Effects of Working Memory Training on L2 Proficiency and Working Memory Capacity

Gregory Colflesh (gcolflesh@casl.umd.edu)
Valerie Karuzis (vkaruzis@casl.umd.edu)
Polly O'Rourke (porourke@casl.umd.edu)
Center for Advanced Study of Language
University of Maryland, 7005 52nd Avenue
College Park, MD 20742 USA

Abstract

The current study examined the effects of working memory training on working memory capacity and second language ability in adult learners of Spanish. In order to maximize the effect of the training for language learners, the stimuli for the training tasks were Spanish words and sentences. While the training group did not show greater improvements on working memory assessments relative to controls, they did show more native-like patterns in a Spanish self-paced reading task. The combination of second language materials with working memory training may be helping users learn to cope with the increased processing demands associated with learning a new language, even if they are not necessarily improving their working memory.

Keywords: Working memory training, second language acquisition, self-paced reading, n-back.

Introduction

Working memory (WM) plays a role in many higher order cognitive functions, including the ability to learn a second language (L2). Indeed, one prominent theory of the acquisition of L2 morphosyntax proposes that the deficits exhibited by learners are largely due to WM limitations, given the high demand of operating in L2 (McDonald, 2006). Due to the importance of WM in a great many everyday activities, exploring the efficacy of WM training has been a very hot research topic in the last five to ten years. In the current study we examined the impact of WM training on second language processing ability.

Given the connection between working memory capacity (WMC) and second language acquisition (Linck, Osthus, Koeth & Bunting, 2014), the possibility of increasing WMC, and thereby enhancing second language learning outcomes, is an exciting prospect. Evidence from experimental research suggests that it may be possible to increase WMC with computer training in which individuals repeatedly perform tasks that tax attentional and executive control abilities (Baniqued et al., 2014; Holmes, Gathercole & Dunning, 2009; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Morrison & Chein, 2011; Novick et al., 2014; Olesen, Westerberg & Klingberg, 2004). Training typically targets domain-general mechanisms which "control attention, gate the flow of information into and out of WM buffers, reduce interference from irrelevant sources of information, and govern the engagement of domain-specific strategies." (Morrison & Chein, 2011, p. 47). These domain-general mechanisms drive the connection between WM and higher order cognition (Cowan et al., 2005; Lépine, Barrouillet, & Camos, 2005; Morrison & Chein, 2011). Training-related improvements in domain-general executive functions have been shown to be associated with increased general fluid intelligence (Jaeggi, et al., 2008) and language processing ability in the first language (Morrison & Chein, 2011; Novick, et al., 2014). Given the strong connection between WMC and second language learning, improving domain-general attentional and executive control abilities through training could also lead to improvements in second language proficiency. In the current study, we examined the impact of WM training on online syntactic processing and sentence comprehension in adult learners of Spanish. Additionally, in order to maximize the benefits of the WM training, the training tasks all contained Spanish language stimuli, thereby providing participants with increased exposure to the target language during training.

In the current study, the WM training consisted of adaptive versions of four standard WM tasks (Reading Span, *n*-back, Sorter, and Running Span). This multifaceted "kitchen sink" approach was implemented as a means of increasing the likelihood of achieving training-related gains by engaging similar WM mechanisms in a variety of different tasks (Morrison & Chein, 2011). In addition, the inclusion of multiple tasks was intended to increase participants' likelihood of enjoying and completing the training, as performing four tasks for 15 minutes each is more engaging than performing the same (difficult) task for 60 minutes. This is in contrast to studies that used a more targeted approach (Jaeggi, et al., 2008; Verhaeghen, Cerella, & Basak, 2004).

The key innovation in our study, which is one of the first to examine the impact of WMT on L2 proficiency, is that the stimuli in our WM training tasks consist of Spanish words and sentences. The motivation for this innovation is to (1) increase individuals' engagement in the training tasks by making the stimuli relevant to their real world learning goals and (2) to enhance the benefit of training on language learning by requiring participants to interact with L2 materials while performing WM demanding tasks.

We predicted that participants who completed the adaptive training tasks with a WM demand (i.e., WM training) would have greater improvement from pre-training to post-training for Spanish language and WM assessments compared to participants who completed the adaptive training tasks with minimal WM demands (i.e., active control).

