



The Effects of Home-Based Cognitive Training on Verbal Working Memory and Language Comprehension in Older Adulthood

Brennan R. Payne^{1,2*} and Elizabeth A. L. Stine-Morrow^{1,3*}

¹ The Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, United States, ² Department of Psychology, University of Utah, Salt Lake City, UT, United States, ³ Department of Educational Psychology, University of Illinois at Urbana-Champaign, Urbana, IL, United States

Effective language understanding is crucial to maintaining cognitive abilities and learning new information through adulthood. However, age-related declines in working memory (WM) have a robust negative influence on multiple aspects of language comprehension and use, potentially limiting communicative competence. In the current study ($N = 41$), we examined the effects of a novel home-based computerized cognitive training program targeting verbal WM on changes in verbal WM and language comprehension in healthy older adults relative to an active component-control group. Participants in the WM training group showed non-linear improvements in performance on trained verbal WM tasks. Relative to the active control group, WM training participants also showed improvements on untrained verbal WM tasks and selective improvements across untrained dimensions of language, including sentence memory, verbal fluency, and comprehension of syntactically ambiguous sentences. Though the current study is preliminary in nature, it does provide initial promising evidence that WM training may influence components of language comprehension in adulthood and suggests that home-based training of WM may be a viable option for probing the scope and limits of cognitive plasticity in older adults.

Keywords: cognitive training, working memory, language, aging

OPEN ACCESS

Edited by:

Pamela M. Greenwood,
George Mason University,
United States

Reviewed by:

Alexandra B. Morrison,
University of Miami, United States
Jerri D. Edwards,
University of South Florida,
United States

*Correspondence:

Brennan R. Payne
brennan.payne@psych.utah.edu
Elizabeth A. L. Stine-Morrow
eals@illinois.edu

Received: 09 August 2016

Accepted: 19 July 2017

Published: 08 August 2017

Citation:

Payne BR and Stine-Morrow EAL (2017) The Effects of Home-Based Cognitive Training on Verbal Working Memory and Language Comprehension in Older Adulthood. *Front. Aging Neurosci.* 9:256. doi: 10.3389/fnagi.2017.00256

INTRODUCTION

Literacy and effective language comprehension are crucial to maintaining cognitive abilities and learning from text through adulthood (Manly et al., 2003; Stern, 2009; Payne et al., 2012a; Stine-Morrow et al., 2015). However, normative age-related cognitive changes have a profound influence on language understanding, especially for effortful comprehension and memory processes (Wingfield and Stine-Morrow, 2000; Wlotko et al., 2010; Payne and Stine-Morrow, 2016; Stine-Morrow and Payne, 2016). 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 (Stine and Wingfield, 1987; Van der Linden et al., 1999; Borella et al., 2011; Kemper, 2012; Payne et al., 2014a). Although virtually all models of language comprehension include some mechanism to account for WM constraints

(see Pickering and van Gompel, 2006; Caplan and Waters, 2013 for reviews), the degree to which the WM system directly supports comprehension and the role of WM in language understanding is a topic of ongoing debate (see, e.g., Just and Carpenter, 1992; Carpenter et al., 1994; Caplan and Waters, 1999, 2013; Just and Varma, 2002, 2007; MacDonald and Christiansen, 2002).

The majority of research examining the influence of WM on language comprehension has relied on either dual-task paradigms to examine the effects of manipulated WM constraints on language comprehension (Smiler et al., 2003; Fedorenko et al., 2006; Kemper and Herman, 2006), or correlational approaches that test the relationship between individual differences in WM and language comprehension (King and Just, 1991; Just and Carpenter, 1992; Caplan and Waters, 1999; DeDe et al., 2004; Stine-Morrow et al., 2008; Noh and Stine-Morrow, 2009; Caplan et al., 2011; Payne et al., 2014a). In contrast, the current study used home-based cognitive training as an experimental approach to examining the degree to which the verbal WM system underlies language comprehension (cf. Novick et al., 2013; Hussey et al., 2016).

Aging, Working Memory, and Language Comprehension

Working memory limitations have historically been invoked in models of language understanding to explain comprehension difficulties for linguistically complex material (Miller and Chomsky, 1963). Current research activity has focused on performance on complex WM span tasks, which have been argued to underpin performance on a wide range of both complex and everyday tasks (Engle, 2010; Baddeley, 2012). While there are many contemporary models of WM, each of which make slightly different predictions or have slightly different foci (e.g., Cowan, 2000; Engle, 2002, 2010; Kane et al., 2007a,b), most models converge on a similar account that the WM system supports “the ability to simultaneously maintain information in an active and readily accessible state, while concurrently and selectively processing new information...” (Conway et al., 2007; p. 3). Complex WM span measures such as the *reading span* (Daneman and Carpenter, 1980; Wingfield et al., 1988) and the *operation span* (Turner and Engle, 1989) task share the requirement to simultaneously hold information in memory while performing some concurrent processing. This dual-task nature of complex span tasks critically sets them apart from simple STM tasks that are not predictive of higher-order cognition (reviewed in Baddeley, 2012). In contrast, performance on complex WM span tasks predict individual differences in a number of higher-order cognitive abilities including reasoning, episodic memory, attentional control, and intelligence (see Conway et al., 2007 for reviews). WM is also related to comprehension, with meta-analytic correlations ranging between $r = 0.41$ and $r = 0.52$ (Daneman and Merikle, 1996). Theoretical accounts of such relationships center on the reliance on WM for constructing, storing, retrieving, and integrating an incremental representation of the text’s meaning as decoding and parsing of the surface input is ongoing (e.g.,

Just and Carpenter, 1992; Gibson, 1998; Lewis and Vasishth, 2005).

Performance on complex span tasks declines with aging (e.g., Bopp and Verhaeghen, 2005), as does comprehension and memory for language (Kemper, 1987; DeDe et al., 2004; Payne et al., 2014b). For example, a meta-analysis by Johnson (2003) revealed that on average, older adults perform at about the 22nd percentile of the distribution of younger adults in text memory. Similar effect sizes for age-related declines in immediate language memory have been found in a longitudinal study tracking changes in older adults’ auditory discourse memory over a 10-year period (Payne et al., 2014b). Although there is considerable debate regarding the impact of WM deficits on on-line measures of real-time language processing in aging (Caplan and Waters, 1999; Kemper and Liu, 2007; Caplan et al., 2011; Payne et al., 2014a), verbal WM has been found to reliably mediate age-related changes in “off-line” measures of language comprehension and language memory (Kwong See and Ryan, 1995; Van der Linden et al., 1999; Hertzog et al., 2003; DeDe et al., 2004; Stine-Morrow et al., 2008; Borella et al., 2011).

Moreover, age differences in sentence comprehension accuracy are larger for sentences that are more semantically or syntactically complex, and these differences in performance have been found to be dependent upon individual differences in verbal WM capacity (Kemper, 1987; Stine and Hindman, 1994; Stine-Morrow et al., 2000; Christianson et al., 2006; Payne et al., 2014a). For example, “garden path” sentences such as (1) introduce a temporary syntactic ambiguity.

- (1) *The experienced soldiers warned about the dangers conducted the midnight raid.*

Typically, the first verb *warned* is initially (and incorrectly) interpreted as the main verb of the sentence (rather than as the verb of the reduced relative clause), creating difficulty when the reader encounters the second verb *conducted*; resolution thus requires a revision of the initial analysis (Bever, 1970; Clifton et al., 2003), which entails maintaining the alternate parse of the sentence during processing. WM capacity is an important predictor of resolution in garden-path ambiguities in younger (Just and Carpenter, 1992; MacDonald et al., 1992; Just and Varma, 2002) and older (Kemtes and Kemper, 1997; Kemper et al., 2004; Christianson et al., 2006) adults, as well as in other syntactically complex constructions, such as object-relative clauses (Stine-Morrow et al., 2000; DeDe et al., 2004), and long distance dependencies (King and Kutas, 1995; Caplan et al., 2011; Payne et al., 2014a).

Cognitive Training in Aging

Cognitive training has a long history in aging research, dating back over 30 years. Studies have reliably demonstrated that older adults show targeted improvements in trained abilities, including episodic memory, inductive reasoning, task switching, psychomotor speed, and WM capacity (Willis et al., 1981; Baltes and Willis, 1982; Willis and Nesselroade, 1990; Ball et al., 2002; Rebok, 2008; cf. Stine-Morrow and Basak, 2011; Rebok et al., 2014, for a review). Importantly, the demonstration that targeted

training in a cognitive domain can improve performance in that domain in older adults is not trivial considering evidence of age-related declines in plasticity (Lövdén et al., 2010). At the same time, there is considerable debate regarding whether and how cognitive training may produce “far” transfer, that is, improvements on untrained tasks that are distal from the trained ability—with studies demonstrating variable effect sizes for transfer (Melby-Lervåg and Hulme, 2013, 2016; Melby-Lervåg et al., 2016; Simons et al., 2016; but see Karbach and Verhaeghen, 2014; Au et al., 2015, 2016; Greenwood and Parasuraman, 2016). Note that there is considerable variability in cognitive training programs as well as what constitutes a target of transfer across the literature (Kelly et al., 2014; Simons et al., 2016)—the focus of the current study is on whether targeted cognitive training in one domain can produce transfer *across other cognitive domains*. Training effects on other outcomes (e.g., instrumental activities of daily living, self-rated health; e.g., Rebok et al., 2014, see Kelly et al., 2014; Simons et al., 2016 for recent reviews) are beyond the scope of this study and are not discussed in further detail. In the cognitive aging literature, cognitive training has most reliably produced narrow transfer across untrained cognitive domains (see reviews in Stine-Morrow and Basak, 2011; Simons et al., 2016). For example, the ACTIVE trial (Ball et al., 2002; Willis et al., 2006; Rebok et al., 2014), was the largest cognitive intervention study to date ($N = 2,832$) and arguably remains the benchmark cognitive training study, conforming to many of the best practices for intervention research. Healthy older adult participants’ completed 10 sessions of training in either processing speed, episodic memory (targeting strategy use), or inductive reasoning. Although evidence of transfer to measures of functional and clinical outcomes (e.g., instrumental activities of daily living, depressive symptoms, driving mobility, and others) has been reported from ACTIVE (e.g., Willis et al., 2006; Wolinsky et al., 2010; Rebok et al., 2014), the effects of the transfer of cognitive training across cognitive outcomes was narrow, with large and maintained effects of training on measures proximal to the training (e.g., memory training improved episodic memory) with little evidence of transfer across other cognitive domains (e.g., memory training had no impact on processing speed or inductive reasoning).

Some researchers have noted that training regimens that target executive control and WM functions have shown more promise in stimulating cognitive improvements beyond near transfer and practice effects in older adults (Karbach and Verhaeghen, 2014; Greenwood and Parasuraman, 2016). For example, a number of studies have demonstrated that WM training increases performance not only on span tasks that are untrained but proximal to WM, but also some (limited) evidence for transfer to other cognitive domains such as inhibitory control, memory, and reasoning (Buschkuhl et al., 2008; Li et al., 2008; Borella et al., 2010, 2017; Brehmer et al., 2011; Richmond et al., 2011; Zinke et al., 2014, see Karbach and Verhaeghen, 2014 for a recent meta-analysis). On the other hand, there is an active debate regarding whether such WM training can produce reliable broad-based transfer across cognitive domains, such as transfer to fluid intelligence, in younger and older adults, with studies

producing overall inconsistent results (e.g., Shipstead et al., 2012; Harrison et al., 2013; Melby-Lervåg and Hulme, 2013; Simons et al., 2016; but see Karbach and Verhaeghen, 2014; Au et al., 2015, 2016).

One limitation of these reviews and meta-analyses is that there is considerable heterogeneity in the tasks used to train and measure WM, making it difficult to evaluate efficacy in the aggregate (cf. Morrison and Chein, 2011; Shipstead et al., 2012; Melby-Lervåg and Hulme, 2013). At the same time, a full understanding of the effects of WM training has been obscured by a literature that is rife with methodological shortcomings. Calls for improved methodological and quantitative standards in cognitive training research are abundant (e.g., Shipstead et al., 2012; Melby-Lervåg and Hulme, 2013; Walton et al., 2014; Simons et al., 2016). Some of the issues clouding the extant literature include the lack of adequate control groups and very small sample sizes. Moreover, any effect of improved WM on other abilities hinges on the assumption that these constructs rely on overlapping cognitive and neural resources that are engaged across multiple domains (cf. Dahlin et al., 2008; Hussey et al., 2016; Lindenberger et al., 2017) and yet there exists a surprising lack of consideration of theoretical mechanisms of training effects and transfer in the literature (cf. Shipstead et al., 2012). One recent attempt to elucidate the benefits of WM training in aging in the context of substantial heterogeneity of training outcomes came from Borella et al. (2017), who performed an integrative data analysis of four training intervention studies from their group that used an identical complex WM training protocol, the same target outcome measures, and similar samples of healthy older adults (total N across studies = 148). This study showed that, in aggregate, there was evidence for near transfer of WM training that was maintained for at least 6–8 months. In addition, they found evidence for immediate transfer of complex span training to measures of reasoning and processing speed, but also showed considerable individual differences in responsiveness to the training (cf. Payne et al., 2012b).

The Current Study

The current study aimed to capitalize on the principles of vertical transfer to examine the degree to which training-related improvements in WM modulate targeted language comprehension functions that putatively rely heavily on WM. Older adults were randomly assigned to either a cognitive training program targeting complex verbal WM or an active control targeting decision speed, both of which were home-based programs delivered via electronic tablets.

