Does Training to Increase Working Memory Capacity Improve Fluid Intelligence?

By

Clayton Lewis Stephenson

A dissertation submitted to the Faculty of Claremont Graduate University in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology

Claremont, California

2010

Approved by:

Diane F. Halpern

Diane F. Halpern, Ph.D.

We, the undersigned, certify that we have read this dissertation of Clayton Lewis Stephenson and approve it as adequate in scope and quality for the degree of Doctor of Philosophy in Psychology.

Dissertation Committee:

Diane F. Halpern

Diane F. Halpern, Ph.D., Chair

Dale E Berger

Dale E. Berger, Ph.D., Member

Gabriel I. Cook, Ph.D., Member

James C. Kaufman

James C. Kaufman, Ph.D., Visiting Member

Abstract of the Dissertation

Does Training to Increase Working Memory Capacity Improve Fluid Intelligence?

By

Clayton L. Stephenson

Claremont Graduate University: 2010

Although a number of theories of intelligence have been created in the past century, Horn and Cattell's (1966) Theory of Fluid and Crystallized Abilities has reigned as one of the more popular theories of intelligence with the Raven's Progressive Matrices being the most used assessment of fluid intelligence. Fluid intelligence is the ability to solve novel problems with minimal involvement from prior training or strategies retrieved from memory. In general, fluid intelligence cannot be improved through training. But, recent studies have shown that working memory capacity and visuospatial abilities have a strong relationship with fluid intelligence. Based on the knowledge that working memory capacity can improve with training, a recent study by Jaeggi, Buschkuehl, Jonides, and Perrig (2008) claimed to have improved fluid intelligence by having participants complete a four week training program using the dual *n*-back task. The dual *n*-back task is a working memory task that presents auditory and visual stimuli simultaneously. A concern of Jaeggi et al.'s study was whether they improved fluid intelligence as the construct is defined by Horn and Cattell, or if the improvement in test performance was due to improved visuospatial abilities. The current study replicated and expanded Jaeggi et al.'s study by having participants complete variations of the dual *n*-back task as training. Participants were assessed with four tests of G*f* and four cognitive tests.

The current study was successful in replicating Jaeggi et al.'s (2008) results. However, the current study also observed improvements in scores on the Raven's Advanced Progressive Matrices for participants who completed a variation of the dual *n*-back task or a short-term memory task training program. Participants' scores improved significantly for only two of the four tests of G*f*, which raises the issue of whether the tests measure the construct G*f* exclusively, as defined by Cattell (1963), or whether they may be sensitive to other factors. The concern is whether the training is actually improving G*f* or if the training is improving attentional control and/or visuospatial skills, which improves performance on specific tests of G*f*. The findings are discussed in terms of implications for conceptualizing and assessing G*f*.

Dedication

To my family, wife, and friends.

Acknowledgments

The African proverb, "It takes a village to raise a child," can be applied to the completion of a dissertation with the dissertation being a child. It took a number of people for the current dissertation to be completed and I am thankful to everyone who provided there support, time, and efforts.

The current dissertation could never have been completed without guidance and support of Diane F. Halpern at Claremont McKenna College (CMC). Diane handed me the article that I ended up replicating and expanding and knew the topic was my dissertation. It was an ambitious project, but Diane believed in its potential and provided me with the support I needed to complete the project. Dale Berger at Claremont Graduate University (CGU) also provided support that was as valuable as gold in helping me think about the design of the study and assisted me with the statistics with grace and patience. Gabriel Cook at CMC also provided precise feedback that gave the dissertation the "edge" that I really wanted it to have. I also thank James Kaufman at California State University, San Bernardino (CSUSB) for taking the time to be an outside reviewer for my dissertation. I am also grateful to Susanne Jaeggi and Martin Buschkuehl, leaders of the team of researchers that published the study I replicated, who shared their materials with enthusiasm.

I was fortunate to be working as an adjunct professor at CSUSB during the time I was collecting data. Robert Cramer, Stacy Brooks, and Yuchin Chien helped provide lab space for me even though I wasn't a tenure-track professor. My colleagues, John Clapper, Mary Dolan, Lisa Looney, Peter Spiegel, and Diane Swanson, allowed me to recruit participants from all of their classes and I am grateful for letting me take up 10 minutes of their class time. I also thank Dale Berger, Michelle Bligh, and Kathy Pezdek at CGU and Gabriel Cook, Shana Levin, and Ronald Riggio at CMC for allowing me to recruit participants in their courses.

It took approximately one thousand hours to collect data, a task that could not have been done without the help of my dear friends, John LaVelle and Kate Byrd-LaVelle. John and Kate provided me with old computers to use, sat in the lab while I taught classes, and offered support in so many ways that I cannot list them all here. I owe much to Jason Brooks who saved me days of tedious data cleaning by providing me with Excel macros that did the job in a matter of hours. Finally, I am grateful to Paul Lyman, a graduate student in philosophy who provided me with new insights and approaches to psychology.

This dissertation could not have been completed without the people who participated. It is amazing that 120 people were interested enough in my study that each person dedicated ten hours of their time to complete the testing and training.

My family and friends in Colorado provided me with support by understanding why I had to leave for California to work on my Ph.D. and always told me to stick with it. Every year I told them I was closer and they always believed that I was getting closer to completion. The next time I see them, I can say I am done.

I am fortunate to have met my wife, Jenny, as I was starting the dissertation. She provided me with loving support by cooking meals and making sure I ate when I was focused on writing, providing me with excuses to take evening walks and hike on Sundays, and sitting in the lab scoring tests of intelligence while I taught classes. Thank you all so very much.

Table of Contents

List of Tables

List of Figures

List of Appendices

Chapter One

Literature Review

The conceptualization and measurement of intelligence has been recognized as one of psychology's greatest achievements as well as one of its greatest controversies. Intelligence is often thought of as being a unitary construct that is determined by the convergence of multiple cognitive aptitudes. The unitary perspective was developed by Spearman (1904), who coined the term "*g*" to refer to a person's general intellectual ability or "general fund of mental energy" (Spearman, 1914, p. 103). A person's general fund of mental energy determines how well and how quickly that person can process different levels of information. Although Spearman's *g* was an original and, for the most part, a sound theoretical construct based on psychometric testing, other researchers have varied his conceptualization of *g* by dividing it into different components that still load onto the unitary construct known as *g*.

Cattell's Theory of Fluid and Crystallized Abilities (1963) is one example that has varied Spearman's *g* and has become a predominant model of conceptualizing and measuring *g* by dividing it into two different components. Horn and Cattell (1966) provided empirical evidence that *g* can be divided into crystallized intelligence (G*c*) and fluid intelligence (G*f*). G*c* is described as a person's accumulation of knowledge and can improve throughout life with experience, training, and formal education. G*c* enables a person to solve problems based on the specific knowledge acquired during the lifespan. For example, if construction workers are trying to figure out the grade and angles of a roof, then they can use the knowledge gained from previous work experience or, possibly, the knowledge gained in a geometry or shop class. As a cognitive process, G*c* is typically an automatic process guided by heuristics. By contrast, G*f* improves very little over the lifespan, and when it does improve, it is a result of schooling and broad training rather than acquisition of knowledge. G*f* is defined as the ability to solve novel problems without the direct application of prior knowledge or training on the problem (Horn & Cattell, 1966). More specifically, G*f* has been conceptualized as an indication of a person's general ability for sequential and inductive reasoning, performance levels on Piagetian reasoning tasks, and cognitive process speeds for reasoning (McGrew, 1997). For example, solving a novel three-dimensional jigsaw puzzle requires a person to develop problem solving strategies rather than relying on solutions stored in memory.

Researchers have found evidence that G*f* cannot be as easily improved as G*c* because G*f*, for the most part, is influenced more by biological predispositions than environmental factors (Thompson et al., 2001). The evidence for G*f* being mostly influenced by biological predispositions comes from brain imaging studies on intelligence and studies on intelligence in monozygotic (identical) and dizygotic (fraternal) twins. Gray and Thompson (2004) reviewed three different methods that researchers use in determining the biological correlates of *g* and G*f*. First, Gray and Thompson reviewed brain imaging methods that determine the volume of frontal gray matter in the brain, which is associated with higher levels of cognitive function. According to Gray and Thompson, researchers have reported that the volume of frontal gray matter can predict *g* beyond what can be predicted from the total volume of the brain. Furthermore, the volume of frontal gray matter is attributed to a string of genes that determine the development of brain structures, as shown through studies on twins.

Second, brain imaging has shown that the lateral prefrontal cortex is the one region of the brain that was activated while participants' *Gf* was being assessed using items from Cattell's Culture Fair Test and the Letter Sets from the Education Testing Service Kit of Factor-Referenced Tests. Third, studies on twins have shown that 40 to 80 percent of the variability in *g* can be determined genetically. Moreover, a higher correlation was found between monozygotic twins compared to dizygotic twins when determining shared volume of frontal gray matter.

Gray and Thompson (2004) recognized that factors such as fetal environment, nutrition, education, socioeconomic status, motivation, and other environmental factors are important in the development of *g* and *Gf*. However, these external actors typically have an impact on specific cognitive abilities such as verbal and visuospatial abilities, which do not rely completely on general abilities such as *Gf*. Therefore, specific cognitive abilities can be improved through acquisition of knowledge through education, training, and practice; whereas *g* and *Gf* cannot be improved as easily because they are, for the most part, biologically predetermined.

Although some researchers argue that *Gf* is primarily influenced by biological predispositions, Jaeggi, Buschkuehl, Jonides, and Perrig (2008) provided evidence that a cognitive training program designed to enhance working memory capacity (WMC) improved scores on the Raven's Advanced Progressive Matrices, thus showing improvements in *Gf*. The present study replicated and expanded on Jaeggi et al.'s study. Two issues present in Jaeggi et al.'s study motivated the current study. First, the Raven's Progressive Matrices is the most widely used measure of *Gf*; however, researchers have questioned the validity of the Raven's Progressive Matrices measuring *Gf* exclusively because the Raven's Progressive Matrices may require people to rely partially on visuospatial abilities (e.g., van der Ven & Ellis, 2000; Vigneau & Bors, 20008). If the Raven's Progressive Matrices measure visuospatial abilities in addition to G*f*, then Jaeggi et al. might have improved visuospatial abilities rather than G*f*. A review of the literature to be presented later shows that training can improve visuospatial abilities.

The second issue of concern is that WMC has been identified as a major factor in G*f*, and is possibly a causal factor or even the same construct as G*f*. Many WMC tasks (e.g., counting span, rotation span) have a strong visuospatial component to them and the dual *n*-back task as used by Jaeggi et al. (2008) is no exception. Jaeggi et al.'s use of the dual *n*-back task as a training technique may have improved peoples' visuospatial abilities, which could explain the improved scores on tests of G*f*. Jaeggi et al.'s study will be discussed in detail with a critique of the dual *n*-back task. The current study addressed whether improvement in G*f*, as measured using the Raven's Advanced Progressive Matrices, is in part due to improved visuospatial skills resulting from training on the dual *n*-back task.

