IMPROVING INTELLIGENCE BY INCREASING WORKING MEMORY CAPACITY

by

WENG TINK CHOOI

Submitted in partial fulfillment of the requirements

For the degree of Doctor of Philosophy

Dissertation Adviser: Lee A. Thompson, Ph. D.

Department of Psychological Sciences

CASE WESTERN RESERVE UNIVERSITY

May, 2011

CASE WESTERN RESERVE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

We hereby approve the thesis/dissertation of

Weng Tink Chooi

candidate for the Doctor of Philosophy degree *.

(signed) Lee A. Thompson, PhD
(chair of the committee)

Douglas K. Detterman, PhD

TJ McCallum, PhD

H. Gerry Taylor, PhD

(date) March 28, 2011

*We also certify that written approval has been obtained for any proprietary material contained therein.

Dedication

To my prayer warriors:

Your faithful prayers have been

a sweet and steady companion on this beautiful journey.

Thank you for cheering me on!

Be anxious for nothing, but in everything by prayer and supplication with thanksgiving let your requests be made known to God. And the peace of God, which surpasses all comprehension, will guard your heart and mind in Christ Jesus.

(Philippians 4:5-6)

Table of Contents

List of Tables

List of Figures

Acknowledgement

I would like to express my deep gratitude to those who have helped make this dissertation possible. First and foremost, I want to thank my academic adviser, Dr Lee Thompson, who has been a kind, patient and encouraging mentor. Thank you for investing your time in me. I have learned a lot from you. I want to thank Dr Douglas Detterman, a caring grandfather-like figure to me and brilliant professor. I thank Dr TJ McCallum for his support and generosity in allowing me to utilize his lab space for my research. I want to thank Dr Gerry Taylor for believing in me and your encouraging advice on finishing my dissertation in due time.

I also want to thank students who have contributed time and energy to my dissertation study. You have been an enormous help to me. Without your assistance, I would not have been able to achieve what I have in a timely manner. Thank you, Michelle Blair, London Holt, Jasmine Core and Linda Lou. It was a privilege to work with you and honor to be part of your academic growth. My experience with each one of you has made me a better student and teacher.

I want to thank the staff of Department of Psychology – Kori Kosek, Kellie Shaffer and Cynthia Teare. Thank you for all the assistance you have provided me over the past four years. I know I can count on each one of you and your service to the department do not go unnoticed.

Last but not least, I want to thank all the students who committed to participating in my study. Without your commitment, I probably would not have been able to graduate.

Improving Intelligence by Increasing Working Memory Capacity

Abstract

by

WENG-TINK CHOOI

A study by Jaeggi and her colleagues (2008) claimed that they were able to improve fluid intelligence (g_f) by training working memory. Subjects who trained their working memory on a dual n-back task for a period of time showed significant improvements in working memory span tasks and fluid intelligence tests such as the Raven's Progressive Matrices and the Bochumer Matrices Test (BOMAT) after training compared to those without training. The current study aimed to replicate and extend the original study conducted by Jaeggi et al. (2008) in a well-controlled experiment that could explain the cause or causes of such transfer if indeed the case. There were a total of 93 participants who completed the study, and they were randomly assigned to one of three groups – a passive control group, active control group and experimental group. Half of the participants were randomly assigned to the 8-day condition and the other half to the 20-day condition. All participants completed a battery of tests at pre- and post-tests that consisted of short timed tests, a complex working memory span and a matrix reasoning task. Participants in the active control group practiced for either 8 days or 20 days on the same task as the one used in the experimental group, the dual n-back, but at the easiest level to control for Hawthorne effect. Results from the current study did not suggest any significant improvement in the mental abilities tested, especially fluid intelligence and

working memory capacity, after training for 8 days or 20 days. This leads to the conclusion that increasing one's working memory capacity by training and practice did not transfer to improvement on fluid intelligence as asserted by Jaeggi and her colleagues (2008, 2010).

Introduction

Part of the nature versus nurture debate is the issue of the malleability of intelligence (Wahlsten, 1997) – can we modify or increase one's intellectual ability? Interests and efforts to raise intelligence as well as other related cognitive abilities have been around for more than a century (Spitz, 1986). The idea that it may be possible to manipulate intelligence has been very appealing to researchers in education and the behavioral sciences, and the large body of research focused on aspects of the treatment of intellectual impairment provides an excellent example of these efforts.

Mental retardation, now known as intellectual disability, is a life time condition characterized by difficulties in learning, social adjustment and every day functioning. The condition is not just devastating for the individual but also a source of stress for the family (Biasini, Grupe, Huffman & Bray, 1999). Gottfredson (1997) presented convincing support that intelligence (*g*) has a pervasive utility in everyday life. She also showed that higher *g* predicted more success and better outcome in occupational attainment. It is also more advantageous to be endowed with higher *g*. People with higher *g* (IQ above 120) were less likely to live in poverty, bear children out of wedlock, depend on social welfare and be incarcerated (for men). With such a powerful force predicting one's outcome in life, it is no wonder that many would want to raise intelligence due to the assumption that increasing intelligence should improve one's life. Efforts to raise intelligence or treat mental retardation date back to early 19th century with Itard. He spent several years teaching and training a 12-year-old boy, Victor, who apparently grew up in the woods. When Itard first met Victor, the boy was found to be mute. Itard's years of effort had improved Victor's condition but he never gained normal human functioning

(Pinchot, 1948). Projects like Head Start, the Milwaukee Project and the Abecedarian Project, among others, in the last five decades have shown the eagerness of psychologists and educators to cure mental retardation, but the ambitious and well-intentioned projects failed to live up to expectations and achieve their goals (Spitz, 1986).

Head Start Program is a nationally funded intervention program that provides educational, health and nutritional services/assistance to disadvantaged children as part of a plan to fight the War of Poverty started by President Lyndon Johnson. Children were enrolled in Head Start at age 3 and continued until they were 5 years of age. Between the year of its implementation in 1965 and 1989, the government spent on average $1.9 billion annually. For the last 10 years, the average annual spending has increased to $7.6 billion. With such massive resources channeled into the program, a recent national evaluation published in January 2010 reported that the benefits in cognitive and socio-emotional development children received through the program were not sustained when they reached first grade. The children who received the intervention did not show significant improvement compared to children in the control group, who were free to enroll in other early education programs (Muhlhausen & Lips, 2010).

The Milwaukee Project was spear headed by Rick Heber from the University of Wisconsin-Madison in the 1960s. His project targeted children of high-risk mothers who had a mean IQ score of below 80. Each child in the experimental group was started on the intervention program soon after they were born. They were sent to an infant stimulation center where they received education to train and develop their language and cognitive skills. The mothers of these children received rehabilitation and training in home-making and childcare. All the children in both experimental and control groups were tested several times throughout the project. When they reached 6 years of age, Heber reported that children in his experimental condition had a significantly higher IQ scores than those in the control group. His results were highly publicized although they were never published in a peer-reviewed journal and his results had remained controversial (Spitz, 1986).

The Abecedarian Project was conducted as a controlled experiment beginning in 1972 to evaluate the efficacy of early education intervention. The project aimed to prevent mental retardation, and the subjects in the project were children of mothers from families with high risk of developing mental retardation in their offspring. Children in the experimental condition received education taught in the form of games health care, nutritional supplements and social services. These latter services were also provided to the children in the control group. The predictor variable was to determine if intellectual improvements were due to benefits received from the cognitive and language training at the day care center, or to the global improvement in diet and health. At the completion of the project, results showed that there was substantial improvement in IQ scores in the experimental group compared to the control group but the superior performance quickly decreased when the project ended (Spitz, 1986).

Each intervention project briefly overviewed focused on global factors that might influence the development of skills and abilities crucial for educational attainment; and these projects have invested a lot of money and time, not to mention expertise of educators and researchers in their effort to raise intelligence. Various areas in the children's lives involved in the intervention programs were targeted, such as nutrition, health services and early education that emphasized cognitive and language training, yet there was no evidence suggesting that IQ scores or mental abilities were improved as a lasting effect. Perhaps trying to change or influence multiple factors at one time might not be as effective or the effects were confounded by the multitude of variables manipulated concurrently in these intervention programs. An alternative to studying the possibility of increasing intelligence could be targeting specific cognitive skills one at a time.

Over the past several years, the literature has seen a proliferation of research in cognitive training, specifically working memory training (Morrison & Chein, 2011). Early efforts include studies by Klingberg and his colleagues in which the researchers trained children with Attention Deficit and Hypertension Disorder (ADHD) on challenging working memory tasks. They detected positive effects from the training such as parental reports of reduced inattentive behavior (Klingberg et al., 2002; Klingberg et al., 2005). Researchers began to speculate if the training paradigm could be applied to other segments of the general population such as the elderly and patients recovering from brain injuries, perhaps even generalizing the effects of training to other cognitive abilities.

Recently, Jaeggi, Buschkuehl, Jonides and Perrig (2008) reported that they had improved fluid intelligence of subjects in a study by training their working memory. There were 70 participants in this study recruited from University of Bern, Switzerland. Half of these participants were trained on a challenging working memory task for four different durations – 8 days, 12 days, 17 days and 19 days. The other half of the participants were assigned as controls in four different groups that matched the training durations. The researchers in this study employed a dual N-back task as their working memory training task. In this computerized task, a series of blue squares was flashed on the screen at one of eight different locations in random order. Simultaneously, a series of letters was presented through an audio output. Each visual and audio stimulus was presented for 500 ms, and there was an interval of 2500 ms between stimuli. Participants were asked to press the A key when the location of the current square presented matched the location N stimuli back and the L key when the letter currently presented was the same as the letter presented N trials back. If both visual and audio stimuli presented were identical to the ones presented N trials ago, participants press both A and L keys (Figure 1). The N value was the same for both streams of stimuli. The task was designed to adapt to the user's performance so that the task would remain challenging to the user. If the users performed adequately in the current N-back task, i.e., they correctly identified matched stimuli in both modalities more than 90% of the time, the program would increase the difficulty level to N+1. However, if the user's performance dropped below 70% the program would reduce the difficulty level to N-1. The difficulty level remained unchanged if users did not meet either condition. The ability of the task to adapt to users' performance so that it remained demanding to the user was purported to have increased working memory capacity that translated to improved measured fluid intelligence (Jaeggi et al., 2008).

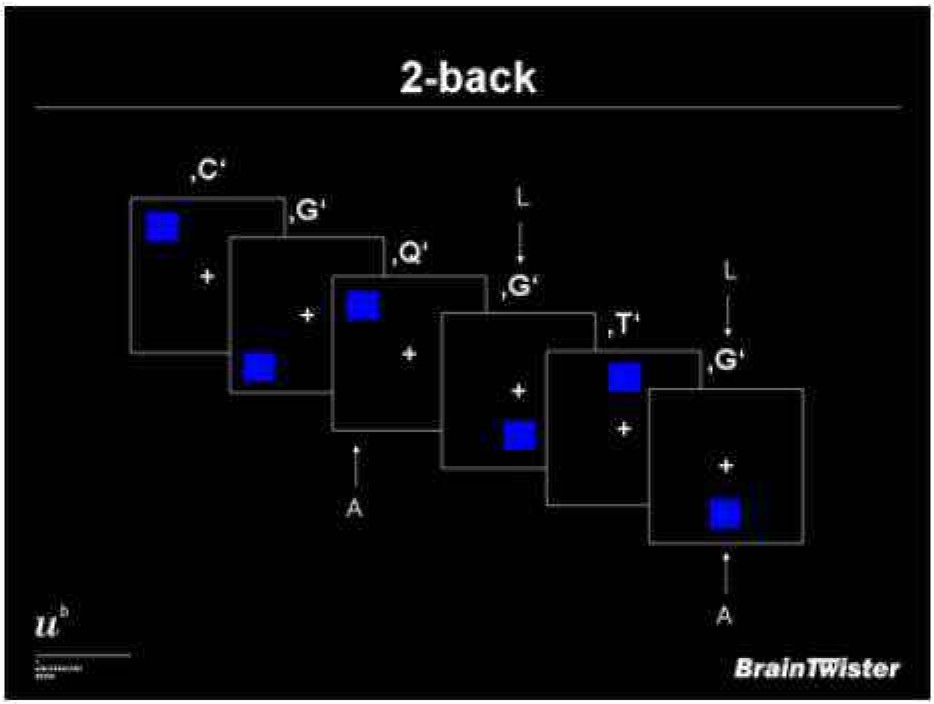

Figure 1: A screenshot of the instruction page that participants viewed prior to each block. In this figure, the block that was to be administered when participants proceeded from this page would be a 2-back task. Participants would need to respond if the current stimuli presented match the stimuli 2 trials back.

The participants in the Jaeggi et al (2008) study were asked to complete 20 blocks of the dual N-back task consisting of 20+n trials in each training session. Participants in the 8-day training group were trained for eight sessions, and those in the 12-day, 17-day and 19-day training groups were trained for 12, 17 and 19 sessions respectively. Before participants in the experimental group started their working memory training, they participated in a pre-test where they were administered a test that measured fluid intelligence for 10 minutes. To control for individual differences in gains in working memory capacity, the researchers also included a digit span task and a reading span task in the pre-test session. After participants completed their training as specified, they completed a post-test where they took the same IQ test in 10 minutes but answered different questions from the ones given in the pre-test. They also completed a digit span task and a reading span task. All the participants in the control group participated in the pre- and post-test sessions – they answered as many questions as they could in the IQ test in 10 minutes, and they completed the working memory span tasks (Jaeggi et al, 2008).