Our goal was to address three key issues. First, we were interested in the extent to which WM, as a critical underpinning for language, is plastic and responsive to training-related improvements. Second, we wanted to test the causal hypothesis that WM capacity is a critical resource for language comprehension and memory. Manipulating WM capacity through training and examining its effects on language outcomes afforded the opportunity to directly examine the causal link that is often assumed based on correlational results (Daneman and Merikle, 1996). A number of studies have examined language comprehension as an outcome of cognitive training interventions

(Chein and Morrison, 2010; Shiran and Breznitz, 2011; Carretti et al., 2013a,b; Novick et al., 2013; Karbach and Verhaeghen, 2014; Hussey et al., 2016) but nearly all of this work has focused on healthy young college adults or child populations with specific reading difficulties (e.g., Shiran and Breznitz, 2011; Karbach et al., 2015). Karbach et al. (2015) found evidence in children that adaptive WM training benefited standardized measures of reading comprehension but not measures of math performance or executive control (e.g., inhibition, task switching), suggesting a potentially unique pathway of WM training to comprehension. To our knowledge, only one study, by Carretti et al. (2013b), has specifically examined the effects of WM training on language outcomes among older adults. The training in this study consisted of multiple components, which included not only complex span tasks but also retrieval tasks incorporated into text processing tasks. As such, the improvements observed in language performance may have derived from direct instruction in components of reading comprehension rather than WM processes. In other words, the substantial overlap between the training and transfer task in the Carretti colleagues experiment makes it difficult to evaluate the isolated effects of WM improvement on language. Thus, there has not as yet been a definitive test of the hypothesis that WM training can modulate language performance in older adults, to our knowledge.

Finally, our goal was to develop a model of home-based training using technology that would both offer potential for scaling up for wider use and provide a medium for effective placebo control. In fact, home-based training in other domains has demonstrated good adherence and gains comparable to those observed in the laboratory in older adults (Margrett and Willis, 2006; Payne et al., 2012b; Stine-Morrow et al., 2014). Training tasks were designed to closely match the properties of complex verbal WM tasks in a mobile electronic format that was not only appealing for users, but also provided a detailed record of adherence to the study protocol as well as daily performance gains. We contrasted this with an active-component control group that was comparable to the training task in surface features, feedback, and engagement. Outcomes were measures of complex span tasks that were not directly trained as well as tasks that assessed various aspects of language performance. We were specifically interested in the degree to which WM training would impact immediate memory for sentences, comprehension of sentences that differed in their syntactic complexity, and discourse comprehension and memory.

MATERIALS AND METHODS

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 had previously participated in studies at the Beckman Institute.

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 240 individuals were contacted either by phone or e-mail from our recruitment database, or after expressing interest in the study. Of those, 134 did not follow-up or reply to our invitation to participate in the study. A total of 106 individuals were then assessed for eligibility. Participants were required to be 60 years of age or older, native English speakers with no exposure to other languages before the age of five, normal or corrected-to-normal vision (self-reported), no history of cancer treatment, closed head injury, or traumatic brain injury, no history of Alzheimer's disease, Parkinson's disease, Schizophrenia, or other neurological or psychiatric disorders, not currently taking any psychoactive medications (e.g., anti-depressant, anti-anxiety, anti-seizure), and had not participated in a physical, social, or cognitive intervention study within the previous 3 years. Of those assessed for eligibility, 39 refused to continue participation after learning more about the study, 22 were excluded for not meeting one or more of the inclusion criteria above ($N = 9$ recently participated in an intervention study; $N = 8$ self-reported history of neurological or psychiatric disease; $N = 4$ currently taking a psychoactive medication; $N = 1$ did not meet the age requirement), and three were excluded for other various reasons (e.g., loss of contact, restrictive scheduling constraints).

Thus, a total of 42 individuals were pre-tested. One participant did not meet inclusion criteria at baseline, based on an inability to complete the pre-test cognitive assessment. Thus 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 in **Table 1**, differences between the two groups in age, $t(39) = 0.29$, education, $t(39) = 0.53$, sex, $\chi^2(1) = 0.005$, MoCA score (a clinical tool used for assessing risk for mild cognitive impairment, Nasreddine et al., 2005), $t(39) = 0.42$, and vocabulary score (ETS extended range vocabulary task administered at baseline only), $t(39) = 0.78$ were negligible. Importantly, we adopted an intention-to-treat analysis approach (Hollis and Campbell, 1999; Gupta, 2011), whereby individuals who did not complete the training were actively recruited to participate in post-testing and were included in all analyses. This method results in a conservative test of the treatment effect by de-confounding any potential treatment effects on outcome measures truly due to non-adherence.

Experimental Design and Overview

A pretest–posttest randomized controlled experimental design with an active control group was used to examine the effects of WM training. Participants were asked to complete a total of five 30-min sessions in each week, for a total of 15 sessions over a 3-week period (or 7.5 h of total training). The interval between pre-test and post-test sessions was held constant across participants such that post-testing occurred no more than 4 weeks from the pre-test dates.

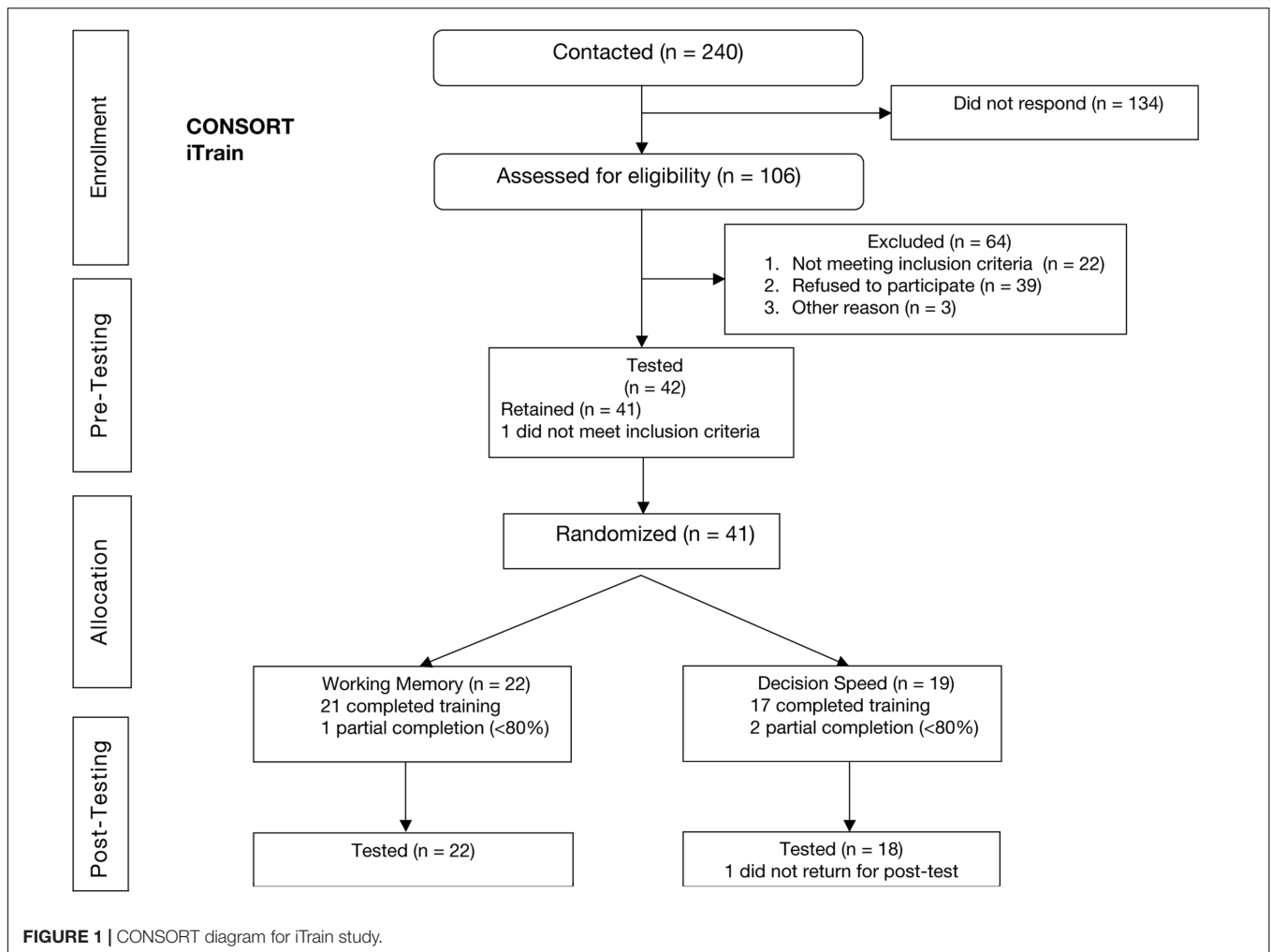


TABLE 1 | Baseline demographics in control and treatment groups.

	Control		Treatment		Diff	95% CI
	M or count	SD or %	M or count	SD or %		
(1) Age	68.11	6.24	67.68	2.77	-0.42	[-2.55, 3.41]
(2) Years of education	17.47	2.38	17.09	2.20	-0.38	[-1.83, 1.07]
(3) MoCA	27.21	2.39	27.77	1.93	0.56	[-0.81, 1.92]
(4) Vocabulary	0.63	0.23	0.68	0.18	0.04	[-0.09, 0.17]
(5) % Female	14	74%	16	73%	0.01	[-0.26, 0.26]

MoCA is Montreal Cognitive Assessment (Maximum = 30, Range = 24–30). Vocabulary is scored as the proportion of accurate items on the ETS Extended Range Vocabulary Test. Diff = mean difference between groups at baseline. There were no significant differences between the treatment and control groups at baseline.

Experimental Groups

Working Memory Training

A novel home-based complex verbal WM training program called iTrain was designed for the study. It was written in Objective-C and implemented for use on iPad tablet computers via the Xcode environment. The program was designed for home-based training to allow participants to complete training sessions without having to make daily visits to the lab while also allowing us to monitor adherence. Prior studies suggest that home-based

cognitive training shows gains on the same order of magnitude as lab-based training (Margrett and Willis, 2006; Stine-Morrow et al., 2014), and also results in high retention rates in healthy older adults in part because participants do not have to travel to the lab daily throughout the course of the intervention.

The three tasks in iTrain – Category Span, Lexical Decision Span, and Sentence Span— were designed to exercise verbal WM by requiring a dual-task load of concurrent language processing and memory storage. In the *Category Span task*, participants were

presented with a semantic category at the top of the screen (e.g., weather) along with a set of single words for which they made validity judgments (e.g., humidity – Yes; chocolate – No). Each trial consisted of the category and target word presented for 4 s. At any point within this duration, participants could decide if the target matched or did not match the category by pressing a “Yes” button or “No” button at the bottom of the screen. Once participants made a decision, the target word would disappear and participants would be presented with accuracy feedback (a green check mark if correct, or a red cross if incorrect) for 1 s. The program would then progress to the next trial within the set. If participants took longer than 4 s to respond, then the target word would disappear to prevent using extra time to develop artificial encoding strategies. However, participants could still respond. If participants failed to respond after a total of 4 s from the target word offset, then the trial would be marked as incorrect and the next trial within the set would begin. At the end of each set, participants were cued to recall each of the words in the order in which they were presented. The cued recall screen consisted of a set of empty text boxes that participants could press and then type their responses via an on-screen keyboard. Participants had no time limit to enter their recall responses at the prompt. Categories and exemplars were drawn from the Van Overschelde et al. (2004) category norms. The final stimulus set included a total of 69 unique categories and over 1500 unique words. Items were drawn randomly such that, within a set, each word had an equal probability of belonging to the presented category or not. Across training sessions, items were rotated through such that all categories had to be selected at least once before a particular category could be repeated again.

In *Lexical Decision Span*, participants were presented with a set of letter strings constituting words (e.g., seek) or non-words (e.g., ceek) and were cued to decide whether or not each string formed a word or not. The letter string was presented for 4 s. At any point within this interval, participants could decide if the letter string was a word or non-word by pressing a “Yes” button or “No” button at the bottom of the screen. Once participants made a decision, the letter string would disappear, and participants were presented with accuracy feedback (a green check mark if correct, or a red cross if incorrect) for 1 s. If participants took longer than 4 s to respond, then the letter string would disappear, but participants could still respond. However, if participants failed to respond after a total of 4 s from offset of the letter string, then the trial would be marked as incorrect and the next trial within the set would begin. Following each lexical decision, an unrelated single letter was presented for 1500 ms for participants to recall at the end of the set. At the end of each set, participants were cued to recall each of the letters in the order in which they were presented. The cued recall screen consisted of a set of empty text boxes that participants could press and then type their responses via an on-screen keyboard. Participants had no time limit to enter their recall responses at the prompt. 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).

Finally, in *Sentence Span*, participants read a series of either semantically congruent sentences or “syntactic prose” sentences (e.g., *As the ship gets better, your child needs to develop this oven*) for which they made sentence acceptability judgments on each sentence (cf. Waters and Caplan, 1996, 2003). Participants had 15 s to read each sentence and make an acceptability judgment by pressing a “Yes” button or “No” button at the bottom of the screen. Once participants made a decision, the sentence would disappear, and participants were then presented with accuracy feedback (a green check mark if correct, or a red cross if incorrect) for 1 s. If participants took longer than 15 s to respond, then the sentence would disappear. If participants failed to respond after a total of 5 s from the offset of the sentence, then the trial would be marked as incorrect and the next trial within the set would begin. At the end of the set, participants were cued to recall the last word of each sentence in the order in which they were presented. The cued recall screen consisted of a set of empty text boxes that participants could press and then type their responses via an on-screen keyboard. Participants had no time limit to enter their recall responses at the prompt. Acceptable sentences were adapted from two sources. The Nelson and Narens (1980) general information question norms provided 244 sentences. The other source was the Manually Annotated Sub-Corpus (MASC) of the Open American National Corpus (Ide et al., 2013), which provided 301 sentences that ranged widely in topic, length, and syntactic structure. In addition, 346 unacceptable sentences were adapted from the “syntactic prose” conditions in earlier studies by Lee and Federmeier (2011) and Payne et al. (2015). Unacceptable sentences have syntactically well-formed sentence frames, but contain no coherent message-level semantics. All sentences ranged between 60 and 90 characters, and all sentence final words were between 4 and 9 characters.