Method

Participants

62 adult learners of Spanish (51 female, age = 20.26 (*SD* = 2.11), years education = 14.31 (*SD* = 1.27)) completed this study. Eligible participants were native English speakers who did not learn any non-English languages at home or during childhood, were currently or recently enrolled in a 200- or 300-level university Spanish class or similar experience, and passed the Spanish proficiency requirements. All participants were neurologically normal, began studying Spanish after age 11, and had not had an immersion experience in a Spanish-speaking country for longer than 6 weeks. Participants were semi-randomly assigned to the training and control groups; they were matched on working memory capacity (as indexed by the Shapebuilder task; see below) and comprehension accuracy of grammatical sentences in the self-paced reading task (see below). Post-hoc analyses confirmed that groups were similar for age, gender, education, self-rated Spanish ability (reading, writing, listening and speaking), and two other measures of Spanish vocabulary knowledge.

Pre-Post Assessment Tasks

Self-Paced Reading Task

In the self-paced reading task, participants read sentences, in Spanish, word-by-word, advancing at their own speed. Words appeared at the subject's prompt within a sentence frame with "-"s as place holders for each letter. Each word reverted to "-"s when the subject advanced to the next word. The sentence stimuli consisted of grammatical sentences and sentences containing subject/verb agreement errors (see sentence 1 below). There were 12 items in each condition. In addition to those 24 sentences, there were 67 additional sentences including fillers and conditions not reported herein. Half of the sentences were followed by a yes/no comprehension question. The timing of the task was based on Sagarra and Herschensohn (2010)'s methods.

1. La investigadora **cree**/***<u>creen</u>** que la ciencia concreta es más importante que la teoría.
   The researcher [**believes**/**believe**] that concrete science is more important than theory.

Working Memory and Executive Function Assessments

N-back

For the *n*-back task used in the current study (Kane, Conway, Colflesh, & Miura, 2007), participants performed blocks of 2-back and 3-back, with lures (±1 and ±2, where appropriate). The *n*-back task indexes the ability to maintain, update and monitor information in the short-term store and resolve the conflict between what is familiar and recollected. While it is a frequently used index of WM ability, its cognitive underpinnings are distinct from those of complex span tasks (Kane et al., 2007).

Reading Span

The automated version of the reading span task (Redick, et al., 2012), a complex span task, was used to measure WMC. Complex span tasks measure the ability to store and recall information in the face of distracting information from an unrelated processing task.

Shapebuilder

For Shapebuilder (Atkins, et al., 2014) participants had to remember the shape and color of items that appeared sequentially on a 4x4 grid. Shapebuilder was used as an additional measure of WMC that did not mimic the reading span, which was also used as a training task.

Simon Task

Participants performed the Simon Task (Simon, 1990) as an index of inhibition control. The Simon Task, while not a WM assessment, was included because it indexes the ability to control attention via the inhibition of a prepotent, but goal-irrelevant response, which underpins the performance of many of the WM tasks.

Training Tasks

Spanish Reading Span

As the name suggests, the Spanish reading span training and control tasks were modeled off the standard reading span task (Daneman & Carpenter, 1980; Kane et al. (2004)). Participants were asked to recall a sequence of letters which appeared one-at-a-time on the screen. In between the presentation of each letter, they saw a sentence in Spanish (e.g., *El médico llegó para ayudar a la gente herido.)* and they had to decide if it was grammatical or not. Participants had to decide among five possible choices, one of which is "No Errors" and the remaining four were possible error types: spelling, accent use, subject/verb agreement, and gender agreement (gender error present in the adjective portion of the example above).

In the training version of the task, participants were presented with a sentence. They had 10 seconds to make a judgment on the sentence. After they made a judgment on the sentence (or did not make a response in the allotted time), they received feedback regarding their selection. Next, a letter was presented. The number of sentence-letter items in a trial started at 1 and increased as participants performed at or above leveling criteria. Leveling criteria was based on sentence and memory performance across four trials. Initially, the four trials were all set size 1. At the next level, three trials were set size 1 and one trial was set size 2. To level up, participants had to have 85% or greater accuracy on the sentences and had to have recalled 85% or greater of the letters presented for a later recall test in the correct serial positions. Participants remained at the same level if they were between 70% and 85% on both, and they went down a level if they were below 70% on either component.