Raven's Progressive Matrices

Raven's Standard Progressive Matrices was first published in 1938 for use by researchers interested in the biological and environmental factors that influence intelligence (Mackintosh & Bennett, 2005; Raven, 2000). The Raven's Progressive Matrices test has become the most accepted and frequently used measure of G*f*. Two variations of the original test have been created: the Raven's Colored Progressive Matrices were created to assess children's intelligence and the Raven's Advanced Progressive Matrices were created as a more difficult version of the Raven's Standard

Progressive Matrices. The original use of the matrices was to test for what Raven referred to as "eductive ability," which is the ability to gain new insights or information from a problem that will eventually result in a solution for the problem (Raven, 2000). Essentially, the description of eductive ability closely matches the description of G*f* in that both descriptions refer to the type of intellectual ability used to reason inductively and solve novel problems. A problem is defined as being novel if solutions cannot be determined by the direct application of experience or knowledge. For example, a problem that relies on G*f* cannot be solved by recalling mathematical formulas or a particular solution from a similar problem, and such a problem would be considered to be solved by a person's crystallized intelligence.

The Raven's tests, and others like it, typically present objects that are organized in matrices. The objects vary throughout the sequence with the last object in the sequence left blank (see Figure 1). The person's task is to choose from several alternatives the one answer that correctly completes the sequencing.

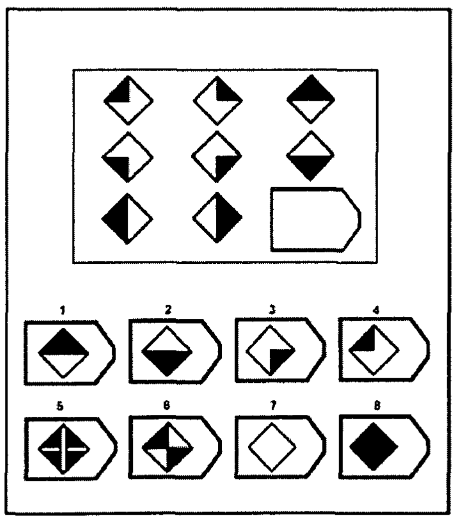


Figure 1: An example from the Raven's Progressive Matrices. This is an example from an earlier version of the test that is no longer in use. The answer is number *8*.

The Raven's Progressive Matrices were designed to exclusively measure the unidimensional construct, G*f* (Raven, 2000; van der Ven & Ellis, 2000). A study by Schweizer, Goldhammer, Rauch, and Moosbrugger (2007) partially supported the assumption that the Raven's Advanced Progressive Matrices is unidimensional by testing its convergent and discriminant validity. Schweizer et al. created models showing the relationship between Raven's scores and reasoning abilities and between Raven's scores and spatial abilities. The first model showed significant convergent validity (.68) between the scores on the Raven's and Horn's (1983) reasoning scale. However, the model showed weaker discriminant validity between the Raven's Advanced Progressive Matrices and Horn's visualization scale (.34), mental rotation scale (.27), and the closure scale (.24). Schweizer et al.'s final model collapsed the visualization scale, mental rotation scale, and the closure scale into one factor labeled as spatial ability. Reasoning was still the better predictor (.64) of scores on the Raven's while spatial ability was nonsignificant (.06). Despite using only one measure each for reasoning and visuospatial abilities, Schweizer et al.'s study provided some evidence that the Raven's Advanced Progressive Matrices convergent validity with reasoning is stronger than the discriminant validity with spatial ability, thus, supporting the idea that the Raven's Advanced Progressive Matrices is a relatively pure measure of G*f*.

Researchers such as Schweizer et al. (2007) argue that the Raven's Progressive Matrices are multidimensional, and they claim that the items require people to engage visuospatial abilities rather than using only G*f* abilities. A corollary of the hypothesis that visuospatial abilities underlie success in Raven's matrices is that if people can improve their visuospatial abilities, their scores on the Raven's Matrices would also increase. Two fundamental questions follow from this reasoning: Does the Raven's measure *Gf* exclusively? And, if the Raven's does not measure *Gf* exclusively, are *Gf* scores improved by training visuospatial abilities? If visuospatial abilities are utilized while working on items in the Raven's tests and if visuospatial abilities are improved, then performance on Raven's tests of *Gf* should increase.

Improving Visuospatial Abilities

Research has shown improvement visuospatial abilities (Castel, Pratt, & Drummond, 2005; Green & Bavelier, 2003), but no one has tested to see if the effects transfer to performance on tests of *Gf*. One way to enhance visuospatial skills is by playing video games that are highly spatial. Although no one has investigated the effects of improving visuospatial abilities to improve scores on tests of *Gf*, it is important to understand what aspects of visuospatial abilities are enhanced by playing video games and how the video games may be similar to training tasks used to enhance WMC.

Castel et al. (2005) addressed whether people who play video games differ in their visuospatial abilities compared to non-video game players by testing the differences in reaction times (RT) between video game players and non-video game players in a stimulus onset asynchrony task and an easy and difficult visual search task. Video game players were defined as people who played action video games at least four times per week and played for at least an hour during each session. Action video games are defined as games that require the player to move characters around in two-dimensional or three-dimensional space. Castel et al. found that video game players had faster RTs in both visual search tasks. Castel et al. suggested that people who play video games do not necessarily use a different cognitive process for attending to visual searches; rather, they are simply faster at mapping the visual stimulus, which leads to a faster response time. Because video game players respond quicker on visual tasks, the next question to answer is whether video-game training improves visuospatial abilities. One of the components of visuospatial abilities is visual attention; more specifically, attention for items over space and over time.

Green and Bavelier (2003) tested the differences between video game players and non-video game players in attentional resources for processing visual stimuli in space and over time and found that video game players had greater visual attention resources and outperformed non-video game players in tasks that tested visual attention for time and space. Green and Bavelier followed up their experiments to test if training the non-video game players by having them play the video game *Medal of Honor: Allied Assault* for an hour each day over a period of ten days would improve their performance on the visual tasks. Green and Bavelier found that after completing the videogame training, non-video game players' performance improved significantly in the visual attention tests when compared to a control group of non-video game players.

The studies by Castel et al. (2005) and Green and Bavelier (2003) are important because they provide two critical research findings. First, video game players outperformed non-video game players on all of the visual tasks that were tested, which could have been a result of self-selection bias. However, Castel et al. suggested that video game players outperforming non-video game players on visual tasks is a result of training (i.e., playing video games) having an effect on visual attention. Green and Bavelier provided support for their hypothesis by having non-video game players play video games as a way of training visual attention. Many other studies have found similar findings showing that playing action video games as training improves mental rotation (Cherney, 2008; Feng Spence, Pratt, 2008; Okagaki & Frensch, 1994), spatial attention (Feng et al., 2008), spatial visualization (Okagaki & Frensch, 1994), spatial resolution of visual processing (Green & Bavelier, 2007), visuospatial attention (Green & Bavelier, 2006a), and the ability to track multiple objects at once (Green & Bavelier, 2006b).

Castel et al. (2005) and Green and Bavelier's (2003) studies suggest that action video games can be used to train visual attention. If visual attention can be improved and if items on the Raven's Progressive Matrices require some visuospatial skills such as visual attention, then visuospatial training could improve scores on the Raven's. However, higher scores on the Raven's Progressive Matrices may not necessarily reflect improved G*f*, especially based on some definitions of G*f* (e.g., Horn & Cattell, 1966). In other words, if the Raven's is multidimensional such that the items tap into visuospatial abilities in addition to G*f*, training programs that strengthen visuospatial skills and *not* G*f*, as defined by Horn and Cattell, may improve scores on the Raven's Progressive Matrices, but not improve actual G*f* abilities. Although Green and Bavelier improved participants' performance on visual tasks after video-game training, playing video games in an everyday context may not necessarily improve visuospatial abilities or any other cognitive abilities. There are, however, WMC tasks that could improve visuospatial abilities because visuospatial abilities are a component of WMC.

Working Memory Capacity

Working memory is a cognitive ability that allows a person to hold relevant information in temporary storage while manipulating other information at the same time (Heitz, Unsworth, & Engle, 2004). WMC refers to the amount of information that people can store and process in working memory. Although there are numerous models of working memory, Baddeley and Hitch's (1974) model has become one of the most widely accepted models and is the most relevant model regarding the relationship between WMC and G*f*. Baddeley and Hitch's model has come to be known as Baddeley's Model of Working Memory, a model which Baddeley (2000) recently updated. Baddeley's original model included three components: a central executive and two subsystems referred to as the visuospatial sketchpad and the phonological loop.

The central executive is the control mechanism for the two subsystems. The visuospatial sketchpad holds and processes visual and spatial information while the phonological loop holds and processes verbal (i.e., acoustic) information. Many terms are used interchangeably when talking about the central executive. The central executive in Baddeley's model is often referred to as the supervisory attentional system (Norman & Shallice, 1986), executive attention (Engle, 2001), and attentional control (Ackerman, Beier, & Boyle, 2005; Engle, 2002; Engle, Tuholski, Laughlin, & Conway, 1999; Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004). For the purposes of the current study, the term "attentional control" will be used when referring to the process of focusing attention or training the component of working memory that controls the manipulation of information.

WMC and G*f*

A link between WMC and G*f* has been made based on their conceptual definitions and the way the two constructs are measured. WMC is an ability to store and process incoming information while ignoring irrelevant information, and items in the tests G*f* are constructed so that the test taker has to identify rules in the problem by recognizing patterns needed for solution while ignoring other patterns or distractions embedded in the problem. It was not until Kyllonen and Christal (1990) published a pioneering study that a strong empirical relationship ($r = .76$) was found between reasoning abilities (i.e., *g)* and WMC. Reasoning abilities were assessed by Kyllonen and Christal using five different tests from the Education Testing Service's Kit of Reference Tests for Cognitive Factors, two subtests from the Armed Services Vocational Aptitude (ASVAB) test, and eight other various tests that measured mathematical, verbal, and analytical reasoning abilities. Processing speed was also measured using the coding speed test and the numerical operations test, which are subtests of the ASVAB. The coding speed test assesses a person's ability to look up a number in a table and associate a letter with the number code. The numerical operations task assesses how many simple arithmetic problems can be completed in ten minutes. Kyllonen and Christal used a number of tests to measure WMC including the ABCD Grammatical Reasoning test, the ABC Numerical Assignment test, the Digit Span test, the Mental Arithmetic test, the Alphabet Recoding test, and the Adjacent Letters test.

Kyllonen and Christal (1990) found strong correlations (.80 to .90) between WMC, in accordance to Baddeley's (1974) model, and reasoning abilities. WMC was also found to be positively correlated with processing speed (.37). Since Kyllonen and Christal's findings, there has been a surge of research on the relationship between WMC and *g* or G*f*. As a result, researchers have either rejected the idea that there is a substantial link between WMC and G*f* (Deary, 2000; Kline, 2000) or made claims to the other extreme that WMC and G*f* are isomorphic (Engle, 2002; Jensen, 1998; Kyllonen, 2002). If WMC and G*f* are isomorphic, then it would mean that measurements of G*f* are also measuring WMC. For example, if G*f* is measured using the Raven's Progressive Matrices, then WMC is being assessed as well.