The training was designed to be individually adaptive (cf. Lustig et al., 2009; Karbach et al., 2015), such that for all three tasks, the set size (number of items to recall within a trial) adaptively changed according to current performance. In this way, each participant was always engaging in the task at a level that was matched to his or her current ability. Task difficulty was programmed to follow a step function, such that when recall was perfect on set size n , the set size for the next set was increased to $n+1$. If perfect recall was not achieved at set size n , the set size was reduced to $n-1$. At the end of each set, feedback was presented to participants on both the accuracy of the judgment task (proportion correctly judged) and the proportion of items correctly recalled. The presentation order of the three tasks was randomized across session. The memory set size on the first session began at $n = 2$ for all tasks and subjects. The memory set size at the end of each training session was saved so that participants began the following session at the memory set level from their prior training session. All timing and set size parameters were based on extensive norming and testing of the iTrain software during its development. The source code for the training can be viewed and downloaded in full at: <https://github.com/TALL1532/itrain>.

Active Component-Control Group

A component-control design (Mohr et al., 2009; Boot et al., 2013) was adopted in designing the active control group. In a component-control design, a multi-component intervention serves as the focal treatment and 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 reduces the likelihood of placebo effects or differential expectancies for change (Boot et al., 2013).

Participants in the active control group completed the same three tasks as in the treatment group without the recall component. Thus, in the *Category Task*, participants practiced making speeded category judgments; in the *Lexical Decision Task*, participants practiced making speeded lexical decision judgments; and in the *Sentence Task*, participants practiced making speeded semantic acceptability judgments. The items were identical to the WM training. Importantly, both the treatment and control groups were matched in their exposure to stimuli, the absolute magnitude of time allocated to training (15 30-min sessions over 3 weeks), and the amount and type of linguistic exposure. Thus, findings comparing the treatment and active control groups are controlled for exposure to linguistic stimuli, an important factor given the putative relationship between verbal WM and language experience (cf. MacDonald and Christiansen, 2002; Wells et al., 2009; Payne et al., 2012a, 2014a). 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 maintain continued interest in the task, de-confound memory load from task adaptivity, and reduce the potential for differences in expectancy for training benefits in the two groups (Boot et al., 2013).

In the control training, participants were presented with stimuli in blocks of 15 items and told to make their judgments (lexical decision, category, sentence acceptability) as quickly as possible. The starting presentation times for each task were identical to the presentation times in the WM training task (described above). However, as participants improved in accuracy in the judgment decisions, presentation rates were increased at a rate of 5% across blocks. When accuracy fell below 80%, the presentation rate was decreased, so that task adaptivity followed a similar step function as in the WM training. Participants were encouraged to practice speeded decisions in each of the linguistic tasks while maintaining high accuracy. A “speed level” score, derived from change in presentation rate from the initial training block, was provided after each block, so that participants could monitor their progress from the first block of the first session to the end of the training, as in the WM training protocol.

Assessment Battery

The cognitive battery, administered at pre-test and post-test, was chosen to target both complex WM performance as well as measures of off-line language performance. Language outcome measures included assessments of sentence processing and discourse memory, which were themselves graded in terms of

their reliance on WM (Stine and Wingfield, 1990; Jefferies et al., 2004; Stine-Morrow et al., 2008), as well as a measure of verbal fluency. At post-test only, a survey to gauge group differences in expectations for cognitive change was administered based on a survey designed by Boot et al. (2013).

Complex Verbal Working Memory

Four complex WM tasks were administered using the Psychophysics Toolbox in MATLAB (Brainard, 1997), adapted from the CogToolbox (Fraundorf et al., 2014). In all four tasks, participants made a series of judgments about each item in a set of verbal stimuli and then, after the set, recalled information related to each item within that set. Alternate forms of each task were administered at pre-test and post-test. In the (1) *reading span* and (2) *listening span* tasks (Daneman and Carpenter, 1980; Stine and Hindman, 1994), participants read or listened to a set of simple declarative sentences (e.g., “A book is often found in a library”), and judged whether the sentence was true or false. Additionally, participants were asked to recall the sentence-final words (e.g., library) after each set. The number of sentences per set increased with progress through the task (until eight sentences per set or when the participant could no longer recall each of the target words in a set successfully). If the participant could not recall all items at a particular set size, a second trial was administered. If the participant could not recall all items within the second trial within that set, the test would terminate (cf. Stine-Morrow et al., 2001; Waters and Caplan, 2003; Payne et al., 2014a). The score was the number of target words recalled from 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 used the same administration and scoring, except that the sentences were presented in the auditory modality. For the reading span task, the minimum sentence presentation was 1s and the maximum was 7s. In both the sentence and reading span tasks, the maximum time to make true/false judgments was 2 s. In the (3) *operation span* task (Turner and Engle, 1989; Conway et al., 2005), the participant was cued with a series of three-term math problems (e.g., $[8/2] - 1 = 3$; True), followed by a letter (e.g., c) to hold in memory after each problem. Following each problem-item set, the participant recalled the set of letters in the order in which they were presented. Fifteen sets were presented randomly, with set size ranging between 3 and 7. Because there are individual differences in the amount of time participants take to solve arithmetic problems, the presentation rate during the memory test was set on a subject-by-subject basis by using a baseline calibration period (see Unsworth et al., 2005). Prior to the onset of the memory task, participants completed a practice block solving math problems without the memory task. The average time it took to solve each problem was calculated separately for each subject, and the maximum presentation rate was calculated as the subjects mean calibration time plus 2.5 *SD*. The total score was the total proportion of correct items in the correct position across all sets (Unsworth et al., 2005). In the (4) *Minus-2 span task* (Waters and Caplan, 2003), participants were presented with a string of digits one at a time with an SOA of 1 s and cued to produce the series with two subtracted from each digit (e.g., [8, 4,

3, 9] to [6, 2, 1, 7]). The total score was the total proportion of correct items in the correct positions across all trials (Waters and Caplan, 2003). Each span task was preceded by practice at the lowest set size. Note that the selected WM span measures vary with respect to their overlap with surface features (e.g., secondary judgment task, source of elements to recall, scoring criteria, set size order, cf. Was et al., 2011).

Sentence Memory

An immediate recall task (Zelinski and Lewis, 2003; Stine-Morrow et al., 2008) was administered in which participants read eight 18-word sentences with presentation time self-paced, and immediately recalled each sentence for later transcription and scoring. Recall was scored as the proportion of individual words correctly recalled (e.g., Potter and Lombardi, 1998; Gilchrist et al., 2008), which for these brief single sentences, was found to correlate very strongly with propositional recall scoring ($r = 0.91$) (Kintsch and van Dijk, 1978; Stine-Morrow et al., 2008). Alternate sentence sets of equivalent difficulty were presented at pre-test and post-test.

Syntactic Comprehension

Participants read a series of sentences and answered a simple a yes/no comprehension question after each sentence. Comprehension was assessed for three different types of syntactic complexity that are known to cause comprehension difficulty among older adults, and have been suggested to increase load on WM capacity: (1) *garden-path* syntactic ambiguities, (2) long-distance relative-clause dependencies (Bartek et al., 2011), and (3) object-relative clauses (see **Table 2** for examples)¹. Sentences were counterbalanced across conditions at each testing occasion, so that each sentence was equally represented in the high- and low-demand conditions. At pre-test and post-test, participants read 20 items from each sentence set (10 low complexity, 10 high

¹ Participants read these sentences while having their eye-movements monitored to examine changes in on-line language processing. The eye-tracking data showed no reliable changes over time nor effects of the intervention.

TABLE 2 | Example items in sentence comprehension test as a function of syntactic demand.

Sentence set	Complexity	Sentence
GP	Low	While the man hunted, the deer that was brown and graceful ran into the woods.
GP	High	While the man hunted the deer that was brown and graceful ran into the woods.
SR/OR	Low	The farmer that knew the barber asked for a loan.
SR/OR	High	The farmer that the barber knew asked for a loan.
LDD	Low	The administrator who the nurse supervised scolded the medic for being late.
LDD	High	The administrator who the nurse who was from the clinic supervised scolded the medic for being late.

GP, *garden path*; SR/OR, *subject/object relative clause*; LDD, *long distance dependency*.

complexity), resulting in a total of 60 sentence-question pairs at each measurement occasion.

Discourse Comprehension and Memory

The Nelson-Denny Standardized Reading Comprehension subtest, to assess general reading comprehension ability, consists of eight prose passages and 36 multiple-choice questions. Participants were given 20 min to read the passages and answer the questions. Alternate forms were administered at pre-test and post-test. In the Rivermead Behavioral Memory Task Paragraph recall subtest (Wilson et al., 2003), participants listened to a short narrative for immediate recall. Production was coded and scored for the number propositions correctly recalled.

Verbal Fluency

Verbal fluency was assessed with the FAS phonemic fluency task (Benton and Hamsher, 1978). In this task, participants were given a letter (at pre-test “F”, “A”, and “S”) and asked to produce as many words that they could think of that begin with that letter for 60 s. 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 and McDowd, 2008) as well as predictive language comprehension (Federmeier, 2007) in older adults. An alternate form, the BDT, was used at post-test (Strauss et al., 2006).

Perceptions of Training Benefits

A 14-item survey to assess individuals’ expectations for the effects of training (cf. Boot et al., 2013) was administered at the end of the post-test session. Items probed whether (1) they perceived general improvement in cognition as a function of training (e.g., “I believe that iTrain helped improve my cognition”), and (2) they improved on specific tasks (e.g., for the Listening Span task, “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?”). Participant’s read each statement and then were asked to endorse those statements on a Likert scale from 1 (Strongly Disagree) to 5 (Strongly Agree).

Procedure

At the onset of the study, all participants completed the cognitive battery in a single 3-h laboratory session. Following the pre-test battery, participants were given an iPad 2 tablet computer containing either the complex WM 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, and participants were given the opportunity to practice the tasks in the lab until they understood each task completely. Participants returned to the lab at the end of the training for post-test. The testing was single blind, as testers were aware of the random assignment condition. However, testing sessions were designed to minimize the amount of contact with the participant, and testers were instructed to provide no

identifying information regarding the training program or the study hypotheses.

RESULTS

A series of linear mixed effects models were used to test for the effects of the intervention on each outcome measure. Analyses focused on effect size estimation and quantification of the precision of these effects via confidence intervals (Kelley and Preacher, 2012; Lakens, 2013; Cumming, 2014). Effect sizes and 95% profile confidence intervals of the critical Training Group (Control vs. Training) \times Time (Pre-test vs. Post-test) interactions were estimated via restricted maximum likelihood estimation, with random intercepts specified for subjects, and by-subject random slopes for the within-subject Time factor. For the syntactic comprehension data, the critical interaction was a Group \times Time \times Sentence Type effect and, for these models, the Time \times Sentence Type interaction was additionally modeled as a random slope (see Barr et al., 2013; Bates, under review for discussions on the treatment of random slopes). Note that, following an intention-to-treat protocol, all participants were invited back for post-testing and included in all analyses, regardless of the number of sessions that were completed. Thus, models were fit to all available data for each outcome. Treatment coding was used for all fixed-effects factors and statistical inference was limited to the critical interactions that would provide statistical support for group differences in the change in each outcome from pre-test to post-test. Because sample sizes are small, supplemental non-parametric analyses were conducted using a robust bootstrapping approach as described by Kirby and Gerlanc (2013), to estimate the standardized effect size and precision of group differences in change in the outcome measures.

Perceptions of Training Benefit

Table 3 presents mean rating endorsements for both general improvement in cognitive ability, as well as improvement across specific tasks. Overall, average endorsement rates ranged from neutral to positive. Importantly, there was no difference between the treatment and control groups in expectations that they improved in overall cognition following training ($b = 0.13$; 95% CI $[-0.15, 0.41]$). There was a trend for the WM tasks to show self-reports of greater improvement in the treatment group relative to the control. However, only for one task—the minus-2 span task—did the group difference reach statistical significance, though this effect was quite small ($b = 0.65$; 95% CI $[0.004, 1.29]$). Importantly, the language tasks showed no evidence of differential expectation for improvement between the control and training groups, with the trend going in the direction of greater perceived improvement in language tasks in the control group (see **Table 3**). Thus, it is unlikely that the effects of training on language outcomes reported below could be attributed to differential expectations for improvement in the training group relative to the active control group.

Training-Related Changes in WM Performance

Trial-level performance was collected by the iPad over the course of training, enabling us to compute the average span score for individuals at each session for each span task. Based on participants who completed at least 80% of the training ($n = 21$), **Figure 2** plots the session-to-session effects of WM training on performance gains for each of the three verbal WM tasks. Raw scores on the span tasks were converted to a metric of percent change from baseline assessment (i.e., their average score on the first day of training) in order to assess the relative degree of improvement from baseline on a similar scale for each span task (e.g., Chein and Morrison, 2010). On average, training gains followed a non-linear trajectory, with larger improvements in early sessions. Indeed, the largest improvements across the three tasks occurred from session 1 to session 2. Over the 15 sessions, trainees showed an approximate 60% peak training improvement from baseline on the category and sentence span tasks and more than doubled their span performance on the lexical decision span task relative to their performance on the first day of training.

To examine individual differences in performance, we calculated subject-specific learning curves, expressed as the percent of change from baseline performance. To accommodate the non-linearity in training gains, a natural cubic smoothing spline was fit to the training data for each participant. Following this, the area under the cubic spline curve was estimated over the training period separately for each participant (using an adaptive quadrature algorithm via the MESS package in R; Venables and Ripley, 2002) as a summary index of non-linear training gains in each task for each individual across the 15 sessions. **Figure 3** plots the bivariate scatterplot matrix among training gains for the three span tasks. With correlations above 0.85, it is apparent that training-related improvements on the trained memory span tasks clustered together tightly, suggesting that training-related improvements occurred broadly and to a similar degree across all tasks, and were thus not likely due to isolated to task-specific strategy development (which would attenuate the correlation among the training-related improvements).

Transfer to Working Memory and Language

Table 4 presents pre-test and post-test mean scores and difference scores for the WM and language tasks administered in the cognitive test battery separately for the control group and treatment groups. Note that, following an intention-to-treat approach, these results reflect all available data at pre-test and post-test, regardless of participants' training adherence. **Figure 4** presents summary bootstrapped effect sizes of the group differences in change in each of the tasks (e.g., the Group \times Time interaction) in units of standard deviation change (Cohen's d) separately for the WM and language tasks. Larger values indicate a positive change from pre-test to post-test that was larger for the treatment group than the control group. Appendix A presents correlations between the WM measures and language tasks at baseline.