In the control version of the task, sentences and memoranda were presented separately within a trial, instead of being interleaved. All sentences presented for the grammatical judgment task were in an uninterrupted sequence. Memoranda letters were presented for the recall task in randomly selected sequences of 2-6 letters, independent of the sentence processing task. Participants went up a level if they achieved 85% accuracy on the sentences. As participants leveled up, there were more sentences to be judged in each block.

Spanish *n*-back

The Spanish *n*-back training and control tasks were modeled off of previous *n*-back tasks (Kane et al., 2007; Novick, et al., 2014). In the Spanish *n*-back task, participants saw a series of 20 + *n* words presented on the screen one at a time, with each presentation accompanied by an audio presentation of the word. The stimuli were presented in a mix of Spanish and English, with approximately 80% of the words in a list presented in Spanish. Participants responded to each word as it was presented. Participants pressed the "Yes" button if the current word matched the word *n* items back, either exactly (*manzana* – *manzana*; *apple* – *apple*), or by meaning (*manzana* – *apple*). If the current word did not match the word or translation of the word *n* items back, participants pressed the "No" button.

In the training version of the task, *n* increased as participants' level increased throughout the task. To go from 1-back to 2-back, participants had to meet leveling criteria for two steps: 1) *n* without lures, and 2) $n \pm 2$ lures. Starting with 2-back, each level required three successful steps to achieve the next level: 1) *n* without lures, 2) $n \pm 2$ lures, and 3) $n \pm 1$ lures.

In the control version of the task, participants performed a 1-back task with the Spanish (and English) word stimuli. Participants were given the appearance that they were attaining different levels if they reached leveling up or down criteria, but the task always remained a 1-back task.

Spanish Running Span

The Spanish running span training and control tasks were modeled off of the running span task (Anderson, 1960; Bunting, Cowan & Saults, 2006; Pollack, Johnson, & Knaff, 1959). In the Spanish running span task, participants saw a series of 12 to 20 + K words, where K is the number of items that needed to be recalled, presented on the screen one at a time, accompanied by a simultaneous audio presentation of the word. The stream of stimuli was presented either entirely in Spanish or entirely in English. At the end of the presentation, participants needed to respond to what the last K words were. K started at 1 item and increased as participants leveled up. If the word list was presented in Spanish, participants saw a list of English words – targets and lures. If the word list was presented in English, participants saw a list of Spanish words – targets and lures.

In the training version of the task, participants had to drag the correct words and drop them into the correct positions. In the control version of the task, participants were given the translations of the last K words and only needed to drag the correct words to the appropriate boxes.

Spanish Sorter

The Spanish sorter training and control tasks were modeled off of the letter-number sequencing task used by Sprenger, et al. (2013). In the Spanish sorter task, participants saw a series of 2 words presented on the screen one-at–a-time, accompanied by a simultaneous audio presentation of the word. At the end of the presentation, participants had to drag the words that had just been presented to the correct semantic category, and the words within each category needed to be recalled in serial order and the correct serial position for that category. As participants leveled up, they were presented with more words to recall and sort.

The Spanish sorter control task was the same as the training task, except participants were provided with the translation equivalent of the words in the categories and they needed to drag the correct words to the correct translations. As participants leveled up, they were presented with more words.

Procedure

The study was comprised of 14 sessions. Sessions 1, 2, 13, and 14 were pre- and post-training sessions (1-2 hours long each), and sessions 3-12 were training sessions (45 minutes-1 hour long each). The pre-post sessions included the tasks described above and additional measures not reported herein. Participants completed pre- and post-training tasks on PCs, and training tasks on iPads. All participants completed the training sessions within 10-14 business days, completed session 13 1-5 days after training, and completed session 14 1-2 days after session 13. Multiple training sessions were not allowed to take place within the same day. Participants were compensated $5/session, with a balloon payment of $175 at the completion of session 14. Participants were also entered into raffles for 1 of 10 $200 gift cards following study completion. In addition to the participants analyzed here, 5 participants attritted from the study (7%), and 4 were found to be ineligible after session 1 (5.6%).

Results

WM Assessments

Data points more than 2.5 SD from the sample mean were excluded from analyses. Participants in the training group did not show greater improvements at post-test than the control group on the reading span or Shapebuilder tasks, indicating that WM training did not improve WMC. Similarly, there were no effects of training condition on the Simon task.