Jaeggi and her collegues (2008) reported that participants who were trained on their working memory significantly improved their fluid intelligence (G_f) scores (Cohen's d = 0.65). Although the participants in the control group also showed some improvements in their post G_f scores, the score gain was not significant (Cohen's d = 0.25). Their results also suggested that the more working memory training participants received, the higher were their post G_f scores – participants who trained longer had larger score gains. The mean level of performance in the dual N-back training task for participants in the experimental group increased at the end of the training sessions. These individuals were able to perform higher levels of the working memory task as they trained more. The performance improvement in the working memory training sessions and the score gains in G_f led the researchers in this study to conclude that training in working memory was transferred to an improvement in fluid intelligence, G_f (Jaeggi et al., 2008).

There were some weaknesses in the study by Jaeggi et al. (2008) as identified by Moody (2009) in an article reviewing the study. Participants in the experimental group did not take the same IQ test. Those in the 8-day training group did the Raven's Advanced Progressive Matrices (RAPM) and the rest of the participants did the Bochumer Matrices Test (BOMAT). The researchers in the study reported significant group differences for the 12-day, 17-day and 19-day training groups, which were the groups that answered the BOMAT, in their Analysis of Covariance (ANCOVA) with the pre-test scores as covariate and post-test scores as dependent variable. Participants who performed the RAPM in the 8-day group did not show significant improvement in their IQ scores after training for 8 days. (Jaeggi et al., 2008). The reason behind this observation could either be that longer training produced more score gains, or the nature of the training task itself facilitated better test taking specifically for the BOMAT (Moody, 2009).

Both the IQ tests – the RAPM and BOMAT – employed in the study by Jaeggi and her colleagues (2008) shared some similarities. Each item in both tests was a matrix of geometric or other figures and a spot in the matrix was left blank. The figures in the matrices were arranged according to a pattern, and test-takers would have to identify the pattern, or patterns, unique to each matrix in order to infer a solution from multiple possible answers given at the bottom of the matrices. Another similarity between the two tests was that the difficulty level of the questions increased as the test-taker progressed through either test. A high score would reflect the test-taker's ability to solve the difficult items. Test-takers would also be able to learn how to solve subsequent questions that were progressively harder based on the patterns they inferred from previous items that they had solved. Participants in the Jaeggi et al (2008) study were not given the opportunity to attempt the more difficult questions because the researchers essentially removed the progressive nature of the tests by reducing the allotted time to take the test from 45 minutes to 10 minutes (Moody, 2009). Not only did the researchers fail to provide an explanation or reason to why they administered different tests to measure their dependent variable, they also failed to administer the tests they chose to use appropriately. The only explanation they provided for allowing their participants only 10 minutes to complete as many questions as they could was in a footnote at the end of the study that acknowledged the deviation of their procedure from the standardized administration. In the same footnote, they defended their method of administration by referring to a study (Fearson & Eysenck, 1986) that showed high correlations between speeded and non-speeded version of a test (Jaeggi et al., 2008). Moody (2009) pointed out that the study by Fearson and Eysenck (1986) was not designed to support the conclusion made by Jaeggi and her colleagues (2008). In fact, the study published in 1986 contained its own footnote that referred to an unpublished experiment that used the RAPM with a reduction in administration time of 50% and not almost 80% reduction as in the study by Jaeggi et al. (2008).

One of the differences between the two tests was that items in the RAPM were made up of 3x3 matrices while items in the BOMAT were made up of 5x3 matrices (Moody, 2009). With 15 figures to keep track in one's working memory when one is solving the items in the BOMAT, it should not be surprising that improved working memory capacity after weeks of training enhanced one's ability to solve the early BOMAT items – items that do not challenge one's fluid intelligence as much as the later items. The reduced visual information provided by each item in the RAPM could be the source of challenge to one's fluid intelligence in inferring a solution to the matrix configurations in each item. It is appropriate in this context to mention once again that the researchers did not find significant improvement in IQ scores for the 8-day training group. Jaeggi and her colleagues (2008) claimed that their training task was "entirely different" from their transfer task, which was the IQ test. One of the dual dimensions inherent in the training task – the dual N-back task – was in fact what the researchers claimed to be. The task of keeping track of information through the audio stream was unlike any of the features in either the RAPM or the BOMAT. However, the visual task of keeping track of squares in several different locations on screen did not seem very different from the nature of the transfer task. The visual dimension of the training task almost represented a simplified version of the pattern identification or recognition in a matrix of figures. The exposure and training on the simplified version of the transfer task could explain why scores on the BOMAT improved more substantially than score gains on the RAPM (Moody, 2009).

The study by Jaeggi and her colleagues (2008) should be replicated and extended to validate their findings. After decades of unsuccessful and inconclusive research and efforts to raise intelligence, a study such as the one conducted by Jaeggi et al. (2008) that suggested otherwise should be subjected to further examination by more rigorously designed experiments to rectify the weaknesses in the original study and improve the procedures for more accurate results and interpretation of the results. One of the most immediate changes in a replication and extension of such a study would be the measurement of its outcome. There should be a more comprehensive battery of tests to measure the intellectual ability of each participant to assess the degree and pattern of change across specific component processes not just *g*. The results of the experiment should be compared with well-controlled groups. In other words, the treatment or experimental group should differ from the control group or a set of control groups only by the very element that affected the outcome. A follow-up study would be necessary to determine if the effects of working memory training were sustained after the completion of the experiment.

The central interest of the study by Jaeggi et al. (2008) was intellectual ability and if it could be enhanced by targeting a specific cognitive skill; thus it is imperative that the subject of interest is well-defined and operationalized in order to interpret the results appropriately. A firm understanding of the concept of intelligence is essential in order to comprehend the full implications of the findings presented by Jaeggi and her colleagues (2008) because they deduced that the level of intelligence of the participants in their study was improved based on their observations in the increase in IQ scores of the participants.

Spearman (1904) first noted that all cognitive ability tests were positively correlated with one another. He also discovered through methods of correlational analysis that the correlations among these tests could be accounted for by one factor. This factor has become known as general intelligence, or *g*. Since the conception of a general factor proposed to explain intelligence, there has been a persistent debate challenging the concept of *g*. Most research, however, has produced support for the existence of a higher order general factor that governs broader mental abilities manifested in mental ability tests. Carroll (2003) assembled and analyzed a variety of data sets using factor analysis to test and verify which model of intelligence was better explained by the data sets. His results indicated that they confirmed the "standard multifactorial view", which argued for a higher-order general intelligence influencing lower order broad mental abilities. Johnson et al (2004) demonstrated that the *g* factors from three different test batteries administered to the same sample were almost completely correlated with one another. Their findings were replicated by Johnson, Nijenhaus and Bouchard (2008) with five test batteries administered to the same group of participants. These studies strongly suggested the existence of a unitary higher-order general factor of intelligence that was not affected by how it was measured or assessed by any cognitive and mental ability tests.

In 1943, Cattell published a paper that criticized the lack of unity in the definition and measurement of intelligence at that time. He reviewed available contemporary mental tests since intelligence testing began and claimed that the tests were insufficient and defective. Most of the tests available at the time of Cattell's work used to measure intellectual abilities in adults were derived from tests designed for children. His dissatisfaction with the state of intelligence research led him to propose a theory of intelligence that described two types of "adult mental capacity" – fluid and crystallized abilities. The theory was based on Cattell's work on factor analysis of mental abilities tests and supported by various observations of differences in intelligence measurement in child and adult samples (Cattell, 1943).

Fluid intelligence was characterized by general, innate abilities to solve novel problems. Individuals with high fluid mental abilities could solve problems that they were not familiar with or had not encountered before without much difficulty. They would be able to identify and discriminate relations between objects and abstract elements better than those with lower fluid abilities. This mental ability increased with age in childhood and after adolescence or early adulthood, this ability would begin to slowly decline. Fluid intelligence was suggested as the ability responsible for the inter-correlations among mental ability tests or Spearman's general factor, *g* (Cattell, 1943).

Crystallized intelligence was associated with the capacity to solve problems and discriminate relations between physical or abstract elements in specialized or familiar fields. The ability to solve problems using prior accumulated knowledge was referred to as crystallized mental abilities. Cattell (1943) proposed that individuals depended on their fluid mental abilities to accumulate knowledge and gain information to build their crystallized intelligence. Once the abilities became crystallized, they would be relied upon more than fluid mental abilities except in novel situations where the individuals could no longer rely on their crystallized abilities (Cattell, 1943).

Since the introduction of the fluid-crystallized intelligence model proposed by Cattell in 1943, it has been widely used and accepted as the dominant theory of structure of human intellect in psychometric tradition (Johnson & Bouchard, 2005). Although the theory of fluid and crystallized intelligence was widely accepted, other theories have been suggested to describe the structure of human intellect. Two of the more prominent theories were Vernon's verbal-perceptual model and Carroll's three-stratum theory. In their 2005 publication, Johnson and Bouchard compared the three models – Cattell's, Vernon's and Carroll's – in an analysis of structural equation modeling using data from 42 mental ability tests (Johnson & Bouchard, 2005). The authors challenged that the theories have not been subjected to modern confirmatory factor analysis for deeper empirical scrutiny, and yet these theories have been the building blocks and underlying assumptions of much research in various fields such as neuroscience, epidemiology, behavior genetics, cognitive psychology and aging. They argued that it was crucial for these psychometric models of human intelligence to be examined more rigorously so that these models could provide a more parsimonious and objective framework with high construct and predictive validities to describe individual differences in intellectual abilities (Johnson & Bouchard, 2005).

Vernon emphasized the existence of a general factor or intelligence governing other mental abilities, and he also suggested that after the general factor was removed from the inter-correlations of a set of mental abilities test, the remaining variance in the set of tests could be categorized into two different abilities. Vernon identified the two categories as the *v:ed* for verbal and educational abilities and *k:m* for analytical, spatial and mechanical abilities. He also noted that the *v:ed* factor was well correlated with verbal fluency, vocabulary capacity, divergent thinking and numerical abilities, while the *k:m* dimension was highly associated with perceptual speed, spatial manipulation and mechanical abilities as well as psychomotor and physical abilities (Vernon, 1965).

Based on his exploratory factor analysis of 460 data sets, Carroll developed a three-stratum theory to explain the structure of human intellect (Johnson & Bouchard, 2005). In his third stratum or level in his model was the general factor of intelligence, *g*. The second level of his model consisted of broad mental abilities such as long-term retrieval, short-term memory, processing speed, auditory processing, visual-spatial thinking, comprehension-knowledge, fluid reasoning and mathematics. The first level thus consisted of specific abilities or skills that were narrow behavior manifestations of the broader mental abilities in the second level. Examples of first level abilities included picture recognition, vocabulary and calculation (Carroll, 2003). Many would agree that Carroll's theory was considered as a synthesis of the many ideas on human intellect over the last century (Johnson & Bouchard, 2005).

Johnson and Bouchard (2005) collected data from 436 participants from the Minnesota Study of Twins Reared Apart (MISTRA) who completed 42 mental abilities tests that were obtained from the Comprehensive Ability Battery (CAB) developed by

Hakstian and Cattell (1975), the Hawaii Battery that included the Raven's Progressive Matrices (DeFries et al., 1974) and the Wechsler Adult Intelligence Scale (WAIS). The data collected from this sample was used to test each of the three psychometric models using structural equation modeling techniques. The authors reported that each model suggested the existence of a general intelligence, *g*, and they all provided reasonable descriptions of the data although none met their established fit criteria. Of the three models, Vernon's verbal-perceptual model seemed to have the best fit. In the course of trying to create a model with the best fit criteria, Johnson and Bouchard observed some persistent negative factor loadings from the verbal and spatial ability tests. Further analysis revealed a third factor that captured tests involving spatial skills especially three-dimensional manipulation tests. The resulting model led the authors to propose a model of the structure of human intellect that suggested a general factor, *g*, that influenced verbal, perceptual and mental rotation abilities (VPR) (Johnson & Bouchard, 2005).

Although the literature on psychometric intelligence seems chaotic with the various theories and models of intelligence presented over the century, a hierarchical model of human intelligence has recently started to dominate the landscape of the psychometric intelligence literature (Deary, 2001). The hierarchical model of human intelligence has the unitary general factor, *g*, at the top stratum and group factors generally characterized by broad mental abilities on the next stratum after the general factor. The lower stratum contains specific mental abilities that contribute unique variances to the higher order broad cognitive abilities. The three-strata model of human intellectual abilities proposed by Carroll in 1993 (Carroll, 2003) as mentioned earlier, and the practice of structural equation modeling in the field have been providing converging support that human intelligence could be described in terms of a hierarchical structure. The VPR model proposed and subjected to constructive replication (Johnson & Bouchard, 2005a; Johnson & Bouchard, 2005b; Johnson, Nijenhaus & Bouchard, 2007) offers an excellent example of the current direction and general acceptance of a hierarchical organization of human intelligence.