TABLE 3 | Results from post-test expectation survey.

	Control		Training		Group difference	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>Diff</i>	95% <i>CI</i>
Reading span	3.23	0.21	3.55	0.21	0.31	[−0.22, 0.84]
Listening span	3.15	0.18	3.5	0.18	0.35	[−0.11, 0.81]
Minus-2 span	3.23	0.21	3.89	0.21	0.65	[0.004, 1.29]
Operation span	2.91	0.16	3.47	0.16	0.56	[−0.03, 1.15]
Syntactic comprehension	3.38	0.16	3.23	0.16	−0.16	[−0.69, 0.27]
Text memory	3.23	0.23	3.00	0.23	−0.23	[−0.23, 0.36]
Reading comprehension	3.62	0.29	3.27	0.29	−0.34	[−1.06, 0.38]
Overall cognition	3.34	0.12	3.47	0.10	0.13	[−0.15, 0.41]

Items on a Likert scale with 1 indicating low endorsement of training benefit and 5 indicating high endorsement of training benefit. Both groups rated expectations of improvement slightly above neutral (3) across nearly all survey items. 95% confidence intervals containing 0 indicate differences that are not statistically significant at $p < 0.05$. The only item that reached a significant difference was for the Minus-2 span task.

Working Memory

There were no reliable baseline differences in any WM task between treatment and control groups (all t 's < 1 , see **Table 4**). As can be seen in **Table 4** and **Figure 4**, there was evidence for training-related improvements in verbal WM. All four WM tasks showed effect sizes larger than $d = 0.50$, indicating at least an approximate half standard deviation difference between the treatment and control groups in change in WM. The average effect size of training collapsing across the four tasks was $d = 0.87$. However, 95% confidence intervals were quite large for all of the tasks, suggesting substantial individual differences in responsiveness to the intervention (cf. Payne et al., 2012b). Results from linear mixed-effects models fit to each task showed reliable Group \times Time interactions for Listening Span ($b = 1.32$, 95% CI [0.54, 2.12]), Operation Span ($b = 0.24$, 95% CI [0.08, 0.40]), and Minus-2 Span ($b = 0.10$, 95% CI [0.02, 0.18]), indicating greater improvement in span for the training group relative to the control group. Of the four tasks, Reading Span showed the weakest effects, with a negative lower-bound on the bootstrapped 95% confidence interval. Consistent with the bootstrapped effect sizes, the linear mixed-effects model revealed a non-significant Group \times Time interaction ($b = 0.61$, 95% CI [−0.12, 1.34]) for reading span only.

Language Outcomes

There were no reliable baseline differences between treatment and control groups on any measure (all t 's < 1.3 , see **Table 4**). As shown in **Table 4** and **Figure 4**, verbal fluency and sentence memory showed evidence of improvement from WM training relative to the control group. The Group \times Time interaction was significant for verbal fluency ($b = 6.57$, 95% CI [1.30, 11.79]) and sentence memory ($b = 0.08$, 95% CI [0.02, 0.14]), such that the WM training group showed a larger increase compared to the control group. In contrast, the two tasks assessing discourse understanding and memory, the Rivermead and the Nelson-Denny tasks, showed no evidence of training-related improvements as the Group \times Time interaction was not reliable for either the Nelson-Denny ($b = 0.002$, 95% CI [−0.11, 0.12]) or the Rivermead discourse memory task ($b = -2.85$, 95% CI [−11.61, 5.91]).

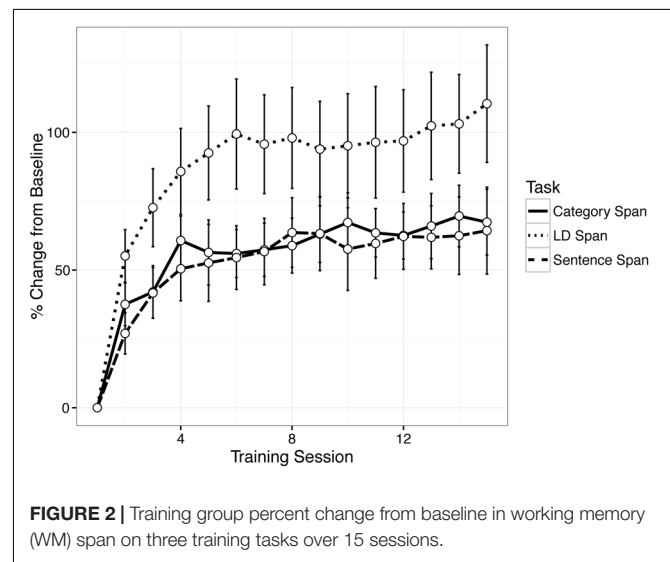


FIGURE 2 | Training group percent change from baseline in working memory (WM) span on three training tasks over 15 sessions.

Finally, WM training showed isolated effects in improving comprehension of ambiguous syntactic forms. **Table 5** presents pre-test and post-test mean scores for the low- and high-demand conditions of each of the three syntactic comprehension sets separately for the control group and treatment groups. For the garden-path sentences, a reliable ambiguity effect was observed at baseline ($b = 0.14$, 95% CI [0.06, 0.23]), such that ambiguous sentences had poorer accuracy than unambiguous sentences. In the model testing training effects on ambiguity resolution, a reliable Syntactic Demand \times Time \times Treatment interaction was found ($b = -0.18$; 95% CI [−0.34, −0.01]), indicating that there were training group differences in the change in accuracy from pre-test to post-test that differed for the ambiguous sentences compared to the unambiguous sentences. This interaction is depicted in **Figure 5**. As can be seen, accuracy was high across both training groups for syntactically unambiguous sentences at pre-test and post-test. For the ambiguous sentences, however, sentence comprehension is poorer at baseline and only the WM training group showed improvement in accuracy from pre-test to post-test.

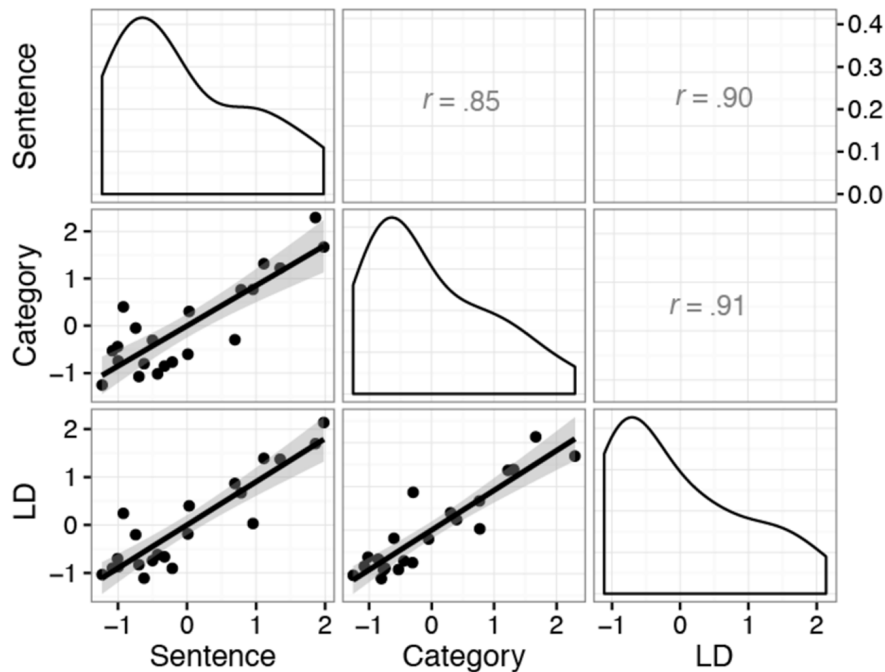


FIGURE 3 | Scatterplot matrix of the relationship between training improvements over 15 training sessions across each of the three WM training tasks. The lower diagonal cells are bivariate scatterplots and best-fit linear functions of the relationship between WM training gains across each pair of tasks. The upper diagonal presents the corresponding Pearson correlation coefficient for each pair of training tasks. The diagonal cells plots the estimated probability density functions of the individual training improvements in each task. Note that training improvements were z-score standardized, so that 0 represents average average training improvement, and a 1 unit increase represents an approximate 1 SD improvement.

TABLE 4 | Pre-test, post-test, and change in working memory and language measures in control and treatment groups.

	Control						Treatment					
	Pre-test		Post-test		Δ		Pre-test		Post-test		Δ	
	M	SE	M	SE	M	95% CI	M	SE	M	SE	M	95% CI
Verbal working memory												
(1) Reading span	2.74	0.25	2.98	0.34	0.22	[-0.20, 0.65]	2.85	0.20	3.70	0.22	0.84	[0.28, 1.40]
(2) Listening span	3.50	0.37	2.88	0.31	-0.64	[-1.31, 0.03]	3.68	0.16	4.40	0.19	0.72	[0.24, 1.18]
(3) Operation span	0.40	0.07	0.44	0.06	0.02	[-0.04, 0.09]	0.39	0.05	0.63	0.05	0.30	[0.19, 0.41]
(4) Minus-2 span	0.58	0.05	0.63	0.06	0.01	[-0.03, 0.05]	0.63	0.04	0.76	0.02	0.12	[0.03, 0.21]
Language transfer												
(5) Nelson-Denny	0.87	0.02	0.82	0.04	-0.06	[-0.13, 0.01]	0.86	0.02	0.81	0.05	-0.05	[-0.11, 0.02]
(6) Verbal fluency	41.00	2.47	42.71	2.07	0.71	[-3.15, 4.56]	43.82	2.22	51.36	3.00	7.55	[3.96, 11.13]
(7) Sentence memory	0.66	0.04	0.71	0.03	0.04	[-0.01, 0.07]	0.65	0.02	0.77	0.02	0.12	[0.08, 0.16]
(8) Discourse memory	0.46	0.04	0.44	0.04	-0.02	[-0.86, 0.43]	0.52	0.03	0.47	0.03	-0.05	[-0.11, 0.01]

Reading span and listening span scores are the number of target words recalled from the highest set with no errors, plus a fraction reflecting the proportion of correctly recalled words on the set with an error. Operation span, Minus-2 span, Nelson-Denny, sentence memory, and discourse memory are scored as proportion correct. Verbal fluency is total number of words produced. Δ is the difference between pre-test and post-test. 95% Confidence Intervals containing 0 indicate differences that are not statistically significant at $p < 0.05$. There were no significant differences between Control and Treatment groups on any measure at baseline.

For the subject and object relative sentences, a reliable OR-cost effect was observed at baseline, such that object-relative sentences had poorer accuracy than subject-relative sentences ($b = 0.13$; $0.04, 0.22$). In the model testing training effects on SR/OR comprehension, a small Syntactic Demand \times Time \times Treatment interaction was observed ($b = -0.18$; 95% CI $[-0.35, -0.007]$).

However, the nature of this interaction was difficult to attribute strictly to training-related improvements in the costs associated with object-relative processing, as the interaction was driven by the WM training group showing a relative improvement in object-relative comprehension from pre-test to post-test (as predicted), but a corresponding decline in improvement in the

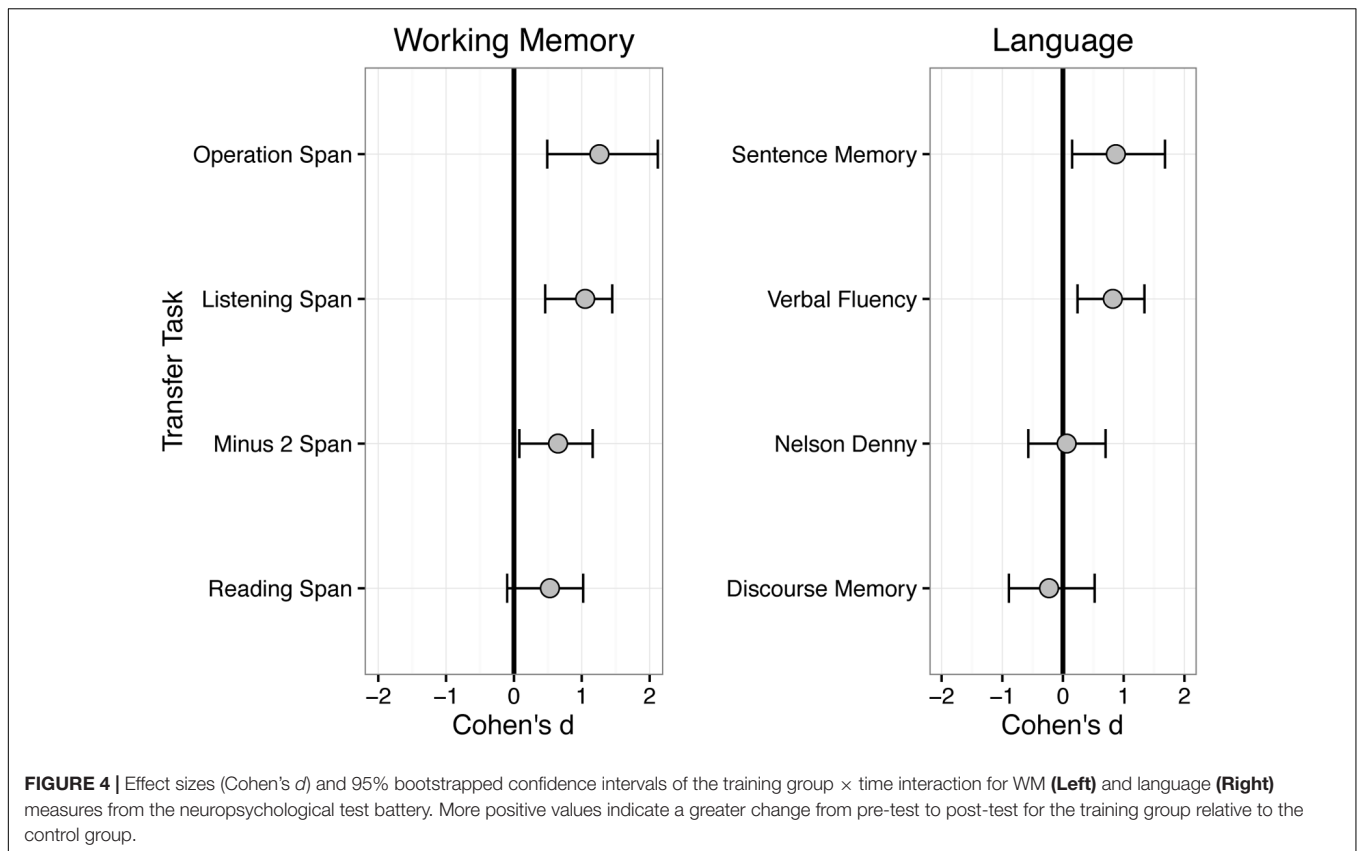


TABLE 5 | Syntactic comprehension in control and treatment groups at pre-test and post-test.