Ackerman et al. (2005) used meta-analytic procedures to determine that WMC and G*f* are not isomorphic. The authors calculated effect sizes using corrected and uncorrected correlations for 57 studies showing correlations between WMC and G*f*. Ackerman et al. found that when the unreliability in the measures for WMC and G*f* was accounted for, only 22.9% of the variance was shared between WMC and G*f*. While showing that they are related, this is a value that is too small to support the claim that they are the same construct. Kane et al. (2004) reported similar findings with WMC accounting for 30% to 40% of the variance in G*f* and Kane, Hambrick, and Conway (2005) found 50% of the variance was accounted for by WMC. Although WMC and G*f* are not identical or isomorphic, it is reasonable to conclude that they are related.

WMC and Gf in Structural Equation Models.

The relationship between WMC and G*f* has been studied with structural equation modeling (SEM). SEM provides information for how observed variables (i.e., various cognitive tasks) load onto different factors (i.e., WMC, STM, verbal comprehension, G*f*). Different models are constructed in order to find a best fit model. There are two types of latent variables in SEM: latent exogenous variables and latent endogenous variables. Latent exogenous variables are constructs that have causal influences in SEM. For the models showing a relationship between WMC, STM, and G*f*, WMC and STM represent latent exogenous variables. Latent endogenous variables have no causal influences on other variables in SEM. Because G*f* is a theoretical component of intelligence, it is used as the latent endogenous variable in SEM procedures used to determine the relationship between WMC and G*f*. Models are constructed by creating causal paths between endogenous and exogenous variables. These paths are typically specified by the researcher on the basis of evidence that supports a theory.

SEM procedures were used in research conducted by Engle et al. (1999) who found a relationship between WMC – more specifically attentional control – and G*f*. Engle et al. argued that STM was a subordinate of WMC, and if the shared variance of STM and WMC was accounted for, then the residual variance could be condensed into single factor that could theoretically be attentional control. A path could then be created between attentional control and G*f* to determine if there is a relationship between the two constructs and the strength of the relationship if it exists. In order to construct their model, Engle et al. gave participants a battery of tasks that measured STM and WMC. G*f* was measured with the Raven's Advanced Progressive Matrices and Cattell's Culture Fair Test. Engle et al. found that their best fit model contained WMC and STM as separate factors that loaded onto the factor that could be considered attentional control and they labeled it "common." Furthermore, the path between the "common" factor and G*f* was significant and resulted in diminishing the path between STM and G*f* to a nonsignificant level. The path between "common" and G*f* did not, however, nullify the path between WMC and G*f*; it only decreased its value.

Engle et al.'s (1999) study is not the only one showing that attentional control has a strong path leading to G*f*. Kane et al. (2004) found similar results even when WMC was teased apart in accordance to Baddeley and Hitch's (1974) model. That is, Kane et al. broke WMC into verbal storage, visuospatial storage, and the central executive as three separate latent exogenous variables with paths leading to verbal reasoning, spatial reasoning, and *Gf* as the three latent endogenous variables. Kane et al. found that the relationship was stronger between verbal storage and verbal reasoning than between verbal storage and *Gf*. Furthermore, the relationship between the central executive and *Gf* was stronger than the relationship between the central executive and verbal reasoning or spatial reasoning. However, one of the more interesting findings by Kane et al. was that the relationship between spatial storage and *Gf* was stronger than the relationship between spatial storage and spatial reasoning. In addition, the relationship between spatial storage and *Gf* was as strong as the relationship between the central executive and *Gf*. Taken together, Kane et al.'s findings suggest that *Gf* relies mostly on the central executive and visuospatial storage, but does not rely as much on verbal storage. Although SEM is not completely predictive of behavioral phenomena, the models built by Engle et al. and Kane et al. provide further support of Kyllonen and Christal's (1990) finding that WMC is a strong indicator of or influential factor in *Gf*.

The evidence from the SEM models constructed by Engle et al. (1999) and Kane et al. (2004) support the notion that WMC and *Gf* are strongly related such that the attentional control component of WMC influences *Gf*. Furthermore, based on Kane et al.'s study, visuospatial storage has just as much of an influence on *Gf* as attentional control. If the models are correct, then it could be possible that cognitive training designed to improve the attentional control's ability to store, update, and process information while dismissing irrelevant information could lead to an increase in scores on tests of *Gf*. However, there is also evidence of a strong connection between visuospatial abilities and *Gf*. Could training designed to increase visuospatial abilities lead to an increase in scores on tests of *Gf* such as the Raven's Matrices? The evidence provided in the literature reviewed thus far indicates that it may be quite possible to improve scores on the Raven's Progressive Matrices because items on the test have been shown to have a strong visuospatial component (van der Ven & Ellis, 2000) and, according to Engle et al. and Kane et al., attentional control *and* visuospatial storage influence G*f*.

Improving Gf

In a recent study, Jaeggi et al. (2008) claimed to have devised a method that can improve G*f* by training designed to enhance WMC. Jaeggi et al. had participants complete a training program using a complex WMC task known as the dual *n*-back task as the training material. The four-week training program was designed to have training sessions five days out of the week with each training session lasting 20 minutes per day. Jaeggi et al. observed an increase in participants' scores on the Raven's Advanced Progressive Matrices after completing the dual *n*-back training program over a span of 8, 12, 17, and 19 days. These scores were compared to a control group who had no training between the pretest and posttest.

A question raised from Jaeggi et al.'s (2008) finding is, "Why were Jaeggi et al. able to enhance G*f* with a relatively short training program?" Part of the answer is that Jaeggi et al. used a dual processing task with the intention of improving WMC. By definition, dual processing is the processing of two cognitive functions simultaneously. In the case of WMC, the two cognitive processes are 1) holding information in short-term storage while 2) processing current information. Jaeggi et al. created a variation of the *n*-back task that was originally used to determine which brain regions activated while using working memory (Braver, Cohen, Nystrom, Jonides, Smith, & Noll, 1997). The variation of the *n*-back task was then applied to a theoretical account that WMC influences G*f*.

The N-Back Task.

Originally, the *n*-back task was designed to measure brain activity using functional magnetic resonance imaging (fMRI; Braver et al., 1997). Braver et al. measured activation levels in the prefrontal cortex while participants completed the *n*-back task to determine the link between the prefrontal cortex and working memory in humans. Stimuli in an *n*-back task can be visual or auditory. The visual stimuli in an *n*-back task are typically blue squares or circles appearing on a computer screen in different locations and the auditory stimuli are typically spoken consonants heard over headphones. A trial represents one presentation of a stimulus and an interstimulus interval (ISI), which is the presentation of a blank screen with a cross in the center. The task for the participant is to indicate if a stimulus that was immediately presented (the target) had been presented on a previous trial exactly *n* steps (i.e., trials) back. Participants are told how many trials back they are supposed to keep in memory. For example, a participant in a two *n*-back task would be instructed that a series of stimuli will be presented and that the goal is to respond by pressing a key on a keyboard if a stimulus is the same as the stimulus presented exactly two trials previously.

The stimuli in an *n*-back task are pseudorandomized and presented serially. Pseudorandomization means that predetermined seed state is used to start a sequence. For example, the number "2" is the seed state that is used to start a pseudorandomized sequence in a two *n*-back task. The stimulus that appears third in a sequence may be randomly assigned to act as a target or not act as a target. If the stimulus is selected to be a target, the pseudorandomization process starts such that the same stimulus that was presented third in the sequence must appear fifth in the sequence (i.e., two steps in the sequence) in order to maintain its status as a target.

Pseudorandomization is used in the *n*-back task because the nature of the *n*-back task demands that a stimulus acts as a target and, at the same time, acts as a foil for other targets that are different (see Figure 2). Thus, a specific algorithm is used to allow partial randomization of the stimuli while still fulfilling the *n*-back requirements that a stimulus appears in the appropriate order to be a target and act as a foil at the same time. The following is a description of the sequence of events in Figure 2 regarding a stimulus acting as a foil and a target at the same time. The sequence in Figure 2 starts on the left hand side. The sequence begins with a trial that presents a blue square in the upper left hand corner for 500 ms. An ISI follows the stimulus and lasts for 2500 ms. The ISI's are not shown in Figure 2 but are set to appear between each stimulus. The stimulus that is presented in the first trial is also presented in same position in the third trial of the sequence, thus it is a target. The stimulus in the fourth position of the sequence acts as a foil but is repeated in the sixth position in the sequence, thus, it becomes a target.

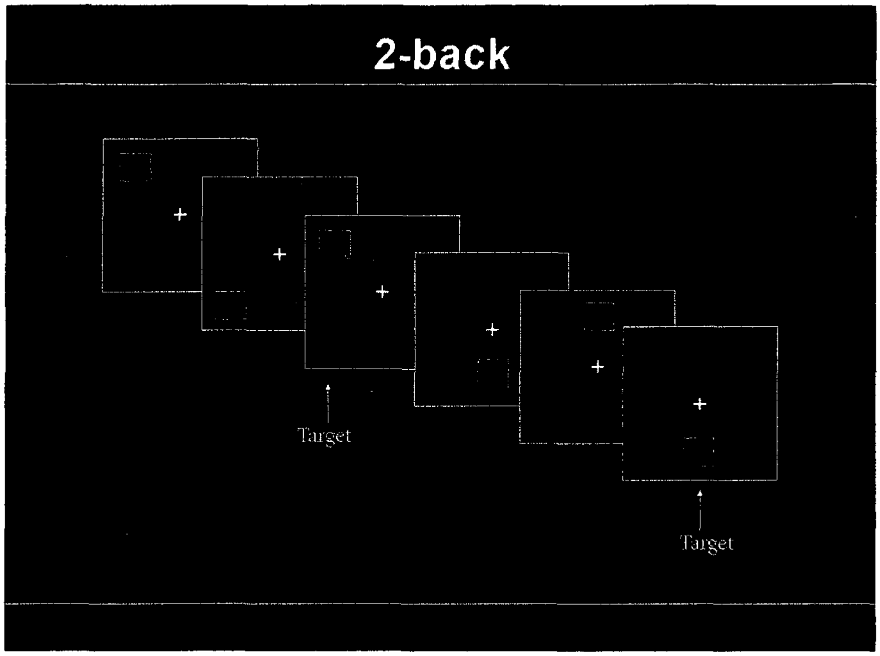


Figure 2: Illustration of a two *n*-back task. Participants are told to respond when a stimulus appears in the same position that it appeared exactly two trials earlier. Pseudorandomization allows for a stimulus to appear in the same position as it did one, two, three, or *n*-back trials. Note that stimuli marked as targets are in the same position as they were exactly two trials earlier. Figure 6 is a screen shot from Buschkuehl, M., Jaeggi, S. M., Kobel, A., & Perrig, W. J. (2008). *BrainTwister: A collection of cognitive training tasks*. University of Bern, Switzerland: Department of Psychology.

Braver et al. (1997) used the *n*-back task to test whether the prefrontal cortex was an important region of the brain during the activation of WMC. The authors found a positive linear relationship between the increase in working memory load and brain activity in the prefrontal cortex. The *n*-back task was assumed to be a task that tapped into WMC in Braver et al.'s study and increased the load on WM as the *n*-back task increased in difficulty (i.e., one *n*-back vs. three *n*-back). The dual nature of the *n*-back task was not empirically tested until a study by Watter, Geffen, and Geffen (2001) was conducted.