The fluid-crystallized intelligence model proposed by Cattell (1943) and further elaborated by Horn (1976) was the dominant theory of human intelligence since the 1970s (Johnson & Bouchard, 2005) before the hierarchical model was introduced in the mid to late 1990s (Deary, 2001). Although the theory proposed by Cattell and Horn did not support the unitary general factor of intelligence (Deary, 2001), the distinction between fluid and crystallized abilities inherent in Cattell and Horn's theory has played a critical role in understanding and describing individual differences in human intelligence despite the theory not being well supported by analysis of structural equation modeling (Johnson & Bouchard, 2005).

Johnson and Bouchard (2005) have acknowledged that their VPR model based on the Vernon model lacked an obvious memory factor. Although they found that the Vernon model had a better fit compared to the fluid-crystallized model and Carroll's three-strata model, they were able to improve Vernon's model by adding a separate memory factor to the initial model of just verbal and perceptual factors. However, as described earlier, Johnson and Bouchard (2005) identified some quirks in their analysis that led them to drop the memory factor they recognized as content memory (shaded in Figure 2) to a lower stratum loading on the perceptual factor and added a third factor, mental rotation, to the third-stratum together with the verbal and perceptual factors. All three factors – verbal, perceptual and mental rotation – were observed to load very highly to a higher-order factor, *g* (Figure 2).

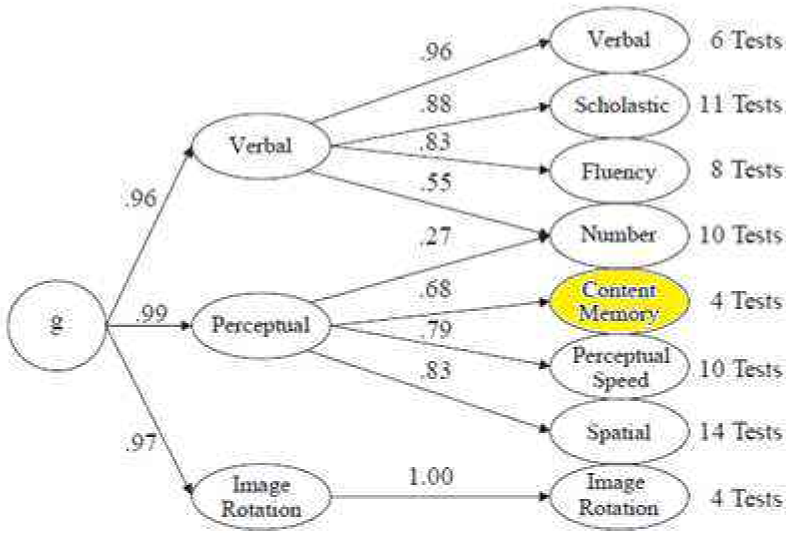

Figure 2: The VPR (verbal, perceptual and mental rotation) model postulated by Johnson and Bouchard (2005). Figure replicated from Johnson and Bouchard (2005).

It has been noted that memory assessment has played an integral role in the measurement of human intelligence. However, most of the tests used to measure memory were rote memory tests that measured one's memory span or short term memory (Ackerman, Beier and Boyle, 2002). Short term memory, or STM, has been shown in latent variable analyses to have no unique relationship with general intelligence, *g* (Engle et al., 1999; Conway et al., 2002). In latent variable analyses, working memory was shown to be a good predictor of *g*. Results from these latent variable analyses provide support and explanation to the high correlations between working memory and general intelligence established in the literature (Conway, Kane & Engle, 2003; Ackerman, Beier & Boyle, 2005).

The memory tests administered in the Johnson and Bouchard's (2005) study were Memory Span, Associative Memory, Meaningful Memory, Immediate Visual Memory and Delayed Visual Memory. The names of these tests suggested that they tapped into one's short-term memory more than one's working memory capacity. Indeed, the brief descriptions of each of the memory tests provided by the authors confirm the recall and rote memorization nature of the tests. Since short-term memory, STM, was not a predictor of *g* as suggested by latent variable analyses (Engle et al., 1999; Conway et al., 2002), it made sense that the memory tests in Johnson and Bouchard's (2005) study did not produce a separate third-stratum factor that explained the structure of human intellect.

The latent variable analyses conducted by Engle and his colleagues (1999) and Conway and his team (2002) strongly suggest that working memory capacity (WMC) plays an important role in predicting general intelligence. In their framework of analysis, WMC was understood as the ability to maintain active representations of information in one's mind in the presence of interference. Representations of task-relevant information could be held in an active state because the process was governed by "executive attention". Although active maintenance of representations could happen and be useful in many instances, controlled attention only came to play in situations rich with interference. This was because task-relevant information could be easily retrieved from long-term memory only in interference-free situation, whereas executive attention must be employed to maintain task-relevant information in an active state when there was high likelihood of retrieving incorrect information and there was a need to inhibit responses in situations with interference (Kane & Engle, 2002). Therefore, WMC was conceptually defined as domain-general executive attention which manages active representations of stored domain-specific information. This conceptual definition was consistent with the working memory model proposed by Baddeley and Hitch (1974). They postulated that there are three components of working memory model (Baddeley, 2000). There are two “slave systems” – the phonological loop and the visuo-spatial sketch pad – that act like short-term memory storage. The phonological loop maintains phonological information and prevents the information from decaying through rehearsal. The visuo-spatial sketchpad stores visual and spatial information. The third component, the central executive, coordinates the information from the two slave systems by directing attention to relevant information and suppressing irrelevant information. In 2000, Baddeley added a fourth component as a third slave system, the episodic buffer, to integrate information from the other two slave systems and information from long-term memory if necessary (Baddeley, 2000).

Given the important contributions of WMC in predicting general intelligence, working memory may be another group factor in addition to the verbal, perceptual and mental rotation group factors proposed by Johnson and Bouchard (2005). Deary (2001) reported a hierarchical model based on all the subtests in the Wechsler Adult Intelligence Scale-III (WAIS-III), and his structural equation modeling results suggested four group factors driven by a general factor, *g* (Figure 3). These four factors were verbal comprehension, perceptual organization, working memory and processing speed, and they bore much resemblance to an extension of Johnson and Bouchard’s VPR model. The perceptual organization and processing speed group factors in Deary’s (2001) model corresponded to Johnson and Bouchard’s (2005) mental rotation and perceptual factors. The addition of working memory as a separate group factor could produce a more stable hierarchical model of human intellect that better described the variances of mental abilities contributing to general intelligence, g.

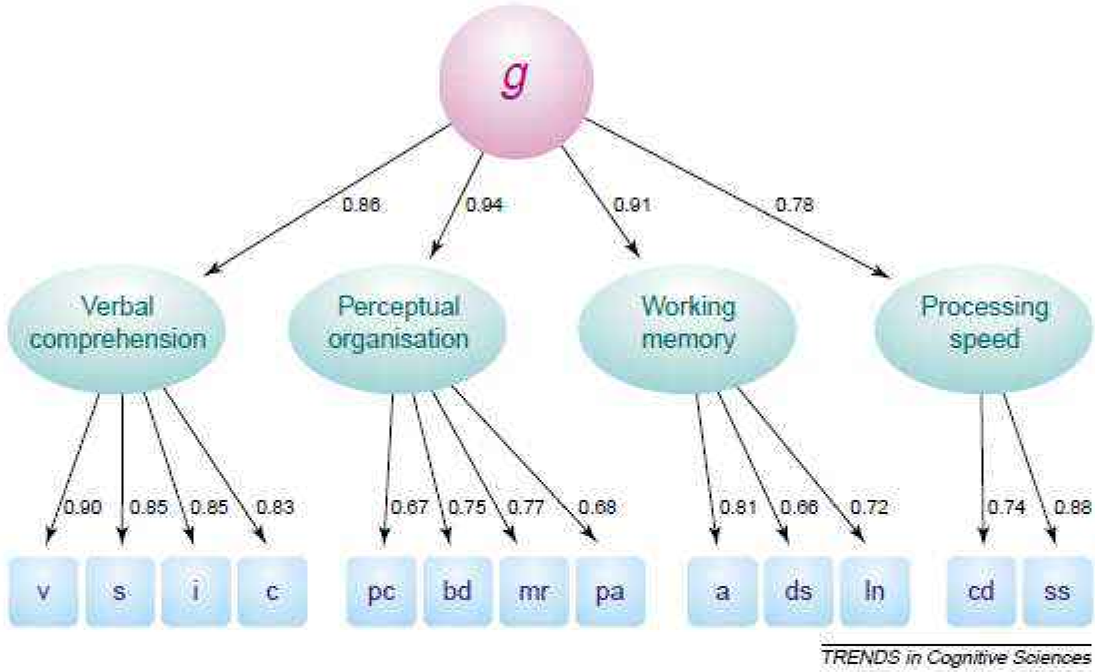

Figure 3: A hypothesis of the hierarchical structure of mental abilities obtained from structural equation modeling on all subtests from Wechsler Adult Intelligence Scale-III (replicated from Deary, 2001). Abbreviations: v=vocabulary; s=similarities; i=information; c=comprehension; pc=picture completion; bd=block design; mr=matrix reasoning; pa=picture arrangement; a=arithmetic; ds=digit span; ln=letter-number sequencing; cd=digit-symbol coding; ss=symbol search.

The latent variable analyses conducted by Engle et al. (1999) and Conway et al. (2002) suggested that STM is a subset of WMC as supported by the definition of WMC which includes both executive functions and the storage of information. In their analyses, after they had removed the shared variance between STM and WMC, the remaining variance in WMC was still a significant predictor of general intelligence, *g*. Engle and his colleagues (1999) reported a correlation of 0.49 ($p<0.05$) between *g* and the unique variance of WMC independent of STM. This observation led them to suggest that the residual in WMC after short-term memory has been parsed out reflected executive attention, which was what predicted *g* in mental abilities tests.

There were other studies in cognitive psychology and neuroscience that tried to explain the relationship between working memory (WM) and intelligence. Halford, Cowan and Andrews (2007) suggested that WM and reasoning shared related capacity limits, and that the common thread between the two functions was the shared requirement to bind elements to slots of a hypothetical coordinate system in one's memory. The process of maintaining the bindings between elements required attention that was essential to WM and reasoning abilities. The authors believed that "the common demand for attention when binding elements into slots is a possible explanation for common capacity limitations in WM and reasoning" (p. 236). These limits might be based on a restrained capacity to form and preserve bindings of elements in memory. Research in the field of neuroscience has suggested that attention is a function of the prefrontal cortex. Gray, Chabris and Braver (2003) suggested that the relation between fluid intelligence (g_f) and working memory was mediated by activities in the lateral prefrontal and parietal regions. Kane and Engle (2002) reported that the dorsolateral prefrontal cortex could have a role in working memory especially related to attention control. Conway, Kane and Engle (2003) supported the hypothesis by Gray and colleagues (2003) that working memory span tasks activate regions in the prefrontal cortex when the executive-control mechanism is recruited to combat interference during the maintenance and manipulation of information.

The assumption that working memory and intelligence may share the same neural network and mental resources could be the grounds that Jaeggi and her colleagues (2008) operated on when they hypothesized that training to improve one's working memory could transfer the improvement to intelligence. The adaptive nature of the dual N-back task used for training was intended to engage the executive attention at all times so that automatic responses could not develop. It was suggested that under consistent format and information conditions, practice would lead to automatic responses and less mental or attentional resources would be employed. However, in variable information or inconsistent contexts, controlled processing that utilized mental resources would still have been taken place even after practicing on the task for a considerable period of time (Ackerman, 1987). Jaeggi and her colleagues (2008) were essentially targeting participants' executive attention in their working memory training with the implementation of the adaptive feature of the training task. Therefore, an increase in working memory performance at the end of training could mean an increase in attention span.

If working memory was related to general intelligence as conceptualized from the model presented by Deary (2001) in Figure 3, and if an increase in working memory capacity (WMC) could lead to improved *g*, it seemed reasonable to postulate that increasing verbal performance, perceptual skills and mental rotation abilities by way of training could all potentially lead to improvements in *g*. In this scenario, each factor would have uniquely contributed to the multi-faceted construct of *g*. The studies mentioned above suggested that the WMC factor could have specifically influenced the attentional component in *g*, while there was not as much support for unique contributions of verbal, perceptual and mental rotation factors to the expression of *g* (Figure 4).

