ABSTRACT

Title of Document: WORKING MEMORY ASSESSMENT AND TRAINING.

Sharona M. Atkins, PhD., 2011

Directed By: Associate Professor Michael R. Dougherty, Department of Psychology

Working memory, the ability to maintain and manipulate information, is a core cognitive function important for everyday life. The capacity of working-memory differs across individuals, with working-memory capacity a reliable predictor of general fluid intelligence, verbal and mathematical abilities, and classroom achievement. However, research has been inconclusive on whether working-memory is a unitary domain-general construct, or multi-component domain-specific construct. Most theories had until recently thought that working-memory was a fixed ability; however, recent research suggests that working-memory is malleable and can be improved through cognitive training. These training-induced improvements have also been shown on untrained cognitive tasks, such as general fluid intelligence, attention, reading, and math. My research examines the structure of working-memory, validates newly designed web-administered working-memory assessments, and investigates the malleability of domain specific working-memory training.

WORKING MEMORY ASSESSMENT AND TRAINING.

By

Sharona M. Atkins

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2011

Advisory Committee:
Associate Professor Michael R. Dougherty ,Chair
Professor Bradley Hatfield
Assistant Professor Donald J. Bolger
Assistant Professor Thomas A. Carlson
Dr. Michael F. Bunting

Acknowledgements

"I believe that simple flight at least is possible to man and that the experiments and investigations of a large number of independent workers will result in the accumulation of information and knowledge and skill which will finally lead to accomplished flight." Wilbur Wright, 1899

Similar to flight, experimental research builds on those who precede them and relies upon others for support. Throughout my research, I relied upon many people who were gracious enough to assist and support me.

I would like to thank my advisor, Mike Dougherty, for putting up with me through the years, and allowing me go in directions neither of us originally anticipated. In addition, I would like to thank DJ Bolger, for serving as my unofficial adviser. I appreciate your willingness to listen, instruct, and guide me in matters whose complexities were beyond my imagination. Similar thanks go to Mike Bunting, for his insight, his guidance, and his support of my research. I would also like to thank Tom Carlson, for taking the time to serve on both my proposal and dissertation committee, and Bradley Hatfield for serving on my dissertation committee as the Dean's representative.

I would like to thank Isaiah Harbison, for listening and improving upon my crazy ideas, and continuing to collaborate with me, and Michael Donovan, who programmed my requests and anticipated my programming needs. Similar thanks go to members of the DAM lab: Amber, Tracy, Erika, Susan and Jeff; members of the LNRL lab: Kristen, Alice, Say Young, and Lesley; members of the CASL Working Memory Training team; Yaron Shlomi and Susan Campbell, who assisted with discussions and ideas, and my many research assistants throughout the years.

I would like to thank Sarah Jane Kim, Cate Stasio, Natasha Belfor and Mike Merzenich, for believing and supporting me before I even started on this journey, and Ariela Gigi who set me on this path.

I would like to thank Carol Gorham and Pam Komarek who guided me through complex bureaucracies, and have always had time to listen. Similar thanks go to Tony Chan and Enamul Haque, to whom I could always turn for assistance.

I would like to thank the Psychology Department, the Center for Advanced Study of Language, and the Neuroscience and Cognitive Science Program for funding me and my research. In addition, I would like to thank the Center for Functional and Molecular Imaging and the Center for the Study of Learning for allowing me to use their facilities for my research.

Special thanks go to friends who assisted in ways unquantifiable. Specifically, I thank Ann, Aya and Basma, who have heard me moan, bitch and cry and have understood, comforted, and provided perspective and the occasional swift kick in the rear when needed; Amir and Heather, who, like family, have provided guidance and support and believed in me; and my friends in Israel, the Bay area and Maryland who have understood and accepted my absences and have always been there for me.

I also need to thank my sizeable and vivacious family, for in many ways they are to blame for me being who I am: my mother, Joyce, for driving me insane enough to try and figure out what I could do with my madness; my sister, Shane, for being the most awesome sister a graduate student could ever dream of (you have read my research more times than any rational person should); my siblings, Shimie, Renana, Oriel, Moshe, Ayelet, and Aviva, for being there when I needed and not bugging too much when I was too busy to be around; my in-laws, Iris and Joe Levy, for being who they are and treating me like their own; and Spice, Ginger and Oreo, for always providing warm, cuddly, and furry love with no strings attached.

And most importantly, I would like to thank Niv, for standing by me through the stress, the laughter, the smiles, and the tears and the many many miles, for always trusting me to find a solution, for never letting me slack off, and for always daring me to find and conquer the next challenge. I have no idea what I would do without you. Thank you.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

The construct of working memory (WM) has been the focus of much research over the last couple decades. Many cognitive models regarding WM processing and the organization of WM exist in the literature (see Miyake & Shah, 1999). The prevailing view from most models is that WM involves both processing and storage and is limited in capacity. However, the nature of the processing component and its underlying structure are still debated.

WM and the limitations to its functioning is commonly assessed through complex span tasks (Daneman & Carpenter, 1980; Engle, Tuholski, Laughlin, & Conway, 1999; Unsworth & Engle, 2005), which allows one to estimate the WM *capacity* (WMC) of that individual (for review see Conway et al., 2005; Oberauer, 2005). The use of complex span measures, along with a variety of other cognitive tasks that tap various component's of cognitive functioning have helped identify the structure and function of WM (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000; Ackerman, Beier, & Boyle, 2002; Kane et al, 2004). Additionally, these complex span tasks have proven to be useful individual difference measures. For example, WM span measures have been shown to be reliable predictors of performance on a variety of tasks and abilities, including tasks that assess general fluid intelligence (gFI) (Conway, Cowan, Bunting, Therriault, & Minkoff, 2002; Conway, Kane, & Engle, 2003; Unsworth & Engle, 2005), SAT performance (Engle, Tuholski, Laughlin, & Conway, 1999) , visual spatial ability (Kane et al., 2004), attention (Bleckley et al., 2003), inhibition (Miyake, Friedman, Emerson, Witzki, & Howerter, 2000), reading ability (Daneman & Carpenter, 1980; Friedman & Miyake, 2004), verbal ability (Kane et al., 2004), mathematical achievements (Ashcraft & Kirk, 2001; Conway et al., 2005; D'Amico & Guarnera, 2005; Bull, Espy & Wiebe, 2008; Kyttälä & Lehto, 2008), and decision making (Dougherty & Hunter, 2003).

Complex span tasks, as well as other WM and cognitive tasks, have also been used as training tasks aimed at improving WM ability (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Olesen, Westerberg & Klingberg, 2004; Chein & Morrison, 2010; Atkins et al., under review). Although previously considered a stable function, recent research has suggested that WM is malleable throughout one's lifetime. The malleable nature of WM has been demonstrated using cognitive training procedures. Prior work has shown the effectiveness of WM training across multiple age groups ranging from early childhood (Thorell, Lindqvist, Bergman, Bohlin, & Klingberg, 2009) to elderly adults (Mahncke et al., 2006), with improvements on the both the trained WM tasks, and the untrained WM measures (Olesen, Westerberg & Klingberg, 2004; Chein & Morrison, 2010). Many studies have shown that the training induced improvements transfer to untrained cognitive measures, such as measures of gFI (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008), measures of WM (Olesen, Westerberg & Klingberg, 2004; Chein & Morrison, 2010) and measures of inhibition (Atkins, et al. under review). However, not all cognitive training experiments have led to improvements on untrained tasks (Owen et al., 2010; Shipstead, Redick, & Engle, 2010). Therefore, the precise nature in which WM is malleable remains subject to debate.

The purpose of my research is twofold. First, my research investigated the structure of WM, examining whether WM is a unitary-construct or multi-component construct. As part of this work, I validated two newly designed assessments of working memory capacity. Second, my research examines the malleability of working memory and the transfer of training-induced improvements to untrained measures of cognitive ability.

Working Memory

WM is a core cognitive process that handles the processing and manipulation of information. Multiple theoretical perspectives exist regarding WM and its limited capacity (see Miyake & Shah, 1999). One of the most debated aspects is whether WM is a unitary, domain general process or a domain specific process consisting of verbal WM (vWM) and visual-spatial WM (vsWM).

In an influential paper, Baddeley and Hitch (1974) proposed their multi-component WM model, which consisted of a domain general central executive component and two domain specific slave systems for the storage of information: the *visual spatial sketchpad* for visuo-spatial information and the *phonological loop* for verbal information. This model is not unitary, and allows for capacity limits for each component (Baddeley & Logie, 1999). More recent versions added the episodic buffer to integrate information from subsystems and long-term memory (Baddeley, 2000), whereas other versions (cf. Logie, 1995) added a processing component to the visual spatial sketchpad. This later model considers the visual-spatial sketch pad more broadly as visual-spatial WM (vsWM).

In contrast, Engle, Kane and Tuholski (1999) theorize WM to be capacity-limited controlled attention, which is assumed to be a domain free unitary process that uses multiple domain specific stores. They define WM capacity (WMC) as the domain general component of the WM system (Conway & Engle, 1996; Engle, 2001; Engle, 2002).

Baddeley's WM model exemplifies the multi-component perspective, whereas Engle's attentional control model exemplifies the unitary-component perspective, but not all models take a clear stance. Cowan's embedded processes model of WM (1995, 2005) views WM as a single cognitive process that maintains information in an unusually accessible state. His model defines the focus of attention as a region of privileged and immediate access, which is embedded in the activated component of memory, which itself is embedded in long-term memory (storage). The information held in the focus of attention is highly accessible conscious information, but the amount of information held there is limited. The short-term system is not limited in capacity, but information in this state can be forgotten due to interference and/or decay. Attentional control processes are required for the manipulation of WM contents and for the focusing, updating, switching, and inhibiting of that content.

Central to the evaluation and testing of these and other theories of WM are the set of procedures for measuring WM capacity. The prevailing approach is to use so-called complex span tasks to measure WM. These tasks involve interleaving a to-be-remembered letter or image with a secondary task, after which the person is required to recognize the to-be-remembered items in the order in which they were presented. For example, in the automated reading span (adapted from Daneman & Carpenter, 1980), the to-be remembered letters are interleaved with sentences that must be read and classified on whether they make sense or not. The most commonly administered WMC tasks maintain the dual task design and apply different stimuli: math operations are used in operation span (Turner & Engle, 1989; Unsworth & Engle, 2005), auditory sentences in listening span, and symmetry decisions in symmetry span (Kane et al., 2004). The interleaving of the processing and memory tasks creates competition regarding to which task the participant should be allotting their attentional resources. Participants are instructed to both maintain the to-be-remembered items, while performing well on the secondary task, for which they are constantly reinforced. Although complex WM spans provide a single unitary score for both the processing and the memory task, research has shown that the domain of the storage (verbal or visual) has greater influence than the domain of the processing. That is, a mixed domain complex span task with spatial to-be-remembered items and verbal processing items will correlate more strongly with spatial tasks, and a mixed domain complex span task with verbal to-be-remembered items and spatial processing tasks will correlate more strongly to verbal tasks (Shah & Miyake, 1996).

Efforts have been made in recent years to automate complex WM span tasks. In prior research, researchers administered the complex span tasks in one-on-one setting for each participant, a time consuming prospect. Although the current complex span tasks have been adapted for automatic computerized administration, this presentation still requires laboratory setting and is labor intensive. An additional difficulty with the automatic complex span task administration is the individualized speed parameter, in which the time allotted for responding is individualize, and therefore different for each participant. Also the dual tasks nature of the complex span tasks can be confusing to participants. Experiment 1 presents and validates two newly designed measures of WM that tap into vsWM and vWM. These new WM measures are fully automated, programmed using Flash and constructed using a single task. Furthermore, these tasks were designed to permit internet administration.

Neurobiology of Working Memory

Complex span tasks are not the only way to measure WMC. Vogel and Michizawa (2004) used electroencephalography, the recording of the electrical voltages along the scalp caused by neuronal firing, to assess vsWM on a delayed-match-to-sample task. Participants were shown a center fixation with an arrow pointing right or left indicating which side to remember, and then an array of colored squares on both sides of the screen. After a delay participants were asked to indicate whether the subsequent array was the same or different. Vogel and Michizawa (2004) observed a large negative voltage over the contralateral hemisphere to the memorized array, primarily over the lateral occipital and posterior parietal regions, which persisted from ~200msec after presentation until the end of the retention interval. Most importantly, they found that amplitudes of the negative voltages were based on the participant's individual ability to maintain the information. For example, a participant with a WMC of four items would show an increase in negative voltage amplitude, when the number of items to remember increased from two to three and from three to four. However, increasing the number of items beyond four did not elicit an additional increase in the negative voltage amplitude, as it exceeded the participants' WMC. These results suggest a neurological capacity limit for the maintenance of information, which corresponds to the individual's WMC.

Using a paradigm similar to that used by Vogel and Michizawa (2004), Todd and Marios (2004) utilized functional magnetic resonance imaging (fMRI) to examine the blood-oxygen-level-dependent (BOLD) activation in the brains of participants partaking in a delay-match-to-sample task. BOLD, while not a direct measure of the neuronal activity, is a measure of the metabolic properties of the neurons and has been shown to be reliably correlated with neuronal activity (Huettel, Song, & McCarthy, 2009). Todd and Marios showed that vsWMC was related to activation in the posterior parietal cortex, the same region implicated in the Vogel and Michizawa study. These results support the hypothesis that the posterior parietal area is actively involved with vsWM tasks and implies a relationship between the neuronal activation in the parietal regions and the limit of WM capacity.

However, the parietal area is not the only brain region implicated in WM. Prefrontal (PFC) regions have long been implicated in WMC tasks (Goldman-Rakic, 1987; D'Esposito, Postle, & Rypma, 2000; Fuster, 2001; Kane & Engle, 2002; Curtis & D'Esposito, 2003; for review see D'Esposito, 2007). For example, Curtis and D'Esposito (2003) suggest that the PFC is involved in the maintenance of information by directing attention to storage regions in the parietal cortex. They review findings from different fMRI experiments showing the involvement of PFC in WM tasks manipulating of the delay between presentation and response and the memory load presented. They suggest that the PFC does not itself store the memory representation of the future response, but instead directs top down attention to the stored representation in the parietal regions.

The results from D'Esposito and colleagues (D'Esposito, Postle & Rypma, 2000; Curtis & D'Esposito, 2003; D'Esposito, 2007) along with the results from Vogel and Michizawa (2004) and Todd and Marios (2004) suggest that multiple brain regions are involved in the maintenance of information in WM. Consistent with this notion, Smith and Jonides (1998) argued for multi-component model of WM. In their model, Smith and Jonides (1998) argue for the presence of domain-specific storage (vsWM: inferior parietal lobe (IPL), vWM: left posterior parietal) and rehearsal components (vsWM: superior parietal (SPL), vWM: inferior frontal gryus (IFG)), and a domain-general executive processing component (PFC). Similarly, Thomason et al., (2009) found differential patterns of BOLD activity where the vsWM activated bilateral occipital, right IPL, right SPL, and right IFG more than vWM, whereas vWM activated left IFG and left mid temporal more than vsWM. In a review paper, D'Esposito (2007) highlights the PFC as the source for active manipulation of information in both vWM and vsWM. According to D'Esposito (2007), a network of brain regions, among them PFC, is critical for the active maintenance of representations necessary for goal directed behavior, where the PFC directs top-down attention to the stored representations in the parietal regions (cf. Cowan, 1995). In addition, Klingberg (2006) illustrated the importance of both the frontal and parietal regions, as well as the white matter connections between them, in the development of vsWMC. He reviews research relating vsWMC to increases in BOLD activation in the intraparietal cortex and the posterior part of the superior frontal sulcus (collected during vsWM tasks), and shows that fractional anisotropy (a measure of the myelination of the axon) is positively correlated to the BOLD activation in these frontal-parietal regions.

Aside from frontal and parietal cortical regions, recent research has shown that the basal ganglia and the anterior cingulate are activated during WMC tasks. For example, McNab and Klingberg (2008) adapted the delayed match-to-sample task (used by Vogel & Michizawa, 2004; and Todd & Marios, 2004), and added a cue prior to the presentation of the first array which indicated which items were to-be-remembered and which items were to-be-ignored. The second array presented a question mark in a prior array location, and participants responded yes or no to whether a target was in that location in the prior array. They found that performance of the cued-delay match to sample was related to activation in the prefrontal cortex and the basal ganglia. They also found that memory storage is related to activity in the parietal cortex. They hypothesized that the prefrontal cortex and basal ganglia wield attentional control over vsWM storage located in the parietal cortex. These findings implicate a network of brain regions related to WMC.

Malleability of Working Memory

While much work has focused on testing theoretical and neurocognitive accounts of WM, there is a long-standing debate regarding whether WM is a stable individual difference ability, or whether it is open to change. The traditional view is that WM, like other cognitive abilities, is largely immutable (Neisser et al., 1996). However, recent research has led to a shift in the perception of WM as stable. Research has shown that it is possible to train WM and increase a person's WMC (for review see Morrison & Chein, 2011). The potential for WM training and for the transfer of training-induced improvement is not limited to children or impaired populations but exists throughout the lifespan of a typically achieving person (Greenwood, 2007; Mercardo, 2008). For example, Mahncke et al. (2006) demonstrated that elderly participants who underwent cognitive training improved on measures of auditory WM assess using the Repeatable Battery for the Assessment of Neuropsychological Status (RBANS). Olesen, Westerberg, and Klingberg (2004) demonstrated improvements in Span board, a vsWM measure, following WM training in normal adults. Jaeggi, Buschkuehl, Jonides, and Perrig (2008) extended these finding of the training improvements to assessments of gFI, as measured by the Ravens and Bomat tasks.

Of great importance is that not only has WM training shown improved performance on the trained tasks but also on other non-trained tasks. Klingberg, Fossberg, and Westerberg (2002) showed that vsWM training in children with attention deficit hyperactivity disorder (ADHD) led to improvements on untrained measures of gFI (Ravens), vsWM assessments (span board task), and a response inhibition measure (the stroop task). Klingberg, Fossberg and Westerberg (2002) also observed a reduction of head movements in children, a clinically relevant measure of ADHD, following the WM training. Similarly, Thorell, Lindqvist, Bergman, Bohlin and Klingberg (2008) showed that after cognitive training, pre-school children improved on vsWM (spatial span), vWM (word span), and inhibition (continuous performance task). Also, Chein and Morrison (2010) showed that WM training improvements transferred to inhibition, as measured by stroop, and reading comprehension, as measured by the Nelson-Denny reading test.

Although there is no consensus among researchers that WM training generalizes and leads to benefits (Owen et al., 2010; Shipstead, Redick, & Engle, 2010), many studies on WM training have shown transfer of training-induced improvements following individually adaptive training. For example, training has been shown to lead to improvements on untrained measures of gFI (Raven: Klingberg, Fossberg & Westerberg, 2002; Raven: Olesen, Westernberg, & Klingberg, 2004; Raven & Bosmat: Jaeggi, Buschkuehl, Jonides, & Perrig, 2008), WM measures (Span board: Klingberg, Fossberg & Westerberg, 2002; Span Boad: Olesen, Westernberg, & Klingberg, 2004; OSpan & SymSpan: Chein & Morrison, 2010; OSpan, SymSpan, Listening Span, & Rotation Span: Atkins et al., under review), and inhibition measures (Stroop: Klingberg, Fossberg & Westerberg, 2002; Olesen, Westernberg, & Klingberg, 2004; Klingberg et al., 2005; Stroop & Antisaccade: Atkins et al., under review).

However, it is important to note that not all studies involving cognitive training have led to improvements on other cognitive abilities. In fact, Owen et al. (2010) did not show any transfer of improvements among participants who underwent cognitive training administered online. Shipstead, Redick, and Engle (2010), in a review the training literature, are skeptical whether the transfer of improvements following training represents changes in WMC or task learning, and are critical of the methodology of training studies. These inconsistencies raise questions about the robustness of WM training and its generalizability.

In a study examining both WM and inhibition training groups, Thorell, Lindqvist, Bergman, Bohlin, and Klingberg (2009) demonstrated that WM training led to improvements in children's WM abilities, as measured by span board and word span tasks, and their attentional abilities, as measured by go/no-go omission and the continuous performance task . In contrast, participants in the inhibition training group did not show any improvement on the transfer tasks. These findings suggest that the generalization of improvements, and any sustainable benefits, from the training is likely dependent upon the nature of the cognitive training administered.

There are indications that the training-induced improvements are long-lasting. Atkins et al. (under review) showed transfer of training-induced improvements to untrained WM and inhibition tasks, immediately following the training. This improvement persisted when measured three months following the cessation of the training, whereas the control group did not show any sustained benefit. These findings of training persistence are consistent with Klingberg et al. (2005), who found that children with ADHD exhibited improvement on measures of WM and attention as well as a reduction in behavioral symptoms, both immediately after cognitive training and three months after the cessation of the training. Importantly, Holmes, Gathercole and Dunning (2009) not only show persistence of the WM improvements six month after the training, they also at the six-month follow-up, show improvements in mathematical ability (mathematical reasoning from the Wechsler Object Number Dimension). A task which had not shown transfer effects at the post-training assessment.

The persistence of the behavioral assessed improvements over long periods of time implies a permanent change in the underlying cortical structures. Several studies support this possibility. Temple et al. (2003) trained dyslexic children and showed the transfer of improvements to untrained assessments of reading and language. These cognitive improvements were related to increases in BOLD activation in the left IFG, the right temporal and parietal regions, and the anterior cingulate gyrus (areas which previously exhibited deficits). Similarly, Olesen, Westerberg, and Klingberg (2004) found BOLD increases in prefrontal and parietal areas in young adults following cognitive training. Dahlin, Stigsdotter-Nelly, Larsson, Bäckman and Nyberg, (2008) also administered a cognitive training, and found that that the improvements following training only transferred to task that rely on the neural network engaged during the training. They showed that the transfer of improvements was mediated by the striatum activation.

BOLD is not the only imaging technique used to examine the brain following WM training. Takeuchi et al. (2010) examined the structural connectivity among the prefrontal and parietal regions using diffusion tensor imaging (DTI) and showed that there were improvements in the white matter fiber tracks following WM training. Also, McNab et al. (2009) demonstrated cortical restructuring following training on a neurotransmitter level, observing changes in dopamine D1 receptor binding potential following cognitive training, indicating a translocation of the D1 receptor from the basal ganglia regions. This finding is consistent with McNab and Klingberg (2008), which implicated the basal ganglia as the filter for irrelevant information in a WM task.

It is important to note that not all studies show an increase in BOLD activation following an increase in performance. In fact, some studies show a decrease in brain activation following training (Dahlin, Bäckman, Stigsdotter-Neely & Nyberg, 2008). Garavan et al. (2000) found that practice on a vsWM task produced decreases in BOLD activity in frontal and parietal lobes. A meta-analysis by Chein and Schnieder (2005) revealed similar decreases in BOLD activation following task learning, as the brain became more efficient. The nature of changes in the BOLD signal following training is not well understood, as some studies show increases in BOLD activation (Temple, et al., 2003; Olesen, Westerberg & Klingberg, 2004) and others show decreases in BOLD activation following the training (Garavan et al., 2000; Chein & Schneider, 2005). There are many factors related to both experimental design and the physiology that may explain these differences. The critical point, however, is that training has consistently been shown to lead to changes in the brain, both in terms of BOLD activation and in terms of structure.

The presented research indicates WM's importance for everyday functions, and establishes the potential benefits from WM malleability. Much of this research implies that WM is a general process that can be further subdivided into domain specific components, such as vsWM and vWM. The subdivision of WM has both behavioral and neuronal support. Therefore, the goals for this research were (1) to examine whether vWM and vsWM are indeed separate components, and to validate new WM measures designed to tap vWM and vsWM separately, (2) to test whether vsWM can be improved through intensive training, and (3) whether this vsWM training will lead to improvements on untrained cognitive tasks, such as vsWM, vWM, inhibition, gFI, spatial abilities and mathematical abilities. Experiment 1 was designed to investigate the subdivision of WM to vWM and vsWM, while validating the new WM measures. Based on the results from Experiment 1, Experiment 2 was designed to investigate the malleability of vsWM, the neural changes following vsWM training, and the transfer of improvements to untrained cognitive tasks.

Chapter 2: Block Span & Letter Number Sequencing: Validation and Confirmation

The purpose of this experiment was to investigate the nature of WM, and whether it represents a domain general construct or domain specific constructs. Structural equation modeling was employed to examine the structure of the WM latent variable, which required the assessment of multiple tasks presumed to assess the same construct in order to reduce the task specific contribution to derive the latent variable. Therefore, while investigating the underlying properties of WM, this study also establishes the reliability and validity of newly adapted measures of WMC. The two newly redesigned tasks are Block-span, a vsWMC measure (inspired by corsi blocks: Milner, 1971), and Letter-number-sequencing, a vWMC measure (inspired by Letter-number-sequencing from Wechsler Memory Scale: Wechsler, 1997). In contrast to complex WM span tasks, these measures only require the participant to undergo a single task and, therefore, allow those taking the assessment to focus on the task at hand and not divide their attentional resources. The new tasks were designed under the premise that WM is a multi-component construct, and therefore Block-span is viewed as a vsWM task and Letter-number-sequencing as a vWM task. The validation and confirmation of these redesigned, automated, and web deployable measures of WM ability would enable easier assessments and standardization of the WMC administration (Atkins, Harbison, Bunting & Dougherty, in preparation).