The 3-back trials in the *n*-back task were analyzed using overall d', a measure of signal detection. A 2 (training condition) x 2 (assessment time) mixed factor ANOVA analysis of d' for the 3-back trials showed no significant effect of assessment time ($F<1$). There was, however, a significant effect of training condition ($F(1, 51)=5.41$, $p<.05$, $\eta_p^2=.10$), where the training group performed better than the control group. Simple comparisons at pre- and post-test showed no significant difference between training group and control group performance at pre-test (*M*s = 1.97 & 1.28, respectively; $t(57)=.44$, $p>.05$) but at post-test, the training group performed significantly better than the control group, (*M*s = 3.09 & 1.97, respectively; $t(54)=2.11$, $p<.05$). This suggests that WM training may have improved cognitive control ability, especially in the face of interference.

Self-Paced Reading Task

Reading times exceeding 3 SD from a given subject's mean were excluded from analyses, and any subject with a mean reading time over 800 ms per word was excluded from the analysis. This resulted in an N of 46, with 23 participants per training group. Separate 2 (assessment time) x 2 (training condition) x 2 (grammaticality) ANOVAs were conducted on the reading times at each of the four positions relative to the critical word: -1, 0 (critical word), +1, +2. Figure 1 shows the reading time data for the two groups at post.

At position -1, 0, +1 and +2 there were main effects of assessment time (position -1: $F(1,44)=51.1$, $p<.001$, $\eta_p^2=.54$; position 0: $F(1,44)=54.2$, $p<.001$, $\eta_p^2=.54$; position +1: $F(1,44)=32.6$, $p<.001$, $\eta_p^2=.43$; position +2: $F(1,44)=28.7$, $p<.001$, $\eta_p^2=.40$, respectively) such that reading times were faster at post. At position +1, there was also a three-way interaction of grammaticality, assessment time, and training condition ($F(1,44)=4.89$, $p<.05$, $\eta_p^2=.10$). Splitting across training condition revealed a main effect of assessment time ($F(1,22)=8.83$, $p<.01$, $\eta_p^2=.29$) and no effects of grammaticality in the control group. In the training group, there was an effect of assessment time ($F(1,22)=27.4$, $p<.001$, $\eta_p^2=.56$) and an interaction of assessment time and grammaticality ($F(1,22)=6.60$, $p<.05$, $\eta_p^2=.23$). This interaction was driven by increased reading times in the ungrammatical condition at post, evidenced by an effect of grammaticality at post ($F(1,22)=5.38$, $p<.05$, $\eta_p^2=.20$). There were no effects of grammaticality in the training group at pre. At position +2, there was a significant interaction of assessment time and grammaticality, driven by an effect of grammaticality at post ($F(1,44)=9.16$, $p<.01$, $\eta_p^2=.17$). While there were no interactions involving training condition at position +2, Figure 1 suggests the grammaticality effect was larger for the control group. Simple comparisons showed a significant effect of grammaticality in the control group ($F(1,22)=9.11$, $p<.01$, $\eta_p^2=.29$) but not in the training group.

Similar ANOVAs performed on accuracy data for the comprehension questions (Table 1) revealed no effects of assessment time, training condition or grammaticality.

Table 1: Mean Comprehension Accuracy (SD)

	Pre	Post
Control - Grammatical	0.78 (.19)	0.78 (.15)
Training - Grammatical	0.81 (.15)	0.79 (.15)
Control - Violation	0.80 (.12)	0.74 (.16)
Training - Violation	0.80 (.16)	0.76 (.14)

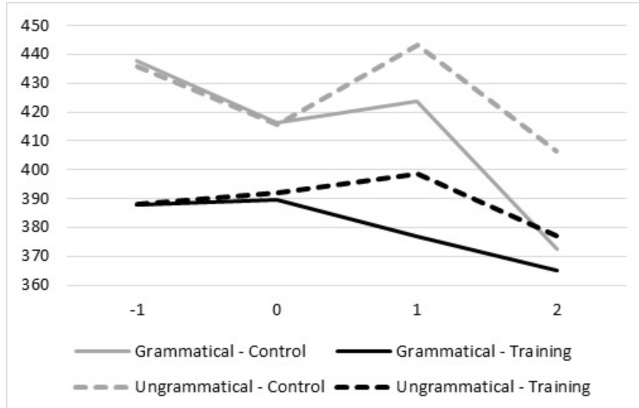

Figure 1. Mean reading times (ms) at post.