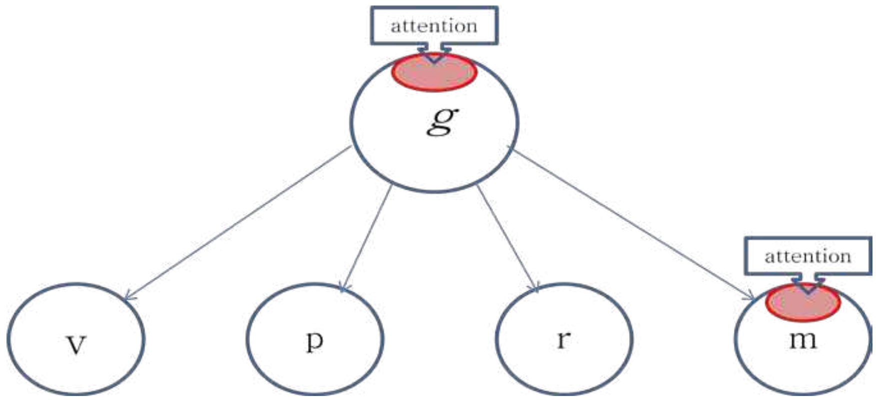

Figure 4: An illustration of how working memory capacity (WMC) labeled as factor **m** in the figure could influence g.

However, the strength of association between WMC and attention has not been successfully quantified, and this association may or may not correspond to the relationship between WMC and *g* as reported in studies of individual differences (Kane, Hambrick and Conway, 2005). Since attention has been proposed to be crucial in linking WMC and reasoning ability or analytic intelligence, and the field of cognitive psychology has not been able to measure attention, perhaps there are other ways to detect the processes of attention. For example, the natural forces of the wind are an experienced phenomenon, and it is not tangible or measurable but its effects can definitely be felt – its activities cause us to experience the breeze in the air and see the leaves rustled. Meteorologists are now able to predict and measure the speed of wind by manipulating its effects on physical matter using an anemometer. Carpenter and her colleagues (1990) suggested a theoretical account of the cognitive processing in one's mind as one solved the problems in Raven's Progressive Matrices. The test is made up of 3x3 matrices of geometric figures arranged in a pattern. The last entry of the matrix is left blank and the goal of the test is to determine the correct entry from a number of possible options given at the bottom of each matrix. Carpenter et al. (1990) reported that the high scoring subjects differed from low scoring subjects by the former's ability to generate and maintain problem solving goals in working memory. The problems in the Raven's Progressive Matrices were designed in such a way that later problems were more difficult than the earlier ones – the level of difficulty increases as the subject progresses through the test. The level of difficulties can be defined by the number of rules that form the arrangement of the matrices. Those who perform well on the Raven's and other analogical reasoning tasks appear to be able to identify steps to solving the problems and keeping them active in their mind as they incrementally process the steps to come to a solution. The authors proposed that it was a "common ability to decompose problems into manageable segments and iterate through them, the differential ability to manage the hierarchy of goals and sub-goals generated by this problem decomposition and the differential ability to form higher level abstraction" (Carpenter et al., 1990, p.429).

If the effects of WMC could be quantified through the analysis of goal management, an increase in WMC could mean better and improved ability to manage representations of information. These effects could be measured by improved performance on fluid ability tests, such as the Raven's, and mental rotation tests where steps of abstractly manipulating the movements of three-dimensional objects must be actively managed in one's mind.

A replication and extension study of Jaeggi et al. (2008) measuring the change of general intelligence as modeled by Johnson and Bouchard (2005) might produce substantial improvements in mental rotation abilities more so than verbal performance and perceptual skills after working memory training. Participants in the famous Shepard and Metzler (1971) study reported that they would imagine rotating the two-dimensional picture of a three-dimensional object in their mind in order to compare if the pictures of the 3-D objects were congruent with one another. The imagined process seemed to resemble the activity of an active representation of information that required executive attention to mentally rotate an imaginary 3-D object. Although the process did not seem to be loaded with interference, the task relevant information did not seem to be easily retrieved from long-term memory. Conway and his colleagues (2002) noted in their latent variable analysis that processing speed did not significantly predict *g*. Given that processing speed loaded quite highly on the perceptual factor in the VPR model (Johnson & Bouchard, 2005) and measures of processing speed did not seem to employ attentional resources to manage abstract representations, an increase in WMC should not affect perceptual abilities. If the verbal performance factor measured by vocabulary tests is a result of accumulated knowledge due to educational experiences, increased WMC should not influence verbal performance.

The effects of increased WMC after training could also be observed in changes on fluid intelligence performance as measured by Raven's Advanced Progressive Matrices, or RAPM (Raven, 1990). Questions in the RAPM consist of matrices of figures arranged based on a pattern or several rules of patterns depending on the level of difficulty, thus the test essentially measured one's reasoning ability in identifying the patterns and deducing the answer based on the patterns identified. As the number of rules and complexity of patterns increases with the level of difficulty, one would have to maintain the discovered patterns while searching for more clues and rules to solve the problem (Carpenter, 1990). The active searching and maintenance of task-relevant information mirrors a general working memory span task; hence, an increase in WMC would most likely benefit one's reasoning abilities assessed through fluid intelligence tests like the RAPM.

Viewed from a purely theoretical standpoint of human intelligence, since WMC is a strong predictor of *g*, specifically fluid intelligence, g_f (Engle et al., 1999; Conway et al., 2002), mental abilities that are directly affected by fluid abilities, such as reasoning ability, mental rotation or spatial manipulation tasks, perceptual organization and processing speed may be most likely affected by increased in WMC. However, the lack of support that processing speed significantly predicted *g* in latent variable analysis (Conway et al., 2002) suggests that the effects of working memory training may target a higher level of cognitive processing and not lower level of cognitive processing such as processing speed. Crystallized abilities such as verbal knowledge may not be affected by changes in WMC. These abilities are based on learned and accumulated knowledge over years of education and are likely stored in long term memory and thus, the information need not be actively held and maintained in one's working memory for successful completion of these tasks.

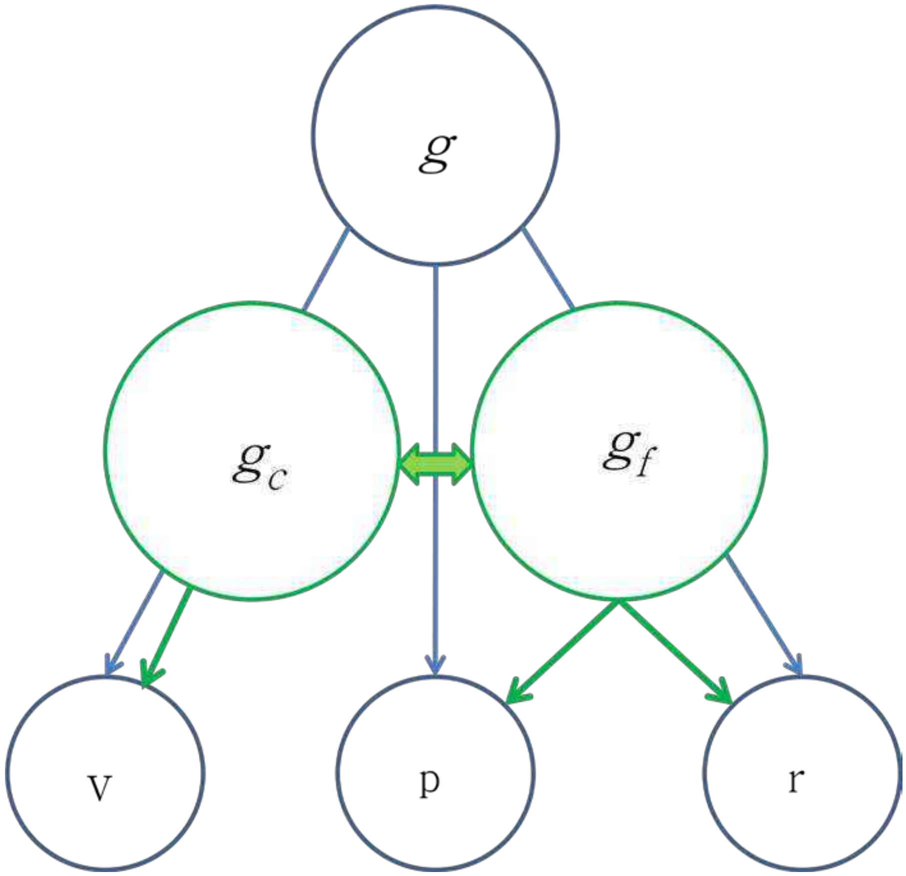

Figure 5: An illustration of the hierarchical of mental abilities that include Cattell and Horn's fluid-crystallized intelligence theory and Johnson and Bouchard's VPR model.

In short, the hypothesis of the current study is that working memory (WM) training increases one's working memory capacity (WMC). As a result, the effects of an increase in WMC can be observed through performance on a battery of tests that measure g as defined by the VPR model (Johnson & Bourchard, 2005) and fluid intelligence, g_f (Cattell, 1943). Specifically,

1. Mental rotation abilities should improve after WM training.
2. Perceptual speed should not be influenced by WM training.
3. Verbal performance should not be affected by WM training.

4. Fluid intelligence abilities such as reasoning ability should improve after WM training.

Methods

The present study aims to replicate and extend the findings reported by Jaeggi et al. (2008). There will be four experimental groups exposed to different training durations on the same working memory task – the dual N-back task and two control groups. Every participant will complete a pre- and post-test session before and after training to assess '*g*' and specific cognitive abilities with the same test battery. A theoretically based, comprehensive battery of cognitive ability tests will be used to measure intelligence to provide insight into specific mental processes influenced and NOT influenced by the working memory training.

The present study will use the RAPM to assess general intelligence in all participants. The standard untimed administration will allow the participants the opportunity to answer the most difficult items on the test. Odd numbered items will be used for the pretest or baseline assessment, and even numbered items will be used for the posttest. If posttest RAPM scores are significantly higher in any or all of the experimental groups when compared to the control groups' posttest scores, then training can be held accountable for the improvement.

If training is shown to improve performance on the RAPM, without the addition of a broad battery of tests, the source of the improvement will not be detectable. As Jaeggi et al. (2008) state, "… tasks that measure G_f are picking up other cognitive skills as well, and perhaps the training is having an effect on these skills even if measures of capacity are not sensitive to them." Johnson and Bouchard (2005) suggest a model of general intelligence, *g*, which is influenced by three factors or dimensions – verbal, perceptual and mental rotation. If WM training can improve RAPM performance, the improvement may be due to global effects or specific effects. The present study will include a series of tests that support the model postulated by Johnson and Bouchard (2005). None of the tests used will possess any similar features with the working memory task used for training.

The present study will incorporate two separate control groups to control for simple practice effects and/or the Hawthorne effect (increased motivation simply because of the attention paid to the participants' task performance). The first control group in the present study will control for simple practice effects as did Jaeggi et al.(2008) The second control group will be exposed to the same experimental format as the first control group with the addition of a filler task designed to be comparable to the training task without its complexity. Participants assigned to the second control group will train for the same amount of time as their counterparts in the experimental training group. The control group filler task has the same modality as the training task but without the adaptive feature inherent in the training task that purportedly contributed to the training effects. The filler task will be a fixed dual 1-back task modified from the adaptive dual N-back task to be used in the experimental group. The adaptive nature of the training task is alleged by Jaeggi et al. to contribute to the improvement of working memory capacity and performance in intellectual abilities, so this allegation could be verified if there are significant differences between the experimental group and both of the control groups.

Participants will be recruited from students enrolled in General Psychology a course at Case Western Reserve University (CWRU). Students can earn research credits to fulfill course requirements through participation in experiments conducted in the

Department of Psychology. Participants will be recruited through an online system that advertises all experiments available for the semester and class announcements.

Participants will be randomly assigned into 3 groups – 1 experimental group with 2 conditions that will receive working memory training and 2 control groups – with approximately equal numbers of participants in each group. Within the experimental group, there will be two training conditions, the amount of training received: 8 days and 20 days of working memory training. These training conditions differ slightly from the original study (Jaeggi et al., 2008), in which the four training durations were 8 days, 12 days, 17 days and 19 days. In the original study, there were no significant improvement in the 8-day condition but significant changes were detected in the 17-day and 19-day conditions; therefore the researchers claimed that improvement were dosage-related – more training led to more gains (Jaeggi et al., 2008). The current study will only compare only two conditions that produced contradictory results, which are the 8-day and 20-day training period.

In the present study, one of the control groups will train on the same filler task for the entire training duration, while the other control group will only participate in the pre- and post-test sessions. The filler task is designed to involve the same modalities employed in the WM training task but with the complexity and adaptive nature of the training task removed. Participants in the filler task control group and WM training group assigned to the 8-day training duration will be trained for 8 times within 2 weeks, and those assigned to the 20-day training durations will be trained 20 times in 5 weeks. At the end of each training duration, the participants in the WM training group and both control groups will complete the post-test, which consists of the same battery of tests in the pre-test session but using different sections or questions of the tests.

The tests administered in the pre- and post-tests include the Mill-Hill vocabulary test, vocabulary tests (part I and II) from the Primary Mental Abilities test battery, Word Beginning and Ending test, Colorado Perceptual Speed Test, Identical Pictures, Finding A's, Card Rotation and Paper Folding from the ETS test battery, Shepard-Metzler Mental Rotation Test (1971) and Raven's Advanced Progressive Matrices (1990).

The **Mill-Hill** vocabulary test is a multiple choice test where subjects choose the option word that is synonymous to the target word.