Method

Participants

Native English speaking participants for the study were recruited from the undergraduate participant pool at the University of Maryland, College Park. Out of the 264 participants who were consented, 244 (148 female, M_{age}=19.45±2.65) completed all three 1.5-2 hour study sessions, and received course credit for their participation.

Assessment Materials

Table 1. Tasks administered for the Block-span and Letter-number-sequencing validation study.

Construct	Measure	Presentation Software
Visual Spatial Working Memory		
	Block-span	Python
	wBlock-span	Flash
	SymSpan	Eprime 1.2
	RotSpan	Eprime 1.2
	NavSpan	Eprime 1.2
Verbal Working Memory		
	Letter-number-sequencing	Python
	wLetter-number-sequencing	Flash
	OSpan	Eprime 1.2
	RSpan	Eprime 1.2
	RunSpan	Eprime 2.0
General Fluid Intelligence		
	RavensO	Eprime 1.2
	RavensE	Eprime 1.2
	WASI	Eprime 1.2
Spatial Abilities		
	AFOQTrb	DirectRT
	VZ2p1	DirectRT
	VZ2p2	DirectRT
Verbal Abilities		
	AFOQTa	DirectRT
	AFOQTrc	DirectRT
	RL3	DirectRT
Math Abilities		
	AFOQTm	DirectRT
	RG1	DirectRT
	RG2	DirectRT

Table 1 presents the twenty-two cognitive assessments administered and the presentation software used to administer the tasks. Both Block-span and Letter-number-sequencing were administered twice per participant.

Block-Span: In this task participants are required to remember the serial order in which a sequence of black blocks appeared in a 4 x 4 grid, where each trial is characterized by a set of 1 to J such sequences, and where each sequence consists of 2 to K blocks ($1 \leq J \leq 5$ and $2 \leq K \leq 4$). Each block within a sequence is flashed for one second, one at a time, in one of the cells within the 4 x 4 grid. The end of one sequence and the start of a new sequence within a set is indicated by flashing the entire grid for 1 second. There is a 1 second delay between the grid being flashed and the presentation of the first block of the next sequence. After the final sequence within a set, participants are prompted to indicate (via mouse click) the spatial location (in serial order) of each block within the first sequence of the set, then spatial location (in serial order) of each block within the second sequence of the set, and so forth for all sequences within the set (see Figure 1). This procedure is then repeated for the next set of sequences for the duration of the task. The Block-span score is based on correctly indicating the location of the serially highlighted blocks (Atkins, Harbison, Bunting, Teubner-Rhodes & Dougherty, 2009).

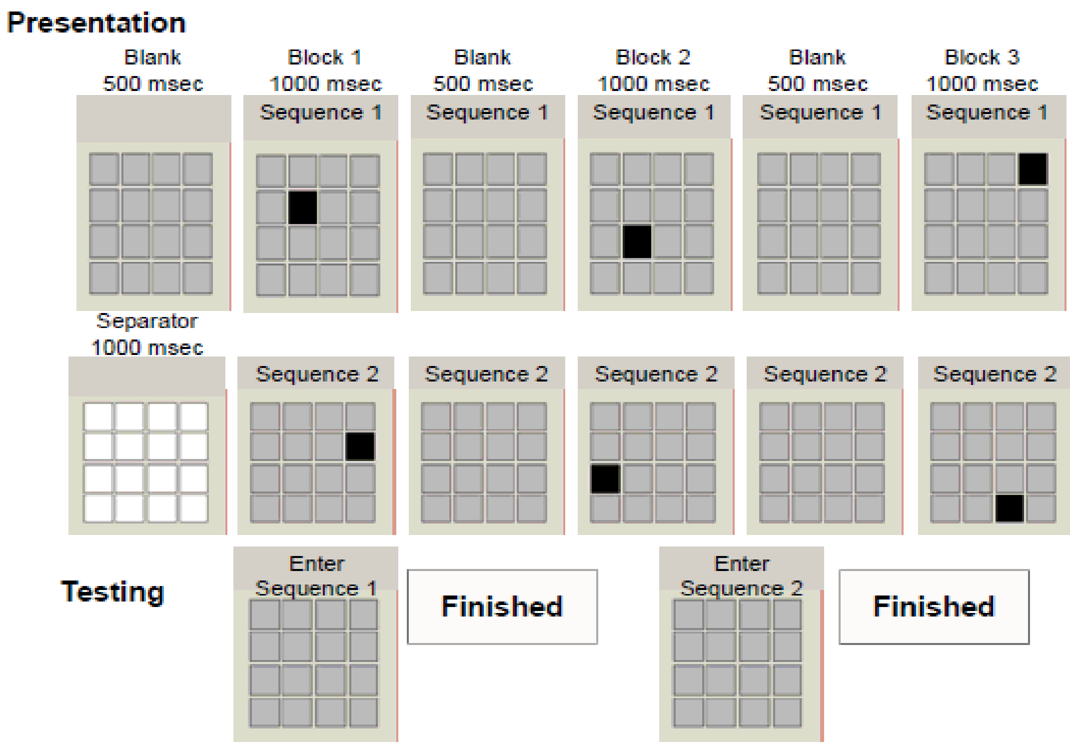


Figure 1. Trial from the Block Span task.

Letter Number Sequencing: In this task, participants were presented with a series of characters (letters and numbers) and were required to remember and restructure the characters, outputting first numbers in ascending order, then the letters in alphabetic order (see Figure 2). The Letter-number-sequencing task consists of a set of 1 to J sequences of characters, where each sequence consists of 2 to K characters (1≤J ≤4 and 2≤K≤8). The task presents each character one at a time in the center of the screen for 500msec, followed by a 500msec blank screen. The top of the screen lists the sequence number being presented. The end of one sequence of characters and the start of a new sequence within a set is indicated by an asterisk presented in the center of the screen for 500msec followed by a 500msec blank screen. After the final sequence within a set, participants are prompted to output the numbers and then letters of the first sequence of the set, then the numbers and then letters of the second sequence of the set, and so forth for all sequences within the set. For example, if presented with "7", then "T", then "1", then an "*", and then "H", then "3", and then "B", the participants should first output "17T" and then "BH3". This procedure is then repeated for the next set of sequences for the duration of the task. Letter-number-sequencing is scored based on correct recollection of the serial reordering of the characters (Atkins, Harbison, Bunting, Teubner-Rhodes, & Dougherty, 2009).

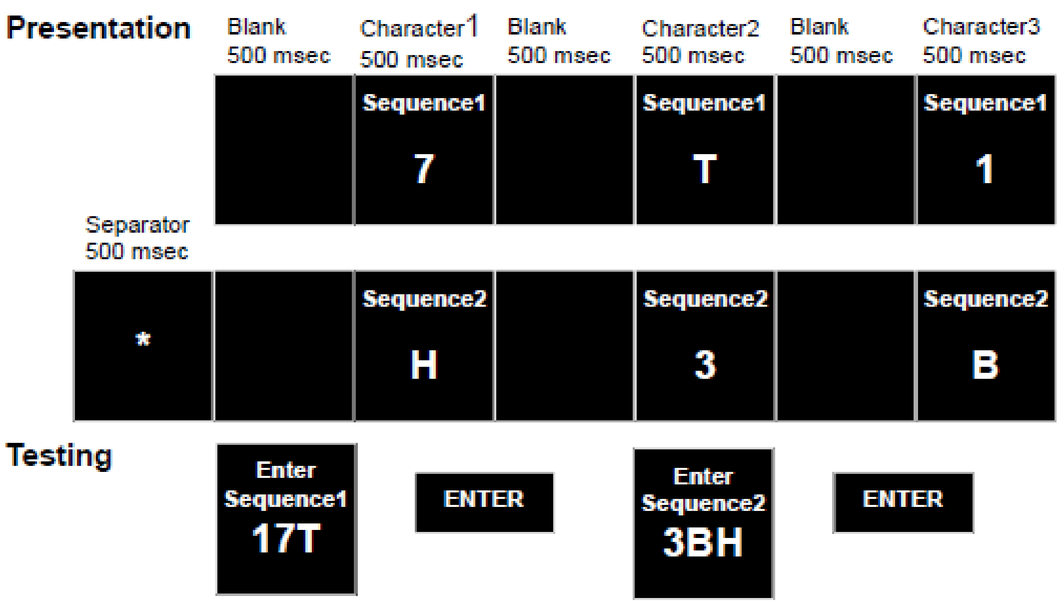


Figure 2. A trial from Letter Number Sequencing.

<u>Verbal WM Tasks:</u>

Automated Operation Span: Participants were asked to recall a series of letters. In between the presentation of the letter, they had to respond via the keyboard whether the presented solution to the math problem is true or false. Following the keyboard response to the problem, a blank screen was presented for 500 msec, followed by a letter for 650 msec. Immediately following the letter, either another math problem appeared, or the recall cue appeared. For the recall cue, participants were presented with a letters and had to recall the letter in the serial order in which they were presented. Set sizes ranged from two to seven math problem-letter displays per trial, for a total of fifteen trials and three practice trials. Correct scores were computed by counting the total number of correctly recognized letters in the correct serial position (Unsworth, Schrock, & Engle, 2004).

Automated Reading Span: Participants were asked to recall a series of letters. In between the presentation of the letters, they had to respond via the keyboard whether the sentence presented on the screen was sensible or not. Following the keyboard response to the sentence, a blank screen was presented for 500 msec, followed by a letter for 650 msec. Immediately following the letter, either another sentence appeared, or the recall cue appeared. For the recall cue, participants were presented with letters and had to recall the letter in serial order in which they were presented. Set sizes ranged from two to seven sentence-letter displays per trial, for a total of fifteen trials, and three practice trials. Correct scores were computed by counting the total number of correctly recognized letters in the correct serial position.

Automated Running Span: Presents auditory sequences of letters (F, H, J, L, N, P, R, T, V, X, Z), with each letter presented for 333 msec. The auditory sequence ends unpredictably and participants are asked to output the last six letters heard. Each auditory sequence contained between 12-20 letters, and participants were presented with 3 practice sequences and 20 trial sequences. Correct scores were computed by counting the total number of correctly outputted letters in the correct serial position (Bunting, Cowan & Saults, 2006; Broadway & Engle, 2010).

Visual Spatial WM Tasks:

Automated Symmetry Span: Participants were asked to recall the location on a 4x4 matrix of a series of red squares presented serially. In between the presentation of the red squares, they had to respond via the keyboard whether a presented image is symmetrical or not along the vertical axis. Following the keyboard response to the presented image, a blank screen was presented for 500 msec, followed by a matrix with a red square for 650 msec. Immediately following the matrix, either another image appeared, or the recall cue appeared. For the recall cue, participants were presented with a matrix and had to indicate the serial order of the location of the red block in the matrix. Set sizes ranged from two to five symmetry matrix displays per trial, for a total of twelve trials and three practice trials. Correct score was computed by counting the total number of correctly recognized arrows in the correct serial position.

Automated Rotation Span: An automated version of the Rotation Span task (Kane et al., 2004), was administered. Participants had to recall a series of short or long arrows originating at the center of the screen, and in between the presentation of the arrows, they had to respond via the keyboard whether the letter presented was normal or mirror-reversed. The letters used were capital G, F, & R, rotated at 0°, 45°, 90°, 135°, 180°, 225°, 270°, or 315°. Participants needed to rotate the letter to respond correctly. Following the keyboard response, a blank screen was presented for 500 msec, followed by a short or long arrow for 1,000 msec. When the arrow disappeared, either another letter appeared, or the recall cue appeared. For the recall cue, participants were presented with two circles of arrows, one long and one short, with each arrow originating from the center and pointing in a direction. Using the mouse, participants recalled the arrows presented, both size and direction, in the serial order of presentation. Set sizes ranged from two to five rotated letter-arrow displays per trial, for a total of twelve trials. Correct scores were computed by counting the total number of correctly recognized arrows in the correct serial position.

Automated Navigation Span: An automated version of the Navigation Span task (Kane et al., 2004), was administered. Participants had to recall a series of the paths in which moving balls moved across the screen, and in between the presentation of the moving balls, participants had to mentally navigate along the edges of a block letter “E” or “H” and indicate whether the next corner would be an inner or outer corner via the keyboard. The starting point and direction of navigation varied among trials. After navigating the entire letter, the participant indicated that he was finished via button press. Following the keyboard response, a blank screen was presented for 500 msec, followed by a ball that journeyed across the screen (varying in starting point and direction of movement). Immediately following the ball’s movement, either another letter appeared, or the recall cue appeared. For the recall cue, participants were presented with paths of movement, varying in movement origin and path, and asked to indicate, using the mouse, the serial order of the presented moving balls. Set sizes ranged from two to five letter navigation-ball movement displays per trial, for a total of twelve trials. Correct score was computed by counting the total number of correctly recognized movement pathways in the correct serial position.

Spatial Ability Tasks:

ETS Paper Folding: A multiple-choice test of spatial reasoning ability from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). Two practice items, followed by ten test items were presented to participants. Each item presented a square of paper being folded along different dimensions between one to four times, after which a hole was punched through the folded paper. Participants had to decide among five options what the paper would look like when unfolded. These items represent ETS VZ2. Part 1 and part 2 were presented on different sessions.

AFOQT Rotated Blocks: Three practice items then ten test items were presented to participants. Each item depicted a three dimensional block at various degree of orientation. Participants had to indicate which of the five presented blocks, was the depicted block from a different orientation. These items were provided for research use from Kane (personal communication) and represent items 332-334, 336-338, 340-342 and 344 from the AFOQT.

Verbal Ability Tasks:

ETS Inference: A multiple-choice test of verbal logical reasoning from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). One practice item and ten test items were presented to participants. Each item presented a passage, one to three sentences in length, about a topic. Participants chose which of the five presented sentences could be inferred from the passage without assuming any additional information or knowledge. These items represent part 1 of ETS RL3.

AFOQT Reading Comprehension: One practice item and fourteen test items were presented to participants. Each item presented a passage, one to three sentences in length, about a topic. Participants chose which of the 5 presented sentences could be inferred from the passage without assuming any additional information or knowledge. These items were provided for research use from Kane (personal communication) and represent items 1-6, 8-11,13,14,16,19,20,22,24,25 from the AFOQT.

AFOQT Verbal Analogies: One practice item and eighteen test items were presented to participants. Each item presented an incomplete analogy. Participants chose which of the five presented words or phrases could best complete the presented analogies. These items were provided for research use from Kane (personal communication) and represent items 1 to 6, 8 to 11,13,14,16,19,20,22,24, and 25 from the AFOQT.

Math Ability Tasks:

AFOQT Math: Two practice items and fifteen test items were presented to participants. Each item presented a math problem. Participants chose which of the five presented answers was correct.

ETS Arithmetic Aptitude Test: A multiple-choice test of arithmetic aptitude from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). One practice item and fifteen test items were presented to participants. Each item presented an arithmetic problem. Participants chose which of the 5 presented options answered the problem. These items represent part 1 of ETS RG1.

ETS Mathematic Aptitude Test: A multiple-choice test of mathematical aptitude from the Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). One practice item and fifteen test items were presented to participants. Each item presented a mathematical word problem. Participants chose which of the five presented options answered the problem. These items represent part 1 of ETS RG2.

General Fluid Intelligence:

Advanced Raven's Matrices: Two practice items and eighteen test items were presented to participants. Each item presented eight black and white figures arranged in a 3 by 3 grid with one figure missing. Participant chose among eight presented options the figure that best completed the pattern (Raven, Raven & Court, 1998). Even and odd items were presented on different visits.

Wechsler Adult Scale of Intelligence Matrices: One practice item and seventeen test items were presented to participants. Each item presented colorful or black and white figures with one figure missing. Participant chose among five presented options the figure that best completed the pattern (Wechsler, 1999).

Design and Procedure

The experiment took up to six hours (allowing for breaks) over three testing sessions. The testing sessions were conducted 1-7 days apart. Participants partook in at least six cognitive tasks per session, a verbal WM span, a spatial WM span, a spatial ability assessment, a verbal ability assessment, a math ability assessment, and a gFI assessment. Block-span and Letter-number-sequencing were always presented first or fourth on the first visits, and in reverse position on the second visit. A subset of the participants also partook in online web versions of Block-span and Letter-number-sequencing on the third session after the completion of the listed tasks. The cognitive assessments were presented to participants in a pseudo-random order based on individualized permutations that counterbalanced the task across participants, such that two span tasks were not presented sequentially and two modality types (verbal or spatial) were not presented sequentially.

All the cognitive assessments were computerized and presented using E-prime 1.2 or 2.0 (Schneider, Eschman & Zuccolotto, 2002), DirectRT (Jarvis, 2006), or Python (www.python.org) and required no researcher administration, beyond initiating each task. The study was administered in groups of up to seven participants.

The dataset for 189 participants is complete for all tasks administered; 55 participants are missing one or more of the data points due to technical administration issues. Due to a programming error, no data were recorded for the WASI and therefore it was excluded from all analysis.

Results of the Laboratory Assessments

The underlying cognitive structure was examined using a latent variable analysis to see whether Block-span and Letter-number-sequencing along with the other WM tasks measure a unitary WM ability or multi-component vsWM and vWM abilities. A latent variable analysis requires multiple tasks that measure the same construct, as each task measures some element of the construct along with other task specific variation. Examining multiple tasks that measure the same construct in LVA allows the task specific variation to be reduced; retaining the construct elements thereby reveals a clearer representation of the construct. This analysis, while requiring multiple tasks per construct of interest, allows for conclusions about the underlying processes and the variable structure, as opposed to making task specific conclusions about the dataset.

To conduct the latent variable analysis, Block-span and Letter-number-sequencing must first be shown to be good psychological measures (internal consistency, test retest reliability), to measure WM at all (construct validity), and to predict performance on other tasks similar to other WM tasks (criterion validity).

Table 2. Descriptive statistics for all the tasks administered in Experiment 1.

Construct	Measure	N	M	Median	SD
Spatial Working Memory					
	Block-span	243	53.16	52.00	13.19
	wBlock-span	58	1476.55	1410.00	597.21
	SymSpan	241	18.85	19.00	9.72
	RotSpan	243	20.20	22.00	10.72
	NavSpan	212	22.08	22.00	9.11
Verbal Working Memory					
	Letter-number-sequencing	242	187.81	191.00	41.94
	wLetter-number-sequencing	58	630.52	665.00	306.18
	Ospan	242	45.81	49.00	19.08
	Rspan	243	35.21	36.00	18.24
	RunSpan	243	48.47	49.17	11.74
General Fluid Intelligence					
	RavensO	240	49.72	50	20.70
	RavenE	244	44.22	38.89	19.61
Spatial Abilities					
	AFOQTrb	241	38.34	40.00	22.11
	VZ2p1	238	53.92	54.55	18.74
	VZ2p2	243	53.54	54.55	15.62
Verbal Abilities					
	AFOQTa	244	58.11	58.33	17.64
	AFOQTrc	243	59.85	64.29	20.81
	RL3	244	58.57	60.00	20.77
Math Abilities					
	AFOQTm	242	20.66	20.00	10.95
	RG1	243	31.00	33.33	12.55
	RG2	241	23.83	21.43	11.81

Internal Consistency

Cronbach's alpha coefficients were computed between the trials of Block-span and also between the trials of Letter-number-sequencing to examine the measurement properties of the tasks. Both Block-span and Letter-number-sequencing were found to be highly reliable tasks (Block-span: 16 items Cronbach $\alpha = 0.76$; Letter-number-sequencing: 102 items Cronbach $\alpha = 0.92$). The next step is to examine whether the Block-span and Letter-number-sequencing tasks measure a stable property.

Test Retest Reliability

Test-retest reliability on Block-span and Letter-number-sequencing was computed based on participants' performance on the first and second session. Both Block-span and Letter-number-sequencing demonstrated high test-retest reliability across sessions, Block-span: $r(242)=0.70$, $p<0.001$ and Letter-number-sequencing: $r(241)=0.73$, $p<0.001$, indicating that both Block-span and Letter-number-sequencing are reliable. Table 3 provides descriptive statistics for Block-span and Letter-number-sequencing (see Table 2 for descriptive statistics for all administered tasks).

Table 3. Descriptive statistics for Block-span and Letter-number-sequencing for both test and retest administrations. LowerQ= lowest quartile; UpperQ= highest quartile.

Measure	N	M	Median	SD	Skew	Kurtosis	LowerQ	UpperQ
Block Span								
Test	243	53.16	52	13.91	0.47	0.1	44	61
Retest	244	53.76	52	14.93	0.25	0.07	44	63.75
Letter Number Sequencing								
Test	242	187.81	191	41.94	-0.65	0.31	164	216
Retest	244	197.86	203.5	45.23	-0.2	0.31	164.25	232

Construct and Criterion Validity

Block-span and Letter-number-sequencing are significantly correlated with all other complex WM span tasks (r's=0.23 to 0.56, all p's<0.0001). In addition, Block-span, being a spatial task, relates more strongly to the spatial WM tasks (r's=0.39 to 0.56), compared to the vWM tasks (r's=0.29 to 0.36), and Letter-number-sequencing, being a verbal task, relates more with the vWM tasks (r's=0.45 to 0.46) than to the vsWM tasks (r's=0.23 to0.43). These results indicated that Block-span and Letter-number-sequencing are valid WM measures as they are strongly correlated to all other WM tasks. The pattern of results shows that Block-span is strongest in relationship to the spatial WM tasks and Letter-number-sequencing is strongest in relationship to the verbal WM tasks, and the implication is that Block-span and Letter-number-sequencing have good construct validity for measuring spatial and verbal WM respectively.

Criterion validity was assessed when examining the relationship between the WM task and Ravens, a measurement of gFI. Block-span and Letter-number-sequencing are as predictive of the gFI measurement (Block-span: r's=0.34 to 0.39, p's<0.001; Letter-number-sequencing: r's=0.27 to 0.37, p's<0.001) as are the other WM assessments in this study (r's=0.22 to 0.37, p's<0.001-0.05). Table 4 shows the correlation between the WM tasks and the gFI measurements.

The gFI measurement, Ravens, was administered twice using odd and even item numbers (RavensO and RavensE). The Ravens test-retest were only correlated at r=0.56, therefore any correlation with gFI would be capped at that level. The block-span and letter-number-sequencing correlations were therefore corrected for the attenuation, such that $r_{corrected} = r_{unattentuated}/sqrt(r_{WMtest\text{-}retest} * r_{RAVENSodd\text{-}even})$. Therefore, the correlations corrected for the attenuation are as follows: RavensEven: block-span: r=0.54 and letter-number-sequencing: r=0.42; RavensOdd: block-span: r=0.62 and letter-number-sequencing: r=0.58.

Table 4. Correlations between Block-span, Letter-number-sequencing, the complex WM span tasks and Ravens. The *n* is in parenthesis. All r's are significant at p<0.05; bolded r's are significant at p<0.001.

	Block -span	Sym-Span	Rot-Span	Nav-Span	Letter-number-sequencing	O-Span	R-Span	Run-Span	Raven O	Raven E
Block-span	-									
SymSpan	**0.47** (240)	-								
RotSpan	**0.39** (242)	**0.38** (240)	-							
NavSpan	**0.56** (211)	**0.41** (210)	**0.48** (211)	-						
Letter-number-sequencing	**0.42** (241)	**0.28** (239)	**0.23** (241)	**0.43** (210)	-					
OSpan	**0.36** (241)	**0.29** (239)	**0.24** (241)	**0.46** (210)	**0.45** (240)	-				
RSpan	**0.33** (242)	**0.41** (240)	**0.20** (242)	**0.4** (211)	**0.46** (241)	**0.63** (241)	-			
RunSpan	**0.29** (242)	0.15 (240)	0.17 (242)	**0.26** (211)	**0.46** (241)	**0.38** (241)	**0.42** (242)	-		
RavenO	**0.39** (239)	**0.31** (237)	**0.31** (239)	**0.37** (208)	**0.37** (238)	**0.24** (238)	**0.23** (240)	**0.24** (239)	-	
RavenE	**0.34** (243)	**0.33** (241)	**0.28** (243)	**0.31** (212)	**0.27** (242)	**0.22** (242)	**0.24** (243)	**0.24** (243)	**0.56** (240)	-

A further analysis of criterion validity assumes a domain specific WM component and examines the relationships between the domain specific reasoning tasks and the new WM measures, Block-span and Letter-number-sequencing (see

Table **8** for the full correlation matrix). As expected Block-span is as strongly correlated to the spatial ability tasks as are the other spatial WM tasks (see Table 5). Similarly Letter-number-sequencing is as strongly correlated to the verbal ability tasks, as are the other verbal WM tasks (see Table 6). The results from the correlation with the math abilities tasks were less clear. One of the math tasks, AFOQTm, was only very weakly correlated with the other math tasks and did not correlate with any of the WM tasks. However, the other math tasks, RG1 and RG2, showed weak positive correlations with Block-span and Letter-number-sequencing and similar correlations with the other WM tasks. In fact, the relationship between RG1 and Block-span is the strongest relationship of the all WM measures to the math tasks ($r(241)=0.33$, $p<0.001$) (see Table 7).