	Control				Treatment			
	Pre-test		Post-test		Pre-test		Post-test	
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>
Garden path								
Unambiguous	0.82	0.02	0.83	0.03	0.85	0.02	0.84	0.02
Ambiguous	0.67	0.03	0.66	0.03	0.65	0.04	0.81	0.03
Subject-object relative								
Subject-relative	0.81	0.04	0.79	0.03	0.87	0.03	0.76	0.03
Object-relative	0.68	0.04	0.68	0.04	0.65	0.04	0.74	0.04
Long-distance dependency								
Short-distance RC	0.68	0.03	0.71	0.03	0.70	0.04	0.71	0.03
Long-distance RC	0.69	0.05	0.58	0.04	0.73	0.04	0.70	0.02

simpler subject-relative sentences, which is not the expected pattern following WM training.

For the long-distance dependency sentences, surprisingly, there was no reliable LDD-cost observed at baseline ($b = -0.007$; 95% CI $[-0.11, 0.10]$), as accuracy was approximately equivalent between relative clauses with and without a long-distance dependency introduced between the head noun and the relative-clause verb. Although there was a trend for a Syntactic Demand × Time × Treatment interaction ($b = -0.21$; 95% CI $[-0.41, 0.005]$) that did not reach statistical significance, like the SR/OR sentences, the nature of the interaction was difficult to

attribute to training-related improvements accuracy for the more complex long-distance dependency case (see Table 5).

DISCUSSION

It is often assumed that normative age-related changes in basic cognitive functions such as WM capacity compromise the ability to comprehend language and learn from complex texts. However, nearly all of the evidence for a role of the WM system in language comprehension is derived from correlational studies, which are

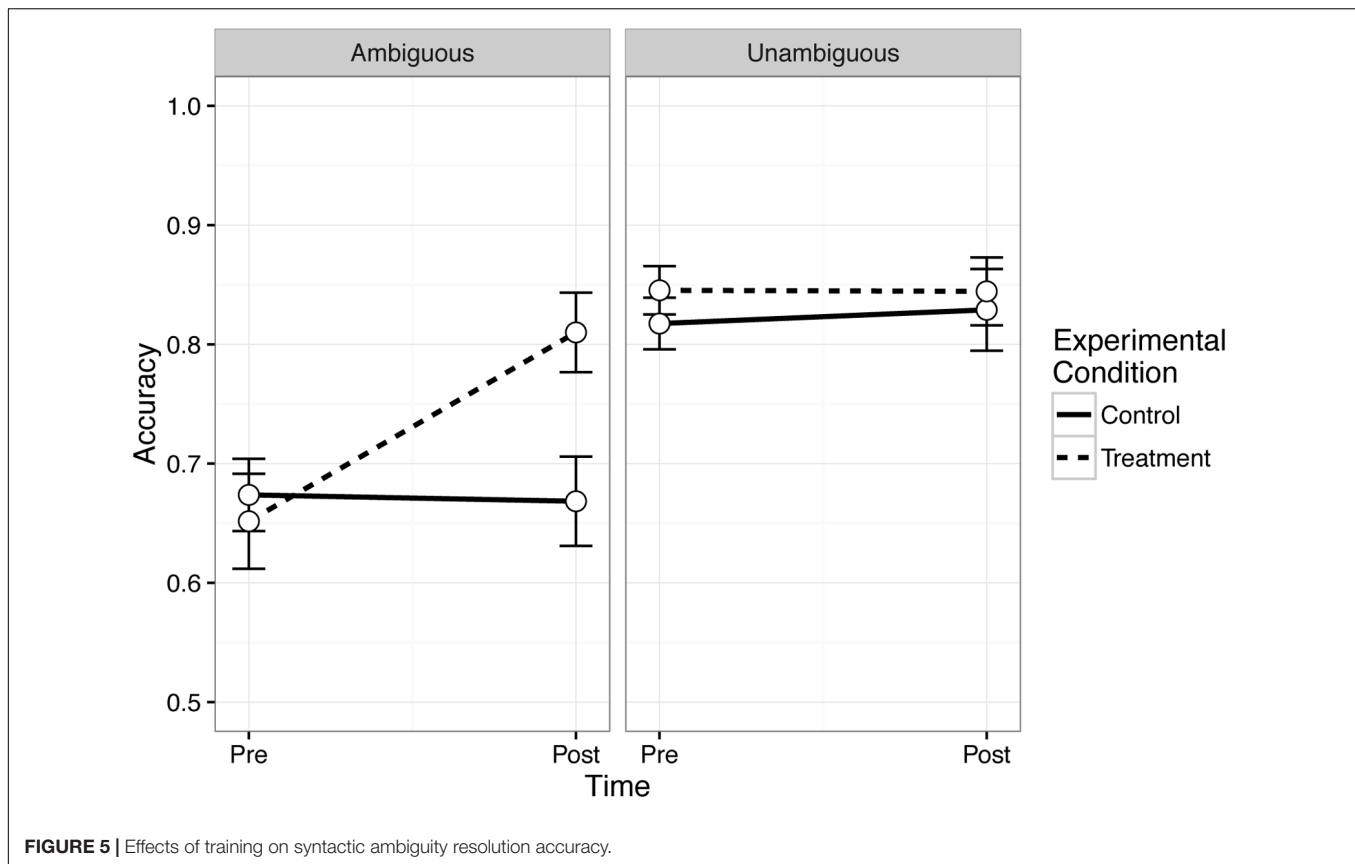


FIGURE 5 | Effects of training on syntactic ambiguity resolution accuracy.

inherently limited. In order to resolve this causal ambiguity, the current study exploited an experimental design to examine the degree to which cognitive training in verbal WM could transfer to aspects of language comprehension among older adults using a novel home-based cognitive training program.

The data presented in the current study yield important insights into both the nature of the verbal WM system and the degree of plasticity in language comprehension among older adults. Specifically, our results indicated that verbal WM is capable of short-term change in adults over the age of 60 through less than 10 h of home-based training over the course of 3 weeks, and that this training showed some evidence of transfer to untrained verbal memory measures as well as measures of language fluency, language memory, and syntactic comprehension. In summary, our findings suggest that WM is plastic in later adulthood, at least in the short-term. These data, while preliminary in nature, are among the first to indicate that selective aspects of WM-dependent language performance can be modified through targeted practice in WM in older adulthood.

Home-Based Working Memory Training

The benefits of home-based training via tablet computers include convenience for the participant and a reduction of resources devoted to weekly testing sessions in the lab. Moreover, lab-based training may lead to biased sampling of study participants who are highly mobile, healthy, and able to allocate substantial amounts of time each week to participating in laboratory sessions.

In contrast, home-based training is likely to lead to more heterogeneous sampling at both ends of the ability distribution (i.e., high-ability adults with substantial time and scheduling constraints that prevent participating, as well as lower ability and lower-mobility adults), as it reduces the burden on the participant to complete daily lab visits 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, which may impact training responsiveness. Very few studies have examined the effects of home-based cognitive training in older adults. In two experiments, Wadley et al. (2006) directly compared training gains in a useful-field-of-view training program among healthy older adults in laboratory and home settings. Both groups showed significant improvements in processing speed relative to a control group that underwent no training. However, 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., 2012b for similar evidence in a home-based reasoning training). 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 (e.g., Payne et al., 2012b). Note however, that Wadley et al.'s (2006) findings could also be attributed to the adaptive nature of the in-lab training, which was not replicated in the

home version. Nevertheless, the data from the current study indicate that self-administration of the WM training is feasible.

A key test of the effectiveness of the training program was the assessment of the degree to which training led to improvements in untrained complex verbal WM span. There was positive evidence for improvement 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, three of the four tasks reaching statistical significance, and a pooled effect size of $d = 0.87$. Thus, the evidence from the current study suggests that home-based training of WM can be effective in improving both trained and untrained complex verbal WM span in the short-term. Note however, that the effects of training on the reading span task did not reach statistical significance despite the fact that this measure appeared to have the highest overlap in surface features to one of the training tasks (the sentence span task) while the same outcome in the auditory modality (listening span) did show reliable evidence for near transfer. One reason for the reduced effect size of training on reading span relative to listening span may be due to the self-paced nature of the reading span measure. Prior studies have found that self-paced administration influences estimated WM capacity, as well as the validity of the measure for predicting higher-order cognition (Friedman and Miyake, 2004; Lépine et al., 2005; Clair-Thompson, 2007; Barrouillet et al., 2008). It may be the case that self-pacing in the processing component of the task influences the magnitude of transfer as well.

Working Memory and Language Understanding

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 in older adults. Adults in the WM training group showed differentially larger improvements in both sentence memory and verbal fluency relative to the active control group. It is perhaps unsurprising that short-term sentence memory showed transfer, as sentence memory performance is highly related to WM (Stine-Morrow et al., 2008; Lewis and Zelinski, 2010; Payne et al., 2012a), and, at least for the reading span task, overlaps to some degree in task demands (MacDonald and Christiansen, 2002; cf. Roberts and Gibson, 2002). However, demonstrating training-related transfer to sentence memory in older adults is critical for at least two reasons. First, although verbal WM and sentence memory 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 upward of 50% of the same variance (Engle, 2010), and yet evidence for transfer of WM training to fluid intelligence has been inconsistent (see Shipstead et al., 2012; Melby-Lervåg and Hulme, 2013, 2016, for reviews). Second, sentence memory shows some of the largest effect sizes for age-related declines among measures of language comprehension and episodic memory (Johnson, 2003; Stine-Morrow et al., 2008). Demonstrating reductions in age-related deficits in language memory is thus

quite valuable for future applications in memory remediation in older adulthood.

The demonstration that WM training transferred to verbal fluency indicates that training can lead to transfer to tasks that share very little overlap with the tasks involved in the training. At the same time, interpreting training effects on verbal fluency are complicated 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 functions including executive functioning (Mayr and Kliegl, 2000), semantic processing efficiency (Troyer et al., 1997), frontal-lobe mediated generative language production (Federmeier, 2007), and lexical knowledge (Nagels et al., 2012). Future work should focus on the role that WM training plays in improving executive control components related to aspects of language production and semantic processing. Nevertheless, given the strong relationship between fluency and language comprehension and production in older adulthood (Federmeier et al., 2002, 2010; Federmeier, 2007; Wlotko et al., 2012), such findings are promising from both applied and basic perspectives.

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., 2014b). While this may be surprising given that prior work has shown that reading comprehension and discourse recall are correlated with WM, one explanation is that age-related declines in discourse understanding are actually quite rare (Stine and Wingfield, 1990; Radvansky, 1999; Radvansky and Dijkstra, 2007). To some extent, this may be due to the reliance of discourse comprehension on the establishment of a situation model, a level of understanding that is robust to cognitive aging (Radvansky and Dijkstra, 2007; Stine-Morrow and Radvansky, in press). Under this account, older adults can rely on situational representations as a compensatory mechanism in order to maintain comprehension despite reduced memory resources. However, for the context-independent sentence memory task, where it is less likely that a situational representation can be established, WM effects are larger, and effects of training are found.

Finally, we tested the degree to which syntactic comprehension accuracy was modulated by WM training. Results were mixed. Sentences that were unambiguous but more syntactically complex (e.g., SR/OR and LDD sentence sets) did not produce the expected pattern of training-related improvements. Only in syntactically ambiguous garden path sentences was there positive evidence for WM-specific-improvements in comprehension of more syntactically difficult sentences. Both the treatment and control groups showed the canonical garden-path ambiguity effect in comprehension (Christianson et al., 2001, 2006) at baseline. At post-test, only the WM training group showed evidence for reduced ambiguity effects on comprehension. This effect was driven by a selective increase in comprehension for the more demanding syntactically ambiguous items. Note that these findings are similar to those of Novick et al. (2013) and Hussey et al. (2016) who have found evidence that younger adults trained on the *n*-back task

with lures showed improvements in comprehension of similar garden-path ambiguities.

One explanation is that WM affords the capacity to maintain multiple alternative syntactic representations of ambiguous phrases, which can be directly accessed at the point of disambiguation. Low-span readers are unable to maintain multiple syntactic representations, and therefore commit to one interpretation, causing subsequent difficulties when they must revise their incorrect interpretation (MacDonald et al., 1992; Kemper et al., 2004). Consistent with this account, Christianson et al. (2006) showed evidence for a robust negative correlation between verbal WM span and the probability of incorrectly interpreting garden path sentences in older adults. These findings suggest that older adults with low WM have particular difficulties in revising an initially incorrect interpretation (see also Payne et al., 2014a for similar evidence in syntactic attachment ambiguities). The training data presented here corroborate prior correlational results and extend these by suggesting that the WM system subserving ambiguity resolution is plastic and is responsive to memory training.