The **Vocabulary** tests consist of target words that have similar meanings to one of the four words given as options. Test-takers must identify the correct answer by circling the word that are synonymous or has the closest meaning to the target word.

In the **Word Beginning and Ending** test, subjects were asked to generate as many words as they could that begin and end with letters specified by the test.

The **Colorado Perceptual Speed Test** (multiple-choice) is a test where subjects choose the exact copy of the target group of characters, for instance, 'vgef' or '9c6d'.

Identical Pictures shares a similar format but instead of characters as target stimuli, this test utilizes pictorial objects as target.

The **Finding A's** test consists of multiple columns of words, and subjects are asked to identify and cross out words that contain the letter 'a'.

In **Card Rotations**, a target shape is presented next to eight versions of the target. These versions are either rotated or flipped or both. Subjects must identify if each of the eight rotated shapes is on the same side or the flipped side of the target shape.

In **Paper Folding** (multiple choice), an illustrated piece of paper is shown to be folded in certain ways. After the paper is folded, a hole is punched at a certain location. Subjects must then visualize where the holes will be after the piece of paper is unfolded back to its original condition.

The **Shepard-Metzler Mental Rotation** is a visualization test of mental spatial manipulation. Participants are asked to compare images of rotated 3-dimensional blocks to a target figure, and identify the two images of the target figure that are rotated at different angles.

There are 36 questions in the **Raven's Progressive Advanced Matrices** (**RAPM**) and the difficulty level increases for subsequent questions; hence the descriptor "progressive" in its name. Each question consists of a 3x3 matrix with the last entry of the matrix left blank. The matrices are formed based on a certain pattern, and the blank spot could be filled in with one of the eight options given once the pattern is identified.

All tests except for the Mill-Hill and RAPM are timed. Participants will be given 90 seconds to complete as many items as they can in the Vocabulary tests, Word Beginning and Ending Test, Card Rotation, Paper Folding and Shepard-Metzler Mental Rotation tests. They will be asked to complete as many problems as they can in 30 seconds for the Colorado Perceptual Speed Test and 45 seconds for the Identical Pictures test. Part I and II of these tests will be administered in the pre- and post-tests respectively.

The odd number questions of the RAPM are administered in the pre-test and the even number questions in the post-test.

A WM span task is included in the pre- and post-test sessions. The Operation Span (OSPAN) will be administered in group format using Microsoft Power Point that was designed by Shelton et al. (2007). This task requires participants to evaluate simple mathematical equation and remember a word that appears after the equation. There are four to seven equations and to-be-remembered (TBR) words for each trial. At the end of each trial, a slide will appear to prompt participants to list all the words shown in the previous trial. Each slide that contains the math equation and TBR word appears for 8 seconds and participants are given 15 seconds to list all the words. There will be a chimed reminder to warn participants that the next trial will begin in 5 seconds.

The training task employed in the present study is the dual n-back task (Jaeggi et al., 2008) administered on a computer. A series of squares is flashed on the screen one at a time on eight (8) different locations. Simultaneously, a series of letters will be presented via an audio output. Participants are asked to identify if the current visual and audio stimuli presented match the stimuli n items back. If the participants perform adequately in the current n-back task, i.e., they correctly identify matched stimuli in both modalities more than 90% of the time, the program will increase the difficulty level to n+1. However, if the participant's performance drops below 70% the program will reduce the difficulty level to n-1. The difficulty level remains unchanged if participants do not meet either condition. The ability of the task to adapt to users' performance so that it remains demanding in terms of mental resources to the user is purported to have increased working memory capacity that translates to improved measured fluid intelligence (Jaeggi et al., 2008).

The filler task is a modified web-based version of the training task – the dual n-back task – publicly available for web users. In the modified version, instead of the program adapting automatically to the user's performance, the user only works on a fixed level of difficulty for the task. In the study, participants in the secondary control group with a filler task complete 20 trials of the dual 1-back working memory task each time they come in for their "training" session.

Planned Statistical Analysis

The dataset will first be screened for outliers and basic descriptive statistics will be computed for all variables in the pre- and post-sessions. A full matrix of inter-correlations among all the test variables in the pre- and post-sessions will be reported. Changes in working memory performance for each participant in the experimental group will be calculated by comparing their baseline performance obtained at the first training session with their performance at the final training session.

Paired t-tests will be conducted on the mean difference between pre-test scores and mean post-test scores for the control and experimental groups. ANOVA will also be performed with the training conditions as between-group factor and gain scores (post-test score minus pre-test score) as the dependent variable to analyze it as a function of training dosage. Analysis of Covariance (ANCOVA) will be conducted with pre-test scores as covariates for the experimental and control groups.

Results

There were a total of 130 participants who came to the pre-test session but only 93 of them came back for the post-test session – 60 of whom were females. The following results were obtained from analyzing data provided from the 93 students who completed both pre- and post-tests.

Independent samples t-test analysis showed no difference in age and SAT scores for those who returned to post-test and those who did not ($t(110) = -1.63$, $p = 0.11$). Table 1 presents the demographic information of the participants in the study, and Table 2 breaks down the demographic information for each separate condition in the study.

	Mean	Std. Dev.
Age		
Pre-Test Only (n=37)	19.97	1.69
Pre- and Post-Test (n=93)	19.84	1.25
Mean SAT score		
Pre-Test Only (n=33)	1308	149.5
Pre- and Post-Test (n=79)	1352	122.5
Mean SAT-Math		
Pre-Test Only (n=19)	652	92.5
Pre- and Post-Test (n=53)	685	89.6
Mean SAT-Verbal Comprehension		
Pre-Test Only (n=18)	672	96.1
Pre- and Post-Test (n=49)	662	67.2

Table 1: Demographic information of the participants in the study

	Control		Filler		Training	
	2-week	5-week	2-week	5-week	2-week	5-week
N (females)	22 (16)	23 (9)	15 (6)	11 (11)	9 (8)	13 (10)
Native Speakers	19	17	11	8	7	10
Mean SAT scores	1335	1351	1334	1368	1327	1397
(SD)	(100.4)	(124.0)	(124.3)	(175.5)	(41.1)	(115.8)
	n=17	n=21	n=15	n=11	n=4	n=11

Table 2: Demographic information of participants broken down according to the group they were randomly assigned to. The number of participants who reported their SAT scores is included for each group.

The mean scores for the separate Math and Verbal Comprehension components on the SAT were not reported in Table 2 because more than half of the participants in each group did not disclose the information. One-way Analysis of Variance (ANOVA) did not show any significant differences in SAT mean scores among the six groups and independent sample t-test comparing scores from the 2-week against the 5-week condition also supported the results from one-way ANOVA ($F(5, 106) = 0.48$, $p=0.79$), which suggested that the average participant in each group behaved similarly.

Gender Differences

Some gender differences were observed. Male participants scored higher on the SAT-Math ($t(70) = 3.18$, $p<0.01$), and they performed significantly better than their female counterparts on the Mental Rotation ($t(128) = 3.33$, $p<0.01$ on pre-test and $t(90) = 3.04$, $p<0.01$ on post-test) and Card Rotation pre-test ($t(124) = 2.92$, $p<0.01$). Female participants performed significantly better on Word Beginning and Ending pre-test ($t(124) = -1.98$, $p=0.05$) and Finding A's ($t(128) = -3.81$, $p<0.01$ on pre-test and $t(92) = -2.87$, $p<0.01$ on post-test). These differences will be taken into consideration in subsequent analyses, particularly in Analysis of Variance where the interaction effects between gender and group will be explored.

Correlations

Table 3 presents a correlation matrix of all the test variables in the pre- and post-tests. The values in the diagonal are reliabilities for each test, in other words, correlations between Part 1 and 2 of the same test used in the pre- and post-test sessions respectively. Values above the diagonal are inter-correlations between variables in pre-test, and values below the diagonal are inter-correlations between variables in post-test. Means and standard deviations for each test variable are also included in the matrix – pre-test values in the upper triangle and post-test values in the lower triangle.

Working Memory Training

There were a total of 22 participants in the working memory training groups, 13 of whom trained for 20 days. The average of their performance on the training task and variance among participants for each session during the 8 and 20 days of training are presented in Figures 6 and 7 respectively. The figures suggest that most of the participants improved their performance on the task as they practiced on it, and the variance among all the participants tended to increase as the number of training periods increased. Some participants were able to make more improvements than others. The final training session and the first training session are highly correlated (r=0.79, $p<0.01$). Most students did not achieve their highest performance on the final training session so a variable that indicated each participant's highest training score was computed. This variable significantly correlated with performance on the final training session (r=0.98,

$p<0.01$) and their first training score (r=0.75, $p<0.01$). Another variable, which measured how much participants had improved over the course of their training, was created by subtracting the first initial training score from the highest training score. This variable, named WM-Improve, will be used in the Multiple Regression analysis to predict test variables for participants in the experimental group.

The percentage of improvement for each training condition was calculated by using the following formula:

$$\% \, Improvement = \frac{Avg. Final\ Training\ score - Avg. First\ Training\ score}{Avg. Final\ Training\ score} \times 100$$

Participants in the 2-week (8 days) condition had a 34% improvement and those in the 5-week (20 days) condition improved by 44%. Participants in the Jaeggi et al. (2008) study displayed similar trends. From the data that they published, it can be estimated that participants in the 8-day training condition improved by 34% and those in the 19-day condition improved by 47% (Jaeggi et al., 2008). The numbers suggested that participants from the current study and the original Jaeggi et al. (2008) study showed very similar performance on the training task.

Participants' first WM training score was significantly related to Card Rotation post-test (r=0.51; $p < 0.05$). Their highest WM training score (WM-High) and WM-Improve correlated significantly with Card Rotation post-test at 0.57 ($p < 0.01$) and 0.50 ($p < 0.05$) respectively. Similarly, these variables had significant correlations with RAPM post-test (WM-High: r=0.50, $p < 0.01$; WM-Improve: r=0.51, $p < 0.05$).

	Vocab	WordBE	CPST	Id. Pic	Find A's	Ppaer F	Card R	Mental R	RAPM	Mill-Hill	OSPAN	Mean	Std. Dev
Vocabulary	**.50****	.28**	.28**	.20*	.24**	.02	.07	.01	.06	.56**	.22**	47.81	11.87
Word BE	.28**	**.61****	.08	.17	.10	-.03	.12	-.02	.01	.26**	.20*	8.64	3.54
CPST	.20	.20*	**.49****	.40**	.21*	-.03	.04	.00	-.07	.02	.10	13.56	2.41
Identical Pic	.06	.15	.44**	**.55****	.33**	.05	.13	.06	-.01	-.12	.21*	20.94	4.54
Find A's	-.04	.11	.13	.24*	**.74****	.13	-.03	-.14	-.00	.11	-.00	15.73	3.73
Paper Folding	.23*	.14	.04	.23*	.26*	**.52****	.30**	.35**	.36**	-.03	.24**	4.97	2.28
Card Rotation	.22*	.30**	.29**	.25*	.10	.32**	**.42****	.36**	.41**	.07	.21*	29.48	14.72
Mental Rotation	.12	.01	.24*	.23*	-.03	.28*	.40**	**.39****	.32**	.06	.21*	2.68	1.77
RAPM	.05	.14	.05	.07	-.04	.04	.35**	.47**	**.55****	.17	.29**	12.54	2.65
Mill-Hill	.53**	.37**	.10	-.12	-.05	.15	.29**	.04	.18	**.63****	.17	18.22	4.05
OSPAN	.28*	.35**	.28**	.19	.21*	.01	.39**	.24*	.42*	.37**	**.66****	56.95	6.61
Mean	32.22	7.24	10.71	17.37	14.32	5.28	34.00	2.17	12.30	18.25	57.58		
Std. Deviation	14.84	2.78	2.14	4.14	2.93	1.64	11.42	1.59	2.55	3.33	6.24		

Table 3: Correlation matrix of all the pre- and post-test variables – information above the diagonal are correlations among all pre-test variables and information below the diagonal are correlations among post-test variables.

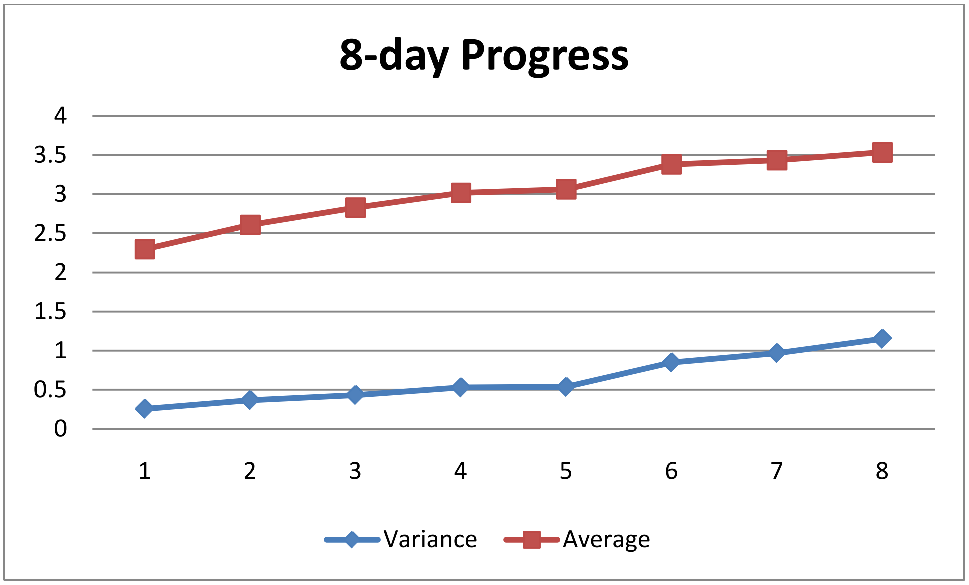

Figure 6: Average performance on the training task and variance among all the participants in the first 8 days of training.