Table 5. Correlations between the spatial reasoning tasks and the vsWM tasks. The *n* is in parenthesis. All r's are significant at p<0.05; bolded r's are significant at p<0.001.

	AFOQTrb	VZ1p1	VZ2p2
AFOQTrb	-		
VZ2p1	**0.42** (235)	-	
VZ2p2	**0.35** (240)	**0.49** (237)	-
Block-span	**0.41** (240)	**0.36** (237)	**0.27** (242)
SymSpan	**0.36** (238)	**0.32** (235)	0.13 (235)
RotSpan	**0.28** (240)	**0.35** (237)	**0.23** (242)
NavSpan	**0.42** (211)	**0.38** (209)	**0.29** (212)

Table 6. Correlation between the verbal reasoning tasks and verbal WM tasks. The *n* is in parenthesis. All r's are significant at p<0.01; bolded r's are significant at p<0.001.

	AFOQTa	AFOQTrc	RL3
AFOQTa	-		
AFOQTrc	**0.50**(243)	-	
RL3	**0.43** (244)	**0.42** (243)	-
Letter-number-sequencing	**0.33** (242)	**0.28** (241)	0.21 (242)
OSpan	**0.39** (242)	0.22 (241)	**0.23** (242)
RSpan	**0.33** (243)	**0.28** (242)	0.19 (243)
RunSpan	**0.38** (243)	**0.30** (242)	0.17 (243)

Table 7. Correlation between the mathematical reasoning tasks and the WM tasks. The *n* is in parenthesis. Underlined r's are not significant, not underlined r's are significant at p<0.01; bolded r's are significant at p<0.001.

	AFOQTm	RG1	RG2
AFOQTm	-		
RG1	0.20 (241)	-	
RG2	0.19 (239)	**0.35** (240)	-
Block-span	*0.12* (241)	**0.33** (242)	**0.24** (240)
Letter-number-sequencing	*0.09* (240)	**0.28** (241)	0.18 (2439
OSpan	*0.09* (240)	0.16(241)	*0.12* (239)
RSpan	*0.01* (241)	*0.12* (242)	0.18 (240)
RunSpan	*0.06* (241)	**0.22** (242)	*0.12* (240)
SymSpan	*0.13* (239)	0.16 (240)	0.18 (239)
RotSpan	*-0.03* (241)	**0.27** (242)	0.14 (240)
NavSpan	*-0.01* (210)	**0.21** (211)	**0.25** (210)

Table 8. Correlations between all the tasks in Experiment 1. The *n* is in parenthesis.

	Block-span	wBlock-span	SymSpan	RotSpan	NavSpan	Letter-number-sequencing	wLetter-number-sequencing	OSpan	RSpan	RunSpan
Block-span	-									
wBlock-span	0.73 (58)	-								
SymSpan	0.47 (240)	0.47 (57)	-							
RotSpan	0.39 (242)	0.26 (58)	0.38 (240)	-						
NavSpan	0.56 (211)	0.31 (57)	0.41 (210)	0.48 (211)	-					
Letter-number-sequencing	0.42 (241)	0.20 (58)	0.28 (239)	0.23 (241)	0.43 (210)	-				
wLetter-number-sequencing	0.32 (58)	0.37 (58)	0.42 (57)	0.53 (58)	0.36 (57)	0.46 (58)	-			
OSpan	0.36 (241)	0.28 (58)	0.29 (239)	0.24 (241)	0.46 (210)	0.46 (210)	0.59 (58)	-		
RSpan	0.33 (242)	0.27 (57)	0.41 (240)	0.20 (242)	0.40 (211)	0.4 (211)	0.40 (57)	0.63 (241)	-	
RunSpan	0.29 (242)	-0.03 (58)	0.15 (240)	0.17 (242)	0.26 (211)	0.26 (211)	0.26(58)	0.38 (241)	0.42 (242)	-
RavenO	0.39 (239)	0.20 (57)	0.31 (237)	0.31 (239)	0.37 (208)	0.37 (208)	0.35 (57)	0.24 (238)	0.23 (240)	0.24 (239)
RavenE	0.34 (243)	0.23 (58)	0.33 (241)	0.28 (243)	0.31 (212)	0.31 (212)	0.19 (58)	0.22 (242)	0.24 (243)	0.24 (243)
AFOQTbr	0.42 (240)	0.27 (58)	0.36 (238)	0.28 (240)	0.42 (211)	0.29 (239)	0.07 (58)	0.27 (239)	0.27 (240)	0.23 (240)
VZ2	0.36 (237)	0.34 (58)	0.32 (235)	0.35 (237)	0.38 (209)	0.36 (236)	0.38 (58)	0.30 (236)	0.27 (237)	0.33 (237)
VZ2p2	0.27 (243)	0.27 (58)	0.13 (241)	0.23 (242)	0.29 (212)	0.25 (241)	0.23 (58)	0.21 (241)	0.13 (242)	0.23 (242)
AFOQTa	0.30 (243)	0.18 (58)	0.20 (241)	0.28 (243)	0.32 (211)	0.33 (242)	0.23 (58)	0.39 (242)	0.33 (243)	0.38 (243)
AFOQTrc	0.25 (242)	0.06 (57)	0.20 (240)	0.22 (242)	0.36 (211)	0.28 (241)	0.02 (58)	0.22 (241)	0.28 (242)	0.30 (242)
RL3	0.21 (243)	0.12 (58)	0.18 (241)	0.20 (243)	0.25 (212)	0.20 (242)	0.27 (58)	0.23 (242)	0.20 (242)	0.17 (243)
AFOQTm	0.12 (241)	0.03 (58)	0.13 (239)	-0.03 (241)	-0.01 (210)	0.09 (240)	0.01 (58)	0.09 (240)	0.08 (241)	0.06 (241)
RG1	0.33 (242)	0.22 (58)	0.16 (240)	0.27 (242)	0.21 (211)	0.28 (241)	0.20 (58)	0.16 (241)	0.12 (242)	0.22 (242)
RG2	0.24 (240)	0.19 (58)	0.18 (239)	0.14 (240)	0.25 (210)	0.18 (239)	0.10 (58)	0.12 (239)	0.18 (240)	0.12 (240)

	RavenO	RavenE	AFOQTbr	VZ2p1	VZ2p2	AFOQTa	AFOQTrc	RL3	AFOQTm	RG1	RG2
RavenO	-										
RavenE	0.56 (240)	-									
AFOQTbr	0.42 (237)	0.34 (241)	-								
VZ1	0.47 (234)	0.41 (238)	0.42 (235)	-							
VZ2p1	0.38 (239)	0.31 (243)	0.35 (240)	0.49 (237)	-						
AFOQTa	0.50 (240)	0.44 (244)	0.41 (241)	0.43 (238)	0.35 (243)	-					
AFOQTrc	0.42 (239)	0.36 (243)	0.32 (240)	0.34 (237)	0.26 (242)	0.50 (243)	-				
RL2	0.35 (240)	0.32 (244)	0.24 (241)	0.26 (238)	0.21 (243)	0.43 (244)	0.42 (243)	-			
AFOQTm	0.14 (238)	0.14 (242)	0.21 (239)	0.11 (236)	0.11 (241)	0.08 (239)	0.03 (241)	-0.01 (242)	-		
RG1	0.19 (239)	0.26 (243)	0.29 (240)	0.23 (237)	0.21 (242)	0.31 (243)	0.14 (242)	0.27 (243)	0.19 (241)	-	
RG2	0.18 (240)	0.24 (237)	0.20 (241)	0.13 (238)	0.19 (235)	0.08 (241)	0.20 (241)	0.13 (241)	0.19 (239)	0.35 (240)	-

Factor Analysis and Model Comparison

Block-span and Letter-number-sequencing showed moderate to strong positive correlations with the other WM measures. They also showed moderate domain specific correlations for the spatial and verbal reasoning abilities. Lisrel 8.8 was used to conduct all factor analyses. An exploratory factor analysis (maximum likelihood, promax rotated) was conducted on all the study tasks[1], producing a three factor solution, with Block-span loading with SymSpan, RotSpan, and NavSpan on one factor, Letter-number-sequencing loading with OSpan, RSpan, and RunSpan on the second factor, and Ravens, AFOQTa, AFOQTrc, RL3, and VZ2 loading on the third factor. AFOQTrb loaded equally on the first and third factors (see Table 9 for exploratory factor loadings). The pattern of factor loadings implies a vsWM factor, a vWM factor, and a factor related to reasoning abilities (see Table 10 for the correlation between the factors). The exploratory factor analyses further strengthens the correlation result suggesting domain specific components for WM and Block-span being a vsWM task and Letter-number-sequencing being a vWM task.

[1] Only one of VZ2 and the Ravens administrations was added into the analysis.

Table 9. Exploratory factor analysis for the WM measures in Experiment 1, *n*=200. Factor loading <.3 are suppressed.

Measure			vsWM	vWM	Reasoning/ Abilities	
	M	SD	Factor1	Factor2	Factor3	Unique Variance
Block-span	52.67	13.37	0.72			0.45
Letter-number-sequencing	187.58	40.54		0.57		0.51
OSpan	45.43	18.93		0.71		0.41
Rspan	34.86	18.28		0.75		0.40
RunSpan	48.54	11.99		0.55		0.64
SymSpan	18.99	9.71	0.60			0.64
RotSpan	20.92	10.20	0.67			0.60
NavSpan	21.93	9.22	0.72			0.41
Ravens	43.69	19.24			0.45	0.65
AFOQTa	6.35	2.47			0.92	0.30
AFOQTrc	6.35	2.47			0.57	0.63
RL3	6.35	2.47			0.51	0.74
AFOQTrb	2.38	1.45	0.45		0.32	0.61
VZ2	3.71	1.44			0.32	0.78

Table 10. Correlation between the factors in the exploratory factor analysis.

	vsWM	vWM	Reasoning
vsWM	-		
vWM	0.65	-	
Reasoning	0.364	0.371	-

Structural equation modeling was used to test whether Block-span and Letter-number-sequencing and the other WM tasks represent a single latent variable or multiple latent variables. Lisrel 8.8 was employed to conduct the path model used to confirm the latent nature of WM. The single unitary WM model consisted of a single latent variable with loading for all the WM tasks. The multi-component WM model consisted of two latent variables, one for the verbal WM with loadings for all the WM tasks involving letters and verbal information (Letter-number-sequencing, OSpan, RSpan, and RunSpan), and one for the visual spatial domain with loading for all the

WM tasks of visual spatial nature (Block-span, SymSpan, RotSpan, and NavSpan) . See Figure 3 for the path models and factor loading.

The unitary WM model does not provide a good fit for the data, $\chi^2(20, N=202) = 96.69$, $p=0.0001$, CFI =0.91, NFI=0.89, NNFI=0.88, GFI=0.87, SRMR=0.08, AGFI=0.77, RMSEA=0.16, as indicated by the model fit indices. For a good model fit, the model fit indices should be as follows: Incremental fit indices: CFI≥0.95, NFI≥0.9 and NNFI≥0.95; Absolute fit indices: GFI≥0.9 and SRMR≤0.8; Parsimonious model fit indices: AGFI≥0.9 and RMSEA≤0.06. As indicated, the unitary WM model is not a good fit for the data, as only one absolute index (SRMR) indicates a good fit, whereas all the other fit indices indicate that the model is not a good fit for the data. In contrast, the multi-component WM model does provide a good fit $\chi^2(19, N=202)=32.4$, $p=0.028$, CFI =0.98, NFI=0.96, NNFI=0.98, GFI= 0.96, SRMR=0.04, AGFI=0.93, RMSEA<0.06, as all the model fit indices indicate that it is a good fit for the data. Furthermore, the multi-component WM model did significantly better than the unitary model of WM, χ^2difference $(1,N=202)=64.29$, $p<0.0001$. This finding suggests that WM can be subdivided to multiple components. It is important to note that the subcomponents of WM are strongly correlated (r=0.69), reflecting the possibility of a general process engaging the subcomponents.

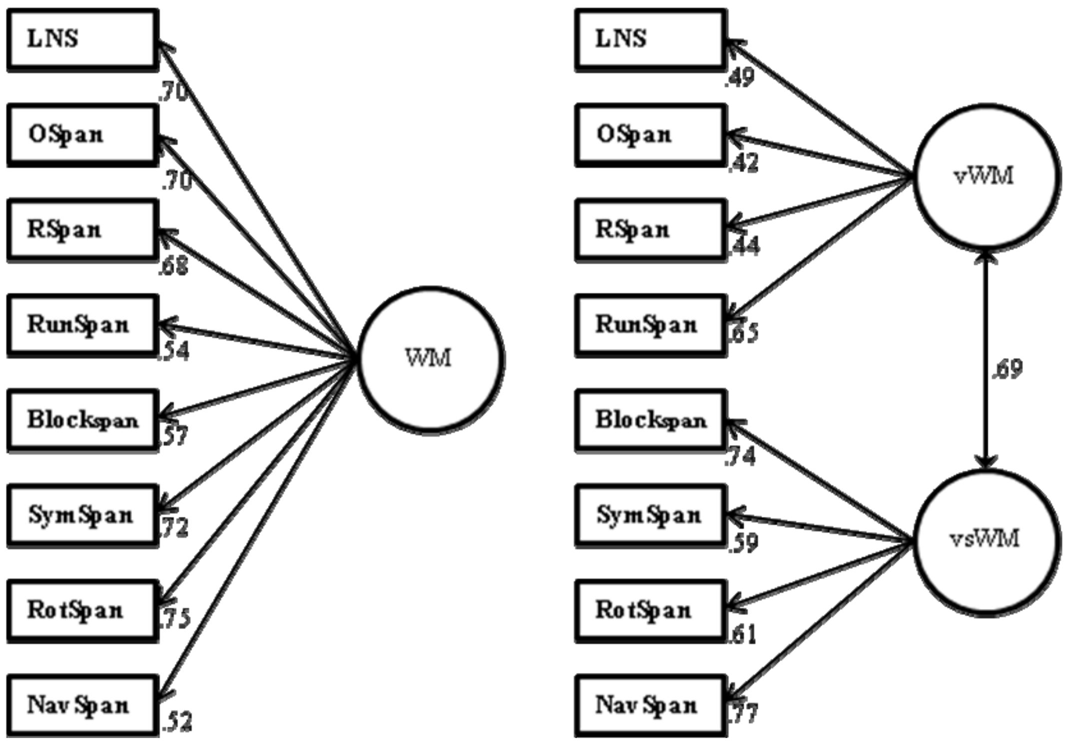


Figure 3. Path models for the confirmatory factor analysis. Panel a: the unitary model of WM. Panel b: Domain specific model of WM.

Results of the Web Assessments

In the previous section, Block-span and Letter-number-sequencing were shown to be valid, reliable assessments of vsWM and vWM when assessed in a laboratory format. The following analysis examines versions of Block-span and Letter-number-sequencing constructed for online, web based administration.

Participants

Fifty-eight participants (26 female, M_{age}=19.41±1.3) completed a web administered version of Block-span and Letter-number-sequencing at the end of their third visit.

Web Assessed Block-span and Letter-number-sequencing

The online Block-span and Letter-number-sequencing versions were programmed in Flash so that they could be administered online within the web browser. The tasks were adapted for web administration, and the differences are described below.

wBlock-span: The task presented a 4 X 4 grid of blue squares and highlighted a yellow square (Figure 4). Participants saw "wait" on the bottom of the screen when they were viewing the block sequences and "go" when they were requested to input the sequences. Immediate feedback was given for performance by a green (correct) or red (incorrect) flash of the pressed block, scores were given for correct responses, and sequential correct responses doubled the score allotted. Participants underwent sixteen trials, four trials in each of the block levels (2, 3, 4, and 5).

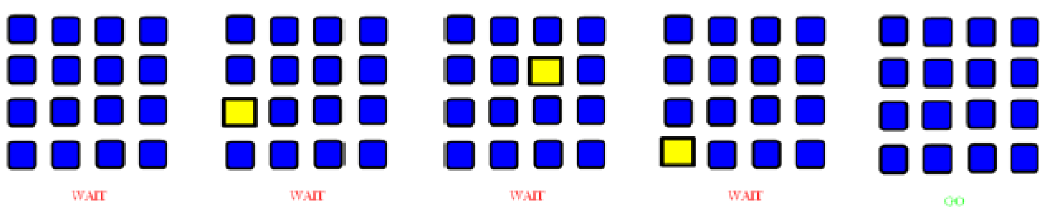


Figure 4. A trial for the web Block Span.

wLetter-number-sequencing: The task presented a series of black characters in a white window presented at the top of a virtual keyboard (Figure 5). Above the white window, there was an indication of the round, sequence, and score. Participants pressed the start button and the characters were presented one at a time in the window. After the presentation a blue square highlighted first the numbers, then the letters section above the keyboard, and participants inputted the number and letters in the sequence. Feedback was presented for each sequence (correct or incorrect in the window), and scores were given for correct responses, and sequential correct responses doubled the score allotted. Participants underwent thirteen trials, seven with one sequence of 2,3,4,5,6,7,and 8 characters respectively and six trials of two sequences of 2, 2, 3,4, 5, 6 respectively.

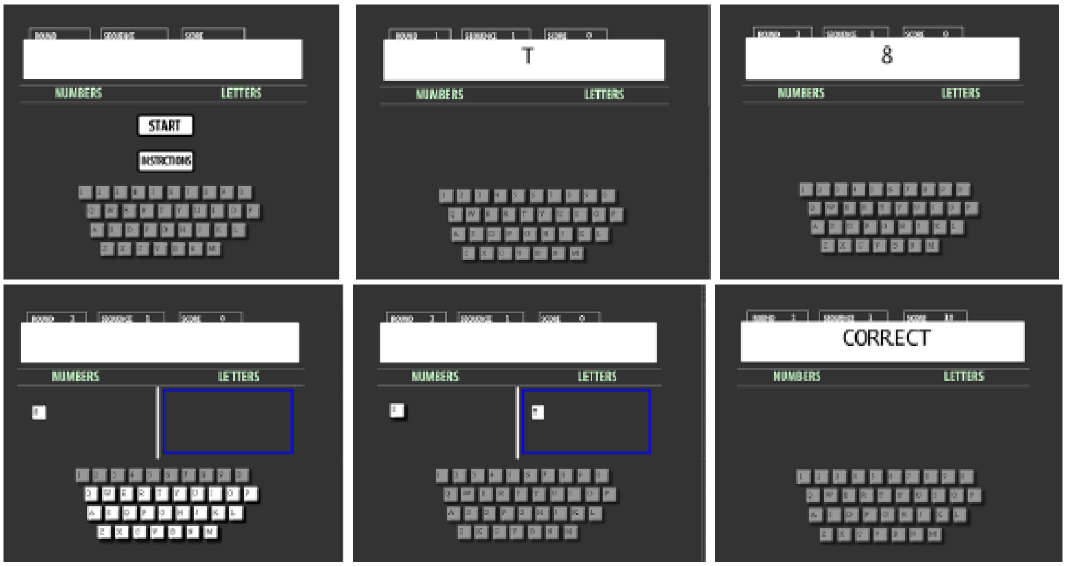


Figure 5. A trial from the web Letter Number Sequencing.

Web Task Reliability

The test retest reliability between the laboratory and the web assessed versions of Block-span and Letter-number-sequencing was computed to examine the relationship between the laboratory administered task and the one designed for web administration. As before, Block-span and Letter-number-sequencing demonstrate strong test-retest reliability between the laboratory and the web administration, Block-span: $r(57)=0.73$ to 0.67, $p<0.001$ and Letter-number-sequencing: $r(57)=0.47$ to 0.59[2], $p<0.001$. These results indicate that the web administrations of the Block-span and Letter-number-sequencing tasks are stable, reliable measures. Table 11

[2] Letter-number-sequencing task show a significant difference between the correlations Test-Retest and the Test-web ($t(55)=3.23$, $p<0.01$). However, the difference between the Retest-Test and the Retest-web is not significant ($t(55)=1.66$, p=n.s). Therefore, the web-Letter-number-sequencing did not differ from the most recent laboratory testing, but from the initial introduction to the Letter-number-sequencing task. A change in the instructions to the web Letter-number-sequencing has been made for clarity purposes.

shows the correlation between all the administrations of Block-span and Letter-number-sequencing.

Table 11. Correlations between the Block-span and Letter-number-sequencing administrations. The *n* is in parenthesis.

Block Span				Letter Number Sequencing			
	Test	Retest	Web		Test	Retest	Web
Test	-			Test	-		
Retest	0.76 (58)	-		Retest	0.74 (58)	-	
Web	0.73 (58)	0.67 (58)	-	Web	0.47 (58)	0.59 (58)	-

Construct and Criterion Validity

The correlations between the wBlock-span and wLetter-number-sequencing show an overall pattern of relationship similar to that of the laboratory versions, where wBlock-span is more strongly correlated with the vsWM tasks and wLetter-number-sequencing is more strongly correlated with the sWM tasks (Table 12, all p's<0.05 except for Block-span-RunSpan). One difference is the lack of a relationship between wBlock-span and RunSpan. At present, I can only speculate about the lack of relationship, but future analysis should look into this relationship, or lack thereof.

Table 12. Correlations between wBlock-span, wLetter-number-sequencing, ravens and the WM tasks. The *n* is in parenthesis.

	wBlock-span	wLetter-number-sequencing
wLetter-number-sequencing	0.37 (58)	-
SymSpan	0.47 (57)	0.42 (57)
RotSpan	0.26 (58)	0.53 (58)
NavSpan	0.31 (57)	0.36 (57)
OSpan	0.28 (58)	0.59 (58)
RSpan	0.27 (57)	0.40 (57)
RunSpan	-0.03 (58)	0.26(58)
Ravens	0.23 (57)	0.35 (57)

Discussion

This experiment introduced two newly designed domain specific assessments of WM ability, Block-span and Letter-number-sequencing, in both laboratory and online web deliverable versions. The new WM tasks were found to have stable, reliable measurement properties and strong positive relationships to other validated vsWM and vWM tasks. Construct validity for the new WM was evident in the correlations with other WM tasks and abilities to which WM is known to relate. The pattern of correlation strengths suggests a multi-components model of WM, as Block-span was more strongly correlated to the vsWM tasks and the visual spatial ability tasks, and Letter-number-sequencing was more strongly correlated to the vWM tasks and the verbal ability tasks. Furthermore, both exploratory and confirmatory factor analysis shows Block-span loading with the other vsWM tasks and Letter-number-sequencing with the vWM tasks.

A latent variable analysis using SEM model comparison showed that that a unitary WM model does not provide a good fit for the data, whereas the multi-component WM model, with two latent variables (vsWM and vWM) provides a very good fit for the data. The significant difference on the comparison between the models provides additional support for the multi-component model of WM. The multi-component WM model showed that Block-span loaded strongly on the vsWM factor and Letter-number-sequencing loaded strongly on the vWM factor, providing further evidence of the construct validity of Block-span and Letter-number-sequencing.

The finding of a two-factor solution (vsWM and vWM) replicates the WM findings of Kane et al.'s (2004), which also producing a two-factor model for vWM and vsWM. Similar to the Kane et al. study, the WM SEM also revealed a strong relationship between the vsWM and vWM latent variables (r=0.69, Kane: r=0.84). Kane et al. argued that the strong relationship between the domain specific WM components reflected a general component which they labeled “attention”, and conducted additional SEM to examine the relationship short-term-memory and WM have to executive attention and storage components.

However, the controlled attention perspective cannot be conclusively supported by results of the current study. In fact, other researchers have shown results inconsistent with this perspective. Friedman and Miyake (2004) argue against the controlled attention perspective, as it has difficulties with accounting for the lack of relationship between response-distractor inhibition and resistance to proactive interference. Additional difficulties lie in the dissociation found between manipulation of vWM and vsWM measures (Shah & Miyake, 1996), as well as the asymmetrical interference patterns, where the visual interference tasks engage more executive processes then verbal interference tasks (Ricker, Cowan & Morey, 2010). The current study’s’ findings are consistent with a growing body of behavioral and neuronal evidence providing support for the multi-component domain specific distinction in WM (Smith, Jonides, & Koeppe, 1996; Smith & Jonides, 1997; Hartley & Speer, 2000; Baddeley, 2000; Klingberg, 2006; D’Esposito, 2007; Tomasi, et al., 2007; Bull, Espy & Wiebe, 2008; Thomason, et al., 2009).

To conclude, the multi-component domain-specific model of WM, with a vsWM and a vWM component, showed a good fit for the data. Block-span loaded with the vsWM measures and Letter-number-sequencing loaded with the vWM measures. Block-span and Letter-number-sequencing are valid, reliable WM measures, with strong construct validity and good criterion validity, and can be successfully deployed online to measure vsWM and vWM respectively.