Discussion

Although many studies have examined the relationship between WMC and L2 acquisition, and many other studies have examined the impact of WM training on WM and general cognition, there has not been much research combining WM training with L2 acquisition. In this study, we explored whether completing WM training in a target language, Spanish in this case, would (1) improve performance on WM measures and (2) improve performance on an online processing task in the target language.

The WM training did have the predicted impact on *n*-back scores, such that the training group performed significantly better than the control group. However, the interaction between training condition and assessment time was not significant, despite a moderate effect size. Thus, further exploration may be warranted. It may be that the L2 WM training did improve some aspect of the WM system; however, it might be that participants in the training group developed better strategies to cope with the increased WM demand of the L2 *n*-back training task.

WM training did not have the predicted impact on the remainder of the battery of WM assessments, which included Reading Span, Shapebuilder, and the Simon task. Previous research found that complex span tasks and *n*-back measure different aspects of the WM system (Kane et al., 2007), so it is not surprising that *n*-back did not follow the same pattern as reading span. The current findings further support the distinction between *n*-back and complex span tasks.

It is important to note that prior research using assessment versions of complex span tasks in the first (L1) and second language have shown strong correlations between L1 and L2 span tasks (Alptekin & Erçetin, 2010; Osaka & Osaka, 1992; Osaka, Osaka & Groner, 1993). This indicates a strong overlap between L1 and L2 versions of the tasks suggesting that the same resources are utilized. The lack of transfer between the L2 reading span and standard version in the current study, therefore, should not be an artifact of task language.

The WM training outcomes in the current study mimic what is generally found in the literature. The studies that have found improvements in WM and a transfer to other abilities tend to utilize the *n*-back task, or a variant of the *n*-back task (Jaeggi, et al., 2008; Jaeggi, Studer-Luethi, Buschkuehl, Su, Jonides, & Perrig, 2010). However, the methodology and outcome of these studies have been greatly scrutinized (Redick, et al., 2013; Shipstead, Redick, & Engle, 2012), specifically the lack of a contact control group. Redick and colleagues (2013) compared dual *n*-back WM training to that of a contact control group, and they did not find that WM training improved WM. It may be that the *n*-back training task allows for the development of strategies to circumvent the WM demands of the *n*-back task and those strategies may transfer to the assessment version of the *n*-back, but not the other assessments. If this is the case, then WM training is not actually leading to improvements in WM, but rather strategies to reduce the reliance on the WM, suggesting that WM is a stable trait.

However, even if WM training does not train working memory, it does not mean that the training is not useful. Perhaps instead of calling it WM training, it should be called cognitive training, as it appears training may improve some cognitive processes, which in turn may improve additional cognitive abilities, such as second language processing.

One interesting effect emerged out of the SPR reading time data: the training group showed an earlier sensitivity (at position +1) to subject-verb agreement violations, while the control group did not show an effect until the last position (+2). In this respect, the training group looked slightly more native-like, as native speakers (and higher proficiency learners) would show effects at the 0 and +1 position (Sagarra & Hernschensohn, 2010). The training, therefore, seemed to contribute toward enhanced online syntactic processing ability. The decomposition of inflected forms (i.e., utilization of morphological information) during online processing is perhaps one of the greatest obstacles for L2 learners. These difficulties are possibly due to reduced ability to engage in combinatorial rule application in L2 (Silva & Clahsen, 2008) or an inability to process the information due to the increased WM demands associated with operating in the L2 (McDonald, 2006). In McDonald (2006)'s account, increasing WMC via training should result in increased sensitivity to L2 morphosyntactic information. The training related changes in L2 morphosyntactic processing that we observed did not correspond to clear pre-post improvements in WMC, thus the findings from the training manipulation do not provide direct support McDonald (2006)'s account. Rather, linguistic exposure and interaction in a speeded and high WM demand context associated with the training may have accelerated the development of participants' morphosyntactic processing ability. Alternatively, the performance of the WM training may have led to improvement in underlying cognitive abilities other than WM that are relevant to language processing. A final possibility is that the training tasks were simply more engaging and, thus, participants were more motivated with respect to processing the Spanish training task stimuli.

In conclusion, working memory training in the second language did not transfer to performance on all of the WM tasks, yet it did lead to enhanced online morphosyntactic processing of the target language. This suggests that although WM training may not serve the laudable purpose of enhancing WMC, it may be useful in improving second language ability and/or other general cognitive abilities which underpin second language processing.