Watter et al. (2001) were the first researchers to confirm the dual nature of the *n*-back task using behavioral and event-related-potential (ERP) data. The authors suggested that the ERP called P300 would peak during the presentation of an *n*-back task. P300 is a wave from an ERP reading that peaks between 300 and 800 ms after a visual stimulus acting as a target has been processed and P300 is particularly sensitive to complex perceptual demands. Watter et al.'s reasoning was based on previous research (Johnson, 1986; 1993) suggesting that as a task's sequence becomes more complex and the probability of a stimulus appearing in a sequence is decreased, then the P300 amplitude would be greater. In other words, when working memory is needed for processing stimuli *and* remembering a sequence of stimuli presented during the time of processing, then the P300 amplitude would increase, thus showing a strain on attentional control. The increase in P300 suggests that more effort is being put into processing a task presumably because dual processing is taking place. The authors' hypothesis was supported in that they found greater P300 amplitudes for participants engaged in a one, two, or three *n*-back task compared to participants engaged in a *n*-back task set as a zero-back task (i.e., participants only had to identify a target when it was presented). Moreover, there were no differences in P300 amplitudes between the one, two, or three *n*-back tasks. Watter et al.'s study provides support for the *n*-back task being a dual processing task; as this is a key element for any task used to tap into WMC.

The Dual N-Back Task.

Since Watter et al.'s (2001) finding, the *n*-back task has been confirmed as a task that requires the simultaneous operation of storing information while processing incoming information. The dual *n*-back task, which was used by Jaeggi et al. (2008) as the training task, however, is not the same as the original *n*-back task. The original *n*-back task presented stimuli for only one modality (visual or auditory) at a time. The word "dual" in front of "*n*-back task" refers to the simultaneous presentation of two modalities: visual and auditory, which is the uniqueness of the dual *n*-back task used by Jaeggi et al. In the case of Jaeggi et al.'s training, the visual stimuli were blue squares that appeared on the screen in one of eight positions: top left corner, top center, top right corner, right center, bottom right corner, bottom center, bottom left corner, left center (see Figure 3).

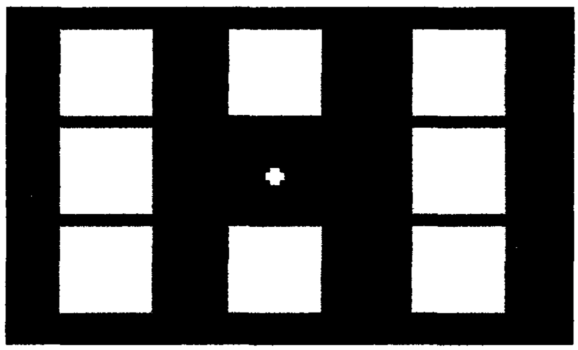


Figure 3: Example of visual stimuli in the dual *n*-back task. Note that only one blue square appears per trial. This figure shows all possible positions.

At the same time that the blue square appeared, a consonant was presented orally through headphones. In the dual *n*-back task, a participant responds if either the visual stimulus *or* the auditory *or* both were presented *n*-trials back (see Figure 4). If both modalities have to be constantly updated and maintained in working memory, then there is likely to be a strain on the processor responsible for attention and control over various cognitive functions. As a result of using two modalities in a WMC task, the dual *n*-back is a unique task that requires a great amount of cognitive resources to perform. Jaeggi et al. (2008) used this task to train WMC, which in turn, led to an increase in scores on the

Raven's Advanced Progressive Matrices. Jaeggi et al.'s (2008) study is one of a kind and pioneering, however, it is not without its flaws.

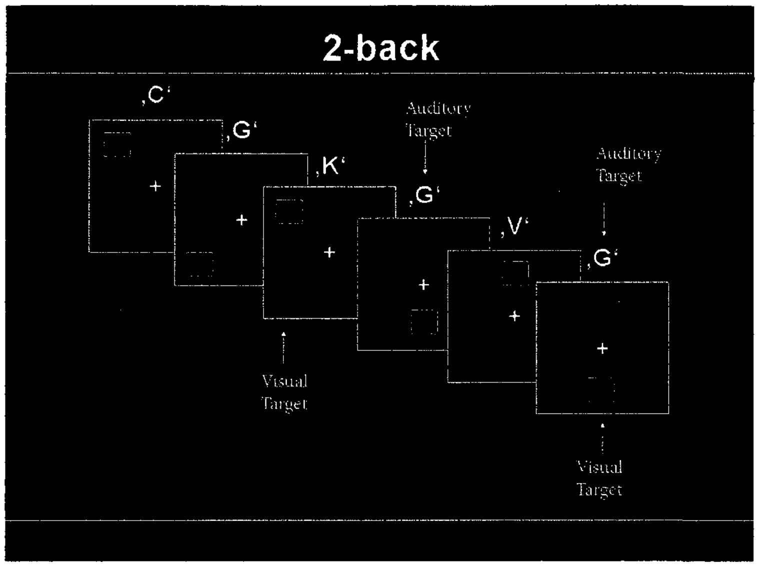


Figure 4: Example of a two back dual *n*-back task with visual and auditory stimuli. Note that stimuli marked as targets are in the same position as in the previous two trials. Figure is a screen shot from Buschkuehl, M., Jaeggi, S. M., Kobel, A., & Perrig, W. J. (2008). *BrainTwister: A collection of cognitive training tasks.* University of Bern, Switzerland: Department of Psychology.

Critique of Jaeggi et al. (2008)

Evidence by Engle et al. (1999) and Kane et al. (2004) has provided support for attentional control being an important factor, possibly even a definitive factor, of G*f*. However, visuospatial abilities are also an important factor of G*f* as well. By using the dual *n*-back task as training for WMC, there is no indication of which cognitive abilities were enhanced in Jaeggi et al.'s (2008) study. One possibility is that the attentional control component of WMC was enhanced, thus influencing G*f*. However, it could also have been that the participants' visuospatial abilities were enhanced resulting in an increase in scores on the Raven's Advanced Progressive Matrices. An important element of Jaeggi et al.'s study is that their entire sample ($N = 70$) of participants was females.

The fact that their entire sample consisted of only females is important because there are known sex differences in visuospatial abilities *and* sex differences on the Raven's Advanced Progressive Matrices that could have been factors that were influential in Jaeggi et al.'s study and need to be considered for the following reasons. First, the effect of training might not exist for males because they already have an advantage in visuospatial abilities. Second, if males do experience an effect of training, they may not experience the same magnitude of increased scores on tests of G*f*.

Sex Differences in Visuospatial Abilities.

The investigation of sex differences in visuospatial abilities is best summarized by three meta-analyses conducted by Linn and Peterson (1985), Voyer, Voyer, and Bryden (1995) and Hyde (2005). Linn and Peterson divided visuospatial abilities into three categories: mental rotation (i.e., rotation of objects by means of mental imagery), spatial perception (i.e., identify spatial relations while distracting information is present), and spatial visualization (i.e., come to a correct solution that requires multiple stages of manipulating visual information). Studies used by Linn and Peterson in the meta-analysis were inclusive across the entire lifespan. Using Hedge's *g* as the effect size indicators, Linn and Peterson found a large effect size of sex differences for mental rotation (.73), a moderate effect size for spatial perception (.44), and a small effect size for spatial visualization (.13), all favoring males.

Voyer et al. (1995) included participants under 18 as well as over 18 and found significant effects for participants over the age of 18 for all three categories of visuospatial abilities, with males outperforming females. Using Cohen's *d* as the effect size indicator, the largest effect size observed was in mental rotation at .66, followed by spatial perception at .48, and an effect size of .23 for spatial visualization.

Two primary conclusions about sex differences in visuospatial abilities can be derived from the meta-analyses by Linn and Peterson (1985) and Voyer et al. (1995). First, the largest effect size for sex differences exists in mental rotation, followed by spatial perception, and the smallest effect size in spatial visualization with males outperforming females in all categories. Second, Voyer et al. divided participants into age categories (i.e., under 13, between 13 and 18, and over 18) and found that age is an important factor when determining sex differences for two of the three categories in visuospatial abilities. Age of participants was not important for the mental rotation task because there were differences in performance across all age groups. For spatial perception, Voyer et al. did not find sex differences for participants under the age of 13 but did so for ages 13 and above. Finally, Voyer et al. found no statistical significance between females and males in spatial visualization prior to the age of 18 years.

Hyde (2005) conducted a meta-analysis of other meta-analyses and found that males and females are more similar when considering multiple facets of a person's life such as cognitive abilities. Hyde also found that the largest difference that exist between males and females regarding cognitive abilities was in the visuospatial abilities category (i.e., Cohen's *d* ranging from .13 to .73). Moreover, there was a moderate effect size for sex differences observed in the Raven's Progressive Matrices for adults (.30), a small effect size for participants between ages 15 and 19 (.16), and a near nonexistent effect size for participants between 6 and 14 years (.02). Overall, based on meta-analyses spanning two decades, there is a robust effect size in sex differences in visuospatial abilities, especially for people older than 18.

The previously reviewed meta-analyses clearly indicate that females do not perform as well on visuospatial tasks as males. The evidence of females' lower performance level on visuospatial tasks may suggest that Jaeggi et al.'s (2008) findings of improved scores on the Raven's Advanced Progressive Matrices were a result of improving visuospatial abilities. The Raven's is suspected of being multidimensional such that visuospatial abilities are needed to solve the items. If the Raven's tests do require visuospatial abilities, then Jaeggi et al.'s results are in question because the increase in the participants' scores – who were all females – on the Raven's Advanced Progressive Matrices in Jaeggi et al.'s study could have been an artifact of increasing visuospatial abilities and not necessarily G*f* as defined by Horn and Cattell (1966).

Sex Differences in Raven's Matrices.

There is evidence of sex differences in visuospatial abilities between males and females (Voyer et al., 1995). Furthermore, there is evidence based on SEM models showing that visuospatial abilities are closely associated to G*f* (Kane et al., 2004). However, there are also strong correlations between what could be considered attentional control and G*f* (Engle et al., 1999; Kane et al., 2004). On the one hand, the dual *n*-back training used by Jaeggi et al. (2008) is a WMC task and could be training attentional control to manage information more efficiently, thus improving G*f* scores through that cognitive process. On the other hand, the dual *n*-back also has a strong visuospatial component with visual stimuli presented in different locations; thus, it could be improving visuospatial abilities leading to the increase in scores on the Raven's through that cognitive process. Research on sex differences in the Raven's Advanced Progressive Matrices provides evidence for the latter.

Since its inception, the Raven's Progressive Matrices have been accepted as a test of intelligence that is not biased toward males or females (Court, 1983; Jensen, 1998; Mackintosh & Bennett, 2005). However, a recent meta-analysis of 22 studies using the Raven's Standard and Advanced Matrices between 1939 and 2002 has shown that males outperform females with the difference in IQ points averaging between 3.2 points and 5 points (Irwing & Lynn, 2005). In a separate study, Lynn and Irwing (2004) found that males outperformed females on the Raven's Advanced Progressive Matrices by an average of 3.4 IQ points, which translates to an effect size of .23. Lynn and Irwing offered various theoretical explanations but no explanations with empirical evidence.