Chapter 3: Malleability of WM.

The second aspect of WM that this research investigates is the malleability of WM. While some studies have shown WM improvements and the transfer of training-induced improvements to other cognitive tasks following WM training, not all studies of WM training have shown transfer. This study addresses the malleability of WM within the domain specific framework derived from the previous experiment, and devised vsWM training tasks.

As in the previous experiment, the training tasks were designed to be administered online and were adapted from two measures of vsWM: Block-span and ShapeBuilder. Most studies of WM training administer training, in individual or group settings with the researcher or with parent present. This type of administration is resource demanding. One of the goals of this study was to test whether online WM training, where task performance is self-motivated and performance competes with real-world distractions, can lead to cognitive improvements.

The effectiveness of WM training is measured not only by improvements in the trained tasks but also by transfer to other tasks. The goal of WM training is to demonstrate improvement on both the trained and untrained tasks. I hypothesize that transfer will be limited to behavioral tasks that have an overlapping neural network with the trained tasks, so called process-specific cognitive tasks (Dahlin, Stigsdotter-Neely, Larsson, Bäckman, & Nyberg, 2008). Tasks that rely on different neural networks should show no benefit from the training.

In this Experiment the process-specific tasks are examined through behavioral measures[3]. However, the neuronal networks engaged via the assessed tasks are used to conceptualize process-specificity. Therefore, prior literature examining task specific brain activation, as well as latent variable analyzes on behavioral measures, serve to infer the relationship between the neural network of the trained task and the behavioral cognitive tasks.

The vsWM training was designed to engage the frontal-parietal network, which has been implicated in a multitude of WM and inhibition tasks (Edin et al., 2007; McNab et al., 2008; Klingberg, 2010). For that reason, vsWM tasks and visual inhibition task are unambiguously process-specific tasks. However, the factor analysis conducted in the first experiment shows domain-specificity for WM abilities, and separates the vsWM, from the vWM, and from general abilities (Ravens and the Verbal and Spatial abilities). Therefore, the definition of process-specificity for some tasks can be unclear, thus process-specificity will be treated as a continuum based on the strength of the relationship, not a dichotomy.

This experiment assessed cognitive ability, behaviorally and neuronally, using fMRI for the trained task, before and after the online WM training, and examined whether online WM training led to neuronal changes in the brain and to behavioral process-specific transfer.

---

[3] A larger set of data exists to examine process-specific neuronal transfer. That data will be reported elsewhere.

Method

Participants

Participants were recruited from Georgetown University and the surrounding community via the Georgetown research volunteer program and flyers placed around campus. Participants were right-handed individuals, aged 18-30, native English speakers, with normal or corrected-to-normal vision, who had no personal history of neurological, neuropsychiatric, and/ or psychiatric disorders and or learning disabilities, and were not taking medication related to neuropsychiatric and/ or psychiatric disorders and or learning disabilities. Other restrictive criteria included that participants not have metal in their body, and that female participants were not pregnant, as confirmed by a pregnancy test.

The study duration was seventeen experimental hours, separated into three and a half hours for each pre and post assessment sessions and ten hours of online computer training, for which participants were compensated $215. Participants also entered a raffle for every new high score achieved on the training; the raffle grand prize was a $200 gift card. Out of the forty-five participants who consented to be in this study, thirty-six completed the study. One of the participants who completed the study was ambidextrous, and was removed from the analyses. Participants were randomly assigned either to the vsWM training group (N=18; 12 female; M_{age} = 2223.11±3.83 years) or to the placebo control group (N = 17; 11 female; M_{age} = 22.88±3.33 years).

Pre-training and Post-training Assessments

The following assessments were administered both pre-training and post-training (see Table 14 for reliability coefficients and Table 15 for descriptive statistics):

wLetter-number-sequencing: described in Experiment1

OSpan: described in Experiment1

SymSpan: described in Experiment1

Ravens : described in Experiment1

ShapeBuilder: Participants need to remember and reproduce the serial order, shape, and location in which a sequence of shapes appears in a 4 x 4 grid, when each trial is characterized by a set of two to four shapes. Each shape within a sequence is flashed for one second, one at a time, in one of the cells within the 4 x 4 grid. After the presentation of a sequence, participants are prompted to reproduce the sequence (via mouse click and drag from a palette surrounding the grid) with respect to the serial order, spatial location, color, and shape type of each presented shapes. This procedure is repeated for the next set of sequences for the duration of the task. Shape Builder is scored based on correct recollection of the serial location of shapes, with extra points for the shape type and color.

Verbal Fluency: Participants were given a minute to generate all the words they can think of that match the presented instruction. The instructions were: Words beginning with "F"; Words beginning with "A", Words beginning with "S", Things you can eat or drink, People's first name, Animals. The number of correct items generated per category was collected.

Mental Math: One practice item and sixty test items, ten trials for each of six levels of mathematical difficulty, were presented to participants. Participants had to compute the answer to the mathematical problem, when items were presented sequentially in the center of the screen. Measures of accuracy and reaction time were collected.

ModMath: Eight practice items, followed by eighty test items, forty problems for each of two levels of mathematical difficulty, were presented to participants. Participants had to indicate whether the solution of the mod math problem on the screen was true or false (Beilock & Carr, 2005). Measurements of accuracy and reaction time were collected.

Mental Rotation: Participants were presented with two two-dimensional shapes presented simultaneously and were asked to indicate whether the two shapes are same or different. A same classification meant that they were rotated on the picture plane, whereas a different classification meant that they were mirror images. Participants classified 150 shape-pairs, half were mirror images and half were rotated images. Equal numbers of trials were presented for 0°, 45°, and 135° orientations. Accuracy and reaction time were collected.

Stroop: Participants were asked to indicate, via button press, the ink color of the series of characters presented on the screen. The series of characters was presented in Green, Blue, Red or Yellow ink, and was constructed from the words Blue, Green, Yellow and Red for the congruent and incongruent trials, and from a series of three, four, five or six asterisks for the baseline trials. The series of characters remained on the screen until participant response. A 750 msec fixation was presented between the character series. Participants went through a practice session of eight congruent and four baseline trials. The task consisted of 24 baseline trials, 24 incongruent trials, and 144 congruent trials. The accuracy and reaction time for the correctly identified congruent, incongruent and baseline trials answered correctly were collected.

Posner Cueing: Participants were presented with an auditory (administered with headsets), then visual cueing tasks (adapted from Facoetti et al., 2005) in separate blocks. A cue (white noise or smiley face) was presented to the left or right field, and following a short delay (auditory 60msec, visual 100msec) or a long delay (auditory 210msec, visual 250msec) a stimulus was presented on either the left or right field. The stimulus (40msec) was either a go stimulus (highbeep or green dot) or a nogo stimulus (low buzz or red dot). For the go stimulus participants were instructed to press the left or right arrows keys to indicate the stimulus position(as opposed to a single space bar response used in the Facoetti et al. study), for the no go stimulus they were instructed to not respond. The resulting design was a 2 (Short or Long) X 2 (Go or NoGo) X 2 (Congruent or Incongruent). The 96 trials, per modality, were evenly distributed across the conditions and were randomly presented. The trial started with a 500msec fixation in the center of the screen, then a cue was presented for 40msec, then a delay (60-250msec), followed by the stimulus (40msec) and the 1500msec response period. The accuracy and reaction time for the correctly identified congruent and incongruent trials were collected.

Gray Oral Reading Test 4 (GORT4): Stories 11, 12, 13 and 14 were presented to the participant to read aloud (Wiederholt & Bryant, 2001). Auditory coding of pronunciation errors was conducted by two raters.

Word Identity (WI) and Word Attack (WA): Word Identity, consisting of 36 words, and Word Attack, consisting of 29 nonwords from the Woodcock Johnson III were administered (Woodcock, McGren & Mather, 2001). Auditory coding of pronunciation errors was conducted by two raters.

Both groups underwent the same assessment battery, at pre-training and post-training (Table 13). AB task versions were used when available. Two versions of the task administration order were created, and used alternately between assessment sessions (Appendix A). The assessment battery was composed of behavioral cognitive assessments and cognitive tasks administered in the MRI scanner. One of the vsWM training tasks, ShapeBuilder, was administered as a behavioral assessment, and the other vsWM training task (Menmosyne/Block-span) was modified and administered in the MRI scanner.

Table 13. Pre-training and post-training tasks administered behaviorally and in the MRI scanner and the relationship between the task and the vsWM training.

Type of Transfer	Pre/Post MRI Tasks	Type of Transfer	Pre/Post Behavioral Tasks
Training Specific	SimonSays (Block-span)	Training Specific	Shape Builder
Process Specific	Guess Dot	Process Specific	SymSpan
	Colorful Dots		OSpan
			Letter-number-sequencing
			Mental Rotation
			Posner Cueing
			Stroop
			Mental Math
			Modular Math
			Ravens
		Process Non-Specific	Grey Oral Reading Test
			Word ID & Word Attack
			Verbal Fluency
			Picture, letter, digit naming

Simon Says:

SimonSays is a modified version of Block-span, as is Memnosyne. While in the MRI, participants were presented with a sequence of blue or black dots, from two to seven in length, presented on a 4 X 4 grid. When blue dots were presented, participants were instructed to reproduce the locations of the dots in serial order (memory load conditions). When black dots were presented, participants were instructed to press randomly on the top row (control position). Trials began with a jittered presentation (1000, 2000, 3000 or 4000msec) of a center fixation, followed by red grid (250msec). Dots were sequentially presented (800msec), with a red grid (200msec) between dot presentations. A black grid (200msec) was presented after the final dot in the sequence, ending the sequence. A green grid was then presented (6,000 or 12,000 msec), and participants indicated their responses using a MRI compatible joystick (Figure 6). Each MRI run consisted of twenty-four dot sequences, three in each of the eight conditions (six load conditions: 2, 3, 4, 5, 6, 7 blue dots & two control conditions: 4 & 7 black dots). Four runs of 7.3 minutes were administered per MRI session.

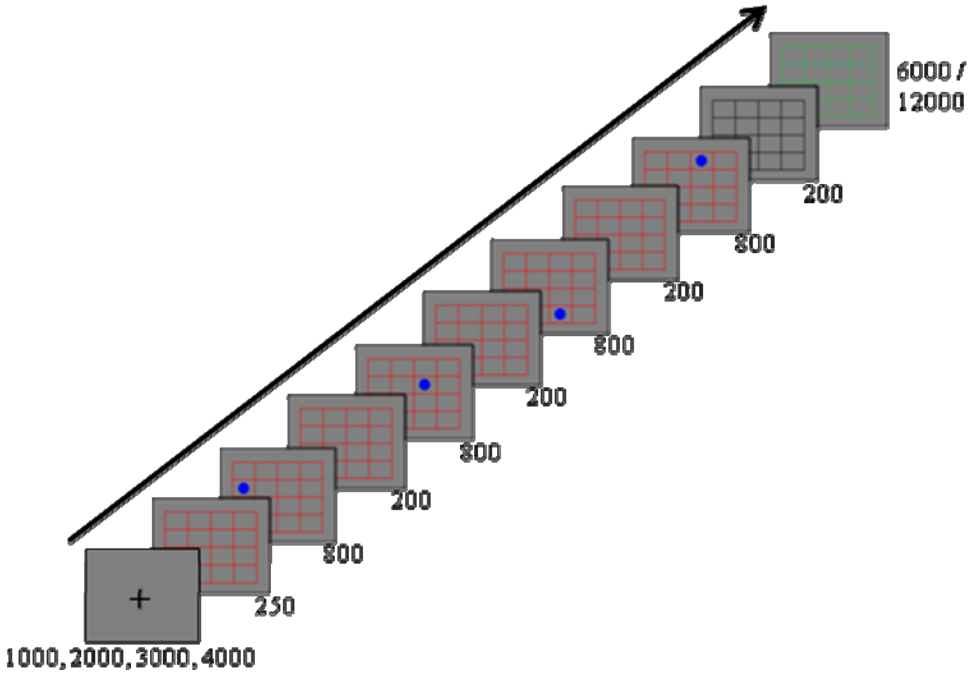


Figure 6. SimonSays memory load trial for 4 dots; duration is presented next to the image (msec). The memory loads trials varied from 2-7 blue dots. In the control trial the dots would be black but otherwise the presentation is identical.

Training Tasks

The online training was administered though a web-site belonging to Prof. Dougherty, http://www.thehygeneproject.org/damlabbeta/index.php/increaseintellect.

Each participant received a login and password, was instructed to train for a maximum of an hour a day, and was allowed three weeks in which to complete the ten hours of online training. For both groups the training alternated between the two training tasks, so that during a fifteen minute session participants engaged in one task for eight minutes (Memnosyne or Sentencical) and the other for seven minutes

(ShapeBuilder or NumberPiles). Before and after each session, participants could view a progress chart for that training task, which illustrated that tasks' scores by session, showed a star on their overall high score, and a line for their average score (Figure 7). In addition, each training task listed the score on the screen at all times to enhance performance. Adherence to the training was monitored remotely by the researcher, who sent reminder emails if the participant had not trained for 3 days.

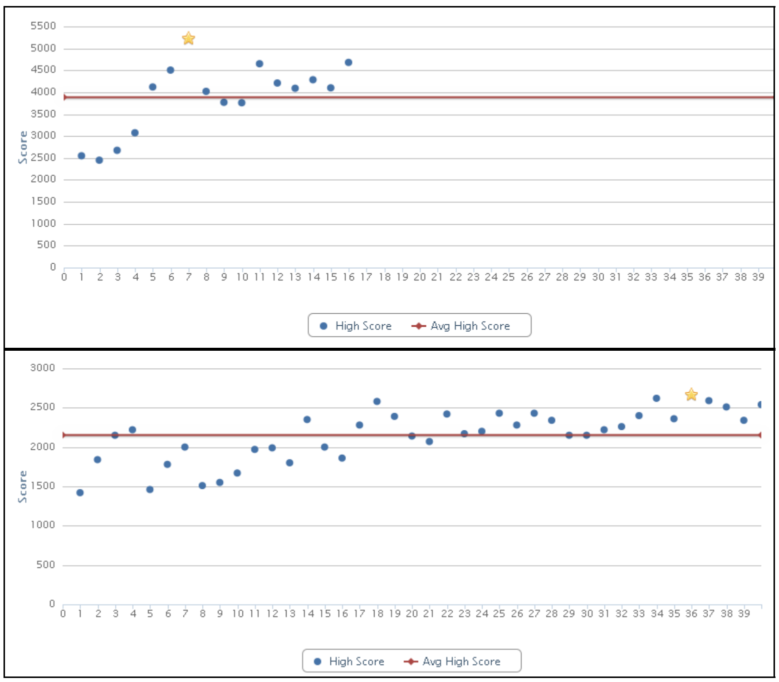


Figure 7. Progression charts of the online training, showing how the session score is indicated, how the highest score is indicated and how the average score is indicated..

<u>vsWM Training Tasks and Assessments</u>

Memnosyne. Memnosyne is the vsWM training task adapted based on the Block-span assessment of WMC (Atkins et al., 2009). Participants were presented with a series of yellow blocks highlighted on a 4 X 4 blue block grid and asked reproduce the location of the yellow blocks in serial order (Figure 8, panel a). Memnosyne adapted in degree of difficulty to the performance level of each participant. Both the number of blocks and speed of presentation increased with performance. Points were awarded for correct block locations in the correct serial position. Points increased for sequential correct identification, and a difficulty bar indicated the level of difficulty (Harbison, Dougherty & Atkins, patent IS-2009-052).

Shape builder. ShapeBuilder-training requests participants to remember and reproduce the serial order of colored (red, blue, yellow, or green) shapes (diamond, triangle, square, or circle) presented sequentially on a 4 x 4 grid (Figure 8, panel b). The ShapeBuilder training adapted to the performance level of each participant by increased or decreased difficulty. The number of shapes and speed of shape presentation adapted with performance. Points were awarded for correct locations, shapes, and colors in the correct serial position. Partial credit was awarded for correct location and color but incorrect shape, or correct location and shape but incorrect color. Points increased for sequential items correctly identified (location, color and shape) (Dougherty, Atkins & Dowling, patent IS-2009-053).

A)

B)

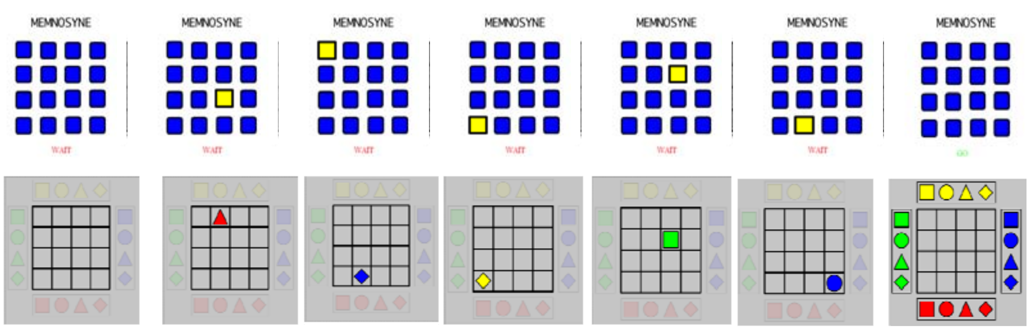


Figure 8. Memnosyne and ShapeBuilder training tasks. Participants are presented with a series of block or shapes and have to reproduce them in serial order.

Placebo Control Training Tasks

Sentencical task. The participant was asked to read sentences that were presented on the screen one character at a time in rapid succession. Once the entire sentence was presented and the participants had indicated that they had read the sentence by clicking the continue button, they were then presented with a yes/no question regarding the presented sentence. An example of a sentence would be "The graduating student promised to bring in cake.", and the question would be "Did the student promise to bring in cake?" The sentences and subsequent questions were randomly selected from a bank of 2110 sentences and questions[4] and included simple and complex comprehension questions, general knowledge questions, and trivia questions. Participants received points for correctly answering the yes/ no question regarding the sentences. Points were awarded for correctly answering the yes/no question. Points increased for consecutive correct answers (Figure 9).

Number Piles. The NumberPiles training requested participants to sum two digits to a stated target number. The task started with the bottom two rows of blocks with digits presented on the center of the screen and a target number presented on the right side of the screen (Figure 9). Participants were requested to click and highlight two digit blocks that sum to reach the target number. If correctly summed and highlighted, the block will explode and disappear from the screen and the participants will be awarded points and assigned a new target number. If the highlighted blocks do not sum to the target number, their sum is presented briefly in both boxes. During the entire task additional digit blocks slowly fall from the top of the screen. Points were awarded for every target number reached (Dougherty & Atkins, patent IS-2009-055).

---

[4] Sentencical sentences and questions are available upon request to S.M. Atkins.

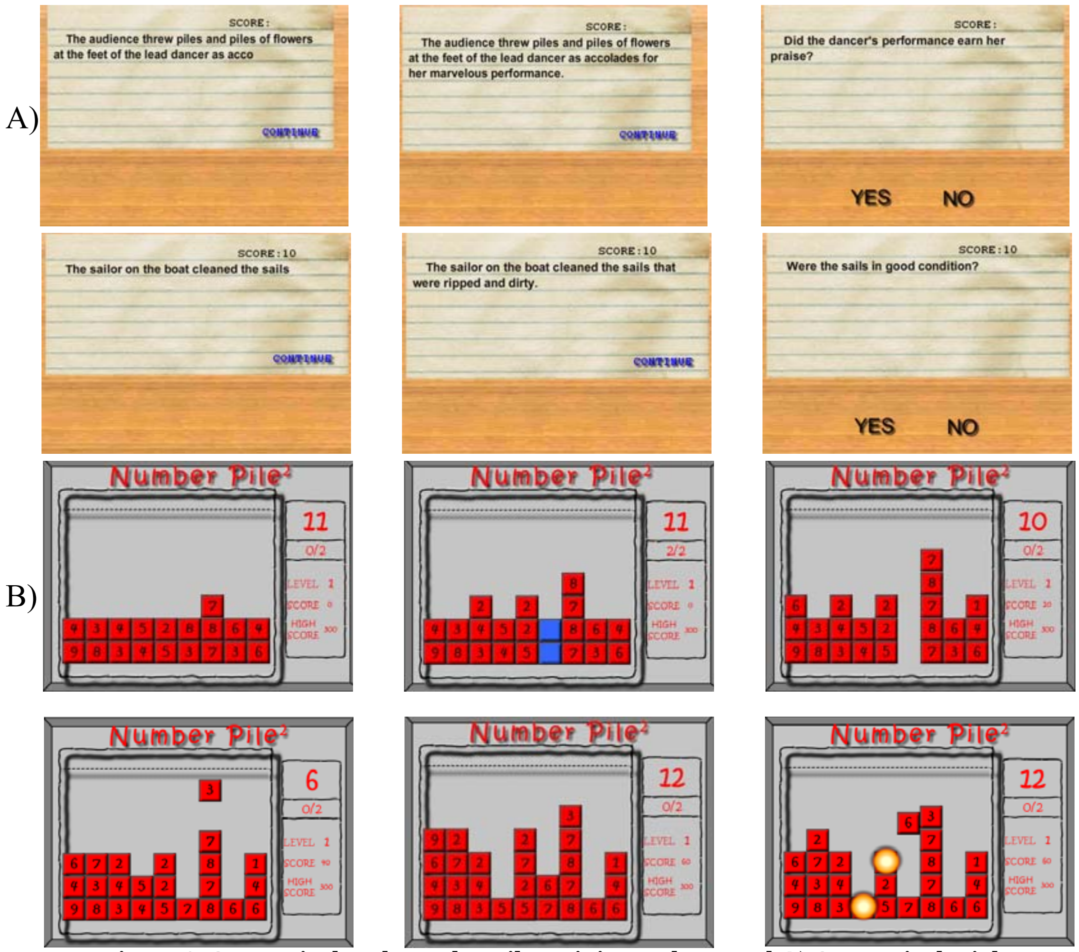


Figure 9. Sentencical and NumberPile training tasks. Panel A) Sentencical trials. Panel B) Number piles trials.

Design and Procedure

Participants came to the MRI lab four times, twice for the pre-training assessment and twice for the post-training assessment. One of the visits for each assessment time was behavioral and one was conducted in the MRI scanner. The post-training assessments were conducted one to seven days from the completion of the online training. The behavioral assessments were administered using E-prime (1.2 & 2.0 professional), DirectRT, and Prof. Dougherty's web site for online assessments, http://www.thehygeneproject.org/damlabbeta/index.php/increaseintellect. E-prime

(2.0 professional) was used to present the stimuli for the cognitive tasks administered in the MRI scanner.

MRI Data Acquisition:

Images were acquired with a 3T Siemens scanner, using a standard 12-channel head coil. Head movement was minimized with the use of cushions. Visual stimuli were projected onto a screen via a mirror attached to the outside of the head coil. Participants' responses were recorded using an optical joystick positioned in the participants' right hand. The BOLD functional images were acquired using the echo planar imaging (EPI) method. The following parameters were used for scanning: time of echo (TE)= 30ms, flip angle= 90^o, field of view (FOV)= 205x205mm, slice thickness = 3.2mm, gap=0.8mm, voxel size=4x4x4mm, number of slices=33 (whole brain: bottom to top); time of repetition (TR)= 2000ms. Four runs of 187 repetitions each were administered. In addition, structural T1 weighted 3D (MPRAGE) images were acquired (TR= 1900ms, TE= 2.52ms, flip angle= 9^o, voxel size= 1x1x1mm, 176 axial slices) using an identical orientation as the functional images.

MR image analysis:

Data analysis was performed using Statistical Parametric Mapping (SPM8, http://www.fil.ion.ucl.ac.uk/spm). The functional images were time series corrected, spatially realigned to the mean, corrected for head movements, co-registered with the anatomical image, and normalized to the standard T1 Montreal Neurological Institute (MNI) template volume. Data was then smoothed with an 8 mm isotropic Gaussian kernel. A high pass filter with a cutoff period of 128 seconds was applied. The preprocessing of the data was conducted on the four pre-training runs and the four post-training run simultaneously.

Behavioral Results

Training Improvements

The training performance in both groups was normalized to the reflect improvements from the first session of training in that task ($Z_i=(X_i-X_1)/SD_1$). The length of the training was used to compare the two groups, where each session of Sentencical and Memnosyne was eight minutes long, and each session of ShapeBuilder and NumberPiles was seven minutes long. The difference in performance improvements is then computed by subtracting the average performance of the beginning sessions (the average of session two and three) from the average performance of the end sessions (the average of session thirty-nine and forty). A 2 X 2 X 2 mixed repeated measure design for assessment time (Pre, or Post), task (train8min, or train7min) and group (vsWM training or Placebo Control) showed a significant main effects for time of assessments ($F(1,33)= 64.76, p<0.001$ $\eta_p^2=.66$) and for task ($F(1,33)= 84.50, p<0.001$ $\eta_p^2=.72$), and a significant task by group interaction ($F(1,33)= 55.09, p<0.001, \eta_p^2=.63$) and a significant time by group interaction ($F(1,33)= 63.51, p<0.001, \eta_p^2=.66$). The three-way interaction was not significant, nor was it expected to be. The improvements on the training tasks were significantly different between the training groups. Improvements on the Memnosyne training task were significantly greater than on the Sentencical training task ($t(33) =5.24$, $p<0.01$). Likewise, improvements on Shape Builder training were significantly greater than improvements on the Number Piles task (t(33) =8.41, p<0.001). Figure 10 shows the normalized performance improvement scores for each session.