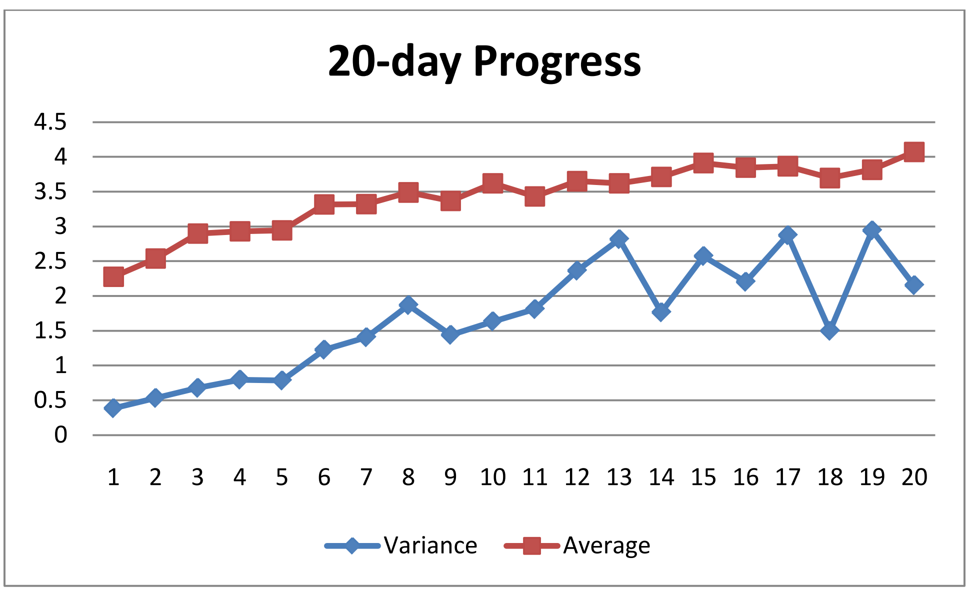

Figure 7: Average performance on the training task and variance among participants in the 20-day condition.

Paired T-tests

Paired t-test analyses revealed several trends. First, working memory capacity and intellectual ability as measured by Raven's Advanced Progressive Matrices (RAPM) did not show significant improvement on post-test performance compared to pre-test (Table 5). Second, there was also no significant improvement in the spatial ability tests (Table 6). Third, data reported in Tables 7 and 8 suggested worse performance on verbal fluency and perceptual speed tests; however, a comparison of data in a pilot study where data of all test variables in both pre- and post-tests was collected from the same administration instead of two showed similar trend. This trend seemed to indicate that part 2 of these tests, which were used in the post-test, were significantly more difficult than part 1 of these tests. The negative t-values for spatial ability tests such as paper folding and card rotation suggested improvement albeit insignificant. Positive t-values indicate negative gain (post-test scores lower than pre-test scores) and negative t-values indicate positive gain (post-test scores higher than pre-test scores).

	2-week		5-week	
	Paired T-Test	*p*	Paired T-Test	*p*
Working Memory task: OSPAN				
Training	$t(8) = -0.73$	.48	$t(12) = -0.53$	.60
Active Control (Filler task)	$t(12) = -0.12$	.91	$t(10) = 0.19$	.85
Passive Control	$t(21) = 0.03$	.97	$t(21) = -0.08$	.94
Matrix Reasoning test: RAPM				
Training	$t(8) = 0.16$	.88	$t(12) = 0.22$	.83
Active Control (Filler task)	$t(13) = 0.00$	1.0	$t(10) = 1.30$	.22
Passive Control	$t(19) = 2.12$	.05	$t(22) = 0.24$	.81

Table 4: Paired t-tests for working memory span task and matrix reasoning test. Negative t-values that indicated improvement are italicized.

	2-week		5-week	
Spatial Ability tests	Paired T-Test	*p*	Paired T-Test	*p*
Paper Folding				
Training	$t(8) = -1.44$	.19	$t(12) = -1.31$	.22
Active Control (Filler task)	$t(14) = 0.89$	.39	$t(10) = -1.37$	.20
Passive Control	$t(20) = 0.52$	.61	$t(23) = -0.36$	.72
Card Rotation				
Training	$t(8) = -2.78$	**.02**	$t(12) = 0.21$	.83
Active Control (Filler task)	$t(14) = -1.14$	.27	$t(10) = -1.28$	.23
Passive Control	$t(17) = -0.38$	.71	$t(23) = -2.14$	**.04**
Mental Rotation				
Training	$t(8) = 1.95$	.09	$t(12) = -0.27$	.79
Active Control (Filler task)	$t(14) = 2.04$	.06	$t(9) = 0.34$	.74
Passive Control	$t(21) = 0.13$	.89	$t(22) = 1.70$	.10

Table 5: Paired t-tests for spatial ability tests.

	2-week		5-week	
Verbal Fluency tests	Paired T-Test	*p*	Paired T-Test	*p*
Mill-Hill Vocabulary Test				
Training	$t(8) = -0.78$	.46	$t(12) = 0.84$	.42
Active Control (Filler task)	$t(13) = 1.83$	.09	$t(10) = 0.12$	.91
Passive Control	$t(18) = 0.66$	.52	$t(22) = -0.49$	.63
Vocabulary (ETS)				
Training	$t(8) = 1.81$	.11	$t(12) = 5.53$	**.00**
Active Control (Filler task)	$t(14) = 4.13$	**.00**	$t(10) = 7.00$	**.00**
Passive Control	$t(21) = 6.63$	**.00**	$t(21) = 5.29$	**.00**
Word Beginning & Ending				
Training	$t(8) = 0.46$	.65	$t(12) = 5.53$	**.01**
Active Control (Filler task)	$t(13) = 0.50$	.62	$t(10) = 1.73$	.11
Passive Control	$t(20) = 2.41$	**.03**	$t(23) = 2.35$	**.03**

Table 6: Paired t-tests for verbal fluency tests.

	2-week		5-week	
Perceptual Speed tests	Paired T-Test	*p*	Paired T-Test	*p*
Colorado Perceptual Speed Test				
Training	$t(8) = 1.84$	.10	$t(12) = 5.55$	**.00**
Active Control (Filler task)	$t(14) = 4.93$	**.00**	$t(8) = 4.20$	**.00**
Passive Control	$t(20) = 6.11$	**.00**	$t(23) = 6.79$	**.00**
Identical Pictures				
Training	$t(8) = 3.24$	**.01**	$t(12) = 2.27$	**.04**
Active Control (Filler task)	$t(14) = 2.06$	.06	$t(10) = 4.14$	**.00**
Passive Control	$t(20) = 3.35$	**.00**	$t(23) = 3.05$	**.01**
Finding As				
Training	$t(8) = 1.90$	.09	$t(12) = 0.80$	.44
Active Control (Filler task)	$t(14) = -0.34$	.74	$t(10) = 0.40$	.69
Passive Control	$t(21) = 4.35$	**.00**	$t(23) = 3.84$	**.00**

Table 7: Paired t-tests for perceptual speed tests.

ANOVA

Analysis of variance (ANOVA) was conducted using gain scores (post-test minus pre-test scores), and these analyses provided additional support to results obtained so far. Particularly, in the 2-week condition only the perceptual speed test Finding As, $F(2, 43) = 4.09$, $p=0.024$, had significant changes across the three groups. Upon further investigation, ANOVA planned comparisons suggested a curvilinear trend for Finding A's test (Figure 8), with participants in the active control group having the highest gain which is significantly different from the other two groups. In the 5-week condition, one-way ANOVA did not show any significant changes in gain scores for the test variables.

3x2 factorial ANOVA was conducted on the gain scores for each test variable, with the group (controls vs. experimental) and condition (8-day vs. 20-day) as independent variables. There were neither significant main effects nor interaction effects produced by the analysis.

One-way ANOVA was also conducted on all the pre-test variables before conducting an analysis of covariance (ANCOVA). This analysis showed that there were differences in initial level ability across groups for the Mental Rotation test, $F(5, 87) = 3.16$, $p=0.011$.

Two-way ANOVA was conducted to test for gender effects and their interaction with group effects for tests that displayed gender effects as reported earlier, such as Mental Rotation, Card Rotation and Finding A's. This analysis showed that there were no significant interaction effects between gender and group (Table 8). In the 5-week condition, there was one less degree of freedom for between groups because there were no male participants in the active control group.

	2-week		5-week	
	F	*p*	*F*	*p*
Mental Rotation	$F(2, 39) = .03$	.97	$F(1, 41) = .00$	.94
Card Rotation	$F(2, 36) = .08$	.92	$F(1, 42) = .33$	.57
Finding A's	$F(2, 40) = .06$	.94	$F(1, 42) = 1.06$	.31

Table 8: *F* and *p* values of interaction effects (gender*group) for tests that indicated gender differences.

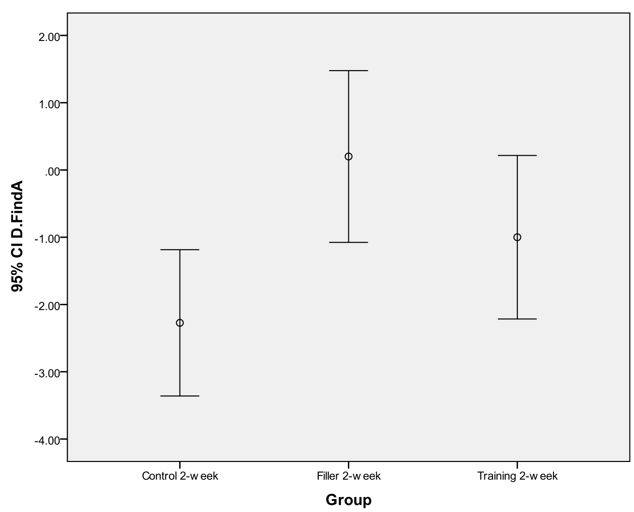

Figure 8: The curvilinear trend for gain scores in Finding As

ANCOVA

Analysis of covariance (ANCOVA) was conducted to determine if the outcome at post-test was influenced by one's initial ability as measured at pre-test. In the 2-week condition, all the pre-test variables significantly predicted the outcome on all corresponding post-test variables except for the Mental Rotation test. Performance on the Mental Rotation post-test was not influenced by performance at pre-test. A quick reference to the correlation matrix in Table 3 suggests a low test-retest reliability index for this test; in fact, it has the lowest reliability index among all the test variables. Only Paper Folding showed a significant difference across groups after controlling for pre-test performance, $F(2, 41) = 3.51$, $p < 0.05$, *partial* $\eta^2 = 0.15$ and Raven's $F(2, 39) = 3.89$, $p < 0.05$, *partial* $\eta^2 = 0.16$. In the 5-week condition, all the pre-test variables significantly predicted the outcome on all corresponding post-test variables except for the Card Rotation test, and this test had the second lowest test-retest reliability index according to Table 3.

Multiple Regression

Multiple regressions were conducted to explore the contribution of SAT scores used as a proxy for general intelligence (Frey and Detterman, 2004), to gain scores for all test variables. After controlling for general intelligence, the analyses investigated if working memory predicted any improvement. This was implemented by conducting a hierarchical regression that employed SAT scores and initial working memory (OSPAN scores at pre-test) to predict gain scores between pre- and post-test variables. This analysis was run separately with data from each of the six groups.

Regression analyses suggested that achievement on SAT contributed significant variance to only a few of the test variables. It affected gain scores in the Mill-Hill Vocabulary Test for participants in the 2-week active control group (R^2=0.36). SAT scores also affected gain scores in perceptual speed tests such as the Colorado Perceptual Speed Test for participants in the 5-week passive control (R^2=0.36) and experimental group (R^2=0.50), as well as the Finding A's test for those in the passive control group (R^2=0.30). SAT scores predicted significant variance in Card Rotation gain scores for participants in the 5-week passive control (R^2=0.26) and experimental groups (R^2=0.54).

After controlling for general intelligence, working memory contributed significant variance to tests such as Vocabulary in the 2-week active control group (R^2=0.37), Mental Rotation in the 2-week active (R^2=0.51) and 5-week passive (R^2=0.30) control groups, Mill-Hill (R^2=0.36) and OSPAN (R^2=0.31) in the 2-week passive control group and Identical Pictures in the 5-week passive control group (R^2=0.26).

A simple regression with improvement in N-back task as predictor was conducted using data from participants in the 5-week training group. This predictor did not significantly contribute to any variance in gain scores for all the variables in pre- and post-tests.