Limitations and Future Research

The primary limitation in the current study is that the small sample size limited our power to detect small-to-moderate effect sizes. The issue of small sample sizes is widespread in the WM training literature. This issue is largely driven by the severe resource constraints associated with conducting adequately powered cognitive training studies in special populations due to issues with staffing, recruitment, retention, maintenance of intention-to-treat protocols, and additional costs of conducting longitudinal randomized controlled trials with large sample sizes. One way in which sample sizes may be increased without substantially increasing costs and resources is through home-based training and assessments, as these approaches require fewer resources to be allocated to each individual subject for daily laboratory visits. Thus, one goal of this work is to illustrate that home-based training is a feasible and valid option for future studies and may be able to help move toward scaling up studies to optimally powered sample sizes to detect more nuanced and reproducible effects of training. Because the current study is our first attempt at targeting language comprehension in older adults using home-based WM training, the results should be interpreted with caution. Future planned work will aim to replicate these results with larger and more diverse samples and continue to follow best practices for cognitive intervention, including pre-registration and examining follow-up and maintenance effects (cf. Simons et al., 2016).

Despite our limited sample size, 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 an appropriate control group in the context of a literature in which inadequate control groups negatively impact many studies. Because treatment and no-contact 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-test. 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. In addition, an intention-to-treat approach was used, in order to downwardly bias effect sizes with differential drop from the training (Hollis and Campbell, 1999). However, because the home-based training resulted in such high retention, the issue of differential drop-out causing the observed training benefits is not plausible.

Despite the relative breadth of the measurement battery for assessing language, it was designed to primarily tap into comprehension processes and not language production. However, there is a growing literature posing a strong relationship between WM, WM limitations (e.g., through aging and brain damage), and language production mechanisms (Acheson and MacDonald, 2009; Martin and Slevc, 2014). Indeed, two tasks in the neuropsychological battery—the verbal fluency and the sentence memory tasks—involved verbal production and also showed the strongest evidence of training benefits, despite production *per se* not being the critical theoretical component of these measures. Thus, future work may benefit from more thoroughly targeting language production outcomes.

CONCLUSION

The contributions of this study are two-fold. First, based on a research design that minimized the role of expectancy effects (Boot et al., 2013), our results suggest that verbal WM among older adults is responsive to home-based training, at least in the short-term. With fewer than 10 h of home-based practice with tasks exercising the simultaneous management of verbal operations and storage over the course of 3 weeks, training effects transferred to untrained verbal WM measures. Second and most importantly, WM training lead to selective improvements in measures of language fluency, sentence memory, and syntactic ambiguity resolution, implying that WM may be a critical resource for these aspects of language performance in older adulthood. These findings are among the first to indicate that selective aspects of language performance can be modified through targeted home-based practice in WM in older adulthood. This is not only of theoretical import in defining the cognitive architecture of language processing across the lifespan (e.g., Just and Carpenter, 1992; Stine-Morrow and Payne, 2016), but also suggests applications for improving cognitive functioning among older adults in significant ways (Stine-Morrow and Basak, 2011). Because age-related declines in language comprehension and memory can have far-reaching effects as adults navigate the ordinary demands of work, family, and health (e.g., Morrow et al., 2006; Chin et al., 2015), the development of pathways to mitigate such deficits offers promise for promoting late-life well-being.

This study is motivated by a specific model of vertical transfer in which a skill (language comprehension) was improved by exercise of a component theorized to be a core process constraining this skill (verbal WM). As described above, the cognitive training literature has generally shown relatively narrow transfer of cognitive training across cognitive domains in near transfer tasks. Yet, it is common for training programs to adopt a very broad cognitive battery and predict broad-based changes in cognition, under the rationale of “use it or lose it,” without a specific model of what is exactly being used or what capability will not be lost by using it. Given what is known from existing literature (cf. Simons et al., 2016), it is plausible that training programs will lead to very specific improvements across cognitive domains in transfer tasks that are subserved by the core mechanisms being exercised in the training tasks. A goal of future work must be to develop sound theories of the cognitive architecture of meaningful activities, as well as credible training and transfer tasks to operationalize those theories. Only in this way can cognitive training be used as a method to target specific mechanisms and test mechanistic accounts of theoretical models (Baltes and Kliegl, 1992; Hussey and Novick, 2012; Lindenberger et al., 2017). Thus, although transfer of cognitive training is a controversial area (see open-letter statements by A Consensus on the Brain Training Industry from the Scientific Community, 2014 and cognitivetrainingdata.org), our view is that the current study contributes to a literature aimed at using training-related cognitive plasticity as a tool for examining basic questions about cognitive architectures and functions rather than as a tool to reverse or slow generalized cognitive decline or substantially alter intellectual functioning (Simons et al., 2016; National Academies of Sciences, Engineering, and Medicine, 2017).

ETHICS STATEMENT

University of Illinois Institutional Review Board. Participants had completely voluntary participation and chose to sign up for the study of their own volition. All appropriate procedures were followed to ensure subject confidentiality and privacy, including

REFERENCES

- A Consensus on the Brain Training Industry from the Scientific Community (2014). *Max Planck Institute for Human Development and Stanford Center on Longevity*. Available at: <http://longevity3.stanford.edu.ezp-prod1.hul.harvard.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community/> [accessed November 22, 2015].
- Acheson, D. J., and MacDonald, M. C. (2009). Verbal working memory and language production: common approaches to the serial ordering of verbal information. *Psychol. Bull.* 135, 50. doi: 10.1037/a0014411
- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., et al. (2001). The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann. Intern. Med.* 134, 663–694. doi: 10.7326/0003-4819-134-8-200104170-00012
- Au, J., Buschkuhl, M., Duncan, G. J., and Jaeggi, S. M. (2016). There is no convincing evidence that working memory training is NOT effective: a reply to Melby-Lervåg and Hulme (2015). *Psychon. Bull. Rev.* 23, 331–337. doi: 10.3758/s13423-015-0967-4

voluntary participation, freedom to withdraw, and informed consent. Prior to participation in the study, informed consent was obtained from participants, which involved verbally explaining the research, risks and benefits, and confidentiality. Risk level for this study was deemed to be no more than minimal risk, meaning that the probability and magnitude of harm or discomfort were not greater than those ordinarily encountered in daily life.

AUTHOR CONTRIBUTIONS

BP and ES-M designed the research. BP performed the research and analyzed the data. BP and ES-M wrote the paper.

FUNDING

This research was funded by the University of Illinois Campus Research Board. The first author was supported by a National Institutes of Health training grant (T32-HD055272) and a Beckman Institute Graduate Fellowship at the University of Illinois at Urbana-Champaign, with funding provided by the Arnold and Mabel Beckman Foundation. Portions of this research were completed by BP in partial fulfillment of the requirements for the degree of Doctor of Philosophy at the University of Illinois and were presented at the Cognitive Aging Conference, 2014. We would like to thank Thomas Deegan and Andy Battles for their programming assistance, and Kyle Payne and Sneha Gummuluri for their recruitment and data collection efforts. We'd like to thank Kiel Christianson, Duane Watson, and Kara Federmeier for their helpful comments on previous versions of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnagi.2017.00256/full#supplementary-material>

- Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuhl, M., and Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychon. Bull. Rev.* 22, 366–377. doi: 10.3758/s13423-014-0699-x
- Baddeley, A. (2012). Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* 63, 1–29. doi: 10.1146/annurev-psych-120710-100422
- Ball, K., Berch, D. B., Helmers, K. F., Jobe, J. B., Leveck, M. D., Marsiske, M., et al. (2002). Effects of cognitive training interventions with older adults: a randomized controlled trial. *J. Am. Med. Assoc. (JAMA)* 288, 2271–2281. doi: 10.1001/jama.288.18.2271
- Balota, D. A., Yap, M. J., Cortese, M. J., Hutchison, K. I., Kessler, B., Loftis, B., et al. (2007). The English lexicon project. *Behav. Res. Methods* 39, 445–459. doi: 10.3758/BF03193014
- Baltes, P. B., and Kliegl, R. (1992). Further testing of limits of cognitive plasticity: negative age differences in a mnemonic skill are robust. *Dev. Psychol.* 28, 121. doi: 10.1037/0012-1649.28.1.121
- Baltes, P. B., and Willis, S. L. (1982). “Enhancement (plasticity) of intellectual functioning in old age: Penn State’s Adult Development and Enrichment Project (ADEPT),” in *Aging and Cognitive Processes*, eds F. I. M. Craik and S. E. Trehub (New York: Plenum Press), 353–389. doi: 10.1007/978-1-4684-4178-9_19

- Barr, D. J., Levy, R., Scheepers, C., and Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: keep it maximal. *J. Mem. Lang.* 68, 255–278. doi: 10.1016/j.jml.2012.11.001
- Barrouillet, P., Lépine, R., and Camos, V. (2008). Is the influence of working memory capacity on high-level cognition mediated by complexity or resource-dependent elementary processes? *Psychon. Bull. Rev.* 15, 528–534. doi: 10.3758/PBR.15.3.528
- Bartek, B., Lewis, R. L., Vasishth, S., and Smith, M. R. (2011). In search of on-line locality effects in sentence comprehension. *J. Exp. Psychol.* 37, 1178–1198. doi: 10.1037/a0024194
- Benton, A. L., and Hamsher, K. S. (1978). *Multilingual Aphasia Examination*, Rev. Edn. Iowa City, IA: Department of Neurology, University of Iowa Hospitals.
- Bever, T. G. (1970). “The cognitive basis for linguistic structure,” in *Cognition and the Development of Language*, ed. J. R. Hayes (New York, NY: John Wiley & Sons).
- Boot, W. R., Simons, D. J., Stothart, C., and Stutts, C. (2013). The pervasive problem with placebos in psychology: why active control groups are not sufficient to rule out placebo effects. *Perspect. Psychol. Sci.* 8, 445–454. doi: 10.1177/1745691613491271
- Bopp, K. L., and Verhaeghen, P. (2005). Aging and verbal memory span: a meta-analysis. *J. Gerontol.* 60, 223–233. doi: 10.1093/geronb/60.5.P223
- Borella, E., Carbone, E., Pastore, M., De Beni, R., and Carretti, B. (2017). Working memory training for healthy older adults: the role of individual characteristics in explaining short- and long-term gains. *Front. Hum. Neurosci.* 11:99. doi: 10.3389/fnhum.2017.00099
- Borella, E., Carretti, B., Riboldi, F., and De Beni, R. (2010). Working memory training in older adults: evidence of transfer and maintenance effects. *Psychol. Aging* 25, 767–778. doi: 10.1037/a0020683
- Borella, E., Ghisletta, P., and de Ribaupierre, A. (2011). Age differences in text processing: the role of working memory, inhibition, and processing speed. *J. Gerontol.* 66B, 311–320. doi: 10.1093/geronb/gbr002
- Brainard, D. H. (1997). The psychophysics toolbox. *Spat. Vis.* 10, 433–436. doi: 10.1163/156856897X00357
- Brehmer, Y., Riekmann, A., Bellander, M., Westerberg, H., Fischer, H., and 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
- Buschkuhl, M., Jaeggi, S. M., Hutchison, S., Perrig-Chiello, P., Dapp, C., Muller, M., et al. (2008). Impact of working memory training on memory performance in old-old adults. *Psychol. Aging* 23, 743–753. doi: 10.1037/a0014342
- Caplan, D., DeDe, G., Waters, G., Michaud, J., and Tripodis, Y. (2011). Effects of age, speed of processing, and working memory on comprehension of sentences with relative clauses. *Psychol. Aging* 26, 439–450. doi: 10.1037/a0021837
- Caplan, D., and Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behav. Brain Sci.* 22, 77–94. doi: 10.1017/S0140525X99001788
- Caplan, D., and Waters, G. S. (2013). Memory mechanisms supporting syntactic comprehension. *Psychon. Bull. Rev.* 20, 243–268. doi: 10.3758/s13423-012-0369-9
- Carpenter, P. A., Miyake, A., and Just, M. A. (1994). “Working memory constraints in comprehension: evidence from individual differences in aphasia, and aging,” in *Handbook of Psycholinguistics*, ed. M. A. Gernsbacher (New York, NY: Academic Press), 1075–1122.
- Carretti, B., Borella, E., Fostinelli, S., and Zavagnin, M. (2013a). Benefits of training working memory in amnesic mild cognitive impairment: specific and transfer effects. *Int. Psychogeriatr.* 25, 617–626. doi: 10.1017/S1041610212002177
- Carretti, B., Borella, E., Zavagnin, M., and De Beni, R. (2013b). Gains in language comprehension related to working memory training in healthy older adults. *Int. J. Geriatr. Psychiatry* 28, 539–546. doi: 10.1002/gps.3859
- Chein, J., and Morrison, A. (2010). Expanding the mind’s workspace: training and transfer effects with a complex working memory span task. *Psychon. Bull. Rev.* 17, 193–199. doi: 10.1162/jocn_a_00657
- Chin, J., Payne, B. R., Gao, X., Stine-Morrow, E. A. L., Morrow, D. G., Conner-Garcia, T., et al. (2015). Knowledge influences comprehension and memory for health information among older adults: distinguishing domain-general and domain-specific knowledge effects. *Memory* 23, 577–589. doi: 10.1080/09658211.2014.912331
- Christianson, K., Hollingworth, A., Halliwell, J. F., and Ferreira, F. (2001). Thematic roles assigned along the garden path linger. *Cognit. Psychol.* 42, 368–407. doi: 10.1006/cogp.2001.0752
- Christianson, K., Williams, C. C., Zacks, R. T., and Ferreira, F. (2006). Younger and older adults’ “good-enough” interpretations of garden-path sentences. *Discourse Process.* 42, 205–238. doi: 10.1207/s15326950dp4202_6
- Clair-Thompson, H. L. (2007). The influence of strategies on relationships between working memory and cognitive skills. *Memory* 15, 353–365. doi: 10.1080/09658210701261845
- Clifton, C., Traxler, M. J., Mohamed, M. T., Williams, R. S., Morris, R. K., and Rayner, K. (2003). The use of thematic role information in parsing: syntactic processing autonomy revisited. *J. Mem. Lang.* 49, 317–334. doi: 10.1016/S0749-596X(03)00070-6
- Conway, A. R., Jarrold, C., Kane, M. J., Miyake, A., and Towse, N. (eds) (2007). *Variation in Working Memory*. Oxford: Oxford University Press.
- Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilheld, O., and Engle, R. W. (2005). Working memory span tasks: a methodological review and user’s guide. *Psychon. Bull. Rev.* 12, 769–786. doi: 10.3758/BF03196772
- Cowan, N. (2000). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav. Brain Sci.* 24, 87–185. doi: 10.1017/S0140525X01003922
- Cumming, G. (2014). The new statistics: why and how. *Psychol. Sci.* 25, 7–29. doi: 10.1177/0956797613504966
- Dahlin, E., Neely, A. S., Larsson, A., Backman, L., and Nyberg, L. (2008). Transfer of learning after updating training mediated by the striatum. *Science* 320, 1510–1512. doi: 10.1126/science.1155466
- Daneman, M., and Carpenter, P. A. (1980). Individual differences in working memory and reading. *J. Verbal Learning Verbal Behav.* 19, 450–466. doi: 10.1016/S0022-5371(80)90312-6
- Daneman, M., and Merikle, P. M. (1996). Working memory and comprehension: a meta-analysis. *Psychon. Bull. Rev.* 3, 422–433. doi: 10.3758/BF03214546
- DeDe, G., Caplan, D., Kemtes, K., and Waters, G. (2004). The relationship between age, verbal working memory and language comprehension. *Psychol. Aging* 19, 601–616. doi: 10.1037/0882-7974.19.4.601
- Engle, R. W. (2002). Working memory capacity as executive attention. *Curr. Dir. Psychol. Sci.* 11, 19. doi: 10.1111/1467-8721.00160
- Engle, R. W. (2010). Role of working-memory capacity in cognitive control. *Curr. Anthropol.* 51, 17–26. doi: 10.1086/650572
- Federmeier, K. D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491–505. doi: 10.1111/j.1469-8986.2007.00531.x
- Federmeier, K. D., Kutas, M., and Schul, R. (2010). Age-related and individual differences in the use of prediction during language comprehension. *Brain Lang.* 115, 149–161. doi: 10.1016/j.bandl.2010.07.006
- Federmeier, K. D., McLennan, D. B., Ochoa, E., and 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, 133–146. doi: 10.1111/1469-8986.3920133
- Fedorenko, E., Gibson, E., and Rohde, D. (2006). The nature of working memory capacity in sentence comprehension: evidence against domain-specific working memory resources. *J. Mem. Lang.* 54, 541–553. doi: 10.1016/j.jml.2005.12.006
- Fraundorf, S. H., Diaz, M. I., Finley, J. R., Lewis, M. L., Tooley, K. M., Isaacs, A. M., et al. (2014). *CogToolbox for MATLAB [computer software]*. Available at: <http://www.scottfraundorf.com/cogtoolbox.html>
- Friedman, N. P., and Miyake, A. (2004). The reading span test and its predictive power for reading comprehension ability. *J. Mem. Lang.* 51, 136–158. doi: 10.1016/j.jml.2004.03.008
- Gibson, E. (1998). Linguistic complexity: locality of syntactic dependencies. *Cognition* 68, 1–76. doi: 10.1016/S0010-0277(98)00034-1
- Gilchrist, A. L., Cowan, N., and Naveh-Benjamin, M. (2008). Working memory capacity for spoken sentences decreases with adult ageing: recall of fewer but not smaller chunks in older adults. *Memory* 16, 773–787. doi: 10.1080/09658210802261124
- Greenwood, P. M., and Parasuraman, R. (2016). The mechanisms of far transfer from cognitive training: review and hypothesis. *Neuropsychology* 30, 742. doi: 10.1037/neu0000235
- Gupta, S. K. (2011). Intention-to-treat concept: a review. *Perspect. Clin. Res.* 2, 109–112. doi: 10.4103/2229-3485.83221