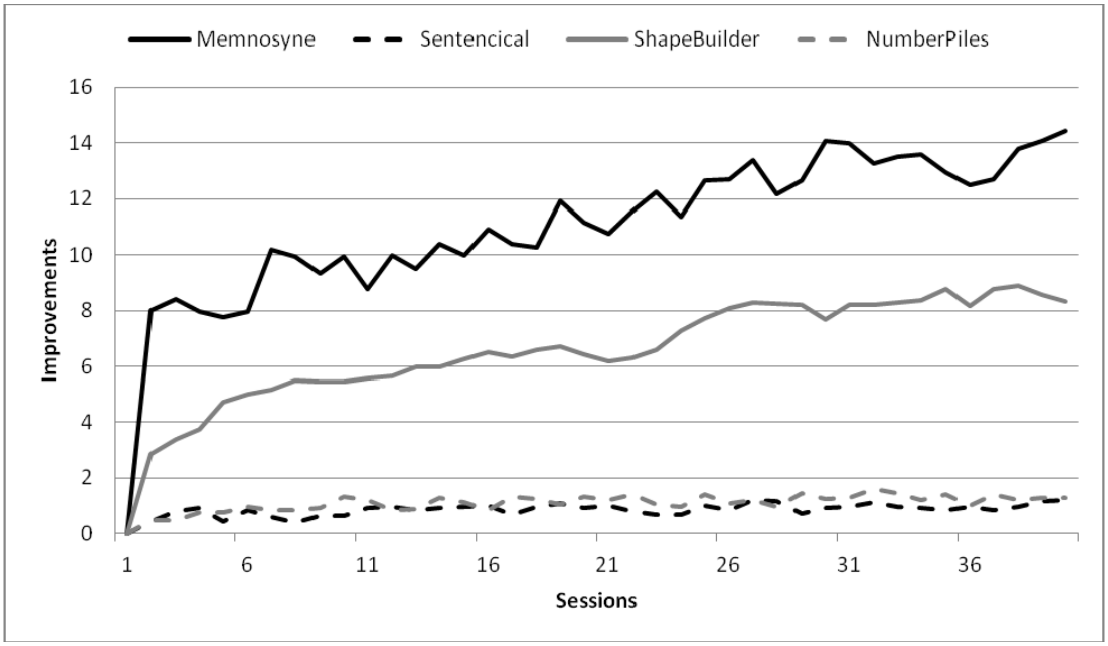


Figure 10. The mean group improvements by training tasks for each session. The raw scores for each session were normalized to the first session, to reflect improvements.

Individual Differences

Although the Memnosyne and ShapeBuilder improvements are strongly related (r(17)=0.46, p<0.05), participants in the vsWM training group did not improve to the same degree on both tasks (Figure 11).

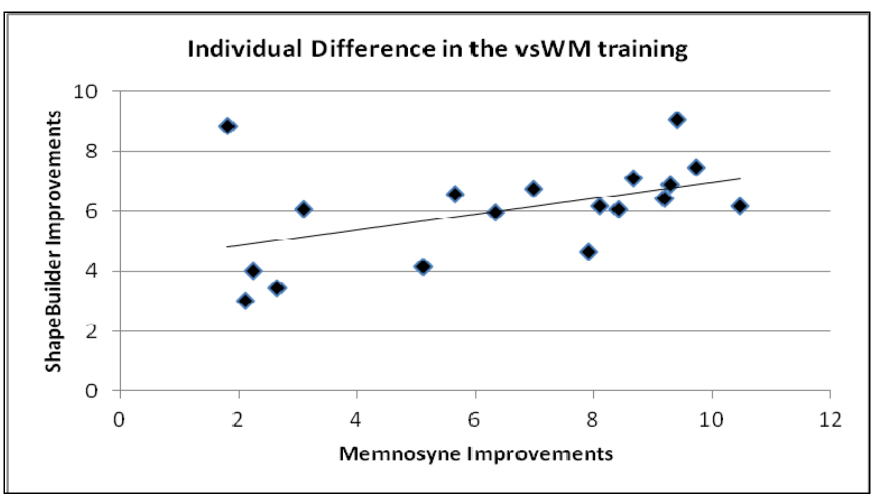


Figure 11. Individual differences among the vsWM training participants on the training task improvements.

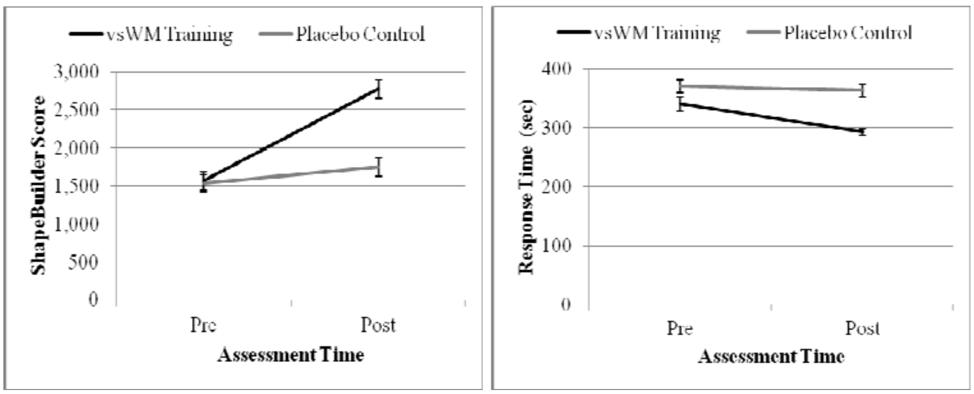
Training Specific Transfer

The training improvement curves show that participants in the vsWM training group improve on the trained tasks. However, do these training effects transfer? The first step in addressing this question is to see whether the improvements transfer to non-adaptive versions of the task. I refer to this as training-specific transfer.

Training specific transfer was evident both on the assessment version of ShapeBuilder and on SimonSays (modified Block-span/Memnosyne).

A mixed 2 X 2 repeated measures ANOVA for assessment time (Pre, Post) by training group (vsWM Training, Placebo Control) was employed for both training specific tasks. ShapeBuilder showed a significant main effect for time of assessments ($F(1,33)= 81.15, p<0.001$ $\eta_p^2=.71$) and a significant time by training condition interaction ($F(1,33)= 39.37, p<0.001, \eta_p^2=.54$). Indeed, as shown in Figure 12 (panel A), there is no difference between the groups at pre-training, but there is a difference at post-training with participants in the vsWM group showing greater performance then the placebo control group ($t(33)=6.7$, $p<0.01$). Furthermore, the vsWM group showed an increase in performance ($t(17) =10.23$, $p<0.01$) but not the placebo control group ($t(16) =2.08$, p=n.s.).

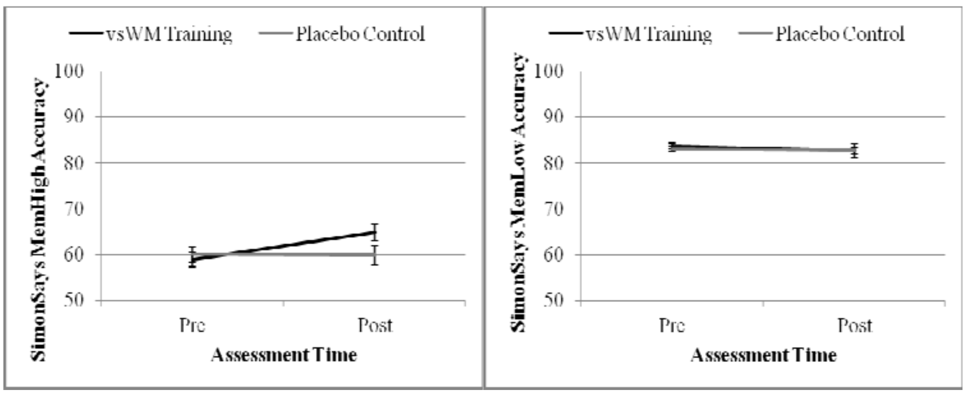
Similar results are found when examining the time it took participants to complete the ShapeBuilder assessment (Figure 12 panel B). The mixed 2 X 2 repeated measures ANOVA with the response time for ShapeBuilder showed a significant main effect for time of assessments ($F(1,33)= 12.06, p<0.01$ $\eta_p^2=.27$), and a significant time by group interaction ($F(1,33)= 6.57, p<0.05, \eta_p^2=.17$). Again no differences on the response time for the ShapeBuilder task exist at pre-training, however post-training there is a significant difference between the groups ($t(33)=6.06$, $p<0.001$). The vsWM training group significantly decreased in the response time for ShapeBuilder from pre-training to post-training, ($t(17)=3.98$, $p<0.01$), yet the placebo control group did not differ in response time for ShapeBuilder ($t(16)=0.71$, p=n.s.).



A. ShapeBuilder Score **B. Response time for ShapeBuilder**

Figure 12. Pre and post performance on (A) the ShapeBuilder Assessment and (B) the response time for ShapeBuilder, for both the vsWM training and the placebo control groups. Error bars reflect standard errors of mean.

The transfer of training-induced improvements to the accuracy for the SimonSays task is also training specific, as SimonSays and Memnosyne are both modified versions of Block-span. SimonSays accuracy was examined separately for the high memory load (MemHigh) and the low memory load trials (MemLow). Each participant's score was given by percent correct. The percentage data underwent arcsine transformation, and entered into a 2 X 2 X 2 mixed repeated measure design for assessment time (Pre, or Post), trial type (MemHigh, or MemLow) and training group (vsWM Training, Placebo Control). The results show a significant main effect for trial type ($F(1,33)= 760.08, p<0.001\ \eta_p^2=.96$), and a significant time by task interaction ($F(1,33)= 13.16, p<0.01, \eta_p^2=.29$), and a significant time by task by group interaction ($F(1,33)=11.92, p<0.01, \eta_p^2=.27$). The pre-training MemHigh trials showed no difference between groups, yet the vsWM training group improved in performance on the MemHigh trials from pre-training to post-training, ($t(17)=2.90$, $p<0.05$) and the placebo control group did not ($t(16)=0.2$, p=n.s.) (Figure 13 panel A). There were no group, or time of assessment differences on the SimonSays MemLow trials (Figure 13 panel B).



A. MemHigh SimonSays Trials. B. MemoLow SimonSays Trials.

Figure 13. Pre and post accuracy for (A) the MemHigh trials and (B) the MemLow trials in the SimonSays task. Error bars reflect standard errors of mean.

Table 14. Reliability coefficients for the behavioral assessments.

Assessment	Cronbach α	Data Source
ShapeBuilder	0.72	Current Study
SymmetrySpan	0.86	Kane et al., 2004
OperationSpan	0.78	Unsworth, Heitz, Schrock & Engle 2005
LetterNumberSequencing	0.67	Current Study
Mental Rotation	0.94	Current Study
PosnerCueing Aud	0.93	Current Study
PosnerCueing Vis	0.98	Current Study
Stroop	0.75	Current Study
Mental Math	0.86	Current Study
Mod Math	0.89	Current Study
Ravens	0.84	Current Study
Verbal Fluency	0.84	Current Study
GORT	0.85-0.99	Wiederholt, J.L. & Bryant, B.R. (2001).
Word ID	0.94	Woodcock, R. W., McGrew, K. S., & Mather, N. (2001).
Word Attack	0.87	Woodcock, R. W., McGrew, K. S., & Mather, N. (2001).

Table 15. Descriptive statistics for the assessments in Experiment 2.

	vsWM training					Placebo Control				
		Pre		Post			Pre		Post	
Assessment	N	M	SD	M	SD	N	M	SD	M	SD
Shape Builder										
Score	18	1,566	491	2,780	476	17	1,535	472	1,753	510
RT	18	339,889	47,297	293,000	23,988	17	370,353	43,564	363,294	42,634
SimonSays										
HML Accuracy	18	58.84	6.95	64.83	7.94	17	60.08	7.05	59.84	8.68
SymSpan										
Score	18	17.67	8.50	22.56	8.30	17	17.00	11.76	24.47	9.88
SymTime	18	6,171	3,758	3,783	2,499	17	5,568	2,879	3,959	1,820
SymQuestRT	18	713	96	668	114	17	792	133	778	202
SymProbRT	18	1,621	587	1,209	598	17	1,732	846	1,340	507
OSpan										
Score	18	53.78	16.28	47.44	16.50	17	45.41	19.11	50.53	17.35
MathTime	18	5,597	2,259	4,975	1,492	17	5,093	2,292	4,754	2,952
OperMathRT	18	994	154	907	132	17	1,101	175	1,067	202
OperProbRT	18	2,061	631	1,817	318	17	2,077	784	1,701	389
Letter-number-sequencing										
Score	18	532	244	603	230	17	581	240	729	166
RT	18	338,444	58,535	299,000	43,804	17	360,529	62,557	349,765	111,032
Mental Rotation										
Accuracy	18	66.49	17.59	73.83	10.59	15	68.38	17.78	70.57	20.47
RT	18	1,964	270	1,751	331	15	1,939	330	1,808	166
Posner Cueing										
Visual Short	16	80.21	38.24	88.54	28.85	17	78.92	39.76	63.73	45.73
Visual Long	16	81.77	36.92	90.10	24.19	17	80.39	35.59	62.25	46.60
Auditory Short	16	67.71	24.70	85.94	15.43	17	77.45	17.12	85.78	16.34
Auditory Long	16	70.31	21.07	89.06	12.06	17	77.45	16.61	83.33	16.67
Stroop										
IncongRT	17	1,158	190	1,046	157	17	1,073	278	990	283
CongRT	17	784	103	733	134	17	704	144	675	148
BaselineRT	17	854	123	806	143	17	791	164	740	161
Mental Math										
Accuracy	18	61.76	7.85	65.46	5.65	15	60.11	8.85	64.22	7.81
RT	18	1,740	244	1,639	226	15	1,866	501	1,895	291
Modular Math										
RT High	18	5,268	1,401	4,647	1,381	16	6,009	2,074	4,923	1,908
RT Low	18	2,078	429	1,723	373	16	2,140	256	1,711	343
Score High	18	16.72	2.61	17.25	1.39	16	17.06	1.68	16.84	1.64
Score Low	18	19.25	1.86	19.56	0.38	16	19.43	0.50	19.16	0.81
Ravens										
Accuracy	18	58.95	5.88	62.04	23.04	17	58.82	23.16	59.48	18.19
GORT										
Accuracy	18	64.44	15.80	65.28	17.86	17	67.65	9.21	72.94	11.60
Errors	18	1.74	2.51	1.46	1.79	17	2.09	1.66	1.88	1.22
RT	18	60,146	13,581	60,176	5,734	17	62,784	7,402	63,117	8,671
Word ID										
Score	18	33.83	2.28	33.67	1.91	17	32.65	3.06	31.94	3.63
RT	18	10,074	4,027	8,052	4,005	17	9,192	1,988	7,549	2,319
Word Attack										
Score	18	27.61	2.70	27.21	2.09	17	26.76	2.70	27.06	1.78
RT	18	8,595	3,520	6,631	3,689	17	7,737	1,589	6,134	1,647
Verbal Fluency										
FAS	18	17.23	4.07	18.56	3.27	17	16.41	5.62	17.63	4.86
Semantic	18	28.22	5.67	29.33	4.92	17	25.96	5.76	27.76	5.39

Process Specific Transfer

Process-specific transfer should occur when the training task and the assessed task have overlapping processing networks, and the degree of overlap should determine the degree of process-specific transfer. The analyses examining process-specific transfer were conducted under the hypothesis that vsWM training would lead to *improvements* and therefore employed one-tailed tests.

For the present study, all cognitive tasks that relate to vsWM are considered process-specific, even tasks whose relationship to the training tasks is ambiguous. While it is conceivable that not all process-specific tasks show training-induced improvements (due to the strength of the overlapping network with the training task), process non-specific tasks should not show any improvements or it would call to question whether participants increase in general due to Hawthorne effects.

Indeed, not all the process-specific tasks showed training-induced improvements in this study. Process-specific improvements were not evident in the Letter-number-sequencing[5] score. Similarly no differences were evident in OSpan. Letter-number-sequencing and OSpan are vWM tasks, whose lack of transfer could be attributed to loading on the verbal as opposed to visual spatial domain.

However, no process-specific transfer was found for SymSpan, another vsWM task. SymSpan's process-specificity was not ambiguous and was expected to show process specific training improvements. Therefore, this finding was unexpected, as previous training studies had showed process-specific transfer to SymSpan following twenty hours of training (Atkins et al., under review). One possible

[5] The letter-number-sequencing task had issues with instruction presentation.

explanation is that the present study only had participants train for 10 hours, as opposed to the prior study that used 20 hours of training.

Similarly, Ravens did not show significant training-induced transfer (Figure 14). This finding is consistent with the factor analysis in Experiment 1, where Ravens loaded on the ability factor and not with the vsWM factor. This finding is also consistent with Chein and Morrison (2010) lack of transfer to Ravens. However, this finding is not consistent with other training studies that do show transfer to Ravens (Jaeggi, Buschkuehl, Jonides, & Perrig, 2008). Looking at the mean accuracies at both times shows that the placebo-control group does not change from pre to post (less than 1%) whereas the vsWM training group increased by over 3%. Needless to say, that slight increase is not an indication of training-induced transfer in this study; however, it does have implications for future studies.

Verbal fluency in a 2 X 2 X 2 mixed repeated measure design for assessment time (Pre or Post), task measure (Letters or Semantic retrieval), and group (vsWM training or Placebo Control) showed a main effect for time of assessment ($F(1,33)=154.67$, $p<0.001$, $\eta_p^2=.82$) and a main effect for the measures ($F(1,33)=9.23$, $p<0.001$, $\eta_p^2=.22$). No between group differences were observed. Research has shown the verbal fluency task to be related to WMC (Rosen & Engle, 1997), but the task is verbal in nature and does not show transfer.

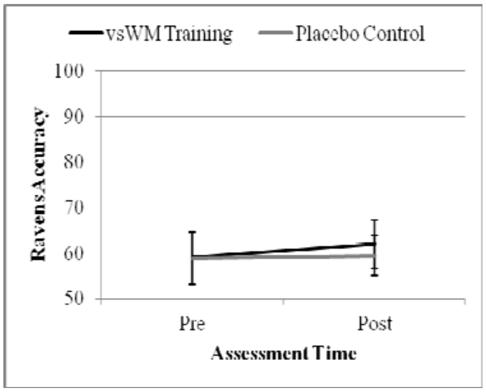


Figure 14. The pre-training and post-training Ravens accuracy. Error bars reflect standard errors of mean.

The incongruent Stroop trials also did not show any process-specific transfer, a 2 X 2 mixed repeated measure design for assessment time (Pre or Post), and group (vsWM training or Placebo Control) showed a main effect for time of assessment ($F(1,32)=8.16$, $p<0.01$, $\eta_p^2=.2$), but no group differences. Stroop is a response inhibition task based on the conflict of reading the word and naming the ink color. Previous studies have been inconsistent regarding the transfer of training-induced improvements to Stroop task. With some training studies showing transfer of improvements to the Stroop task following training (Klingberg, Fossberg, & Westerberg, 2002; Atkins et al., under review), yet other studies show no transfer of improvements on the Stroop task following training (Dahlin, Stigsdotter-Nelly, Larsson, Bäckman & Nyberg, 2008; Thorell, Lindqvist, Bergman, Bohlin, & Klingberg, 2009).

Process-specific transfer was evident in the Posner cueing task. The mixed 2 X 2 X 2 X 2 repeated measure design for assessment time (Pre,or Post), modality (Visual or Auditory), delay (Long or Short) and group (vsWM or PlaceboControl), showed a significant assessment time by group interaction ($F(1,30)=5.28$, $p<0.05$, $\eta_p^2=.15$) (see Figure 15) and a significant assessment time by modality interaction ($F(1,30)=5.18$,

$p<0.05$, $\eta_p^2=.15$). As can been seen when examining the modalities separately (Figure 16), the visual Posner cueing task showed improvements for the vsWM training group between pre-training and post-training, whereas the placebo control group did not improve (Figure 16 panels A & B). The vsWM training group improvement in the post-training minus pre-training difference score was evident in both the long delay ($t(31)=2.27$, $p<0.05$) and short delay trials ($t(31)=1.89$, $p<0.05$). The auditory Posner cueing (Figure 16 panels C & D) showed significant improvements for the post-training minus pre-training difference in the long delay trials ($t(31)=1.90$, $p<0.05$). Whereas the short delay suggests similar improvements the difference score was not significant ($t(31)=1.45$, p=n.s). No other main effects or interactions were significant.

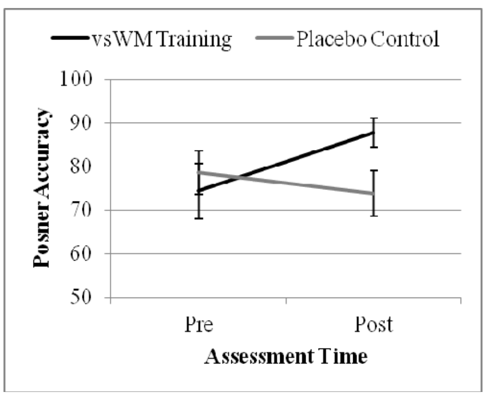
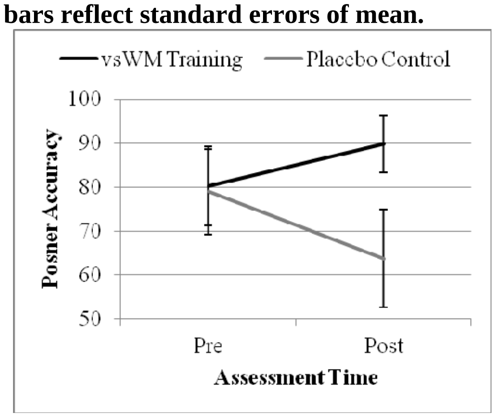
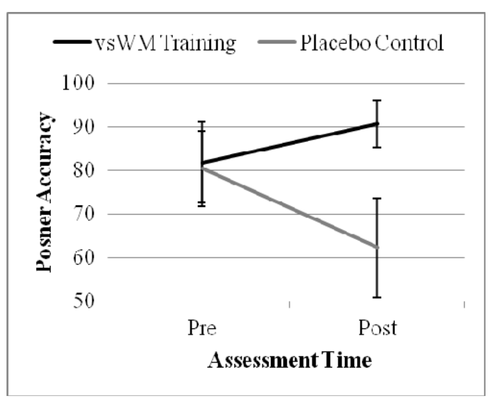


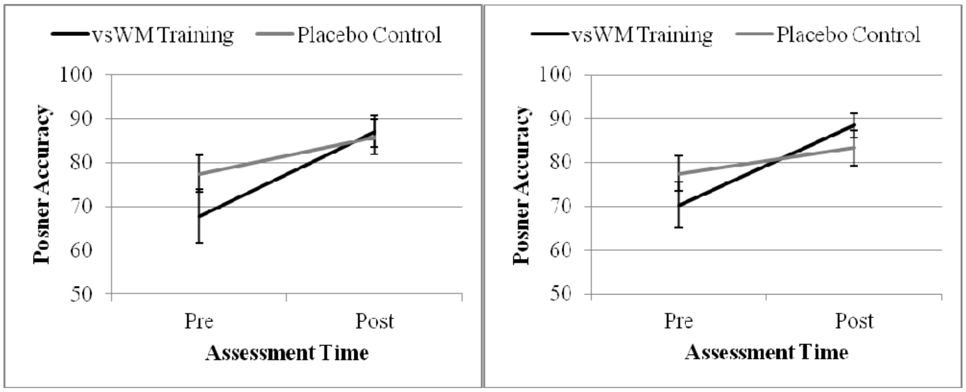
Figure 15. Pre-training and post-training overall accuracy for the Posner cueing task. Overall accuracy was computed by averaging accuracy across delay length for both modalities. Error bars reflect standard errors of mean.



