iTrain program



*Figure A4.* Sentence Reading Span Training in the iTrain program

### **Hyperlinks to video demonstrations of iTrain**

1. Category Task Demo: <http://tinyurl.com/categoryTask>
2. Lexical Decision Task Demo: <http://tinyurl.com/lexDecTask>
3. Reading Task Demo: <http://tinyurl.com/readSpanTask>
4. Control Demo (Lexical Decision): <http://tinyurl.com/controlTask>

## Appendix B: Expectation Survey and Results

*Below you will read a number of statements that describe experiences and beliefs about the training intervention in which you took part. Please read each of the statements below and rate, for each one, the degree to which you agree with the statement.*

### RATING SCALE

Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree
1	2	3	4	5

1. I believe that iTrain helped improve my cognition.

1	2	3	4	5
---	---	---	---	---

2. I believe that iTrain helped improve my memory.

1	2	3	4	5
---	---	---	---	---

3. I believe that iTrain helped improve my reading ability.

1	2	3	4	5
---	---	---	---	---

4. I believe that iTrain helped improve my ability to quickly respond to things.

1	2	3	4	5
---	---	---	---	---

5. I believe that iTrain helped improve my attention.

1	2	3	4	5
---	---	---	---	---

6. I believe that iTrain helped improve my knowledge of words.

1	2	3	4	5
---	---	---	---	---

7. I believe that iTrain helped improve my ability to do more than one thing at the same time.

1	2	3	4	5
---	---	---	---	---

**Now we are going to ask you to rate some statements about some of the specific tasks that you completed at pre-testing and post-testing. Please read each of the statements below and rate, for each one, the degree to which you agree with the statement.**

8. You completed a task called "Reading Memory". In this task, you were shown a series of sentences to read aloud and you were asked to judge if the sentences were true or not. You were also asked to remember the last word of each of the sentences in that section in order.

Do you believe that iTrain helped lead to better performance on this task?

1	2	3	4	5
---	---	---	---	---

9. You completed a task called "Listening Memory". In this task, you heard a series of sentences and you were asked to judge if the sentences were true or not. You were also asked to remember the last word of each of the sentences in that section in order.

Do you believe that iTrain helped lead to better performance on this task?

1	2	3	4	5
---	---	---	---	---

10. You completed a task called "Number Memory". In this task, you were shown a group of numbers, one at a time, and asked to read the numbers aloud. You were then asked to remember each number in order and to type each of the numbers you saw in their original order, after subtracting two from those numbers.

Do you believe that iTrain helped lead to better performance on this task?

1	2	3	4	5
---	---	---	---	---

11. You completed a task called "Equations". In this task, you were shown a series of mathematical equations to solve. After solving each of the equations, a letter was displayed on the screen for you to remember. You were asked to type in the letters you saw, in the order you saw them.

Do you believe that iTrain helped lead to better performance on this task?

1	2	3	4	5
---	---	---	---	---

12. You completed a task that involved having your eyes monitored while you read a number of different kinds of sentences and answered comprehension questions about those sentences.

Do you believe that iTrain helped lead to better performance in understanding those sentences?

1	2	3	4	5
---	---	---	---	---

13. You completed a task called "Sentence and Paragraph Recall". In this task, you were shown a series of sentences or longer paragraphs to read silently. After reading each sentence or passage, you were asked to recall aloud as much of the sentence or paragraph as you could remember.

Do you believe that iTrain helped lead to better performance on this task?

1	2	3	4	5
---	---	---	---	---

14. You completed a task called "Reading Comprehension". In this task, you were given some longer passages to read, followed by multiple choice questions to answer. You were asked to read as many of the passages and answer as many comprehension questions as you could in the 20 minutes allotted for the task.

Do you believe that iTrain helped lead to better performance on this task?

1	2	3	4	5
---	---	---	---	---

Figure B1. Group Differences in Perceived Improvement in General Cognition (Items 1-7)