References

Anderson, N. S. (1960). Poststimulus cuing in immediate memory. *Journal of Experimental Psychology*, *60*(4), 216.

Alptekin, C., & Erçetin, G. (2010). The role of L1 and L2 working memory in literal and inferential comprehension in L2 reading. *Journal of Research in Reading*, *33*(2), 206-219.

Atkins, S. M., Sprenger, A. M., Colflesh, G. J. H., Briner, T. L., Buchanan, J. B., Chavis, S. E., . . . & Dougherty, M. R. (2014). Measuring Working Memory Is All Fun and Games. *Experimental Psychology, 61*(6), 417-438.

Baniqued, P. L., Kranz, M. B., Voss, M. W., Lee, H., Cosman, J. D., Severson, J., & Kramer, A. F. (2013). Cognitive training with casual video games: points to consider. *Frontiers in Psychology*, *4*(1010), 1-19.

Bunting, M., Cowan, N., & Scott Saults, J. (2006). How does running memory span work?. *The Quarterly Journal of Experimental Psychology*, *59*(10), 1691-1700.

Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*, 42-100.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450-466.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9-F15.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*(19), 6829-6833.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning—

Implications for training and transfer. *Intelligence*, *38*, 625–635.

Kane, M. J., Conway, A. R., Miura, T. K., & Colflesh, G. J. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(3), 615-622.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, *133*(2), 189.

Lépine, R., Barrouillet, P., & Camos, V. (2005). What makes working memory spans so predictive of high-level cognition? *Psychonomic Bulletin & Review*, *12*(1), 165–170.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*(4), 861-883.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language*, *55*(3), 381-401.

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin & Review*, *18*(1), 46-60.

Novick, J. M., Hussey, E., Teubner-Rhodes, S., Harbison, J. I., & Bunting, M. F. (2014). Clearing the garden-path: Improving sentence processing through cognitive control training. *Language, Cognition and Neuroscience*, *29*(2), 186-217.

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience*, *7*(1), 75–79.

Osaka, M. & Osaka, N. (1992). Language-independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society*, *30*, 287–289.

Osaka, M., Osaka, N. & Groner, R. (1993). Language-independent working memory: Evidence from German and French reading span tests. *Bulletin of the Psychonomic Society*, *31*, 117–118.

Pollack, I., Johnson, L. B., & Knaff, P. R. (1959). Running memory span. *Journal of Experimental Psychology*, *57*(3), 137-146.

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*, 164-171.

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., . . . Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*, 359–379.

Sagarra, N., & Herschensohn, J. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. *Lingua, 120*(8), 2022-2039.

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*, 628-654.

Silva, R., & Clahsen, H. (2008). Morphologically complex words in L1 and L2 processing: Evidence from masked priming experiments in English. *Bilingualism: Language and Cognition*, *11*(02), 245-260.

Simon, J. R. (1990). The effects of an irrelevant directional cue on human information processing. In R. W. Proctor & T. G. Reeve (Eds.). Stimulus-response compatibility: An integrated perspective (pp. 31–88). Amsterdam, The Netherlands: North Holland.

Sprenger, A. M., Atkins, S. M., Bolger, D. J., Harbison, J. I., Novick, J. M., Chrabaszcz, J. S., . . . Dougherty, M. R. (2013). Training working memory: Limits of transfer. *Intelligence*, *41*, 638-663.

Verhaeghen, P., Cerella, J., & Basak, C. (2004). A working memory workout: how to expand the focus of serial attention from one to four items in 10 hours or less. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1322-1337.

Acknowledgements

This research is based upon work supported, in whole or in part, with funding from the United States Government. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the University of Maryland, College Park and/or any agency or entity of the United States Government. Nothing in this article is intended to be and shall not be treated or construed as an endorsement or recommendation by the University of Maryland, United States Government, or the author of the product, process, or service that is the subject of this article. Correspondence concerning this article should be addressed to Gregory Colflesh, Center for Advanced Study of Language, University of Maryland, 7005 52nd Avenue, College Park, MD 20742 (e-mail: gcolflesh@casl.umd.edu). Special thanks to Mike Bunting for supporting this line of research, and to Ashlyn Testut, Grace Goff, Kenneth Mai, and Mari Sakai for their help in setting up the lab for testing, running participants, and cleaning data.