Other researchers have provided empirical evidence that the sex differences in the Raven's tests may be a result of a visuospatial component. For example, Colom, Escorial, and Rebollo (2004) had participants complete the Raven's Advanced Progressive Matrices and the spatial subsection of the Primary Mental Abilities Battery (PMA). Colom et al. found a high correlation (i.e., .34) between the Raven's and the spatial subsection of the PMA, suggesting that the Raven's Advanced Progressive Matrices contains a visuospatial component or the PMA requires G*f*. Furthermore, Colom et al. observed a significant sex difference in IQ (4.3 points) as measured by the Raven's test such that males performed better than females.

A similar result in sex differences on scores on the Raven's test was observed by Colom and García-Lopez (2002). The authors had participants complete the Raven's Advanced Progressive Matrices, Cattell's Culture Fair Test, and the inductive reasoning subsection of the PMA. Colom and García-Lopez found sex differences in the Raven's Advanced Progressive Matrices (i.e., male advantage of 4.2 IQ points), the PMA inductive reasoning subsection (i.e., female advantage of 2.9 IQ points), and no sex differences in Cattell's Culture Fair Test. It can be concluded that the difference between males and females scores on Raven's Advanced Progressive Matrices averages around 4 IQ points across studies and in favor of males. However, some researchers argue that this observed sex difference is not a result of the test requiring visuospatial skills (Abad, Colom, Rebollo, and Escorial, 2004). Abad et al. had participants complete the Raven's Advanced Progressive Matrices to determine if sex differences would be nullified by statistically removing items that were identified as items that required participants to use visuospatial abilities. They found that the male advantage decreased from 4.06 IQ points to 3.32 IQ points, but the difference was still significant.

The evidence is mixed when considering whether or not there are sex differences in the Raven's Advanced Progressive Matrices as a result of the test being multidimensional in that multiple cognitive functions are needed to complete the items. It does appear to require at least *some* spatial reasoning to solve items on the test. Empirical investigations have provided evidence for the possibility of multidimensionality of the Raven's Advanced Progressive Matrices, and evidence of sex differences in the test strengthen the argument for multidimensionality. While no decisive conclusions can be made about whether or not visuospatial abilities are an important aspect in solving problems on the Raven's test, the notion that visuospatial abilities were improved and not *Gf* still needs to be considered as a possible explanation for Jaeggi et al.'s (2008) results.

Sternberg's (2008) Critique

In a critique and review of Jaeggi et al.'s (2008) study, Sternberg (2008) pointed out eight limitations in their study: 1) a single WMC training task was used; 2) only the Raven's Advanced Progressive Matrices was used to measure G*f*, so results may be specific to this measure and not other indicators of G*f*; 3) there was no evidence of transfer of training beyond the Raven's Matrices test of intelligence; 4) there was no way to tell if the power of Raven's Advanced Progressive Matrices for predicting other cognitive abilities increased or decreased as a result of the training program; 5) there was no measure of the transfer of training over time; 6) no alternative task was given to the control group; 7) only one experiment was conducted, and 8) a very specific sample was used. As Sternberg pointed out, the limitations do not diminish the importance of Jaeggi et al.'s study; rather, the limitations set a stage for future research designed to investigate the possibility of improving G*f*. Similarly, the potential confound of sex differences in visuospatial abilities and the Raven's Matrices do not diminish Jaeggi et al.'s study either. Instead, the limitations and potential confounds allow for researchers to investigate important questions about intelligence and sex differences in cognitive abilities.

Summary

Intelligence is a topic of great controversy in psychology including how intelligence is conceptualized, how intelligence is measured, and the extent of biological bases of intelligence. The notion that G*f* can be improved through training WMC, as shown by Jaeggi et al.'s (2008) study, only adds to the controversies of measurement and conceptualization of G*f*. Jaeggi et al.'s study was the first of its kind to report an increase in scores on the Raven's Advanced Progressive Matrices as a result of cognitive training. There are many reasons to be skeptical of the evidence provided by Jaeggi et al. First, the dimensionality of the Raven's Matrices has been in question such that there is evidence that the test taps into visuospatial abilities (van der Ven & Ellis, 2000, Vigneau & Bors, 2008). Evidence that visuospatial abilities strongly influence *Gf* has also been provided in SEM procedures showing as strong of a relationship between visuospatial storage and *Gf* as the relationship between attentional control and *Gf* (Engle et al., 1999; Kane et al., 2004).

Second, the dual *n*-back task that was used as training in Jaeggi et al.'s (2008) study is a WMC task that requires a person to retain information in a short-term store while processing updated information. It is not clear why training worked: it may have improved attentional control component of WMC or it could have improved visuospatial abilities. Again, evidence from SEM models show that the causal mechanism in improving *Gf* could be attentional control or visuospatial storage because both constructs have a strong connection to *Gf* (Engle et al., 1999; Kane et al. 2004). Regardless of which component of WMC was improved by the dual *n*-back task, it is important to remember that WMC and *Gf* are not isomorphic (Ackerman et al. 2005). Therefore, it is questioned if *Gf*, as it is defined by Horn and Cattell (1966), was actually improved or if the increase in scores was the result of improving WMC, which is not the same as improving *Gf* because the two theoretical constructs do not share a one-to-one relationship (Ackerman et al., 2005).

A third reason to be skeptical of Jaeggi et al.'s (2008) study is the fact that their sample contained all females. It is well established that there are sex differences in visuospatial abilities with males having the advantage. If it is true that the dual *n*-back task improved only visuospatial skills that led to the increase in scores on the Raven's Advanced Progressive Matrices, then it could be that this training will not have as strong of an effect on males because of their existing advantage in visuospatial tasks. Although it may be an interesting finding that a training task to improve WMC can improve females' scores on a test of G*f*, the training should generalize to males as well if it is genuinely improving scores on tests of G*f*.

The implications of Jaeggi et al.'s (2008) study are significant at the basic level in the way philosophers, psychologists, and educators think about intelligence. Their study also has implications for the psychometric properties and uses of the Raven's Progressive Matrices Tests. The practical implications lie in what we know about how people learn in an education setting and how people are hired in the workforce when such tests are used during the hiring process. Therefore, a careful replication is needed to substantiate and expand the finding that training WMC actually improves G*f*.

The Current Study

The current study replicated and expanded on Jaeggi et al.'s (2008) study by changing the study's design so that sex differences and generalizability across other measures of G*f* could be empirically tested.

Replication of Results.

The primary limitation of Jaeggi et al.'s (2008) study was that only one study was conducted with no follow up studies. The current study addressed this issue by replicating and expanding Jaeggi et al.'s study to provide evidence whether the design, training, and improvement in G*f* was replicable. To ensure the quality of the replication, the same materials that Jaeggi et al. used were also used in the current study (see Materials section).

Generalizations of Training.

The limitations imposed by the use of a single working memory task (i.e., the dual *n*-back task) *and* no alternative tasks in Jaeggi et al.'s (2008) study was addressed by incorporating a visual *n*-back task, an auditory *n*-back task, and a spatial matrix span task into the design of the current study (see Methods section for a description of each task). The implementation of different training tasks to multiple groups addressed whether training generalizes to variations of the dual *n*-back task or if it is specific to the dual *n*-back task. The spatial matrix span task is a STM task that served as an alternative control task to address the limitation in Jaeggi et al.'s (2008) study that there was no group that was given an alternative task to act as a control.

Generalizations of Gains in Gf.

The limitation in Jaeggi et al.'s (2008) study that only one measure of G*f* was used was addressed in the current study by using multiple measures including the Raven's Advanced Progressive Matrices, which was used in the Jaeggi et al.'s study, the Wechsler Abbreviated Scale of Intelligence (WASI) Matrix Reasoning subset, Cattell's Culture Fair Test, and the BETA III Matrix Reasoning subset (see Materials section for a description of each test). These tests were selected because they are strong indicators of G*f*. Engle et al. (1999) found that Cattell's Culture Fair Test had a .74 loading (i.e., percent variance accounted for) and Raven's was .91 when both tests were used to measure G*f*. Kane et al. (2004) found that the WASI Matrix Reasoning Subset loaded on G*f* at .74, the BETA III Matrix Reasoning subset at .78, while the relationship with

Raven's was lower (.76) than the .91 Engle et al. found. Based on the evidence, all four of these measures are theoretically and statistically valid measures of G*f* and were used in the current study.

Generalizations beyond Gf.

The issue of transfer of training beyond one psychometric test was addressed by giving participants four cognitive tests during the pretest and the posttest session. The four cognitive tests were the Mental Rotation Test, the Paper Folding Test, the Lexical Decision Test, and an Extended Range Vocabulary Test (see Materials section in the second chapter for a description of each test). The cognitive tests were used to test if training transfers to the domain-specific tasks of visuospatial and verbal abilities. The two visuospatial tasks are the Mental Rotation Test and the Paper Folding Test and the two verbal tasks are the Lexical Decision Test and the Extended Range Vocabulary Test. Visuospatial and verbal skills are, for the most part, independent of each other (Halpern, 2000). Therefore, improvement in visuospatial skills should not transfer to verbal skills. Based on this reasoning, verbal abilities were tested to ensure domain specific improvement instead of improvements being made as a result of familiarity with visuospatial tasks (Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). Visuospatial and verbal abilities are quite different from one another and improvement in one domain should not transfer to another.

Generalizations in a Representative Sample.

The issue of Jaeggi et al.'s (2008) single sex sample was addressed by including men *and* women in the current study. Thus, the current study addressed many of the limitations raised by Sternberg (2008) and expanded on Jaeggi et al.'s findings.

Research Questions and Predictions

A number of research questions were created to help address the broad question, "Does training to enhance WMC improve G*f*?" To help make the research question clear, the following section presents each research question separately accompanied with a theoretical rational for the question. Also, each research question will be immediately followed by the corresponding prediction for that question.

Research question #1.

Does cognitive training using the dual *n*-back task improve scores on tests of G*f*? The first question is based on the theory that there is a strong connection between WMC and G*f* such that improving WMC will affect scores on tests of G*f*.

Prediction #1.

Only one published study (Jaeggi et al., 2008) has ever shown gains in scores on a test of G*f* after participants completed training. The prediction for the first research question is that scores on the Raven's Advanced Progressive Matrices, Cattell's Culture Fair Test, the WASI Matrix Reasoning subtest, and the BETA-III Matrix Reasoning subtest (see Materials section for detailed descriptions) would increase significantly for participants who complete the dual *n*-back task training compared to participants in the control group; thus, replicating Jaeggi et al.'s findings.

Research question #2.

Will participants in the dual, visual, and auditory *n*-back training groups and the STM training group experience greater gains on tests of G*f* than the control group? Training using *any* WMC task may improve scores on tests of G*f*, thus, the question is based on the same theoretical constructs as the first research question. Furthermore, it may be that training using a STM task may be effective as well and will act as a good control task. Furthermore, do participants who complete the dual *n*-back task training programs experience greater increases in scores on tests of G*f* compared to groups completing the visual or auditory *n*-back training or the STM training? The second part of the second research question is based on the theory that the more complex dual *n*-back task used to improve WMC and attentional control specifically will result in larger gains on tests of G*f*.