- Harrison, T. L., Shipstead, Z., Hicks, K. L., Hambrick, D. Z., Redick, T. S., and Engle, R. W. (2013). Working memory training may increase working memory capacity but not fluid intelligence. *Psychol. Sci.* 24, 2409–2419. doi: 10.1177/0956797613492984
- Hertzog, C., Dixon, R. A., Hulstsch, D. F., and 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? *Psychol. Aging* 18, 755–769. doi: 10.1037/0882-7974.18.4.755
- Hollis, S., and Campbell, F. (1999). What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ* 319, 670–674. doi: 10.1136/bmj.319.7211.670
- Hussey, E. K., Harbison, J., Teubner-Rhodes, S. E., Mishler, A., Velnoskey, K., and Novick, J. M. (2016). Memory and language improvements following cognitive control training. *J. Exp. Psychol.* 43, 23–58.
- Hussey, E. K., and Novick, J. M. (2012). The benefits of executive control training and the implications for language processing. *Front. Psychol.* 3:158. doi: 10.3389/fpsyg.2012.00158
- Ide, N., Suderman, K., Baker, C., Passonneau, R., and Fellbaum, C. (2013). *Manually Annotated Sub-Corpus Third Release*. Philadelphia: Linguistic Data Consortium.
- Jefferies, E., Ralph, M. A. L., and Baddeley, A. D. (2004). Automatic and controlled processing in sentence recall: the role of long-term and working memory. *J. Mem. Lang.* 51, 623–643. doi: 10.1016/j.jml.2004.07.005
- Johnson, R. A. (2003). Aging and the remembering of text. *Dev. Rev.* 23, 261–346. doi: 10.1016/S0273-2297(03)00009-1
- Just, M. A., and Carpenter, P. (1992). A capacity theory of comprehension: individual differences in working memory. *Psychol. Rev.* 99, 122–149. doi: 10.1037/0033-295X.99.1.122
- Just, M. A., and Varma, S. (2002). A hybrid architecture for working memory: reply to MacDonald and Christiansen (2002). *Psychol. Rev.* 109, 55–65. doi: 10.1037/0033-295X.109.1.55
- Just, M. A., and Varma, S. (2007). The organization of thinking: what functional brain imaging reveals about the neuroarchitecture of complex cognition. *Cogn. Affect. Behav. Neurosci.* 7, 153–191. doi: 10.3758/CABN.7.3.153
- Kane, M. J., Conway, A. R. A., Hambrick, D. Z., and Engle, R. W. (2007a). “Variation in working memory capacity as variation in executive attention and control,” in *Variation in Working Memory*, eds A. R. A. Conway, C. Jarrold, M. J. Kane, A. Miyake, and J. Towse (Oxford: Oxford University Press).
- Kane, M. J., Conway, A. R. A., Miura, T. K., and Colflesh, G. J. H. (2007b). Working memory, attention control, and the n-back task: a question of construct validity. *J. Exp. Psychol.* 33, 615–622.
- Karbach, J., Strobach, T., and Schubert, T. (2015). Adaptive working-memory training benefits reading, but not mathematics in middle childhood. *Child Neuropsychol.* 21, 285–301. doi: 10.1080/09297049.2014.899336
- Karbach, J., and Verhaeghen, P. (2014). Making working memory work: a meta analysis of executive control and working memory training in older adults. *Psychol. Sci.* 25, 2027–2037. doi: 10.1177/0956797614548725
- Kelley, K., and Preacher, K. J. (2012). On effect size. *Psychol. Methods* 17, 137–152. doi: 10.1037/a0028086
- Kelly, M. E., Loughrey, D., Lawlor, B. A., Robertson, I. H., Walsh, C., and Brennan, S. (2014). The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: a systematic review and meta-analysis. *Ageing Res. Rev.* 15, 28–43. doi: 10.1016/j.arr.2014.02.004
- Kemper, S. (1987). Syntactic complexity and elderly adults’ prose recall. *Exp. Aging Res.* 13, 47–52. doi: 10.1080/03610738708259299
- Kemper, S. (2012). “The interaction of linguistic constraints, working memory, and aging on language production and comprehension,” in *Memory and Aging: Current Issues and Further Directions*, eds M. N. Benjamin and N. Ohata (Hove: Psychology Press).
- Kemper, S., Crow, A., and Kemtes, K. (2004). Eye-fixation patterns of high- and low-span young and older adults: down the garden path and back again. *Psychol. Aging* 19, 157–170. doi: 10.1037/0882-7974.19.1.157
- Kemper, S., and Herman, R. E. (2006). Age differences in memory-load interference effects in syntactic processing. *J. Gerontol.* 61, 327–332. doi: 10.1093/geronb/61.6.P327
- Kemper, S., and Liu, C. (2007). Eye movements of young and older adults during reading. *Psychol. Aging* 22, 84–93. doi: 10.1037/0882-7974.22.1.84
- Kemper, S., and McDowd, J. (2008). “Dimensions of cognitive aging: executive function and verbal fluency,” in *Handbook of Cognitive Aging: Interdisciplinary Perspectives*, eds S. M. Hofer and D. F. Alwin (Thousand Oaks, CA: Sage).
- Kemtes, K. A., and Kemper, S. (1997). Younger and older adults’ on-line processing of syntactically ambiguous sentences. *Psychol. Aging* 12, 362. doi: 10.1037/0882-7974.12.2.362
- King, J. W., and Just, M. A. (1991). Individual differences in syntactic parsing: the role of working memory. *J. Mem. Lang.* 30, 580–602. doi: 10.1016/0749-596X(91)90027
- King, J. W., and Kutas, M. (1995). Who did what and when? Using word- and clause-level ERPs to monitor working memory usage in reading. *J. Cogn. Neurosci.* 7, 376–395. doi: 10.1162/jocn.1995.7.3.376
- Kintsch, W., and van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychol. Rev.* 85, 363–394. doi: 10.1037/0033-295X.85.5.363
- Kirby, K. N., and Gerlanc, D. (2013). BootES: an R package for bootstrap confidence intervals on effect sizes. *Behav. Res. Methods* 45, 905–927. doi: 10.3758/s13428-013-0330-5
- Kwong See, S. T., and Ryan, E. B. (1995). Cognitive mediation of adult age differences in language performance. *Psychol. Aging* 10, 458–468. doi: 10.1037/0882-7974.10.3.458
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front. Psychol.* 4:863. doi: 10.3389/fpsyg.2013.00863
- Lee, C., and Federmeier, K. D. (2011). Differential age effects on lexical ambiguity resolution mechanisms. *Psychophysiology* 48, 960–972. doi: 10.1111/j.1469-8986.2010.01158.x
- Lépine, R., Parrouillet, P., and Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychon. Bull. Rev.* 12, 165–170. doi: 10.3758/BF03196363
- Lewis, R. L., and Vasishth, S. (2005). An activation-based model of sentence processing as skilled memory retrieval. *Cogn. Sci.* 29, 375–419. doi: 10.1207/s15516709cog0000_25
- Lewis, K. L., and Zelinski, E. M. (2010). List and text recall differ in their predictors: replication over samples and time. *J. Gerontol.* 65, 449–458. doi: 10.1093/geronb/gbq034
- Li, S.-C., Schmiedek, F., Huxhold, O., Röcke, C., Smith, J., and Lindenberger, U. (2008). Working memory plasticity in old age: practice gain, transfer, and maintenance. *Psychol. Aging* 23, 731–742. doi: 10.1037/a0014343
- Lindenberger, U., Wenger, E., and Lövdén, M. (2017). Towards a stronger science of human plasticity. *Nat. Rev. Neurosci.* 18, 261–262. doi: 10.1038/nrn.2017.44
- Lövdén, M., Bäckman, L., Lindenberger, U., Schaeffer, S., and Schmiedek, F. (2010). A theoretical framework for the study of adult cognitive plasticity. *Psychol. Bull.* 136, 659–676. doi: 10.1037/a0020080
- Lustig, C., Shah, P., Seidler, R. D., and Reuter-Lorenz, P. (2009). Aging, training, and the brain: a review and future directions. *Neuropsychol. Rev.* 4, 504–522. doi: 10.1007/s11065-009-9119-9
- MacDonald, M., Just, M., and Carpenter, P. (1992). Working memory constraints on the processing of syntactic ambiguity. *Cognit. Psychol.* 24, 56–98. doi: 10.1016/0010-0285(92)90003-K
- MacDonald, M. C., and Christiansen, M. C. (2002). Reassessing working memory: comment on Just and Carpenter (1992) and Waters and Caplan (1996). *Psychol. Rev.* 109, 35–54. doi: 10.1037/0033-295X.109.1.35
- Manly, J. J., Touradjji, P., Tang, M.-X., and Stern, Y. (2003). Literacy and memory decline among ethnically diverse elders. *J. Clin. Exp. Neuropsychol.* 25, 680–690. doi: 10.1076/j.jcen.25.5.680.14579
- Margrett, J. A., and Willis, S. L. (2006). In-home cognitive training with older married couples: individual versus collaborative learning. *Aging Neuropsychol. Cogn.* 13, 173–195. doi: 10.1080/138255890969285
- Martin, R. C., and Slevc, L. R. (2014). “Language production and working memory,” in *The Oxford Handbook of Language Production*, eds M. Goldrick, V. Ferreira, and M. Miozzo (New York, NY: Oxford University Press), 437–450.
- Mayr, U., and Kliegl, R. (2000). Complex semantic processing in old age: does it stay or does it go? *Psychol. Aging* 15, 29–43. doi: 10.1037/0882-7974.15.1.29
- Melby-Lervåg, M., and Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Dev. Psychol.* 49, 270–291. doi: 10.1037/a0028228
- Melby-Lervåg, M., and Hulme, C. (2016). There is no convincing evidence that working memory training is effective: a reply to Au et al. (2014) and Karbach