A) Visual Short Delay



B) Visual Long Delay



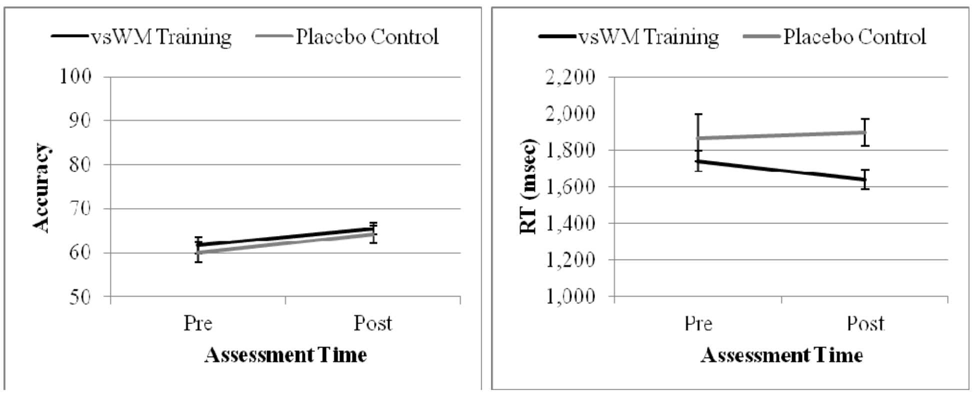
C) Auditory Short Delay **D) Auditory Long Delay**

Figure 16. Pre-training and post-training accuracy scores for the Posner cueing task for Visual- short delay (panel A), Visual-long delay (panel B), Auditory-short delay (panel C), and Auditory-long delay (panel D). Error bars reflect standard errors of mean.

The mathematical tasks are considered process-specific, as research has shown that math calculations requires the involvement of the IPS and PFC (Dehaene, Piazza, Pinel, & Cohen, 2003; Dehaene, Molko, Cohen, & Wilson, 2004; Dehaene, 2009). And the mental math results suggest the existence of process-specific transfer. The mental math levels were combined (see Appendix B for the individual level charts), and accuracy was arcsine transformed, while RT was log transformed[6]. A mixed 2 X 2 X 2 mixed repeated measures design for assessment time (Pre, Post) by measurement type (1-Acc (to align the direction of improvements), RT) by training group (vsWM, Placebo Control) showed a significant main effect for time of assessments ($F(1,33)= 4.76, p<0.05$ $\eta_p^2=.13$) and a main effect for measurement type ($F(1,33)= 108238.14, p<0.05$ $\eta_p^2=1.00$) (Figure 17). The interactions were not significant. Although both training groups improved on their accuracy in the mental math task from pre-training to post-training (vsWM: $t(17)=3.27, p<0.05$; Placebo: $t(16)=3.5$, $p<0.05$), only the vsWM training group improved on their RT from pre-training to post-training (vsWM: $t(17)=2.1$, $p<0.05$; Placebo: $t(16)=0.40$, p=n.s.). The

[6] In all analyses using log transformations the test result did not differ when using raw RT's or log transformed RT's.

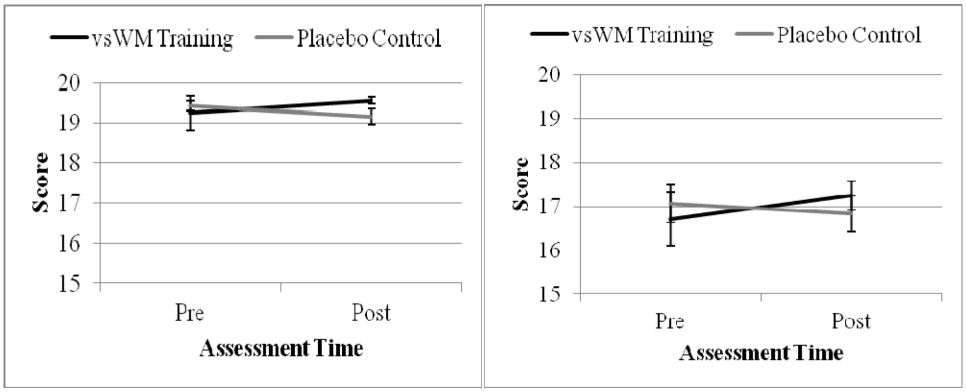
pre-training mental math RT showed no group differences, but post-training showed a significant difference between the training groups (t(33)=2.09, p<0.05), where the vsWM group exhibited faster RT following training (vsWM: M_{pre}=1,740 M_{post}=1,638; Placebo: M_{pre}=1,873 M_{post}=1,835). However, the difference of the post-pre RT was not significant (t(33)=1.28, p=n.s).



A) MentalMath Accuracy **B) MentalMath RT**

Figure 17. Pre training and post training MentalMath accuracy (panel A) and RT (panel B). Error bars reflect standard errors of mean.

The ModMath task results do not show process-specific transfer. The mixed 2 X 2 X 2 design for assessment time (Pre, Post) by trial type (High, Low) by training group (vsWM, Placebo Control) on the arcsine transformed score showed only a main effect for trial type (F(1,33)= 201.96,p<0.001 η_p^2=.86). Interestingly enough, performance on both the high and low difficulty trials improves in the vsWM training group from pre-training to post-training, whereas performance decreases in the placebo control group (Figure 18). This is particularly interesting as the Placebo Control group trained on simple math with the Number Piles training task, whereas the vsWM training group had no mathematical training.



A. ModMath Low Difficulty **B. ModMath High Difficulty**

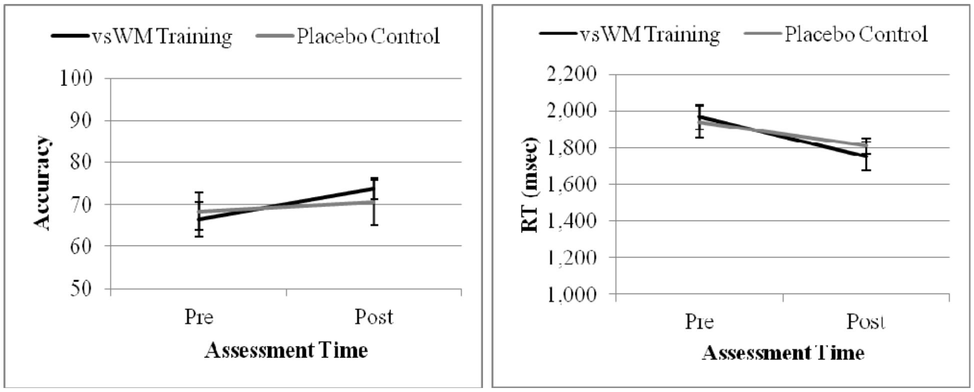
Figure 18. Pre training and post training ModMath scores for the low (panel a) and high difficulty (panel b) trials. Error bars reflect standard errors of mean.

The mental rotation task is unambiguously a process-specific task, as mental rotation has been shown to be related to WM ability, and reliably involves the activation of the parietal cortex, and many studies have also implicated the PFC (meta-analysis: Zacks, 2008). The mental rotation task analysis combined RT[7] and accuracy for all rotated images relative to the images with no-rotation (Rot0) (see Appendix C for individual level descriptives). The RT was Log transformed and the accuracy underwent arcsine transformation. In a 2 X 2 X 2 mixed repeated measure design for assessment time (Pre, Post) by task measure (1-Acc, RT) by training group (vsWM training, Placebo control) on the log transformed RT showed a significant main effect for time of assessment ($F(1,30)= 43561.29, p<0.05$ $\eta_p^2=.99$), and a time by group interaction ($F(1,30)= 5.35, p<0.05$ $\eta_p^2=.15$), and a marginally significant three-way interaction for measure by assessment time by group interaction ($F(1,30)= 3.39, p=0.075$ $\eta_p^2=.10$). The vsWM training group improved significantly from pre-training to post-training in both accuracy ($t(17)=2.12$, $p<0.05$) and RT ($t(17)=1.98$, $p<0.05$), whereas the placebo control group did not show any pre-training to post-

---

[7] Two participants were excluded from this analysis, as they did not respond correctly to any items in the level and therefore did not provide RT data.

training improvements on either measure (t(14)=1.1, p=n.s.; RT: t(14)=1.6, p=n.s.) (Figure 19).



A. Mental Rotation Accuracy **B. Mental Rotation RT**

Figure 19. Pre training and post training MR accuracy (panel a) and RT (panel b). Error bars reflect standard errors of mean.

Process Non-Specific Tasks

Process non-specific tasks are tasks that are not expected to show any transfer of the training-induced improvements. They are included to demonstrate a lack of transfer, so that the transfer shown in the process specific tasks will not be attributed to general motivation on the part of the participant. That is not to say that these tasks will never show training-induced improvements, but that these tasks rely more on acquired knowledge and could benefit from the WM training in the long term, through ease of knowledge acquisition that potentially follows WM training.

The Word ID and Word Attack tasks show significant main effects for time of assessment with the RT measure, where participants in both groups are getting faster at the task ($F(1,33)=47.06$, $p<0.001$, $\eta_p^2=.59$), and a main effect for the differences between Word attack and Word ID for both RT ($F(1,33)=110.1$, $p<0.01$, $\eta_p^2=.77$) and score ($F(1,33)=274.2$, $p<0.01$, $\eta_p^2=.89$). No between group differences were observed (see Table 15 for descriptive statistics).

Similarly, in a 2 X3 X 2 mixed repeated measure design for assessment time (Pre or Post), task measure (Reading Errors, Accuracy or Reading Time), and group (vsWM training or Placebo Control), GORT showed a main effect for time of assessment ($F(1,33)=1938.16$, $p<0.001$, $\eta_p^2=.98$), a main effect for the measures ($F(2,32)=1211.96$, $p<0.001$, $\eta_p^2=.99$), and a main interaction of measure by time ($F(2,32)=1213.17$, $p<0.001$, $\eta_p^2=.99$). No between group differences were observed (see Table 15 for descriptive statistic).

The predictions for the process non-specific tasks were for a lack of transfer, and indeed the null hypothesis cannot be rejected in the above analyses. However, the hypothesis testing employed above cannot provide support for the null hypothesis, therefore Bayesian statistics were employed to examine whether there is support for the null hypothesis for the process non-specific tasks. The Bayes factors (BF) estimate the probability of the null versus the probability of the alternative hypothesis. Assuming equal priors, the BF provides an estimate of the posterior probability of the null hypothesis given the data relative to the probability of the alternative hypothesis given the data (Olejnik & Algina, 2003; Wagenmakers, 2007; Rouder, Speckman, Sun, Morey & Iverson, 2009; Masson, 2011). The current study evaluated the alternative hypothesis (interaction effect from the repeated measure design of each task) compared to the null hypothesis (Masson, 2011), and produced Bayes factors in favor of the null for all process non-specific tasks (see Table 16), providing supporting evidence for the lack of transfer effects to these tasks.

Table 16. Bayes Factors for the process non-specific assessments. Bayes factors above 1 indicate odds in favor of the null hypothesis, whereas bayes factors under 1 indicate odds in favor of the alterntaive hypothesis.

Assessment	Bayes Factors
Word ID	5.66
Word Attack	5.81
GORT	
Accuracy	5.27
Reading Time	5.91
Reading Errors	5.91

Imaging Results

The analysis of the fMRI data was conducted on the SimonSays (modified Block-span/Memnosyne) task. Behavioral improvements on the SimonSays task were shown earlier as part of the training-specific transfer of improvements. The changes in neural activity while partaking in the SimonSays tasks reveal the neural network engaged during the training. The fMRI analysis of the images rendered while partaking in SimonSays was first conducted as a whole brain analysis, examining the activation of the entire brain, and then as a region of interest analysis (ROI), examining the IPL and PFC regions that are part of the targeted network.

Whole Brain Analysis

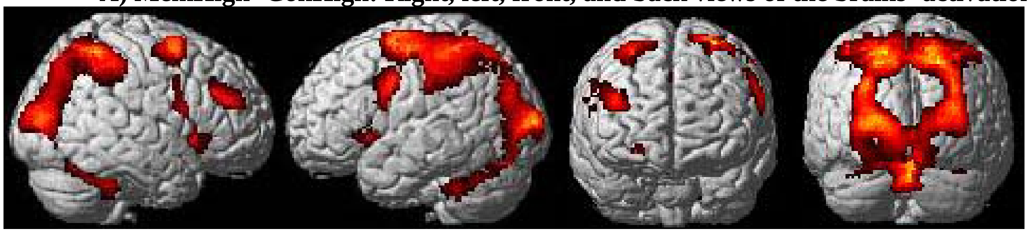
Whole brain analysis allows us to examine the task specific pattern of brain activation. This allows for confirmation of the task activation, and provides a referencing for the ROI analysis. The fMRI analysis was conducted on the sustained response for the low memory load trials (MemLow), containing 2, 3 or 4 blue dots, in contrast to the sustained response for the low control trials (ConLow) containing 4 black dots, and on the sustained response for the high memory load trials (MemHigh), containing 5, 6, or 7 blue dots, in contrast with the sustained response for the control high trials (ConHigh), containing 7 black dots. The MemLow and ConLow contrasts did not yield any significant findings, on the pre-training or post-training assessments.

Therefore the focuses of this presentation are the MemHigh and ConHigh contrasts. The pre-training fMRI data showed no group differences in the brain activation during MemHigh>ConHigh contrast or the ConHigh>MemHigh contrast. Table 17 shows brain areas that survived FDR $p<0.05$ for the MemHigh>ConHigh and the ConHigh>MemHigh contrasts for both groups. Table 17. Pre-training brain activation for the MemHigh>ConHigh and the ConHigh>MemHigh contrasts, for both groups, $p<0.001$uncorrected, k=10. Volume information is presented for the first listed cluster and identified by letter for subsequent areas belonging to that cluster.

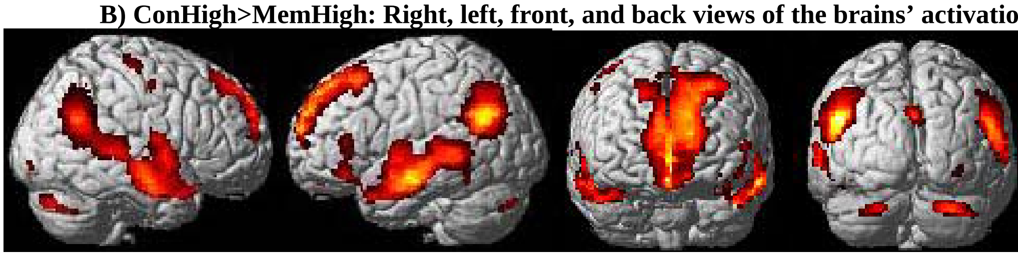
Region	BA	x	Y	z	vol	t
MemHigh > ConHigh						
Parietal						
R. Superior parietal/ Precuneus	7	18	-64	58	28,018a	11.77
L. Superior parietal/ Precuneus	7	-16	-64	50	a	10.77
R. Inferior parietal	40	38	-48	54	a	9.24
L. Inferior parietal	40	-34	-42	42	a	9.63
Frontal						
L. Middle frontal gyrus	6	-28	-4	58	a	10.58
R. Middle frontal gyrus	6	24	-2	48	3,871b	8.99
R. Middle frontal gyrus	11	24	48	-10	47	5.92
R. Middle frontal gyrus	10	42	42	24	481	5.49
R. Inferior frontal gyrus	47	34	26	-4	b	5.15
Sub-cortical						
R. Putamen		24	4	16	b	5.9
R. Putamen		30	18	6	b	5.59
R. Thalamus		24	-28	12	b	5.89
R. Thalamus		20	-12	18	b	5.67
ConHigh>MemHigh						
Parietal						
L. Superior/ Inferior parietal	7	-36	-74	46	1,327c	7.03
L. Supramarginal	40	-64	-46	26	c	3.76
L. Precuneas/ Posterior Cingulate	29	-4	-52	10	17,409d	6.05
R. Supramarginal gyrus	40	58	-64	26	d	8.75
Frontal						
L. Superior Medial	10	-2	54	0	d	8.1
L. Superior Medial	9	-12	56	30	d	7.49
R. Superior Medial	10	10	60	14	d	7.51
Cingulate						
R. Anterior Cingulate	32	6	28	-10	d	7.34
L. Anterior Cingulate	32	-4	38	-10	d	7.91
R. Posterior Cingulate	31	2	-48	32	1,778	6.59
L.Middle Cingulate	24	-2	-8	38	268	5.42
Temporal						
R. Middle Temporal	21	54	-16	-12	d	7.26
L. Middle Temporal	21	-58	-12	-16	c	8.16
L. Temporal Lobe/Angular Gyrus	39	-48	-74	30	c	10.56

The MemHigh>ConHigh contrast shows brain areas more active during the memory trials (Figure 20, panel a), whereas the ConHigh>MemHigh (Figure 20, panel b) shows activity in areas commonly associated with the activation of the brain’s default mode (Raichele, et al., 2001; Buckner, Andrews-Hanna, & Schacter, 2009) that are more active during the control trials. The default mode is a network of brain areas active during rest, or non demanding tasks. Therefore, an increase in default network activity during the MemHigh trials at post-training would indicate that the MemHigh trials are less demanding.

A) MemHigh>ConHigh: Right, left, front, and back views of the brains' activation.



B) ConHigh>MemHigh: Right, left, front, and back views of the brains' activation.



C) MemHigh>ConHigh (red) and ConHigh>MemHigh (blue).

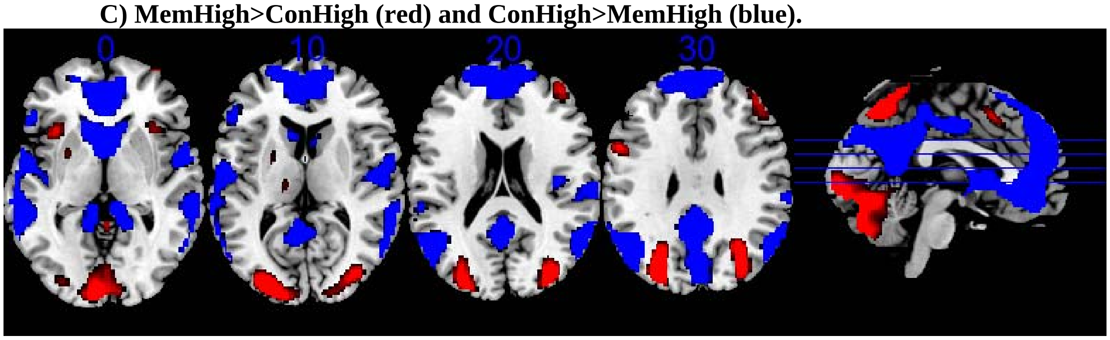


Figure 20. Pre-training brain activity combined for both groups, p<0.001uncorrected, k=10.

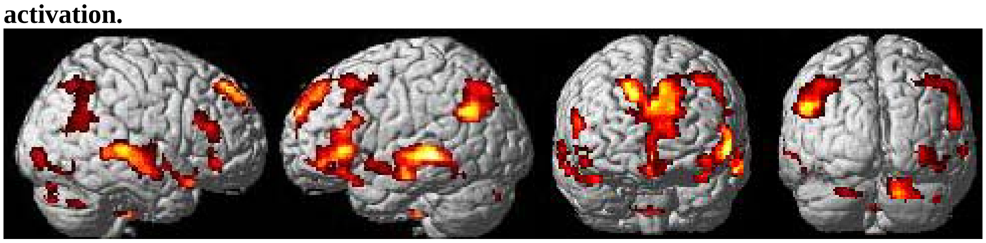
The neural effects of the training are evident in the post-training minus pre-training data, where both the MemHigh>ConHigh and the ConHigh>MemHigh show group differences in the brain activation (Table11).

Table 18. Post-training minus pre-training brain activation, for vsWM training group> Placebo-control and the Placebo Control>vsWM training group, p<0.005uncorrected, k=10.

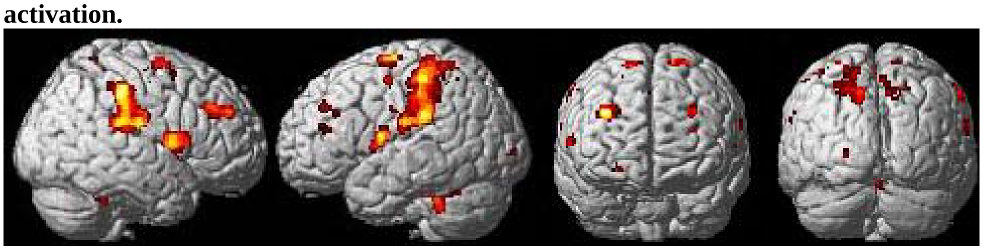
	Region	BA	x	y	z	vol	t	Interaction
	vsWM Training> Placebo Control							
Parietal	R. Inferior parietal/ Supramarginal	40	62	-30	24	60	3.35	
	L. Inferior parietal	40	-30	-30	32	634a	3.73	3-way
	R. Postcentral		46	-16	24	178b	3.95	2-way
Frontal/Insula	L. Frontal		-24	6	30	a	4	3-way
	L. Frontal		-30	-14	32	a	4.6	3-way
	R. Frontal	44	62	6	8	67	4.01	
	R. Postcentral		32	-20	28	b	3.36	
	R. Insula	13	38	8	-12	11	3.05	
	R. Insula	22	46	10	-4	27	3.07	2-way
	L. Insula	13	-36	0	-8	112c	3.97	2-way
Cingulate	R. Middle Cingulate		12	-2	34	46	3.87	2-way
	L. Middle Cingulate		-16	-10	36	a	3.37	3-way
	R. Middle Cingulate	31	16	-28	36	93	4.67	3-way
	L. Middle Cingulate		-10	2	34	a	3.13	2-way
	R. Posterior Cingulate	23	6	-26	18	43d	3.18	3-way
	L. Posterior Cingulate		-14	-46	16	21	3.62	
Temporal	L. Superior Temporal	22	-58	-2	10	247e	3.69	
	L. Superior Temporal	22	-56	6	0	e	3.33	
	L. Parahippocampus		-38	0	-22	161f	3.67	
	L. Hippocampus		-30	-14	-14	f	3.62	2-way
	L. Fusiform	19	-36	-50	-10	31	3.76	2-way
Sub-cortical	L. Putamen		-34	-12	-2	c	3.3	2-way
	R. Thalamus		22	-30	16	24	3.33	
	L. Thalamus		-6	-24	20	d	3.36	3-way
	Placebo Control >vsWM Training							
Parietal	L. Inferior parietal	40	-56	-48	46	21g	3.16	
	L. Inferior parietal	40	-52	-60	44	g	2.96	3-way
	L. Superior parietal	7	-34	-76	44	119h	3.44	3-way
	L. Superior parietal	7	-40	-66	50	h	3.29	3-way
	L. Precuneas	39	-48	-68	36	h	3.16	3-way
Frontal	L. Superior frontal	10	-16	58	30	326i	4.67	3-way
	L. Superior frontal	9	-10	54	42	i	3.51	3-way
	R. Superior frontal	9	16	58	36	769j	3.54	3-way
	R. Superior frontal	8	22	42	48	107k	4.45	3-way
	L. Middle frontal	10	-38	56	14	i	3.16	
	L. Middle frontal	10	-28	54	22	i	2.94	
	L. Middle frontal	9	-48	24	40	39	3.53	
	L. Inferior frontal Tri.	46	-46	48	4	26	3.38	
	L. Medial frontal Sup. Motor	8	-6	20	50	61	4.15	
	R. Superior Medial frontal	8	8	52	46	k	3.3	3-way
	R. Superior Medial frontal	10	10	62	24	j	3.59	
	R. Medial frontal orbital	11	8	52	-10	J	4.14	
Cingulate	R. Anterior Cingulate		0	48	6	j	3.14	3-way
	L. Anterior Cingulate	32	-2	42	-10	j	3.2	3-way
	L. Anterior Cingulate	32	-4	34	30	91	3.48	
Temporal	R. Superior Temporal	22	62	-56	20	10	3.59	3-way
	L. Middle Temporal		-54	-46	-8	20	3.4	3-way

The activation difference between the post-training and the pre-training on the MemHigh>ConHigh contrast is presented for both the vsWM Training group>Placebo Control group (Figure 21 panel A) and the Placebo Control group>vsWM Training group (Figure 21 panel B). The vsWM training groups increased in activation in areas associated with the default network and decreased in activation in parietal and frontal regions, in comparison to the placebo control group (Figure 21 panel C).

A) vsWM Training>PlaceboControl: Right, left, front, and back views of the brains' activation.



B) PlaceboControl>vsWM Training: Right, left, front, and back views of the brains' activation.



C) vsWM Training>Placebo Control (blue) and PlaceboControl>vsWM Training (red).

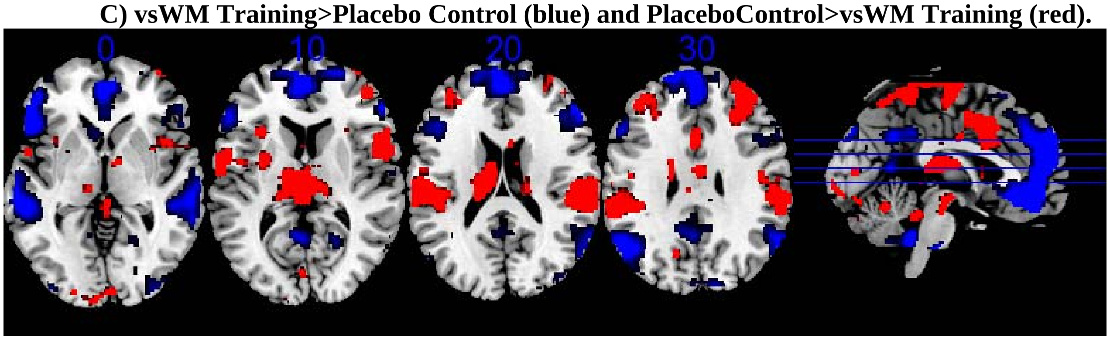
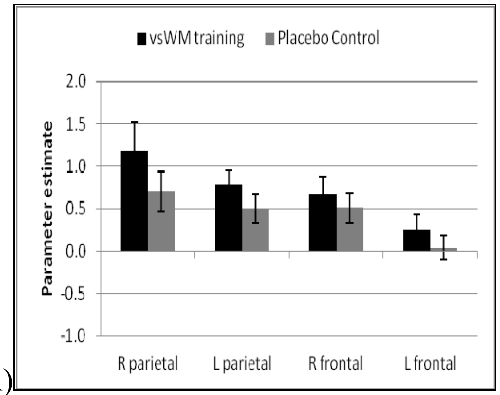


Figure 21. Group differences on the difference between post-training and pre-training in the MemHigh>ConHigh contrast; p<0.05 uncorrected, k=10.