Prediction #2.

When each training group is compared to the control group, the dual and visual *n*-back training groups are expected to experience greater gains and the auditory *n*-back and STM training groups will experience marginal or nonsignificant gains. As previous research has shown, there is a strong positive relationship between visuospatial abilities and G*f* (Engle et al., 1999; Kane et al., 2004) and that visuospatial abilities are needed to solve items on the Raven's Progressive Matrices (van der Ven & Ellis, 2000). However, there is also a strong positive relationship between attentional control and G*f* as shown by Engle et al. and Kane et al. Based on these findings, participants who complete the training in the dual *n*-back task condition should show greater increases on measures of G*f* than participants who train with other methods. Even though the visual *n*-back task has the same visuospatial component as the dual *n*-back task, participants in the dual *n*-back task should have greater increases in scores because attentional control would be receiving more training as a result of monitoring two modalities instead of one. In addition to monitoring two modalities, the dual *n*-back task also trains both attentional control *and* visuospatial abilities, which are both important factors in G*f*. Participants who complete the training using the visual *n*-back task are also predicted to have greater increases in scores than participants who complete the auditory *n*-back task or the STM task. In addition, the increase in scores for participants who complete the auditory *n*-back task would be only slightly more compared to participants who complete a STM task because the attentional control component is trained in the auditory *n*-back task.

Research question #3.

Will participants' performance on the Extended Range Vocabulary Test, the Lexical Decision Test, the Mental Rotation Test, and the Paper Folding test (see Materials section for descriptions) improve as a result of the training? The third research question is based on the theory that WMC is not domain specific and that improving WMC may transfer to cognitive abilities other than G*f*.

Prediction #3.

Similar to the second prediction, no research has tested for improvement on the Extended Range Vocabulary Test, the Lexical Decision Test, the Mental Rotation Test, and the Paper Folding test after completing an extensive training program involving the *n*-back task or a STM task to use a base for predicting. There was no reason to suspect that any of the training conditions would improve performance on the Lexical Decision Test or the Extended Range Vocabulary Test because vocabulary is derived from crystallized intelligence, which is developed through experience and education (Horn & Cattell, 1966). However, RTs and accuracy should improve on both the Mental Rotation Test and the Paper Folding Test (Terlecki & Newcombe, 2005; Terlecki, Newcombe, & Little, 2008) because both WMC and STM training regimens have a strong visuospatial component, which may lead to an increase in attentional control for visuospatial information.

Research question #4.

If participants improve on the Mental Rotation Test and the Paper Folding Test, will there be differences in their improvement based on the type of training they received? The theoretical basis for the fourth research question is similar to the theoretical basis of the third research question with the exception that training attentional control using visual stimuli may have advantages over training without visual stimuli for performance on visuospatial tests.

Prediction #4.

Participants in the dual *n*-back and the visual *n*-back tasks should improve their performance (i.e., higher percentage correct and faster RTs) on the Mental Rotation Test and the Paper Folding Test because both training regimens have strong visuospatial and attentional control components, which should improve the ability to mentally manipulate visual information. Furthermore, participants in the auditory *n*-back task condition and the STM task condition should not have any significant increases in the Mental Rotation Test or the Paper Folding Test because auditory *n*-back task lacks a strong visuospatial component and the STM task does not train attentional control as strongly when compared to the dual and visual *n*-back training tasks.

Research question #5.

Will both females and males experience an improvement in visuospatial abilities? Furthermore, will female's improvement be greater than males in visuospatial abilities and tests of G*f*? The fifth research question is based on the theory that males have an advantage over females in visuospatial tests and that training may benefit females more than males because males may not have as much room to improve as females.

Prediction #5.

Previous research has shown that males outperform females by approximately 4 IQ points on the Raven's Advanced Progressive Matrices (Colom & García-Lopez, 2002; Lynn & Irwing, 2004; Irwing & Lynn, 2005) and that males perform better on visuospatial tests than females (Linn & Peterson; 1985; Voyer et al., 1995). Furthermore, previous research on sex differences for improving visuospatial abilities have shown that females experience greater gains from training than males (Cherney, 2008; Feng, Spence, & Pratt, 2007; Terlecki et al., 2008; Wright et al., 2008). Based on these findings, females' gains in scores between pretest and posttest on the four tests of G*f* should be greater than males. The greater gains by the females do not mean males will not experience gains, only that females will experience greater gains. However, the greater gains experienced by females also does not mean they would necessarily surpass males' performance. Instead, the sex difference in tests of G*f* would be reduced, perhaps to be nonsignificant or eliminated. Greater gains on tests of G*f* were expected for the participants in the dual *n*-back condition. For the visuospatial *n*-back condition, females are expected to still experience greater gains than males; however, these gains would not be as great as for females and males who were in the dual *n*-back task condition. Finally, there should be no difference in gains on scores of tests of G*f* between males and females who were in the auditory *n*-back task and the STM task conditions.

Similar patterns of improvement among different training groups and between females and males were made for the cognitive tests as were made for tests of G*f*.

Females in the dual *n*-back condition should improve their percentage correct and reduce RTs for items on the Mental Rotation Test and the Paper Folding Test. Males in the dual *n*-back condition should also experience gains but not as much as females. The prediction for females experiencing greater gains is based on previous research that found females gains' were greater but performance did not surpass males (Cherney, 2008; Feng et al., 2007; Terlecki et al., 2008; Wright et al., 2008). Participants in the visual *n*-back task would experience improvement in the Mental Rotation Test and the Paper Folding Test but the improvements would not be as strong as for participants in the dual *n*-back task. The previous prediction was made because the additional complexity of processing two modalities in the dual *n*-back task is expected to train attentional control more than it would in a single *n*-back task; thus, training the attentional control to be more efficient and lead to faster processing. Significant improvements on the Mental Rotation Test and the Paper Folding Test were not expected for participants in the auditory *n*-back task condition or the STM task condition because the auditory *n*-back task lacks a strong visuospatial component and the STM task does not train attentional control.

No improvements on the Extended Range Vocabulary Test were expected for participants in any of the conditions. Slightly faster RTs were expected for the Lexical Decision Test from participants who were in all *n*-back task conditions because attentional control was trained and may improve access to long-term memory. No sex difference in the Lexical Decision Test was predicted. No improvement in the Lexical Decision Test was expected for participants in the STM task condition.

Chapter Two

Method

Participants and Design

One of the goals of the current study was to obtain a diverse young adult sample with a range of cognitive capabilities. Thus, 136 participants (71 females, 65 males) were recruited at California State University, San Bernardino, which is a general admission, and public university and the Claremont Colleges, which are private and highly selective institutions. The majority of participants received course credit or extra credit for participating in the study. Participants were also entered in a raffle to win a $15 gift card. The raffle occurred once per week for each training group. A total of 110 people completed the training sessions and 26 people were in the control group (see Table 1 for more detailed description of participant numbers per cell).

Table 1

Number of Participants per Training Group by Sex

Training Group	Females	Males	Total
Dual *N*-Back	14	14	28
Visual N-Back	15	14	29
Auditory N-Back	13	12	25
Spatial Matrix Span	15	13	28
No Training	14	12	26
Total	71	65	136

Participants mean age was 22.48 (5.83) with no differences between men and women, t (134) = -.53, p = .60 and no statistically significant difference among training groups, F (4, 131) = 2.11, p = .08. Participants' education was also calculated such that education included kindergarten through twelfth grade (i.e., 13 years) plus years of college completed. Participants mean education was 15.57 (2.23) years with no difference between men and women, t (134) = .29, p = .77, and no differences among training groups, F (4, 131) = .63, p = .64. One male and two females were excluded from the data. The male was excluded because his daily average performance levels were double group's average. The two females did not go beyond *2*-back throughout the entire training program even though they were instructed each day on how to complete the task successfully, thus, they were excluded from the analysis.

Participants were randomly assigned to different training conditions regardless of the school in which they were enrolled. A loglinear test was conducted to determine if there was a significant three-way interaction between sex, training, and the university students were enrolled. There was no significant difference in the number of females and males from each school per training condition, χ^2 (13) = 2.99, p = .99. There was a similar concern for equal distribution among training groups. A loglinear test also showed that there were no differences in the distribution of participants to training groups based on the university they were enrolled in, χ^2 (4) = 1.39, p = .85.

A 2 (participants' sex; females and males) X 2 (repeated testing; pretest and posttest) X 5 (type of training; dual *n*-back, visual *n*-back, auditory *n*-back, spatial matrix span, and control) mixed design was implemented. Participants' sex and type of training were between subjects and repeated testing was within subjects. Participants were randomly assigned into one of five groups. Four of the groups received one type of cognitive training: a) the dual *n*-back task, b) the visual *n*-back task, c) the auditory *n*-back task, or d) the spatial matrix span, which is a STM task (see Materials section for a description of each type of training). The fifth group received no training and acted as a strict control group. To keep in accordance with Jaeggi et al.'s (2008) design, participants were trained five days a week (i.e., Monday through Friday) for four weeks for approximately 20 minutes each day. Participants were also divided into groups based on sex with approximately equal numbers of males and females in each group. For repeated testing, all participants' G*f* and cognitive abilities were assessed two times: one pretest and one posttest. The posttest was taken within four days of completing the final training session.

Materials

Training session tasks.

The key indicator that a task requires WMC instead of STM is the need to actively process information while maintaining information about the target (Kane et al., 2004). The dual *n*-back task, the visual *n*-back task, and the auditory *n*-back task were used as the three WMC training programs. The spatial matrix span task was used as the STM training program. The dual *n*-back task was used to fulfill the requirement of replicating Jaeggi et al.'s (2008) study. An *n*-back task that utilizes only the visual modality was used to test the possibility that the increase in scores on tests of G*f* was an artifact of improving visuospatial abilities. The auditory *n*-back task was used to control for the visuospatial component of the *n*-back task while still using the *n*-back paradigm. The auditory *n*-back task was used to partially determine how much of an influence the visuospatial component of the *n*-back task had on improving scores on tests of G*f*. Finally, the spatial matrix span was used as a STM training regimen and served as a control training regimen. All cognitive training regimens were presented using BrainTwister 1.0.2 (Buschkuehl, Jaeggi, Kobel, & Perrig, 2008). Each task is described in more detail.

Dual n-back task.

The dual *n*-back task is a complex task with multiple elements. First, the letter "n" in the term "*n*-back" refers to the number of trials back that a participant must remember that the target was presented. For example, a *2*-back task requires the participant to remember whether the presentation of a target stimulus was presented two trials prior to the current stimulus presentation. Thus, participants must continually update their memory to accommodate a new stimulus as a potential item to be remembered in the future.