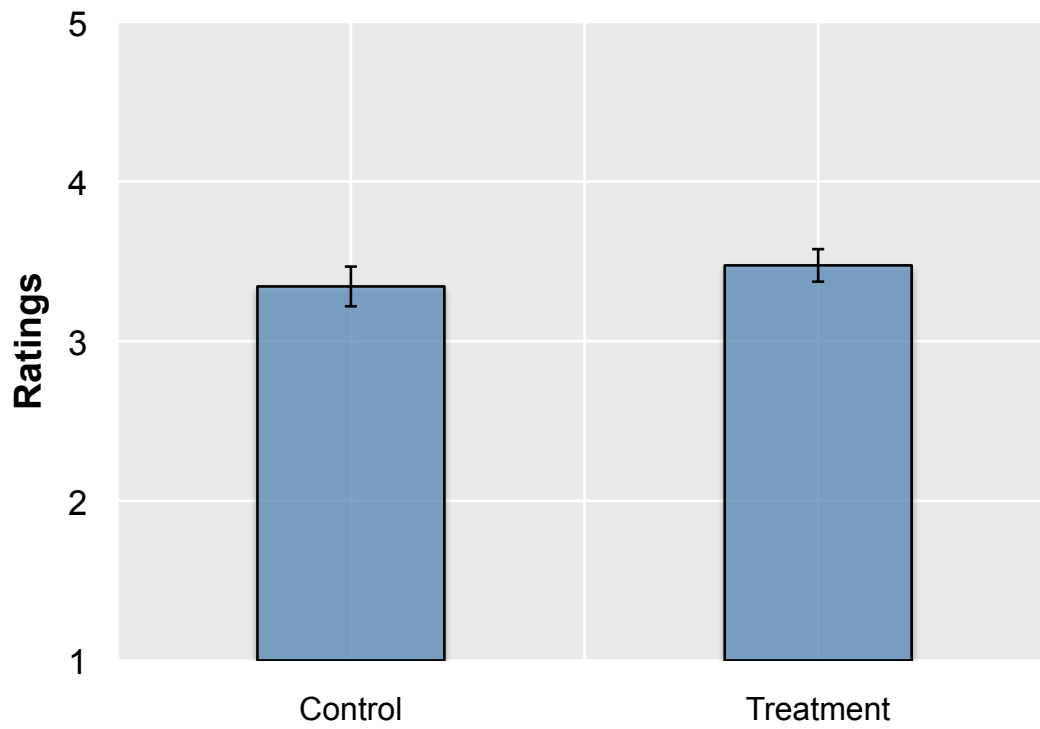
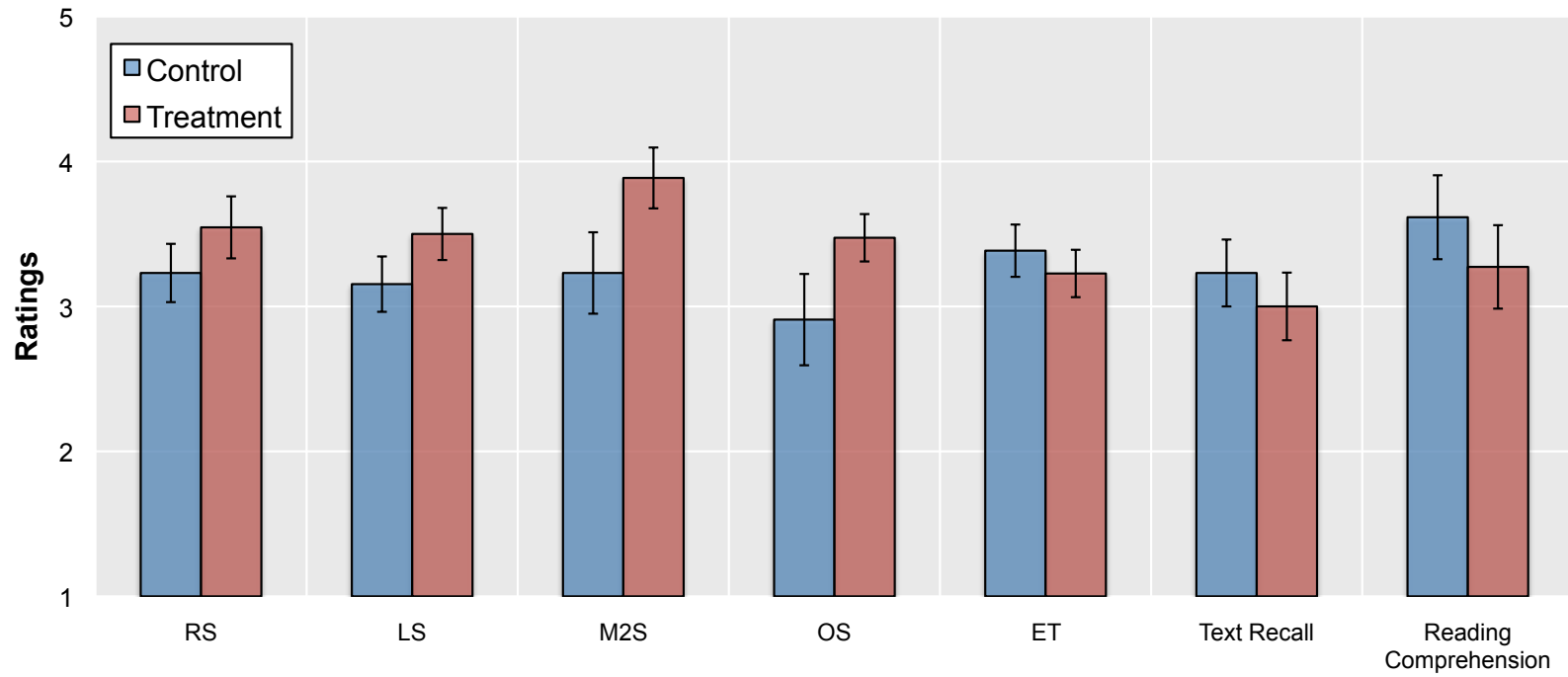




Figure B2. Group Differences in Perceived Improvements in Specific Tasks.



*Table B1.* Estimates and 95% Confidence Intervals of Group Differences Across Specific Expectation Items

<b>Item</b>	<b>diff</b>	<b>95% CI</b>
Reading Span	.31	[-.22, .84]
Listening Span	.35	[-.11, .81]
Minus 2 Span	.65	[.004, 1.29]
Operation Span	.56	[-.03, 1.15]
Eye-Tracking	-.16	[-.69, .27]
Text Recall	-.23	[-.23, .36]
Reading Comprehension	-.34	[-1.06, .38]