	2-week Passive Control						2-week Active Control					
	OSPAN (R^2=0.36)			Mill-Hill (R^2=0.31)			Mill-Hill (R^2=0.36)			Vocabulary (R^2=0.37)		
	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Step 1												
Constant	−29.33	19.02		2.64	12.55		20.97	9.88		−23.00	48.85	
SAT	.02	.01	.38	−.00	.01	−.07	−.02	.01	−.61*	.00	.04	.04
Step 2												
Constant	−6.19	17.70		−9.91	11.54		23.70	18.74		−161.06	71.76	
SAT	.02	.01	.42	−.00	.01	−.15	−.02	.01	−.62*	.02	.03	.15
OSPAN	−.47	.17	−.56*	.28	.11	.60*	−.04	.23	−.05	2.08	.89	.62*

Table 9: Multiple Regression analyses with gain scores as dependent variable. *$p<0.05$

	5-week Passive Control									5-week Training		
	Identical Pictures (R^2=0.26)			Finding A's (R^2=0.30)			CPST (R^2=0.36)			CPST (R^2=0.50)		
	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Step 1												
Constant	7.14	13.86		9.70	4.71		12.15	5.07		−21.47	5.96	
SAT	−.01	.01	−.18	−.01	.00	−.55*	−.01	.00	−.60*	.01	.00	.71*
Step 2												
Constant	−28.63	20.06		3.22	7.66		15.29	8.50		−18.15	7.18	
SAT	−.01	.01	−.16	−.01	.00	−.54*	−.01	.00	−.61*	.01	.00	.75*
OSPAN	.59	.26	.51*	.11	.10	.23	−.05	.11	−.10	.07	.08	−.21

Table 10: Multiple Regression analyses with gain scores as dependent variable. $^*p<0.05$

	5-week Passive Control						5-week Training			2-week Active Control		
	Mental Rotation (R^2=0.30)			Card Rotation (R^2=0.26)			Card Rotation (R^2=0.54)			Mental Rotation (R^2=0.51)		
	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Step 1												
Constant	.57	3.78		−80.31	38.82		−155.47	47.36		−2.28	7.24	
SAT	−.00	.00	−.09	.06	.03	.51*	.11	.03	.73*	.00	.00	.05
Step 2												
Constant	10.80	5.41		−53.33	64.98		−146.62	59.31		−26.07	9.43	
SAT	−.00	.00	−.10	.06	.03	.50*	.11	.04	.75*	.00	.00	.18
OSPAN	−.17	.07	−.55*	−.45	.85	−.12	−.19	.70	−.07	.36	.12	.72*

Table 11: Multiple Regression analyses with gain scores as dependent variable. *$p<0.05$

Discussion

There were 93 participants who completed the study, and two thirds of them were females. They averaged slightly younger than 20 years of age. There were more participants in the passive control group compared to the active control or experimental group because those assigned to the latter two groups were given the option to withdraw from the portion where they were asked to work on the filler or training task for 8 or 20 days. Most who withdrew were willing to return to complete the post-test, and these participants were assigned to the passive control group. Some gender differences were observed – male participants in the study achieved better SAT Math scores (Cohen's d = 0.76). They also performed better on Mental Rotation (Cohen's d = 0.60) and Card Rotation (Cohen's d = 0.52) tasks. Female participants, on the other hand, scored significantly higher than their male counterparts on Word Beginning and Ending tests (Cohen's d = 0.35). They also did better on the perceptual speed test Finding A's at both pre- and post-tests (Cohen's d = 0.67 and 0.60 respectively). In this test, participants went through long lists of words and crossed out words that have the letter A. It could be due to the language component of this test that females found significantly more words with the letter A than the males. This observation is supported by education studies where males are generally superior in math and spatial abilities, while females have stronger verbal abilities (Hyde, 1981). The size of mean difference between the gender groups suggests an overall medium effect (Cohen, 1988).

When pre-test measures were correlated, as expected, Vocabulary proficiency is significantly correlated with Word Beginning and Ending and the Mill-Hill Vocabulary Test, and they both had significant correlations with each other. All three perceptual speed tests – Colorado Perceptual Speed Test, Identical Pictures and Finding A's – were significantly correlated with each other. Spatial ability tests (Paper Folding, Card Rotation and Mental Rotation) significantly correlated with one another and also with the Raven's Advanced Progressive Matrices (RAPM). It seems reasonable that working memory as measured by the OSPAN significantly correlated with all spatial ability tests and the RAPM because these tests share the similar characteristics of requiring the test-taker to hold and manipulate representations of the information presented in each item in their mental workspace to find the solution. Vocabulary correlates significantly with all perceptual speed tests, perhaps since vocabulary is a measure of crystallized information and the ability to identify as many words as possible in a given time constraint taps into perceptual speed. Together with Word Beginning and Ending, they were considerably related to working memory probably due to shared verbal content among the tests.

Correlations among post-test variables showed a similar pattern to pre-test variables, where all verbal fluency, perceptual speed and spatial ability tests correlated amongst themselves significantly. Although training may have influenced the means of the test variables, the correlations among the test variables should not be affected by training. At post-test, Paper Folding did not correlate with RAPM or OSPAN, but together with Card Rotation correlated with Vocabulary. OSPAN correlated with all test variables except for Paper Folding and Identical Pictures.

Results from dependent sample t-tests suggested no significant improvement overall after training. There were significant declines in performance on some of the verbal fluency and all of the perceptual speed tests, but this trend is consistent with a previous study suggesting that the items used in the post-test are potentially more difficult than the items used in the pre-test. Looking at the results for participants who trained for 2 weeks, they did not show any significant decline in performance, which may suggest that they actually improved on speed after training. These participants also improved significantly on the Card Rotation test, and they showed some improvement on the Paper Folding test even though it was not significant. The same cannot be said of those who trained for 5 weeks. They showed significant declines in performance on the verbal fluency and perceptual speed tests except for Finding As. They did show improvement on Paper Folding and Mental Rotation albeit insignificant. Interestingly, those who did not train took the post-test 5 weeks after pre-test improved significantly on Card Rotation.

In the passive control group for the 5-week condition, SAT predicted gain scores on Finding As and Colorado Perceptual Speed Test in the negative direction. Reduction in SAT by slightly more than half a standard deviation (Table 9) increased gain scores in perceptual speed tests by one standard deviation, which indicated that higher SAT scores reduced the variability of the difference between pre- and post-test scores on these tests. Dependent sample t-test analyses showed that in general, participants performed significantly worse on the post-test; yet, multiple regression suggested that performance on part 2 of these tests did not deteriorate as badly for those who achieved higher SAT scores. However, this trend was reversed for participants who trained their working memory. Higher SAT scores seemed to increase the variability between pre- and post-tests scores; however, it is impossible to discern whether working memory training reduced speed of processing or if the effect was just an artifact.

SAT scores also predicted variance in Card Rotation, and the increase of one standard deviation in gain score was predicted by a half standard deviation in the passive control and about three quarters of a standard deviation in the training group. After controlling for variance attributed to SAT, working memory was seen to have significant influence on Vocabulary and Mental Rotation for participants in the 2-week active control groups. OSPAN predicted a reverse trend for OSPAN in 2-week passive control group as well as Identical Pictures and Mental Rotation in 5-week passive control group.

One of the limitations in the study is the small sample size, which may have lowered the power of analyses conducted. However, when Jaeggi and her colleagues (2010) repeated the study with 25 students who trained on the Raven's Advanced Progressive Matrices (RAPM) for 20 days, they obtained an effect size (Cohen's *d*) of 0.98. With such a large effect size, the analysis of paired t-test could achieve a power of 0.80 with 10-12 participants. Except for the group that trained for 8 days (n=9), all the other groups had at least 10 participants.

Another limitation is that the group administered OSPAN may not have good discriminating validity because every participant is given the same amount of time (8 seconds) to evaluate the math equation and remember the word presented. Participants who were better and quicker at math would have had more time to remember and rehearse the words presented in the list, so these participants would probably have scored higher on the OSPAN.

Results from the current study did not suggest improvement in general intelligence, specifically fluid intelligence as measured by RAPM, after repeated training on a challenging working memory task. This observation is in contrast to the conclusion published by Jaeggi and her colleagues in 2008 and replicated recently in 2010. Participants in the current study showed no significant improvement in solving items on the RAPM even after training for 20 days. Participants in the current study and those in the studies conducted by Jaeggi et al., (2008, 2010) took the test under different administration procedures. RAPM was administered with no time constraint in the current study as recommended by the test provider, so participants were allowed to solve as many items as they could under no time pressure. Jaeggi and her colleagues administered their transfer tasks, the RAPM and BOMAT, with a time constraint – participants in their studies only had 10 minutes to solve as many items as they could (Jaeggi et al., 2008). In their first study, those in the 19-day training group answered about 4 more items on the BOMAT correctly at post-test (Jaeggi et al., 2008) and in their second study, the 20-day training group correctly answered 3 additional items in 16 minutes at post-test (Jaeggi et al., 2010). In their replication study, participants answered 2 additional items on the RAPM in 11 minutes after training for 20 days (Jaeggi et al., 2010). There was inconsistent usage of transfer tasks in the original study, where the researchers used the RAPM in the 8-day condition and not in the other training conditions. Participants who trained for 8-days showed no significant improvement on the RAPM at post-test (Jaeggi et al., 2008).

Upon closer inspection of the data reported by Jaeggi et al. in 2008, the authors collapsed post-test scores for all training groups and concluded a significant improvement in performance on intelligence tests after training based on an increase of about 2.5 points.

This is misleading and inappropriate since not all participants took the same test for the purposes of detecting transfer effects.

In the original study, there were 7 participants who trained for 19 days and took the BOMAT at pre- and post-tests (Jaeggi et al., 2008). When Jaeggi and her colleagues replicated their original study, they recruited 50 participants to train their working memory for 20 days. Half of them took the BOMAT and the other half took the RAPM (Jaeggi et al., 2010). In both studies, they claimed that fluid intelligence was improved after working memory training. The current study had 13 participants who took the RAPM before and after training for 20 days, and they showed no significant increase in scores. Although there were differences in sample size and test administration procedures, there was enough power in the current study with the effect size detected from the replication study (Jaeggi et al., 2010) and most participants in the current study on average only needed about 10 minutes to attempt the items on the pre- and post-tests. The pre- and post-tests in the current study were conducted in an hour. In the pre-test, after the administration of the working memory task OSPAN and timed tests, participants had between 15-20 minutes before the end of the session to complete the Mill-Hill Vocabulary Test and RAPM. Participants were told that there were no time constraints and they could take as much time as they wanted to complete the items on both tests. Similarly, participants were given 15-20 minutes at the beginning of post-test session to work on the Mill-Hill and RAPM before the timed tests and OSPAN were administered. In essence, participants in the current study had as much time as those in the studies carried out by Jaeggi et al. (2008, 2010) with the added advantage of no time pressure exerted on the participants.

Jaeggi and her colleagues (2008, 2010) argued that the speeded administration of the transfer tasks was comparable to the non-speeded administration, and they decided to administer the tests with time limits to avoid ceiling effects. Results from the current study did not suggest any ceiling effects of any for the tests administered (RAPM pre-test mean = 12.5, SD = 2.65; RAPM post-test mean = 12.3, SD = 2.55; there were 18 items on pre- and post-tests).

Participants who trained their working memory in the current study improved on the training task just as much those in the original study (Jaeggi et al., 2008). This improvement suggests that participants in the current study were just as motivated and committed as participants in the original study conducted by Jaeggi et al. A comparison of improvement in the training task with participants in the replication study conducted in Taiwan (Jaeggi et al., 2010) was not included because the researchers modified the training task to accommodate the Chinese-speaking participants. The auditory modality of the dual n-back task was modified to syllables in the Mandarin phonetic system and instead of 20 blocks in each training session (~30 minutes), there were only 15 blocks which took about 17-20 minutes to complete (Jaeggi et al., 2010).

The current study was designed to replicate and extend the original study by Jaeggi and her colleagues (2008); thus, it was designed not only to detect an increase in scores but also to determine how the increase in performance arose should there be any. A series of short timed mental abilities tests were also administered at pre- and post-tests in addition to a working memory span task and the transfer task. The tests were selected based on Johnson and Bouchard's (2005) proposed model of intelligence. Their confirmatory factor analyses of various psychometric tests led them to a hierarchical model of intelligence that suggests verbal, perceptual and mental rotation (VPR model) abilities are what drive intellectual capacity (Johnson & Bouchard, 2005). There were three tests for each construct postulated in the VPR model, and exploratory factor analysis conducted on the test variables at pre-test (N=117) in the current study support the model (Figure 9). The test scores were extracted using the principal axis factoring method, and the factor loadings reported in Figure 9 were rotated using the oblique method Direct Oblimin because the factors were assumed to be correlated with one another. Three factors were extracted, and each factor corresponded to each construct in the VPR model – the verbal fluency tests loaded on the first (verbal) factor, the perceptual speed tests loaded on the second (perceptual) factor and the spatial ability tests loaded on the third (mental rotation) factor. If general intelligence were improved after working memory training, it is imperative to know what underlying ability(ies) specifically was improved leading to an increase in general intelligence. The additional tests administered in the current study provided additional information should there was an increase in the transfer task, RAPM. However, results from the current study suggested no improvement overall in each of the three abilities.