The second important element to the *n*-back task is the presentation of stimuli. Participants are first presented with a white cross as a fixation point on a dark background for 2500 ms. Participants then simultaneously see a blue box appear in one of eight positions on the computer screen (i.e., top left, top center, top right, center right, bottom right, bottom center, bottom left, or center left; see Figure 8) and hear one of eight consonants (i.e., c, d, g, k, p, q, t, and v) in headphones. Although it seems that the consonants could be easily confused because of their similarity, Jaeggi, Buschkuehl, Etienne, Ozdoba, Perrig, and Nirkko (2007) and Jaeggi et al. (2008) did not report participants having any difficulty distinguishing between them. The visual and auditory stimuli are presented for 500 ms. The presentation of the stimuli is followed by an inter-stimulus interval (ISI) that shows only the white cross against a dark background, which lasts for 2500 ms. During the 2500ms ISI, the participant is supposed to press the letter "A" on the keyboard to indicate the blue square was presented *n*-back trials or to press "L" to indicate the consonant was presented *n*-back trials. A trial represents one presentation of a stimulus (500 ms) *and* one ISI (2500 ms); thus, one trial equal 3000 ms (see Figure 5).

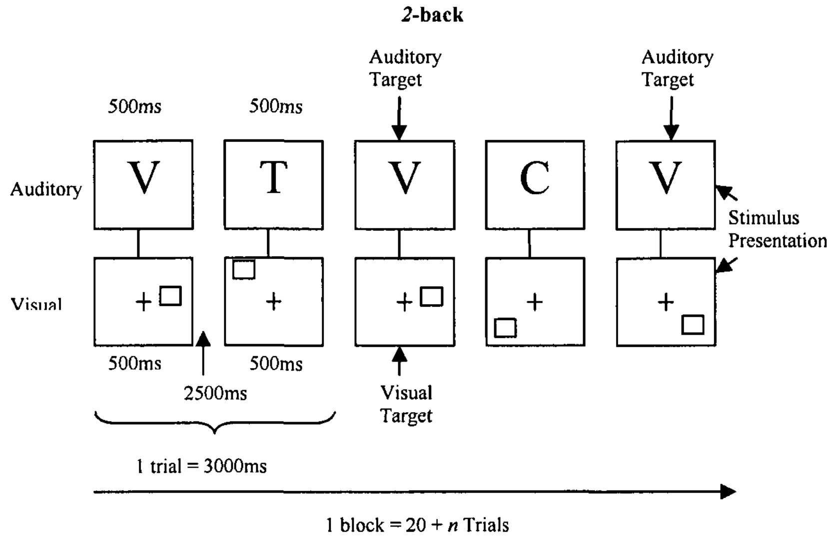


Figure 5: Display of five trials on the dual *n*-back task.

The third element to understanding the *n*-back task is that a block is a set of 20 + *n* trials. The *n* indicates that additional trials may be needed depending on the number of *n*-backs that a participant is asked to remember. For example, if the block is a three *n*-back task, then three additional trials will be needed. The challenging part of the dual *n*-back task is that it requires participants to constantly update their memory storage for two modalities at the same time in order to complete the task successfully.

The dual *n*-back task used by Jaeggi et al. (2008) varied in difficulty for each block depending on the participant's performance. Each training session began with a *2*-back task with a specified number of targets for the participant to remember. If the participant made more than five mistakes, then the next block would be decreased (e.g., to a *1*-back task). If the participant made fewer than three mistakes, then the next block would be increased (e.g., to a *3*-back task). If the participant made three to five mistakes, then the next block would remain at the same level (e.g., a *2*-back task). Jaeggi et al. found that, on average, participants make it to *6*-back on the nineteenth day of training.

All of the cognitive tasks used for training changed in difficulty based on the participant's performance for each block. The adapting feature of the dual *n*-back training was the same for every training regimen as that used in Jaeggi et al.'s (2008) study. All stimulus materials were developed by Jaeggi and her colleagues.

Visual n-back task.

The visual *n*-back task was used as an alternative WMC training regimen. This task was used to test whether the visuospatial component of the *n*-back task was responsible for the increase in scores on tests of G*f* in Jaeggi et al.'s (2008) study. The visual *n*-back task is similar to the dual *n*-back task with the exception that only the visual stimuli (i.e., the blue boxes) are presented. The goal for participants was the same as the dual *n*-back task in that they had to remember whether the target stimulus was presented in the previous one, two, three, or *n*-back trials. The visual *n*-back task adapted to the participants performance using the same rules as the dual *n*-back task.

Auditory n-back task.

The auditory *n*-back task was also used as an alternative for WMC training regimen. The auditory *n*-back task was used to further address two possible confounds in Jaeggi et al.'s (2008) study. First, it was used to determine if the visuospatial component of the dual *n*-back task was necessary for participants' scores on tests of G*f* to be increased. Second, it was also used to determine if the dual nature of a working memory task was responsible for the increase in scores without the need for visuospatial stimuli. The auditory *n*-back task is similar to the dual *n*-back task in that it uses the same consonants. It differs in that it does not present any visual stimuli. The task of the participants was the same as the visual *n*-back task in that the target stimulus was determined as being presented one, two, three, or *n*-back trials. This task also changed in difficulty for each block based on participants' performance.

Spatial matrix span.

The spatial matrix span task was given to participants as a control for the WMC training regimens. This task was used in Kane et al. (2004) and is considered to be a spatial STM task. The spatial matrix span is much like the game of memory and begins with a presentation of a 4 x 4 matrix on the computer screen. A blue circle is used as the stimulus and is presented in one of the sixteen cells at a time (see Figure 6). Each blue circle is presented for 500 ms. The order of cells in which the blue circle is presented is chosen randomly. For example, Figure 6 shows the blue circle appearing in the cell where the third column intersects the second row in the first trial. This is followed by the blue circle appearing in a cell where the second column intersects the fourth row in the second trial. The participants' task is to recall the sequence of blue circles appearing in the cells by moving the cursor and clicking the mouse in the appropriate cells in the correct order. The spatial matrix span training adapts to participant's performance in the same manner as the *n*-back tasks. The task starts with having the participant remember two positions the circle was presented in. If the participant remembers the sequence correctly, then the number of positions to remember increases by one. If the participant makes a mistake, then the number of positions decreases by one. Similar to the other tasks, the spatial matrix span was set on a timer of 20 minutes, but the training would continue past 20 minutes if the participant was in the middle of completing a block.

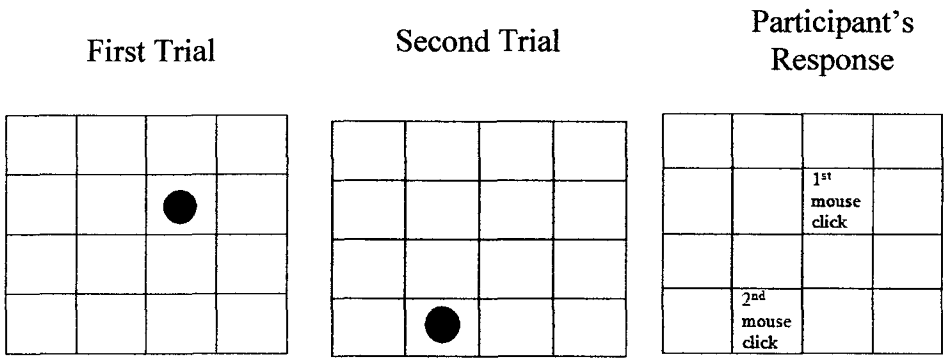


Figure 6: Sequence of events for the Spatial Matrix Span training. The words in the matrix labeled "Participant's Response" do not appear during the test. The words are simply indicating what the participant must do in order to complete the task correctly.

Measures of Fluid Intelligence.

Four measures were used to assess participants' *Gf*: Raven's Advanced Progressive Matrices, Cattell's Culture Fair Test, the Matrix Reasoning subtest for Wechsler's Abbreviated Scale of Intelligence (WASI), and the Matrix Reasoning subtest for the BETA-III. The tests of *Gf* were chosen because of their sound psychometric properties and because they have recently been updated in their standardizations and norms (Kellogg & Morton, 1999; Raven, 2000, WASI Manual, 1999). These tests were also used by Engle et al. (1999) and Kane et al. (2004) to load on *Gf* as a latent exogenous variable in their SEM procedures. More specifically, Engle et al. used the Raven's Advanced Progressive Matrices and Cattell's Culture Fair Test to measure *Gf*. Kane et al. used the Raven's Advanced Progressive Matrices, the WASI Matrix Reasoning subtest, and the BETA-III Matrix Reasoning subtest as measures of *Gf*. Participants were given practice items for each test before taking the actual test.

A time restriction was implemented on all measures of G*f*. The first and foremost reason for collecting response time measures was that Jaeggi et al. (2008) restricted time in their study. If the current study was to be a true replication, then the same procedures in measuring G*f* needed to be followed. The implementation of time restrictions on the Raven's Advanced Progressive Matrices does not reduce the reliability or validity. Jaeggi et al. stated that the correlation between timed and untimed versions of the Raven's Matrices is .95 and Raven, Raven, and Court (1998) stated that intellectual efficiency is more likely to be assessed when people are timed. Second, the Cattell's Culture Fair Test and the BETA-III Matrix Reasoning subtest were originally designed for time-limited administration. The WASI Matrix Reasoning subtest is not typically timed when used with the rest of its subtests, and using a time restriction could make the test less valid. However, Kane et al. (2004) put a 7-minute time restriction on the WASI subtest in their study; thus, a similar time restriction was used in the current study. Based on these reasons, all measures were administered with a time limit to maintain consistency in procedures across all tests.

Raven's Advanced Progressive Matrices.

The Raven's Advanced Progressive Matrices assesses a person's ability to come to a solution of a problem by reasoning inductively (Wilhoit & McCallum, 2003). Items in the Raven's Advanced Progressive Matrices are presented in a box approximately 12 cm x 8 cm. Eight black and white items are presented within the box in a 3x3 matrix with an empty space for the ninth cell. Eight choices, each with a different item, are provided below the box. The participants' task is to choose one of eight possibilities to complete the sequence (see Figure 1).

Items in all versions of the Raven's Progressive Matrices progress in their difficulty. Jaeggi et al. (2008) and Kane et al. (2004) used Set II of the Raven's Advanced Progressive Matrices and so did the current study. The current study implemented the same time restriction as Jaeggi et al. (2008) such that participants were only allowed 10 minutes to solve the 36 problems.

Cattell's Culture Fair Test.

The 1963 version of Cattell's Culture Fair Test, Scale 3, Form A, was used as a measure of *Gf.* Cattell's Culture Fair Test contains four problem subtests that are timed. Each subtest is taken with a time restriction and each subtest progresses in difficulty. The first problem subtest contains 13 items that test a person's ability to complete a series that show a progressive change in an object, shape, or figure. The items are presented in three boxes followed by a fourth box that is blank. The participants' task is to choose the best answer from six choices that will complete the series (see Figure 7). Participants are allowed three minutes to complete Test 1.