ROI Analysis

ROI analyses examine the brains' activation in regions of interest defined a-priori. For this study, the ROI are bilateral parietal and frontal regions (BA 7, 39, 40 & BA 8, 9, 44, 45, 46, 47), which have been implicated many times in WM tasks (Klingberg, 2000; D'Esposito, Postle, & Rypma, 2000; Olesen, Westerberg, &

Klingberg, 2004; Todd & Marois, 2004; D'Esposito, 2007; Tomasi, Chang, Caparelli & Ernst, 2007; McNab & Klingberg, 2008). These regions were anatomically selected using the automatic anatomical labeling atlas. Additional regions to examine are the default network regions, which include BA 24, 9, 10, 32, 29, 30, 23, 31, 39, 40, 21 and the hippocampal regions. These default network regions have been shown to change based on task demands (Raichle, et al., 2001; Buckner, Andrews-Hanna & Schacter, 2008). Some of the general regions (as defined by BA) for the default mode and task activation supposedly overlap. In those cases, the specific activation of the examined cluster was classified based on pre-training activation in the MemHigh or ConHigh trials. Parameter estimates (6mm spheres) were extracted from post-training minus pre-training difference for regions that SPM indicated showed differences in activation on either the vsWM training>Placebo control or the Placebo control> vsWM training contrast. Pre-training parameter estimates were extracted for those regions showed no group differences (Figure 22; all p's>0.1). Therefore, the differences between the groups at post-training and at post-training minus pre-training (Figure 23) are believed to reflect the changes related to the training.

A) 

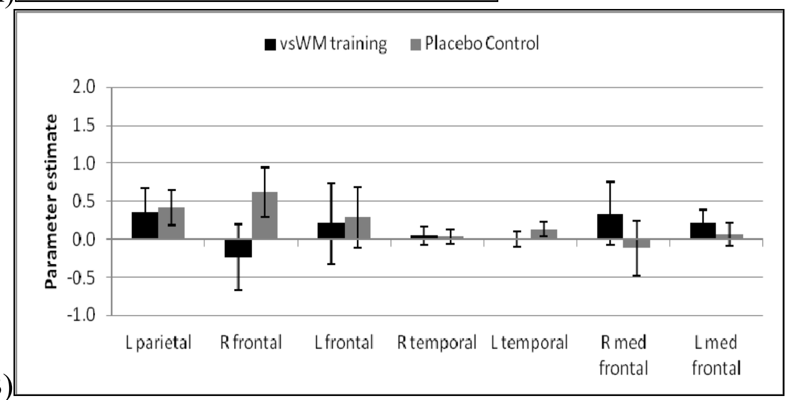
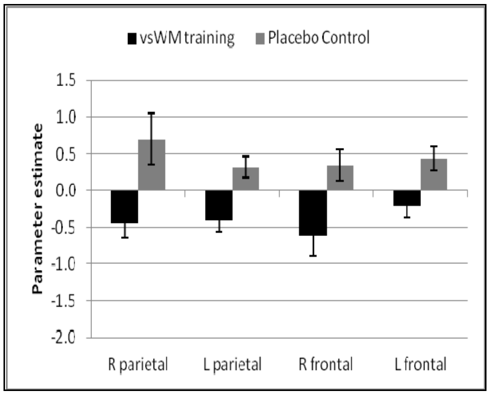
B) 

Figure 22. Parameter estimates for pre-training, showing no group differences. Error bars reflect standard errors of mean. Panel A are regions that showed task related activations, panel B are regions associated with the default mode network. Data for these charts was extracted directly from SPM.

The vsWM training groups increased in activation in areas associated with the default network (Figure 23 panel B) and decreased in activation in parietal and frontal regions (Figure 23 panel A), in comparison to the placebo control group ($t(33)=|4.33$ to $2.65|$, all p's<0.05) .

A) 

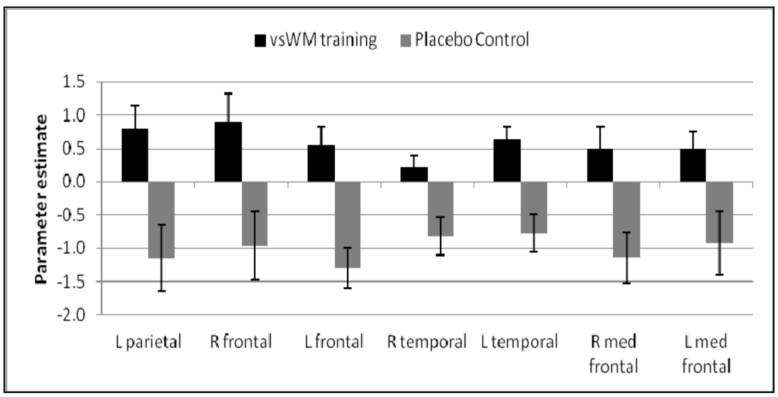
B) 

Figure 23. Parameter estimates for the post-training minus pre-training differences. Panel A is the vsWM training> Placebo Control. Panel B is the Placebo Control>vsWM training. Error bars reflect standard errors of mean. Data for these charts was extracted directly from SPM.

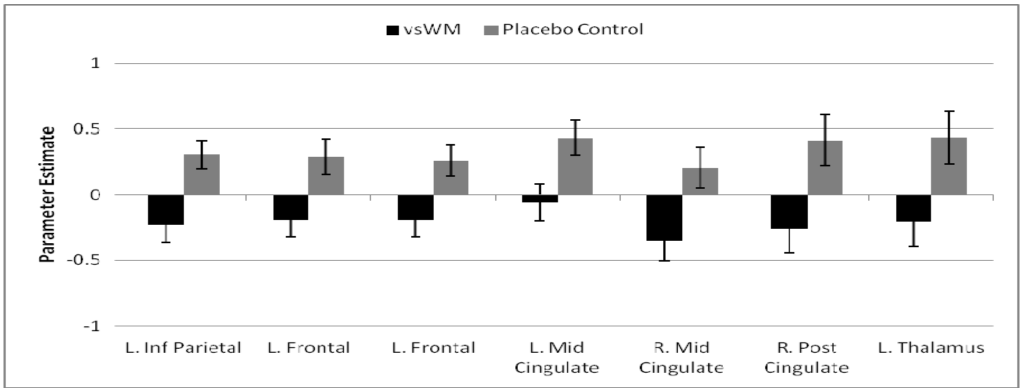
The ROI analysis is constructed as three-way interactions (depicted in Figure 21), between the time of assessment (Pre and Post), the trial type (MemHigh and ConHigh), and the training group (vsWM training and placebo-control). To better understand the source of the differences, the parameter estimates of the brain activation were extracted for the forty-six regions listed in Table 18, for the pre-training MemHigh, the pre-training ConHigh, the post-training MemHigh, and the post-training ConHigh. The extraction used a 6mm sphere from a map with p=1. The threshold was set high in order to capture all the voxels that SPM used for the analysis. The parameter estimates for each region was subjected to a mixed repeated measure ANOVA, with time (pre, or post) and trial type (MemHigh, or ConHigh) as repeated within group variables and the group type (vsWM training or placebocontrol) as the between group variable. Twenty of the forty-six regions showed significant three-way interactions between time, group and type (Table 19), and an additional eight regions showed significant two-way interactions between time and group (Table 20 & Figure 25). The relationship on the parameter estimates between MemHigh minus ConHigh for both groups at both time points is presented in Figure 24, vsWM training> Placebo Control regions (panel A), and for Placebo Control>vsWM training regions (panel B). The vsWM group shows decreases in activation at post-training for regions typically active during demanding tasks and increases in activations in areas associated with the default network activation.

Table 19. Statistics for regions showing a three-way interaction: trial type, assessment time and training group.

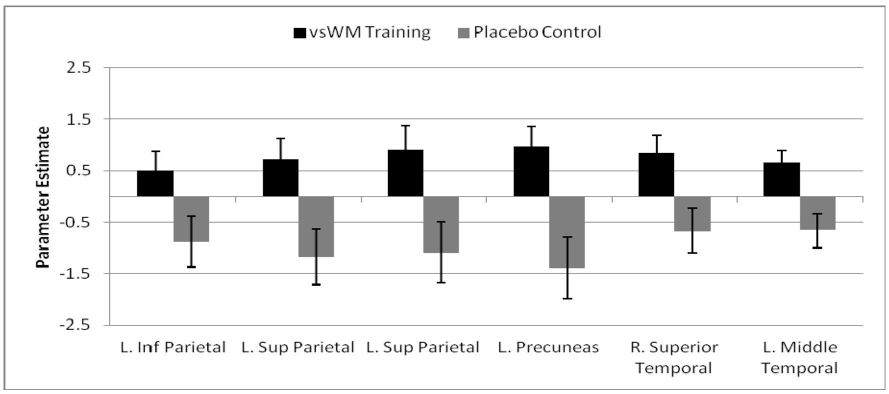
				vsWM Training		Placebo Control		3 way interaction		
Region	x	y	z	Pre β	Post β	Pre β	Post β	F	p	η_p^2
Parietal										
L. Inferior	-30	-30	32	0.52	0.29	0.40	0.71	9.23	0.01	0.22
L. Inferior	-52	-60	44	-1.81	-1.30	-1.07	-1.94	5.10	0.05	0.13
L. Superior	-34	-76	44	-1.42	-0.68	-1.17	-2.36	8.54	0.01	0.21
L. Superior	-40	-66	50	-2.24	-1.33	-0.72	-1.82	7.15	0.05	0.18
L. Precuneas	-48	-68	36	-2.22	-1.24	-1.76	-3.15	11.21	0.01	0.25
Frontal/Insula										
L. Frontal	-24	6	30	0.31	0.12	0.09	0.38	6.74	0.05	0.17
L. Frontal	-30	-14	32	0.32	0.13	0.16	0.42	6.79	0.05	0.17
L. Superior	-16	58	30	-1.59	-0.99	-1.22	-2.68	8.63	0.01	0.21
L. Superior	-10	54	42	-1.83	-0.97	-1.70	-2.92	5.78	0.05	0.15
R. Superior	16	58	36	-1.06	-0.85	-1.05	-2.32	4.58	0.05	0.12
R. Superior	22	42	48	-1.43	-0.70	-0.54	-1.61	9.72	0.01	0.23
R. Superior Medial	8	52	46	-1.29	-0.95	-0.97	-2.01	5.54	0.05	0.14
Cingulate										
L. Middle	-16	-10	36	0.24	0.18	0.11	0.54	6.55	0.05	0.17
R. Middle	16	-28	36	0.17	-0.18	-0.24	-0.03	6.47	0.05	0.16
R. Posterior	6	-26	18	0.31	0.05	-0.13	0.29	6.32	0.05	0.16
R. Anterior	0	48	6	-2.50	-1.55	-2.17	-3.66	6.93	0.05	0.17
L. Anterior	-2	42	-10	-3.19	-1.47	-2.54	-3.81	9.24	0.01	0.22
Temporal										
R. Superior	62	-56	20	-1.59	-0.76	-1.39	-2.06	7.38	0.05	0.18
L. Middle	-54	-46	-8	-1.05	-0.38	-0.44	-1.10	11.41	0.01	0.26
Sub-cortical										
L. Thalamus	-6	-24	20	0.28	0.08	-0.15	0.28	5.56	0.05	0.14

Table 20. Statistics for the regions that showed a time by group interaction.

				vsWM Training		Placebo Control		Time by Group interaction		
Region	x	y	z	Pre β	Post β	Pre β	Post β	F	P	η_p^2
Parietal										
R. Postcentral	46	-16	24	-0.06	-0.36	-0.49	-0.35	9.39	0.01	0.22
Frontal/Insula										
R. Insula	46	10	-4	0.17	-0.39	-0.01	-0.38	4.59	0.05	0.12
L. Insula	-36	0	-8	-56.89	-0.46	-40.87	-0.58	6.33	0.05	0.16
Cingulate										
R. Middle	12	-2	34	0.21	-0.04	-0.02	0.02	4.40	0.05	0.12
L. Middle	-10	2	34	0.35	0.10	-0.08	-0.08	7.71	0.01	0.19
Temporal										
L. Hippocampus	-30	-14	-14	-0.65	-0.56	-0.64	-0.30	6.36	0.05	0.16
L. Fusiform	-36	-50	-10	-0.14	-0.23	-0.18	-0.08	6.43	0.05	0.16
Sub-cortical										
L. Putamen	-34	-12	-2	0.12	0.07	-0.20	-0.14	5.90	0.05	0.15



A)



B)

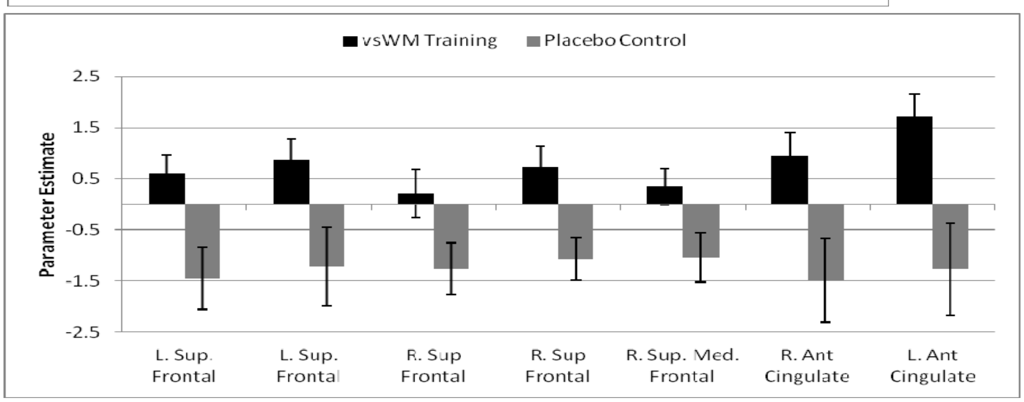


Figure 24. The parameter estimates of activation for the regions that show three-way interactions between time of assessment, trial type and training group. Panel A are the areas associated with decreases in activation, panel B are areas associated with the default mode network. Error bars reflect standard errors of mean.

The post-training decreases in brain activation for task demanding areas, and increases in activation for areas related to the default network imply that the vsWM group is more efficient in performing the high memory load trials and therefore needs to recruit less neural resources to perform the task.

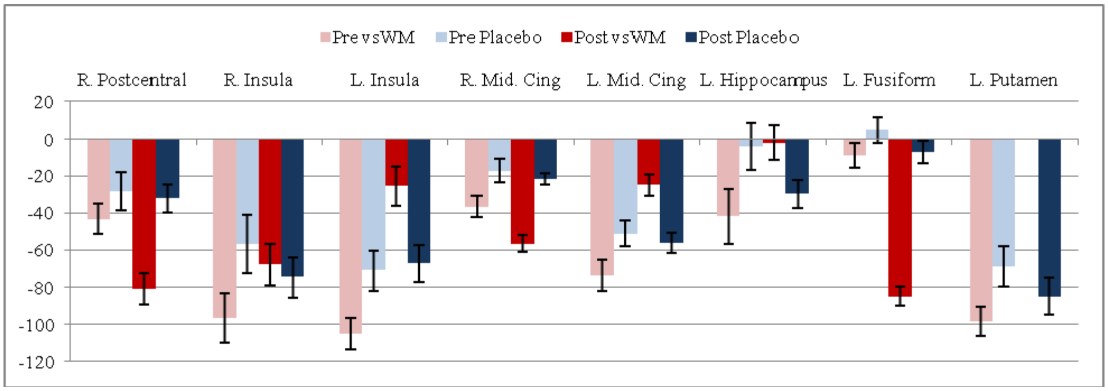


Figure 25. The parameter estimates of activation for the eight brain regions showing a significant two-way interaction between time of assessment and training group. Error bars reflect standard errors of mean.

Brain Behavior Correlations

Correlations between the BOLD activation in the twenty regions with the significant 3-ways interactions (post-minus-pre, MemHigh vs ConHigh, and the vwWM vs PlaceboControl groups), were examined with the pre and post behavioral assessments, letter-number-sequencing, ShapeBuilder, Ravens, and SimonsSays HighMem accuracy. The pre-training data was collapsed across group assignment, as participants had not yet been assigned to a group. The pattern of activation is not consistent across regions within tasks. Future analysis will examine the correlations between the task based BOLD activation and the training and assessment measures.

Letter-number-sequencing: Pre-training there were no significant correlations of the BOLD activation with letter-number-sequencing. Post-training only the vsWM training group displayed significant correlations with the post BOLD activation. Moderate correlation ($r=0.49$ to 0.58, $p<0.05$), for the L frontal, the right posterior cingulate and bilaterally for the middle cingulate.

ShapeBuilder: Pre-training there were no significant correlations of the BOLD activation with Shape Builder. Post-training, the vsWM training group displayed significant correlation with the left inferior parietal and the left middle cingulate (r=0.5 to 0.6, p<0.01), whereas the placebo control group displayed significant correlations with the left precuneas and IPL regions, the right temporal, bilateral anterior cingulate, and bilateral superior frontal regions (r=0.52 to 0.67, p<0.05).

Ravens: Pre-training Ravens score showed significant negative correlations with BOLD activation in the left precuneas, bilateral superior frontal, bilateral anterior cingulate and right medial superior frontal (r=-0.34 to -0.50, p<0.05). Post-training showed only significant correlations in the vsWM training group, with bilateral superior frontal, and the right medial superior frontal (r=-0.48 to -0.65, p<0.05).

SimonSays HighMem: Pre-training SimonSays for the high memory load trials showed significant correlations to the left middle cingulate (r=0.34, p<0.05). Post-training in the vsWM training group, there were moderate to strong correlations (r=-0.5 to 0.65, p<0.05) that switched direction based on region; right superior cingulate and the left thalamus showed a positive relationship, whereas the left superior frontal region showed a negative one. In the placebo control group, the post-training Raven's score positively correlated with the BOLD activation in the left inferior and superior frontal and the right superior temporal (r=0.5 to 0.55, p<0.05).

Discussion

This experiment investigates the malleability of WM, and the transfer of improvements to trained and untrained tasks. The results suggest that WM can be improved via training. The training specific tasks depict the improvement in performance on a non-adaptive version of the trained task. The pattern of results for the process-specific transfer is mixed, as the degree of network overlap of the training with the tasks differs. OSpan, while an assessment of WM, is verbal in nature and loaded on a separate factor than the vsWM in the first study. The same applies to Letter-number-sequencing, verbal fluency and Stroop, as they all involve verbal material. Ravens, although a spatial reasoning task, also did not load with the vsWM tasks in the latent variable analysis; it loaded with the reasoning ability tasks. It was somewhat surprising that the online vsWM training did not lead to transfer on the SymSpan task, as prior training studies have found process-specific transfer to complex span tasks (Atkins et al., under review).

There are many possible reasons for the lack of transfer, one being the duration of training. The cognitive training that showed process specific transfer to the complex WM tasks used twenty hours, whereas the current training is only ten hours. Future research should examine the issues with regard to the lack of predicted process-specific transfer in the complex WM span tasks.

Additional evidence for process-specific transfer to the visual spatial network was found in the Posner cueing improvements. The Posner cueing improvements were primarily in the visual presentation, although transfer was also evident in the long auditory trials. This stronger benefit to the visual modality is consistent with degree of process-specific transfer, as the relationship of the visual Posner cueing to the vsWM training would be stronger than the auditory version. Process-specific transfer was also evident in the mental rotation task, a task known to be related to vsWM (Zacks, 2008). Benefits were also evident in the response time on the math problems in the mental math task, suggesting that the 10 hours of vsWM training did induce process-specific improvements to untrained cognitive tasks. Additionally, no transfer was found to the process non-specific tasks, confirms that there was no Hawthorne effect and that there was no general motivational increase to perform better at the end of the study.

The neural network targeted by these vsWM training tasks is the frontal-parietal network. Many studies have shown this network to be engaged during WM, executive control, and attentional processes (Smith & Johnides, 1997; Olesen, Westerberg, & Klingberg, 2004; D'Esposito, 2007; McNab et al., 2008; Klingberg, 2010). The vsWM training group exhibited decreases in parietal and prefrontal regions related to task performance, compared to the placebo control group. This reduction in brain activity suggests that the participants in the vsWM training group became more efficient at performing the task and, therefore, needed to recruit fewer resources to complete the task.

Research has at times shown that frontal eye fields can predict behavioral task performance (Hayes, Petrov & Sederberg. 2011). Therefore, it is important to note that the task activation regions were not frontal eye-fields regions (x:-24 to -40 or 21 to 40; y:-6 to 1; z: 44 to 51). In fact, in the stated analysis, the HighControl should show greater frontal eye field activation than the HighMemory, as the HighControl contained sequences of 7 dots always, and the HighMemory contained sequences of 5, 6 or 7 dots. Therefore, eye movement is not a factor that needs to be accounted for in the task analysis.

Additional support to this hypothesis comes from the increased activity in the vsWM training group following the training in regions associated with the default mode network. The increase in default mode activity during the memory trials, for the vsWM training group, implies that the vsWM training group does not find the task as demanding as it was prior to the training, again supporting the hypothesis that the vsWM training group is more efficient when performing the demanding memory task.

Chapter 4: General Discussion

The work presented here investigated the underlying structure of WM and its malleability. This research investigated whether WM is a unitary domain general construct or a multi-component domain specific construct, and presented two new measures of WMC, Block-span and Letter-number-sequencing. The presented research also investigated the malleable nature of WM, through vsWM training, and examined the transfer of training-induced improvements to untrained cognitive tasks.

The first major finding from this work is that WM is a multi-component construct. Experiment 1 both verified Block-span and Letter-number-sequencing as valid, reliable measures of vsWM and vWM respectively, and provided supporting evidence for the multi-component domain specific perspective of WM. The latent variable analysis used structural equation modeling to examine the unitary model of WM and the multi-component model of WM. The multi-component model of WM, with a vsWM component and a vWM component, provided a good fit to the data, as indicated by all model fit indices, whereas the unitary domain general model provided a relatively poor, and significantly worse, fit to the data.

This finding replicates the findings of Kane et al.'s (2004, fig3). Indeed, similar to the Kane et al. study, Experiment 1 also revealed a strong relationship between the vsWM and vWM latent variables (r=0.69), although it is weaker than that found in the Kane study (r=0.84). Kane et al. argued that the strong relationship between the domain specific WM components reflected a general component which they labeled "attention". In order to highlight the executive attention variable, Kane et al. (2004) added short-term memory measures into their latent variable analysis. The addition of the short-term memory measures resulted in a three factor solution for executive attention and verbal and visual storages.

The current data cannot replicate the additional analyses conducted by Kane et al. (2004), as short-term memory measures were not collected. While the SEM for the WM tasks suggested a two-factor solution (vsWM and vWM), the full exploratory factor analysis produced a three factor solution, consisting of vWM, vWM and reasoning ability tasks. This is consistent with a growing body of behavioral and neuronal evidence which provides support for the multi-component domain specific distinction in WM (Smith, Jonides, & Koeppe, 1996; Smith & Jonides, 1997; Hartley & Speer, 2000; Baddeley, 2000; Klingberg, 2006; D'Esposito, 2007; Tomasi, et al., 2007; Bull, Espy & Wiebe, 2008; Thomason, et al., 2009).

The second major finding is that WM can be improved through training. This finding adds to the growing literature regarding the malleability of WM (Olesen, Westerberg & Klingber, 2004; Klingberg et al., 2005; Jaeggi, Buschkuehl, Jonides, & Perrig,2006; Dahlin, Stigsdotter-Neely, Larsson, Bäckman, & Nyberg, 2008; Holmes, Gathercole & Dunning, 2009; McNab et al., 2009; Chein & Morrison, 2010; Atkins et al., under review). This study has shown that it is possible to improve vsWM with 10 hours of adaptive training, and show transfer of the improvements to behaviorally assessed, non-trained, cognitive tasks that rely on the same network as the trained tasks. These improvements on the training tasks and the transfer tasks occur only in the vsWM training group, not the Placebo Control group. The training-induced improvement transferred to the mental rotation task, to the Posner cueing task and the results suggest some degree of transfer to the mental math task.