Pattern Matrix[a]

	Factor		
	1	2	3
MillHill1	.917	-.250	
Vocab1	.660	.181	
WordBE1	.268	.146	
IdentPix1	-.120	.894	.135
CPST1		.556	
FindA1	.138	.343	-.114
MentalRot1			.763
CardRot1			.511
PaperFold1			.433

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 12 iterations.

Figure 9: EFA results on pre-test variables – Factor 1 corresponds to the Verbal factor, Factor 2 Perceptual factor and Factor 3 Mental Rotation factor in Johnson & Bouchard's (2005) VPR model.

Over the past several decades, many studies suggested that training or repeated practice on a task with instructional aid only increased task-specific variance and rarely did the improvement transfer to other abilities (Belmont & Butterfield, 1977). As an example, Ferrara, Brown and Campione (1986) trained third grade students on a reasoning ability task (Thurstone Letter Series Completion) and tested if the skill that these students learned in solving the practice task would transfer to more difficult items. They found that students needed more prompts or hints when solving items that indicated far transfer compared to items that merely maintained the skill previously learned. Such maintenance items would be very similar to items exposed to students during training. The number of prompts was significantly correlated to students' level of IQ, but their results still suggested failure to transfer even on a very similar reasoning ability task (Ferrara, Brown & Campione, 1986).

Although repeated practice on a challenging task such as the dual N-back task did not show any significant transfer effects to *g* or related mental abilities comprising *g*, it may not be wise to completely reject the notion that there are no practical benefits to training one's working memory. There are many studies in the literature that suggest positive outcome from working memory training, such as reduced inattentive symptoms in ADHD children (Klingberg et al., 2002; Klingberg et al., 2005), increased memory performance in older adults (Buschkuehl et al., 2008), increased math performance in children with working memory deficits (Holmes et al., 2009), improved short term memory in adolescents with borderline intellectual disability (Van der Molen et al., 2010), reduced cognitive deficits in schizophrenic patients (Wykes et al., 1999), significant reduction in symptoms of cognitive problems in patients with stroke (Westerberg et al., 2007) and improved fatigue symptoms in adults with multiple sclerosis (Vogt et al., 2009). Some of these studies reported no improvement in fluid intelligence (Westerberg et al., 2007; Holmes et al., 2009; Van der Molen et al., 2010) and some reported significant improvement (Klingberg et al., 2002; Klingberg et al., 2005). Anecdotal examples in the current study have been encouraging where some participants in the study claimed that they have improved their focus in general. One participant commented on the post-test questionnaire:

> I feel like sometimes I struggled to keep my attention after 12-15 minutes but have noticed during my voice lessons I can focus more on multiple techniques at once, which takes a lot of focus!

This particular participant also mentioned to the experimenter that she first noticed increase in focus when her voice lesson instructor commented on her improved techniques.

Correlational analyses in the current study indicated that working memory capacity measured by a complex span task, OSPAN, was significantly related to all spatial ability tests and the matrix reasoning test, RAPM. This supported the underlying basis for two of the hypotheses proposed in the Introduction section, where mental rotation abilities and reasoning ability would be improved after working memory training, because engagement in these abilities and working memory depended on controlled attention. The proposed mechanism of transferring improved working memory to enhanced intellectual abilities hinted that transfer could be due to increased in focused attention. Although there were no specific tasks that measured controlled attention, its effects could be detected by simpler tasks that required the involvement of working memory, such as spatial ability tests. Failure to detect any improvement in reasoning and spatial ability tests suggested that training on the dual n-back task either failed to improve controlled attention or improve an underlying ability that could not be identified in the current study.

There are studies that support the conclusions drawn from the current study. One is a study by Colom et al. (2010), where participants in the study were divided into two training groups. Participants in group 1 trained on short term memory and working memory tasks, while those in group 2 trained on processing speed and attention tasks. All participants had higher post-test scores than pre-test scores on four intelligence tests (Advanced Progressive Matrices and the abstract reasoning, verbal reasoning and spatial relations from the Differential Aptitude Test Battery), but factor analyses of these test scores suggested no significant changes in the most *g*-loaded tests. This led the researchers to conclude that general intelligence has a high degree of stability (Colom et al., 2010). A recent study by Owen and colleagues recruited a large sample of participants (N=11, 430) to test the validity of brain training. These participants were recruited online and randomly assigned to one of three conditions. Those assigned to experimental group 1 trained on six tasks that tapped reasoning, planning and problem-solving abilities. Participants in experimental group 2 trained on short term memory, attention, visuo-spatial processing and math abilities tasks that are usually on most of the available commercial brain training programs. Participants in the control group did not practice on any specific task but were asked to answer random questions from different available online resources. Post-test was conducted after six weeks of training, and the researchers observed improvements in all the cognitive tasks that were trained but no transfer effects to untrained tasks, even when the untrained tasks shared similar cognitive functioning with those that were trained (Owen et al., 2010).

So far, most studies on working memory training had been focused on improvements on single tasks rather than general cognitive abilities associated with specific tasks. This field could further develop with more studies investigating changes in latent abilities due to working memory training using more sophisticated methods such as structural equation modeling. One such study has been published recently (Schmiedek, Lovden & Lindenberger, 2010), where younger (age 20-31) and older (age 65-80) adults trained one-hour daily for 100 days on perceptual speed, episodic memory and working memory. They were then tested on six working memory transfer tasks (three for near transfer and three for far transfer), and three tasks for each of the following abilities – fluid intelligence, episodic memory and perceptual speed. In both age groups, these researchers found not only positive changes in individual tasks but also the latent abilities represented by the individual tasks. In addition, they observed similar pattern of changes in the practiced latent factors and the transfer latent factors, thus leading them to believe that the positive changes after training was not due to unspecified sources such as motivation (Schmiedek, Lovden & Lindenberger, 2010). More of such studies should be conducted regarding modification in latent abilities due to training before making any definitive conclusion about cognitive training.

References

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102(1)*, 3-27.

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

Ackerman, P. L., Beier, M. E. & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin, 131(1)*, 30-60.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science, 4(11)*, 417-423.

Baddeley, A. D. & Hitch, G. J. (1974). Working Memory. In Bower, G. A. (Ed.), *Recent advances in learning and motivation* (Vol. 8, pp. 47-90). New York: Academic Press.

Belmont, J. M. & Butterfield, E. C. (1977). The instructional approach to developmental cognitive research. In Kail, R. & Hagen, J. (Eds.), *Perspectives on the Development of Memory and Cognition* (pp. 437-481). Hillsdale, NJ: Lawrence Erlbaum Associates,

Biasini, F. J., Grupe, L., Huffman, L. & Bray, N. W. (1999). Mental Retardation: A symptom and a syndrome. In Netherton, S., Holmes, D. & Walker, C. E. (Eds.), *Comprehensive Textbook of Child and Adolescent Disorders*. New York; Oxford University Press.

Buschkuehl, M., Jaeggi, S. M., Hutchinson, S., Perrig-Chiello, P., Dapp, C., Muller, M., Breil, F., Hoppeler, H & Perrig, W. J. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology and Aging, 23(4)*, 743-753.

Carpenter, P. A. Just, M. A. & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97(3)*, 404-431.

Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and 10 broad factor. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 5-21). Amsterdam: Pergamon.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, *3*, 153-193.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd. ed.). Lawrence Erlbaum Associates.

Colom, R., Quiroga, M. A., Shih, P. C., Martinez, K., Burgaleta, M., Martinez-Molina, A., Roman, F. J., Requena, L. & Ramirez, I. (2010). Improvement in working memory is not related to increased intelligence scores. *Intelligence, 38*, 497-505.

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., Kane, M. J. & Engle, R. W. (2003). Working memory capacity and its relation to general intelligence. *Trends in Cognitive Science, 7(12)*, 547-552.

Deary, I. J. (2001) Human intelligence differences: towards a recent history. *Trends in Cognitive Science, 5(3)*, 127-130.

DeFries, J. C., Vandenberg, S. G., McClearn, G. E., Kuse, A. R., Wilson, J. R., Ashton, G. G., et al. (1974). Near identity of cognitive structure in two ethnic groups. *Science*, *183*, 338-339.

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.

Fearson, W. &Eysenck, H. J. (1986). Intelligence, reaction time (RT), and a new odd-man-out RT paradigm. *Personality and Individual Differences, 7*, 807-818.

Ferrara, R. A., Brown, A. L. & Campione, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies of proximal development. *Child Development, 57*, 1087-1099.

Frey, M. & Detterman, D. (2004). Scholastic assessment of g? The relationship between the scholastic assessment test and general cognitive ability. *Psychological Science*, *15*, 373-378.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence*, *24*, 79-132.

Gray, J. R., Chabris, C. F. & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Nature Neuroscience, 6(3)*, 316-322.

Hakstian, A. R. & Cattell, R. B. (1975). *Comprehensive ability battery*. Champaign, Ill: Institute for Personality and Ability Testing.

Halford, G. S., Cowan, N. & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in Cognitive Science, 11(6)*, 236-242.

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

Horn, J. L. (1976) Human abilities: A review of research and theory in the early 1970's. *Annual Review of Psychology, 27*, 437-485.

Hyde, J. S. (1981) How large are cognitive gender differences? A meta-analysis using ω^2 and *d*. *American Psychologist, 36(8)*, 892-901.

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(19)*, 6829-6833.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y. F., Jonides, J. & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning – implications for training and transfer. *Intelligence, 38*, 625-535.

Johnson, W., Bouchard, T. J., Krueger, R. F., McGue, M. & Gottesman, I. I. (2004). Just one *g*: Consistent results from three test batteries. *Intelligence*, *32*, 95-107.

Johnson, W. & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*, 393-416.

Johnson, W. & Bouchard, T. J. (2005). Constructive replication of the visual-perception-image rotation model in Thurstone's (1941) battery of 60 tests of mental ability. *Intelligence*, *33*, 417-430.

Johnson, W., te Nijenhaus, J. & Bouchard, T. J. (2007). Replication of the hierarchical visual-perceptual-image rotation model in de Wolff and Buiten's (1963) battery of 46 tests of mental ability. *Intelligence*, *35*, 69-81.

Johnson, W., te Nijenhaus, J. & Bouchard, T. J. (2008). Still just 1 *g*: Consistent results from five test batteries. *Intelligence*, *36*, 81-95.

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., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131(1)*, 66–71.

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlstrom, K., et al. (2005). Computerized training of working memory training in children with

ADHD – a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry, 44*(2), 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.

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence*, *37*, 327-328.

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

Muhlhausen, D. B. & Lips, D. (2010). Head Start earns an F: No lasting impact for children by first grade. *Backgrounder*, *2363*, 1-13.

Owen, A. M., Hampsire, A., Grahn, J. A., Stenton, J., Dajani, S., Burns, A. S., Howard, R. J. & Ballard, C. G. (2010). Putting brain training to the test. *Nature, 465*(7299),775-U776.

Pinchot, P. (1948). French pioneers in the field of mental deficiency. *American Journal of Mental Deficiency, 3*(1)*,* 128-137.

Raven, J. C. (1990). *Advanced Progressive Matrices: Sets I, II*. Oxford: Oxford Univ Press.

Schmiedek, F., Lovden, M. & Lindenberger, U. (2010). Hundred days of cognitive training enhance broad cognitive abilities in adulthood: findings from the COGITO study. *Frontiers in Aging Neuroscience, 2*(27), 1-10.

Shepard, R. & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*, 701-703.

Spearman, C. (1904) “General Intelligence”, Objectively Determined and Measured. *American Journal of Psychology*, *15*(2), 201-292.

Spitz, H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence*. Hillsdale, NJ: Erlbaum.

Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I. & Jongmans, M. J. (2010). Effectiveness of a computerized working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research, 54*, 433-447.

Vernon, P. E. (1965) Ability factors and environmental influences. *American Psychologist, 20*, 723-733.

Vogt, A., Kappos, L., Calabrese, P., Stocklin, M., Gschwind, L., Opwis, K. & Penner, I. (2009). Working memory training in patients with multiple sclerosis – comparison of two different training schedules. *Restorative Neurology and Neuroscience, 27*(3), 225-2345.

Wahlsten, D. (1997). The malleability of intelligence is not constrained by heritability. In B. Devlin, S.E. Fienberg., & K. Roeder (pp. 71-87). *Intelligence, Genes, and Success: Scientists respond to The Bell Curve*. New York: Springer.

Westerberg, H., Jacobaeus, H., Hirvikoski, T., Clevberger, P., Ostensson, M. L., Bartfai, A. & Klingberg, T. (2007). Computerized working memory training after stroke – A pilot study. *Brain Injury, 21*(1), 21-29.

Wykes, T., Reeder, C., Corner, J., Williams, C. & Everitt, B. (1999). The effects of neurocognitive remediation on executive processing in patients with schizophrenia. *Schizophrenia Bulletin, 25*(2), 291-307.