- and Verhaeghen (2014). *Psychon. Bull. Rev.* 23, 324–330. doi: 10.3758/s13423-015-0862-z
- Melby-Lervåg, M., Redick, T. S., and Hulme, C. (2016). Working memory training does not improve performance on measures of intelligence or other measures of “far transfer” evidence from a meta-analytic review. *Perspect. Psychol. Sci.* 11, 512–534. doi: 10.1177/1745691616635612
- Miller, G. A., and Chomsky, N. (1963). “Finitary models of language users,” in *Handbook of Mathematical Psychology*, Vol. II, eds D. R. Luce, R. R. Bush, and E. Galanter (New York, NY: John Wiley).
- Mohr, D. C., Spring, B., Freedland, K. E., Beckner, V., Arean, P., Hollon, S. D., et al. (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychother. Psychosom.* 78, 275–284. doi: 10.1159/000228248
- Morrison, A. B., and Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychon. Bull. Rev.* 18, 46–60. doi: 10.3758/s13423-010-0034-0
- Morrow, D. G., Clark, D., Tu, W., Wu, J., Weiner, M., Steinley, D., et al. (2006). Correlates of health literacy in patients with chronic heart failure. *Gerontologist* 46, 669–676. doi: 10.1093/geront/46.5.669
- Nagels, A., Kircher, T., Dietsche, B., Backes, H., Marquetand, J., and Krug, A. (2012). Neural processing of overt word generation in healthy individuals: the effect of age and word knowledge. *Neuroimage* 61, 832–840. doi: 10.1016/j.neuroimage.2012.04.019
- Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., et al. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 53, 695–699. doi: 10.1111/j.1532-5415.2005.53221.x
- National Academies of Sciences, Engineering, and Medicine (2017). *Preventing Cognitive Decline and Dementia: A Way Forward*. Washington, DC: The National Academies Press.
- Nelson, T. O., and Narens, L. (1980). Norms of 300 general-information questions: accuracy of recall, latency of recall, and feeling-of-knowing ratings. *J. Verbal Learning Verbal Behav.* 19, 338–368. doi: 10.1016/S0022-5371(80)90266-2
- Noh, S. R., and Stine-Morrow, E. A. L. (2009). Age differences in tracking characters during narrative comprehension. *Mem. Cogn.* 37, 769–778. doi: 10.3758/MC.37.6.769
- Novick, J., Hussey, E. K., Teubner-Rhodes, S., Harbison, J. I., and Bunting, M. (2013). Clearing the garden path: improving sentence processing through executive control training. *Lang. Cogn. Process.* 29, 186–217. doi: 10.1080/01690965.2012.758297
- Payne, B. R., Gao, X., Noh, S. R., Anderson, C. J., and Stine-Morrow, E. A. L. (2012a). The effects of print exposure on sentence processing and memory among older adults: evidence for efficiency and reserve. *Aging Neuropsychol. Cogn.* 19, 122–149. doi: 10.1080/13825585.2011.628376
- Payne, B. R., Jackson, J. J., Hill, P. L., Gao, X., Roberts, B. W., and Stine-Morrow, E. A. L. (2012b). Memory self-efficacy predicts responsiveness to inductive reasoning training in older adults. *J. Gerontol.* 67, 27–35. doi: 10.1093/geronb/gbr073
- Payne, B. R., Grison, S., Gao, X., Christianson, K., Morrow, D. G., and Stine-Morrow, E. A. L. (2014a). Aging and individual differences in binding during sentence understanding: evidence from temporary and global syntactic attachment ambiguities. *Cognition* 130, 157–173. doi: 10.1016/j.cognition.2013.10.005
- Payne, B. R., Gross, A. L., Parisi, J. M., Sisco, S., Stine-Morrow, E. A. L., Marsiske, M., et al. (2014b). Modeling longitudinal changes in older adults’ memory for spoken discourse: findings from the ACTIVE cohort. *Memory* 22, 990–1001. doi: 10.1080/09658211.2013.861916
- Payne, B. R., Lee, C.-L., and Federmeier, K. D. (2015). Revisiting the incremental effects of context on word processing: evidence from single-word event-related brain potentials. *Psychophysiology* 52, 1456–1469. doi: 10.1111/psyp.12515
- Payne, B. R., and Stine-Morrow, E. A. L. (2016). Risk for mild cognitive impairment is associated with semantic integration deficits in sentence processing and memory. *J. Gerontol.* 71, 243–253. doi: 10.1093/geronb/gbu103
- Pickering, M. J., and van Gompel, R. P. G. (2006). “Syntactic parsing,” in *Handbook of Psycholinguistics*, 2nd Edn, eds M. J. Traxler and M. Gernsbacher (Amsterdam: Academic Press), 455–503. doi: 10.1016/B978-012369374-7/50013-4
- Potter, M. C., and Lombardi, L. (1998). Syntactic priming in immediate recall of sentences. *J. Mem. Lang.* 38, 265–282. doi: 10.1006/jmla.1997.2546
- Radvansky, G. A. (1999). Memory retrieval and suppression: the inhibition of situation models. *J. Exp. Psychol.* 128, 563. doi: 10.1037/0096-3445.128.4.563
- Radvansky, G. A., and Dijkstra, K. (2007). Aging and situation model processing. *Psychon. Bull. Rev.* 14, 1027–1042. doi: 10.3758/BF03193088
- Rebok, G. W. (2008). *Cognitive Training: Influence on Neuropsychological and Brain Function in Later Life*. London: Government Office for Science.
- Rebok, G. W., Ball, K., Guey, L. T., Jones, R. N., Kim, H. Y., King, J. W., et al. (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. *J. Am. Geriatr. Soc.* 62, 16–24. doi: 10.1111/jgs.12607
- Richmond, L. L., Morrison, A. B., Chein, J. M., and Olson, I. R. (2011). Working memory training and transfer in older adults. *Psychol. Aging* 26, 813. doi: 10.1037/a0023631
- Roberts, R., and Gibson, E. (2002). Individual differences in sentence memory. *J. Psycholinguist. Res.* 31, 573–598. doi: 10.1023/A:1021213004302
- Shipstead, Z., Redick, T. S., and Engle, R. W. (2012). Is working memory training effective? *Psychol. Bull.* 138, 628–654. doi: 10.1037/a0027473
- Shiran, A., and 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. *J. Neurolinguistics* 24, 524–537. doi: 10.1016/j.jneuroling.2010.12.001
- Simons, D. J., Boot, W. R., Charness, N., Gathercole, S. E., Chabris, C. F., Hambrick, D. Z., et al. (2016). Do “brain-training” programs work? *Psychol. Sci. Public Interest* 17, 103–186. doi: 10.1177/1529100616661983
- Smiler, A., Gagne, D. D., and Stine-Morrow, E. A. (2003). Aging, memory load, and resource allocation during reading. *Psychol. Aging* 18, 203. doi: 10.1037/0882-7974.18.2.203
- Stern, Y. (2009). Cognitive reserve. *Neuropsychologia* 47, 2015–2028. doi: 10.1016/j.neuropsychologia.2009.03.004
- Stine, E. A., and Hindman, J. (1994). Age differences in reading time allocation for propositionally dense sentences. *Aging Cogn.* 1, 2–16. doi: 10.1080/09289919408251446
- Stine, E. A. L., and Wingfield, A. (1987). Process and strategy in memory for speech among younger and older adults. *Psychol. Aging* 2, 272–279. doi: 10.1037/0882-7974.2.3.272
- Stine, E. A. L., and Wingfield, A. (1990). How much do working memory deficits contribute to age differences in discourse memory? *Eur. J. Cogn. Psychol.* 2, 289–304. doi: 10.1080/09541449008406209
- Stine-Morrow, E. A., Milinder, L. A., Pullara, O., and Herman, B. (2001). Patterns of resource allocation are reliable among younger and older readers. *Psychol. Aging* 16, 69–84. doi: 10.1037/0882-7974.16.1.69
- Stine-Morrow, E. A. L., and Basak, C. (2011). “Cognitive interventions,” in *Handbook of the Psychology of Aging*, 7th Edn, eds K. W. Schaie and S. L. Willis (New York, NY: Elsevier).
- Stine-Morrow, E. A. L., Hussey, E. K., and Ng, S. (2015). The potential for literacy to shape lifelong cognitive health. *Policy Insights Behav. Brain Sci.* 2, 92–100. doi: 10.1177/2372732215600889
- Stine-Morrow, E. A. L., Miller, L. M. S., Gagne, D. D., and Hertzog, C. (2008). Self-regulated reading in adulthood. *Psychol. Aging* 23, 131–153. doi: 10.1037/0882-7974.23.1.131
- Stine-Morrow, E. A. L., and Payne, B. R. (2016). Age differences in language segmentation. *Exp. Aging Res.* 42, 107–125. doi: 10.1080/0361073X.2016.1108751
- Stine-Morrow, E. A. L., Payne, B. R., Hill, P. L., Jackson, J. J., Roberts, B. W., Kramer, A. F., et al. (2014). Training versus engagement as paths to cognitive optimization with aging. *Psychol. Aging* 29, 891–906. doi: 10.1037/a0038244
- Stine-Morrow, E. A. L., and Radvansky, G. A. (in press). “Discourse processing and development through the adult lifespan,” in *Handbook of Discourse Processes*, eds M. F. Schobel, D. N. Rapp, and M. A. Britt.
- Stine-Morrow, E. A. L., Ryan, S. J., and Leonard, S. E. (2000). Age differences in on-line syntactic processing. *Exp. Aging Res.* 26, 315–322. doi: 10.1080/036107300750015714
- Strauss, E., Sherman, E. M., and Spreen, O. (2006). *A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary*. Washington, DC: American Chemical Society.

- Troyer, A. K., Moscovitch, M., and Winocur, G. (1997). Clustering and switching as two components of verbal fluency: evidence from younger and older healthy adults. *Neuropsychology* 11, 138. doi: 10.1037/0894-4105.11.1.138
- Turner, M. L., and Engle, R. W. (1989). Is working memory capacity task dependent? *J. Mem. Lang.* 28, 127–154. doi: 10.1016/0749-596X(89)90040-5
- Unsworth, N., Heitz, R. P., Schrock, J. C., and Engle, R. W. (2005). An automated version of the operation span task. *Behav. Res. Methods* 37, 498–505. doi: 10.3758/BF03192720
- 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 Neuropsychol. Cogn.* 6, 11–24. doi: 10.1076/anec.6.1.32.791
- Van Overschelde, J. P., Rawson, K. A., and Dunlosky, J. (2004). Category norms: an updated and expanded version of the norms. *J. Mem. Lang.* 50, 289–335. doi: 10.1016/j.jml.2003.10.003
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York, NY: Springer. doi: 10.1007/978-0-387-21706-2
- Wadley, V. G., Benz, R. L., Ball, K. K., Roenker, D. L., Edwards, J. D., and Vance, D. E. (2006). Development and evaluation of home based speed of processing training for older adults. *Arch. Phys. Med. Rehabil.* 87, 757–763. doi: 10.1016/j.apmr.2006.02.027
- Walton, C. C., Mowszowski, L., Lewis, S. J., and Naismith, S. L. (2014). Stuck in the mud: time for change in the implementation of cognitive training research in ageing? *Front. Aging Neurosci.* 6:43. doi: 10.3389/fnagi.2014.00043
- Was, C. A., Rawson, K. A., Bailey, H., and Dunlosky, J. (2011). Content-embedded tasks beat complex span for predicting comprehension. *Behav. Res. Methods* 43, 910–915. doi: 10.3758/s13428-011-0112-x
- Waters, G. S., and Caplan, D. (1996). Processing resource capacity and the comprehension of garden path sentences. *Mem. Cognit.* 24, 342–355.
- Waters, G. S., and Caplan, D. (2003). The reliability and stability of verbal working memory measures. *Behav. Res. Methods Instrum. Comput.* 35, 550–564. doi: 10.3758/BF03195534
- Wells, J. B., Christiansen, M. H., Race, D. S., Acheson, D. J., and MacDonald, M. C. (2009). Experience and sentence comprehension: statistical learning and relative clause comprehension. *Cognit. Psychol.* 58, 250–271. doi: 10.1016/j.cogpsych.2008.08.002
- Willis, S. L., Blieszner, R., and Baltes, P. B. (1981). Intellectual training research in aging: modification of performance on the fluid ability of figural relations. *J. Educ. Psychol.* 73, 41–50. doi: 10.1037/0022-0663.73.1.41
- Willis, S. L., and Nesselroade, C. S. (1990). Long-term effects of fluid ability training in old-old age. *Dev. Psychol.* 26, 905–910. doi: 10.1037/0012-1649.26.6.905
- Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., et al. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *JAMA* 296, 2805–2814. doi: 10.1001/jama.296.23.2805
- Wilson, B. A., Cockburn, J., and Baddeley, A. D. (2003). *The Rivermead Behavioural Memory Test*, Second Edn. London: Pearson Assessment.
- Wingfield, A., Stine, E. A. L., Lahar, C. J., and Aberdeen, J. S. (1988). Does the capacity of working memory change with age? *Exp. Aging Res.* 14, 103–107. doi: 10.1080/03610738808259731
- Wingfield, A., and Stine-Morrow, E. A. L. (2000). “Language and speech,” in *The Handbook of Aging and Cognition*, 2nd Edn, eds F. I. M. Craik and T. A. Salthouse (Mahwah, NJ: Erlbaum), 359–416.
- Wlotko, E. W., Federmeier, K. D., and Kutas, M. (2012). To predict or not to predict: age-related differences in the use of sentential context. *Psychol. Aging* 27, 975–988. doi: 10.1037/a0029206
- Wlotko, E. W., Lee, C.-L., and Federmeier, K. D. (2010). Language of the aging brain: event-related potential studies of comprehension in older adults. *Lang. Linguist. Compass* 4, 623–638. doi: 10.1111/j.1749-818X.2010.00224.x
- Wolinsky, F. D., Mahncke, H., Vander Weg, M. W., Martin, R., Unverzagt, F. W., Ball, K. K., et al. (2010). Speed of processing training protects self-rated health in older adults: enduring effects observed in the multi-site ACTIVE randomized controlled trial. *Int. Psychogeriatr.* 22, 470–478. doi: 10.1017/S1041610209991281
- Zelinski, E. M., and Lewis, K. L. (2003). Adult age differences in multiple cognitive functions: differentiation, dedifferentiation, or process-specific change? *Psychol. Aging* 18, 727–745. doi: 10.1037/0882-7974.18.4.727
- Zinke, K., Zeintl, M., Rose, N. S., Putzmann, J., Pydde, A., and Kliegel, M. (2014). Working memory training and transfer in older adults: effects of age, baseline performance, and training gains. *Dev. Psychol.* 50, 304–315. doi: 10.1037/a0032982

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Payne and Stine-Morrow. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.