This study also shows that the WM training administered in an online format can lead to improvements; this is in direct contrast to the results from Owen et al, (2010). Owen et al, (2010) conducted a training study that administered both online training tasks and online pre-testing and post-testing. Owen and colleagues did not find training improvements or training induced transfer in their study, even though they had a sample size of over 11,000 participants. However, their study has several methodological issues. First, they recruited participants from a website intending to debunk the theory of cognitive training ('Band goes the theory'). Second, they included participants who complete between both assessment test and at least **two** training sessions during the 6 week training period. Each participant trained on six training task during a 10 minute session. Therefore, participants who completed two training sessions saw each of the training tasks twice, and trained for a total of 3-5 minutes per task, with a total training time of 20 minutes. Research implies that the degree of training improvement is directly related to the duration of the training (Jaeggi, Buschkuehl, Jonides & Perrig, 2008). Therefore, the duration of training in the Owen et al (2010) study should not lead to improvements and transfer as one does not really train in the task. An additional issue is the training adaptively and task difficulty. The online training in the Owen et al (2010) study consisted of six tasks during 10 minutes, leaving just a couple minutes for each of the training tasks. That brings up the question of whether participants are training in an adaptive or demanding task. Adaptive training brings participants to their threshold performance level, and maintains that difficulty until participants achieve that difficulty level (Mahncke et al., 2006). For example, in my study participants did not hit their threshold until the second or third training session, and each of my sessions was 7 or 8 minutes in duration. Therefore, I doubt that participants, who trained for two sessions, ever reached their threshold performance levels on the trained tasks.

The results from the Owen et al (2010) study highlight the difficult of conducting good training studies. Recruiting participants with a pre-conceived notion of what the study is investigating leads participants to behave based on the predictions (Boot, Blakely & Simons, 2011). In my second experiment, all people were recruited and randomly placed into one of two groups of training. The study did not require or question participants regarding expertise in other types of computer games of any kind. In addition, the assessment battery was designed to relate to tasks trained upon by both groups, maintaining the perception for participants in both training groups that their training expected to yield improvements. The vsWM training group trained on vsWM tasks (Memnoysne and ShapeBuilder), and the placebo control group trained on reading and math tasks (Sentencical and Number piles). The pre-training and post-training assessment battery contained tasks involving vsWM tasks, reading and math tasks, so both groups would be equally likely to expect improvements.

Another pitfall that training studies must avoid is the Hawthorne effect (Shipstead, Redick & Engle, 2010), in which participants might be inclined to improve on their performance to appease the researcher. One way of getting around the Hawthorne effect is to include tasks to which you are not expected to show the transfer. Predicting the lack of transfer allows one to examine whether the participant exhibited a generalized post-testing improvement, or only improvements on the tasks specifically predicted to improvement after training. In the presented training study, process specific tasks were predicted to show training-induced improvements following the vsWM training, whereas, process non-specific tasks were predicted to not show any improvement following training. A- priori predictions regarding the transfer of improvements and including assessment tasks predicted to show no-transfer effects, addresses whether the Hawthorne effect in the training data.

The results from the online vsWM training study present a theoretical framework for distinguishing between training-specific, process-specific, and process non-specific transfer of the training-induced improvements. Examining the transfer of improvements based on processes, as opposed to more surface properties within the near and far transfer framework, provides additional information about the expectations and underlying sources of the transfer of improvements. An assessment task was considered process specific based on the overlap in the brain network active during that assessment task and the frontal-parietal network targeted by the training tasks, not just whether the transfer task and the training task shared surface task properties or were in the same modality. Behaviorally, process specific tasks should load on the same factor as the trained tasks. Therefore, the process specificity classification also takes into account factor analyzes on behavioral data, which typically imply a distinction between modalities (Kane et al., 2004; Experiment 1 of the current study). The modality aspect is therefore only one of the aspects that will eventually define whether a task is process specific or not. At times these expectations make clear prediction regarding the transfer of improvement, as they did with visual Posner cueing, and the mental rotation task. Yet at times, different theoretical perspectives predict different transfer patterns, as was evident in OSpan, a WM task that relies on a similar network as the vsWM task (McNab et al., 2008), but loaded on a different factor in the first experiment.

The third major finding is that WM training produced decreases in BOLD activation in areas associated with task demand, the parietal and frontal regions. These finding are consistent with other finding in the literature regarding a decrease in activation following training (Garavan, et al. 2000; Chein & Schnieder, 2005; Dahlin, Bäckman, Stigsdotter-Neely & Nyberg, 2008). In addition, these finding show increases in activation in areas associated with the default mode network (middle and posterior cingulate, medial frontal and temporal regions), similar to the findings of Olesen, Westerberg & Klingberg (2004). This finding might help clarify the inconsistencies in the literature regarding whether activation should increase or decrease following training, as the direction of the activation change depends on the exact task and definition of the ROI, as default mode will at times activate the parietal regions.

Decreases in the BOLD activation following training are predicted to be the result neural plasticity, the ability of the brain to continuously update based on past experiences. Neural plasticity has been suggested to influence processing by improving the processing of select neurons or by the inhibition of non-selective neurons, and has been shown on occur after short duration (habituation, priming or repetition suppression) or after long durations (learning or training) (Kourtzi & Grill-Spector, 2005). Plasticity on short duration, similar to priming, are thought to reflect stimulus-response learning, where on the subsequent encounter of the stimuli bypasses the response processing stage (Horner & Henson, 2008). However, long term plasticity is thought involve a reduction of the cognitive load created by the task, and therefore leads to a reduction in the extent and magnitude of cortical activation (Chein & Schneider, 2005). The current training study led to long term plasticity evident by decreases in regions of task related activation and increase in default mode activation, support the theoretical perspective of greater efficiency in the recruitment of brain resources following WM training.

Both of the presented experiments showed that it is possible to administer cognitive assessments and training over the web. The advantages of deploying experiments via the web is threefold. First, there is no proprietary software that restricts other researchers' usage. Second, it reduces the time and resources each researcher must devote to face to face meetings and administration of tasks. Third, it allows to administer the training outside of a tightly controlled laboratory setting, Experiment 2 has shown that cognitive training can be effectively removed from the laboratory setting and the requirement of individualized interactions. For cognitive training to have a large scale impact and be deployable to target populations and society in general, it needs to be mass distributable in a format that allows for limited interaction. This research shows that it is possible.

Future Experiments and Analyses

Very few studies provide clear and unequivocal conclusions of research questions. The research presented here is part of a series of studies examining the underlying structure of WM, modeling performance during WM tasks, designing and validating additional WM measures, and furthering our understanding of the nature and process of WM training.

The research on the malleability of WM administered additional cognitive tasks that were designed to investigate process-specific transfer in the brain, as well as allow for usage of functional connectivity to examine the neural network underlying the tasks. Diffusion-tensor-imaging was also collected to allow the examination of the changes in white matter tracks following the vsWM training. Similarly, the high resolution structural images collected can be used as to examine voxel-based-morphology pre-training and post-training, to investigate the volumetric changes to grey matter following the vsWM training.

Future experimental designs should investigate the effects of WM training duration and the relationship between specific training tasks on the improvement following the training, with respect to domain specific versus domain general abilities. Additional research is needed to investigate the differences in the adaptivity of the WM training progression. These are important research questions that will further our understanding of how to design and administer WM training and better create WM assessments and most importantly further our understanding of the core cognitive process called WM.

Appendix A: WM Malleability: Assessment Order

The two versions of the ordering of the cognitive tasks presented at the pre and post training sessions for the second study.

	Order A	Order B
1)	OSpan	ShapeBuilder
2)	SymSpan	Letter-number-sequencing
3)	Letter-number-sequencing	Stroop
4)	ShapeBuilder	GORTa
5)	ModMathB	Word ID & Attack a
6)	Stroop	Ctop &VF
7)	Posner	MentalMath
8)	MentalMath	Mental Rotation
9)	RavensE	Posner
10)	Mental Rotation	SymSpan
11)	Ctop &VF	RavensO
12)	Word ID & Attack b	OSpan
13)	GORTb	ModMathA

Appendix B: WM Malleability: Mental Math Descriptives

Descriptive statistics for the different levels in the Mental Math task.

	vsWM Training (n=18)				Placebo Control (n=15)			
#Level	Pre		Post		Pre		Post	
Example	M	SD	M	SD	M	SD	M	SD
Accuracy								
1 5+4-6	92.22	8.78	93.89	6.98	91.33	11.87	92	10.14
2 12-11+23	48.89	11.32	53.33	9.7	46.67	13.45	54.67	13.02
3 4*3/2	93.33	12.37	94.44	5.11	88	10.82	94	9.1
4 15*12/10	55	8.58	56.11	9.78	57.33	7.04	54.67	6.4
5 15*2/10+4	62.22	15.17	66.11	13.78	56.67	19.15	61.33	19.59
6 10/5*20/2	18.99	13.67	28.89	16.76	20.67	15.34	28.67	20.31
RT								
1 5+4-6	1480.56	333.64	1354.83	278.75	1468.35	349.81	1528.71	339.6
2 12-11+23	1808.46	522.31	1713.16	432.65	1860.31	824.93	2130.4	519.56
3 4*3/2	1266.78	289.07	1162.44	274.29	1229.06	411.06	1202.05	311.35
4 15*12/10	1246.23	347.2	1249.75	264.97	1380.54	501.95	1370.59	338.47
5 15*2/10+4	2139.61	536.26	1932.78	443.58	2512.14	867.77	2374.42	472.44
6 10/5*20/2	2713.67	750.67	2524.43	595.49	2973.17	1089.78	2905.53	683.02

Appendix C: WM Malleability: Mental Rotation Descriptives

Descriptive statistics for the different degrees of rotation in the Mental Rotation task.

	vsWM Training (n=18)				Placebo Control (n=15)			
	Pre		Post		Pre		Post	
	M	SD	M	SD	M	SD	M	SD
Accuracy								
Rot0	91.31	19.36	97.76	5.00	94.67	14.16	92.07	19.52
Rot45	78.19	15.55	77.74	15.62	76.58	17.86	79.31	15.23
Rot135	70.45	16.89	73.17	18.12	74.13	13.70	72.55	18.09
Mir0	80.55	19.43	87.52	9.83	78.44	16.53	86.69	10.41
Mir45	70.13	14.95	74.60	19.93	65.48	16.57	73.97	21.84
Mir135	62.16	18.34	63.57	20.40	62.75	18.53	64.67	21.26
RT								
Rot0	1443.89	248.29	1318.80	237.49	1489.37	279.29	1355.36	222.75
Rot45	1984.41	314.51	1731.79	369.26	1948.44	276.85	1807.85	224.32
Rot135	2214.68	359.19	2006.76	474.88	2193.15	454.78	2072.62	360.78
Mir0	1829.24	302.15	1839.89	642.98	1809.01	340.15	1857.91	703.75
Mir45	2103.88	344.05	1892.23	652.69	2019.16	315.76	1729.95	396.61
Mir135	2208.71	360.24	1946.14	505.93	2174.48	543.63	1966.06	446.14

Glossary

ADHD= attention deficit hyperactive disorder

AFOQTa= Airforce officer qualifying test verbal analogies.

AFOQTm= Airforce officer qualifying test mathematical knowledge.

AFOQTrb= Airforce officer qualifying test block rotation.

AFOQTrc= Airforce officer qualifying test reading comprehension.

AGFI= adjusted goodness of fit index.

BOLD= blood oxygen level dependant.

CFI= comparative fit index.

DTI= diffusion tensor imaging.

FDR= false discovery rate corrected.

fMRI= functional magnetic resonance imaging.

FWE= family wise error rate corrected.

gFI= general fluid intelligence.

GFI= goodness of fit index.

GORT= Gray oral reading test.

IFG= inferior frontal gyrus.

IPL= Inferior parietal Lobe.

MNI= Montreal neurological institute

MRI= magnetic resonance Imaging.

msec= milliseconds

NavSpan= automated navigation span.

NFI=normed fit index.

NNFI= non-normed fit index.

OSpan= automated operation span.

PFC=prefrontal cortex.

Ravens= advanced raven progressive matrices; E for even trials; O for odd trials.

RG1= ETS arithmetic aptitude test.

RG2= ETS mathematic aptitude test.

RL3= ETS inference test.

RMSEA=Root Mean Square Error of Approximation.

ROI= region of interest.

RotSpan= automated rotation span.

RSpan= automated reading span.

RT=reaction time

RunSpan= running span.

Sec=seconds.

SRMR= standard root mean square residuals.

SVC= small volume corrected.

SymSpan=automated symmetry span.

vsWM= visual spatial working memory.

vWM= verbal working memory.

VZ2= ETS paper folding, p1= part 1; p2=part 2.

WASI= Wechsler abbreviated scale of intelligence: matrices.

WM= working memory.

WMC= working memory.

References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a nomological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, 131, 567–589.

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General,* 130, 224-237.

Atkins, S.M., Dougherty, M.R., Harbison, J.I., Novick, J.M., Weems, S.A., Chrabaszcz, J.S., & Bunting, M.F. (Under Review). Persistence and transferability of working memory-training over time.

Atkins, S.M., Harbison, I.J., Dougherty, M.R. *Alpha-Numeric Memory Builder and Structuring Task*, Disclosure No. IS-2009-054: pending.

Atkins, S.M., Harbison, J.I., Bunting, M.F., & Dougherty, M.R. (In preparation). Using Online Games to Assess Cognitive Ability: Validity of two web-deployable measures of cognitive ability.

Atkins, S.M., Harbison, J.I., Bunting, M.F., Teubner-Rhodes, S., & Dougherty, M.R. (2009). Convergent Validity of Automated Block Span and Letter Number Sequencing with Other Working Memory Assessments. Poster presented at Psychonomics. Boston, USA.

Baddeley A.D., & Logie, R.H. (1999). Working memory: The multiple-component model. In Miyake, A., & Shah, P. (Eds.). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4: 417-423.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York, NY: Academic Press.

Beilock, S. L. & Carr, T. H. (2005). When high-powered people fail: Working memory and "choking under pressure" in math. *Psychological Science*, 16, 101-105.

Broadway, J.M. & Engle, R.W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, 42(2), 563-570.

Buckner, R. L., Andrews-Hanna, J. R. & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences,* 1124: 1–38.

Bull, R., Espy, K.A., & Wiebe, S.A. (2008) Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 Years. *Developmental Neuropsychology*,33(3), 205-228.

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

Chein, J.M., & Morrison, A.B. (2010). Expanding the mind's workspace: Training and transfer effects with a complex working memory span task. *Psychonomic Bulletin & Review*, 17, 193-199.

Chein, J.M., & Schneider, W. (2005). Neuroimaging studies of practice-related change: fMRI and meta-analytic evidence of a domain-general control network for learning. *Cognitive Brain Research*, 25, 607-623.

Conway, A. R. A., & Engle, R. W. (1996). Individual differences in working memory capacity: More evidence for a general capacity theory. *Memory*, 4, 577-590.

Conway, A. R. A., Cowan, N., Bunting, M. F., Therriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence,* 30, 163-183.

Conway, A. R. A., Kane, M. J., & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science*, 7, 547-552.

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

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford Psychology Series, No. 26. New York: Oxford University Press.

Cowan, N. (2005). *Working memory capacity*. Hove, East Sussex, UK: Psychology Press.

Curtis, C.E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences*, 7(9), 415-423.

D'Amico, A., & Guarnera, M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences*, 15, 189.

D’Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society B Biological Sciences*, 362, 761-772.

D'Esposito, M., Postle, B.R., and Rypma, B. (2000). Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Experimental Brain Research*, 133, 3-11

Dahlin, E., Stigsdotter-Neely, A., Larsson, A., Bäckman, L. & Nyberg, L. (2008).Transfer of learning after updating training mediated by the striatum. *Science,* 320, 1510-1512.

Dahlin, E., Bäckman, L., Stigsdotter-Neely, A., Larsson, A., & Nyberg, L. (2009).Transfer of the executive component of working memory: Subcortical area mediate transfer effects. *Restorative Neurology and Neuroscience*, 27, 405-419.

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

Dehaene, S. (2009).Origins of mathematical intuitions: The case of arithmetic. *Annals of the New York Academy of Sciences*, 1156, 232-259.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A.J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology*, 14,218-224.

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20:487- 506.

Dougherty & Atkins, S.M., *Number Piles*, Discloser No. IS-2009-055: pending.

Dougherty, M. R. P., & Hunter, J. (2003). Probability judgment and subadditivity: the role of working memory capacity and constraining retrieval. *Memory & Cognition*, 31, 968-982.

Dougherty, M.R., Atkins, S.M., & Dowling, E. *Modular Shape Builder*, Discloser No. IS-2009-053: pending.

Ekstrom, R. B., French, J. W., Harman, M. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Engle, R. W. (2001). What is working-memory capacity? In H. L. Roediger III & J. S. Nairne (Eds.), *The Nature of Remembering: Essays in Honor of Robert G. Crowder* (pp. 297-314). Washington, DC: American Psychological Association.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11, 19-23.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology: General*, 128, 309-331.

Engle, R.W., Kane, M.J., & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In Miyake, A., & Shah, P. (Eds.). *Models of working memory: Mechanisms of active maintenance and executive control.* New York: Cambridge University Press.

Facoetti, A., Trussardi, A.N., Ruffino, M., Lorusso, M.L., Cattaneo, C., Galli, R., Molteni, M., & Zorzi1, M. (2005). Multisensory Spatial Attention Deficits Are

Predictive of Phonological Decoding Skills in Developmental Dyslexia. *Journal of Cognitive Neuroscience*, 22(5), 1011-1025.

Friedman, N. P., & Miyake, A. (2004a). The reading span test and its predictive power for reading comprehension ability. *Journal of Memory & Language*, 51, 136-158.

Friedman, N. P., & Miyake, A. (2004b). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1), 101-135

Fuster, J.M. (2001). The prefrontal cortex–an update: time is of the essence. *Neuron* 30, 319–333

Garavan, H., Kelley, D., Rosen, A., Rao, S.M., Stein, E.A., (2000). Practice-related functional activation changes in a working memory task. *Microscopy Research and Technique*. 51, 54–63.

Goldman-Rakic, P. (1987). Circuitry of primate prefrontal cortex and regulation of behavior by representational memory. In Plum, F. (Ed.), *Handbook of Physiology, the Nervous System, Higher Functions of the Brain.* pp. 373–417, American Physiological Society.

Greenwood., P.M., (2007) Functional Plasticity in Cognitive Aging: Review and Hypothesis. *Neuropsychology*, 21(6), 657-673.

Harbison, I.J., Dougherty, M.R. & Atkins, S.M. *Matrix Memory Builder,* Discloser, No. IS-2009-052: pending.

Hartley, A.A. & Speer, N.K. (2000). Locating and fractionating working memory using functional neuroimaging: Storage, maintenance, and executive functions. *Micorscopy Research and Technique*, 51, 45–53

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision*, 11 (10:10), 1-11.

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), 9-15.

Horner, A.J. & Henson, R.N. (2008). Priming, response learning and repetition suppression. *Neuropsychologia,* 26, 1979-1991.

Huettel, S.A, Song, A.W., & McCarthy, G. (2009). Functional *magnetic resonance imaging*. Massachusetts, USA: Sinauer Associates.

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 of the United States of America*, 105, 6829-6833.

Jarvis, B. G. (2006). *DirectRT* (Version 2006.2.0.28) [Computer Software]. New York, NY: Empirisoft Corporation.

Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual differences perspective. *Psychonomic Bulletin & Review*, 9, 637-671.

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,* 153, 128-217.

Klingberg, T. (2006). Development of a superior frontal-intraparietal network for visuo-spatial working memory. *Neuropsychologia*, 44(11), 2171-2177.

Klingberg, T., Fernell, E., Olesen, P., Johnson, M., Gustafsson, P., Dahlström, K., Gillberg, C.G., Forssberg, H., & Westerberg, H. (2005). Computerized training of working memory in children with ADHD – a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, 44, 177-186.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, 24(6), 781-791.

Kourtzi, Z. and Grill-Spector, K. (2005). fMRI adaptation: a tool for studying visual representations in the primate brain. *In Fitting the Mind to the World: Adaptation and After-Effects in High-Level Vision* (Rhodes, G. and Clifford, C.W.G., eds), pp. 173–188, Oxford University Press.

Kyttälä, M., & Lehto, J.E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education*, 23(1), 77-94.

Logie, R.H. (1995). *Visuo-spatial working memory*. Hove, UK: Psychology Press.

Mahncke, H. W., Connor, B. B., Appelman, J., Ahsanuddin, O. N., Hardy, J. L., Wood, R. A., Joyce, N. M., Boniske, T., Atkins, S. M., & Merzenich, M. M. (2006). Memory enhancement in healthy older adults using a brain plasticity-based training program: A randomized, controlled study. *Proceedings of the National Academy of Science of the United States of America,* 103, 12523–12528.

Masson, M.E.J. (2011). A tutorial on a practical Bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods, 43*, 679-690.

McNab, F., & Klingberg, T. (2008). Prefrontal cortex and basal ganglia control access to working memory. *Nature Neuroscience*, 11:1,103-107.

McNab, F., Leroux, G., Strand, F., Thorell, L., Bergman, S., Klingberg, T. (2008). Common and unique components of inhibition and working memory: An fMRI, within-subjects investigation. *Neuropsychologia*, 46(11): 2668-2682.

McNab, F., Varrone, A., Farde, L., Jucaite A., Bystritsky, P., Forssberg, H., & Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science,* 323, 800-802.

Mercado, E. (2008). Neural and Cognitive Plasticity: From Maps to Minds. *Psychological Bulletin*, 134(1):109–137.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin*, 27, 272-277.

Miyake, A., & Shah, P. (Eds.) (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., & Howerter, A. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49-100.

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

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77–101.

Oberauer, K. (2005). The measurement of working memory capacity. In O. Wilhelm & R. W. Engle (Eds.): *Handbook of understanding and measuring intelligence*. Thousand Oaks, CA: Sage Publications, Inc.

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

Owen, A.M., Hampshire, A., Grahn, J.A., Stenton, R, Dajani, S. Burns, A.S., Howard, R.J., k& Ballard, C.G. (2010). Putting brain training to the test. *Nature*, 465,775-778.

Raven, J. C., Raven, J. E., & Court, J. H. (1998). *Progressive matrices*. Oxford, England: Oxford Psychologists Press.

Ricker, T.J., Cowan, N., Morey, C.C (2010). Visual working memory is disrupted by covert verbal retrieval, *Psychonomic Bulletin & Review*, 17(4), 516-521.

Rosen, V.M. & Engle, R.W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, 126, 211–227.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychological Software Tools.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General*,125, 4–27.

Shipstead, Z., Redick, T.S. & Engle, R.W. (2010). Does working memory training generalize? *Psychologica Belgica*, 50 , 245-276.

Smith, E.E., & Jonides, J. (1998). Neuroimaging analyses of human working memory. *Proceedings of the National Academy of Science of the United States of America*, 95, 12061-12068.

Smith, E.E., Jonides, J., Koeppe, R.A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex*, 6, 11-20.

Takeuchi, H., Sekiguchi, A., Taki, Y., Yokoyama, S., Yomogida, Y., Komuro, N., Yamanouchi, T., Suzuki, S., & Kawahima, R (2010). Training of working memory impacts structural connectivity. *The Journal of Neuroscience*, 30(9), 3297-3303.

Temple, E., Deutsch, G. K., Poldrack, R. A., Miller, S. L., Tallal, P., Merzenich, M. M., & Gabrieli, J. D. E. (2003). Neural deficits in children with dyslexia ameliorated by behavioral remediation: Evidence from functional MRI. *Proceedings of the National Academy of Science of the United States of America*, 100, 2860-2865.

Thomason, M., Race, E., Burrows, B., Whitfield-Gabrieli, S., Glover, G., Gabrieli, J. (2009). Development of spatial and verbal working memory capacity in the human brain. *Journal of Cognitive Neuroscience*, 21:316–332.

Thorell, L.B., Lindqvist, S., Bergman, S., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science*, 12:1, 106-113.

Todd, J.J & Marois, R., (2004). Capacity limit of visual short-term memory in human posterior parietal cortex. *Nature*, 428, 751-754

Tomasi, D., Chang, L., Caparelli, E.C., & Ernst, T. (2007). Different activation patterns for working memory load and visual attention load. *Brain Research*, 1132, 158-165.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory & Language*, 28, 127-154.

Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Operation Span and Raven. *Intelligence*, 33, 67-81.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, 37, 498-505.

Vogel, E. K. & Machizawa, M. G. (2004). Neural activity predicts individual differences in visual working memory capacity. *Nature*, 428, 784-751.

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review*, 14, 779–804.

Wechsler, D. (1997). *The WMS–III administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1999).*Wechsler Abbreviated Scale of intelligence*. The psychological corporation: A Harcourt Brace & Company.

Wiederholt, J.L. & Bryant, B.R. (2001). *Gray Oral Reading Test-IV*. Austin; TX.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of achievement*. Itasca, IL: The Riverside Publishing Company.

Zacks, J.M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, 20 (1), 1-19.