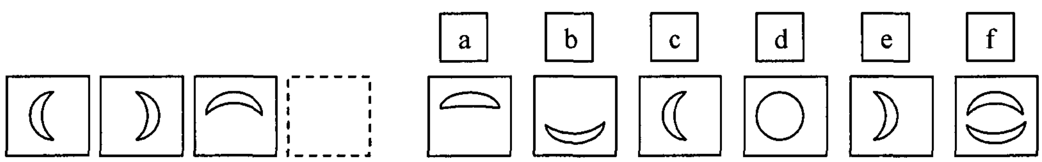


Figure 7: An example of an item on Test 1 in Cattell's Culture Fair Test. This example was created by the author and is constructed based on the same principle used in the original test. The answer is *b*.

The second problem subtest contains 14 items that test a person's ability to classify objects by identifying two objects, shapes, or figures that do not belong in a set of five (see Figure 8). Participants are allowed four minutes to complete Test 2.

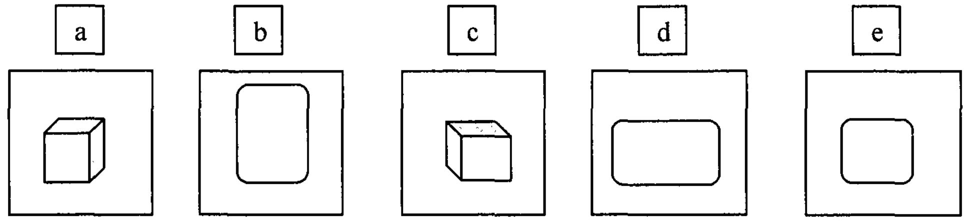


Figure 8: An example of an item on Test 2 in Cattell's Culture Fair Test. This example was created by the author and is constructed based on the same principle used in the original test. The answer is *a* and c.

The third problem subtest test contains 13 items that test a person's ability to complete a matrix with four to nine boxes in each matrix. One block of the matrix is missing and the participants' task is to choose the best figure from six choices to complete the matrix (see Figure 9). Participants are allowed three minutes to complete Test 3.

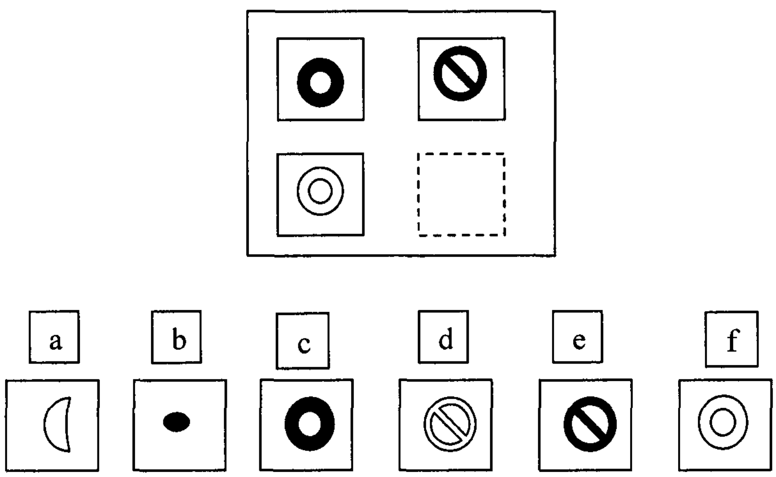


Figure 9: An example of an item on Test 3 in Cattell's Culture Fair Test. This example was created by the author and was constructed based on the same principle used in the original test. The answer is *d*.

The fourth problem subtest contains ten items with dots, lines, and various geometrical shapes in each item. Each item contains a cell on the left hand side that shows the relationship between the lines, dots, and shapes. The relationship of each item is based on rules that determine the placement of the dot among the other shapes and lines. The participants' task is to choose one item of five that shows the same relationship between the dot and the objects. In order to complete this task, the participants must 1) identify the rules governing the relationship between the dot and objects and 2) imagine the dot's position in the five choices to determine which one fits the same rules (see Figure 10). Participants are allowed two and a half minutes to complete Test 4.

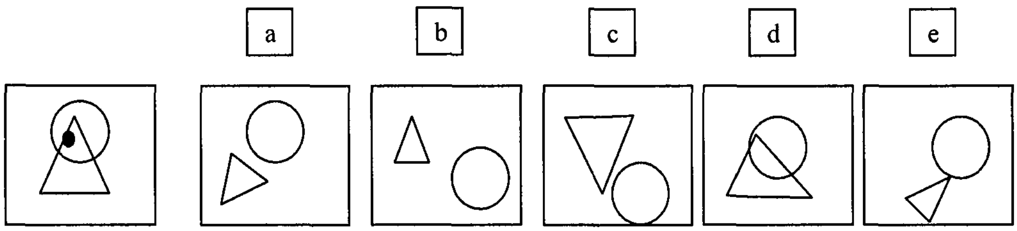


Figure 10: An example of an item on Test 4 in Cattell's Culture Fair Test. This example was created by the author and was constructed based on the same principle used in the original test. The answer is *d*.

WASI Matrix Reasoning subtest.

The Matrix Reasoning subtest is one of four tests in the WASI, which is an abbreviated test of the Weschler Adult Intelligence Scale (WAIS). Only the Matrix Reasoning subtest was given to participants because it is the portion of the test that assesses G*f.* The items are presented in a matrix with all but one of the cells containing colored figures. Below the matrix are five boxes that each contains a figure (see Figure 11).

Similar to the Raven's Progressive Matrices, the WASI Matrix Reasoning subtest presents figures in a matrix where each figure either remains constant or changes in each cell. The participants' task is to choose the answer from one of the five boxes to complete the sequence. A time restriction was implemented such that participants had five minutes to complete 29 items.

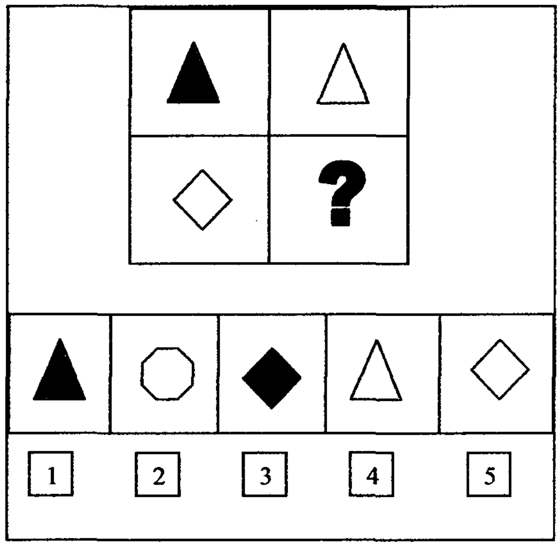


Figure 11: An example of an item on the WASI Matrix Reasoning subtest. This example was created by the author and was constructed based on the same principles as used in the original test. The answer is number *3*.

BETA-III Matrix Reasoning subtest.

The Matrix Reasoning subtest is one of five tests of the BETA-III. Only the Matrix Reasoning subtest (i.e., Test 5) was given to participants. BETA-III problems are presented in a 2 X 2 matrix that contain black and white figures and are very similar to WASI Matrix Reasoning subtest; thus, no figure is reproduced. Just like the WASI Matrix Reasoning subtest, each matrix in the BETA-III shows figures in a sequence with an empty cell. To the right hand side of the matrix are five cells that each contains a different figure. The participants' task is to choose the cell that contains the figure that will complete the sequence. The test contains 25 questions and, based on the test's instructions, participants were given five minutes to complete all questions.

Cognitive assessments.

Four cognitive tests were used to determine if the WMC training transfers beyond a psychometric test of *Gf*. Because tests of *Gf* are thought to have visuospatial components (van der Ven & Ellis, 2000), the Mental Rotation Test (Shepard & Metzler, 1971; Vandenberg & Kuse, 1978) and the Paper Folding Test (Ekstrom, French, Harman, & Dermen, 1976) were used to test visuospatial abilities. If the dual *n*-back task is improving visuospatial abilities, then there should be an improvement of performance on the two visuospatial tests. These two visuospatial abilities tests were used because the Mental Rotation Test taps into mental rotation abilities and the Paper Folding Test taps into spatial visualization abilities (Linn & Peterson, 1985). The Extended Range Vocabulary Test (Ekstrom et al., 1976) and the Lexical Decision Test (Ratcliff, Gomez, & McKoon, 2004) were used because they are tests of verbal abilities and a goal of the current study is to determine if WMC training will transfer to tasks that are not visuospatially based. All cognitive tests were presented on SuperLab 4.0 and were given during the pretest and posttest sessions. Each test is discussed separately.

Mental rotation test.

Vandenburg and Kuse's (1978) Mental Rotation Test was used as a test of participants' ability to mentally rotate an object. The paper-and-pencil version was scanned into a computer and programmed to be presented in SuperLab 4.0 to record response times. The stimuli for the Mental Rotation Test represent three-dimensional objects made of ten small blocks (see Figure 12). Three different objects are presented on the right hand side of the computer screen and it is the task of the participants to match one of three objects with a target object (presented on the left hand side of the computer screen).

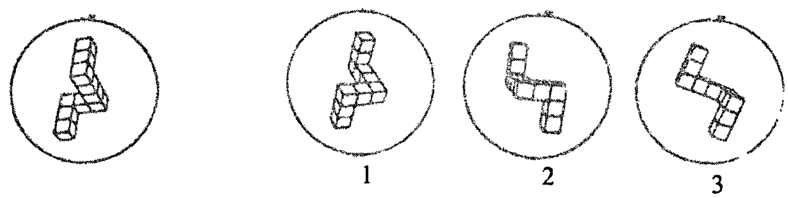


Figure 12: Example of a practice item from the Mental Rotation Test. The answer is number *2*.

Paper folding test.

The Paper Folding Test was adapted from the Educational Testing Service kit (Ekstrom et al., 1976). Similar to the Mental Rotation Test, all stimuli were scanned into a computer and presented on SuperLab 4.0 so RTs could be measured. The Paper Folding Test stimuli consist of a step-by-step presentation of a paper being folded with a minimum of one fold to a maximum of three folds. The final step shows a hole punched in the paper. The task of the participant is to choose from five choices what the punched paper would look like unfolded (see Figure 13).

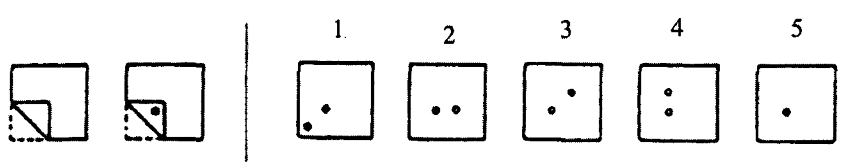


Figure 13: Example of a practice item in the Paper Folding Test. The answer is number *1*.

Extended range vocabulary test.

The Extended Range Vocabulary Test is another test from the Education Testing Service kit (Ekstrom et al., 1976). This test was also presented on the computer instead of paper-and-pencil to collect response times, and is a standard vocabulary test with 48 words. The test is presented such that a target word is given in the middle of the screen. Five different words are presented below the target word and the participants' task is to choose which word best defines the target word by pressing a number on the keyboard that is associated with that word. For example, the target word might be "bantam." The choices for this target are, "1) fowl, 2) ridicule, 3) cripple, 4) vegetable, and 5) ensign." The correct answer is "fowl."

Lexical decision test.

The Lexical Decision Test (Ratcliff et al., 2004) uses presentation of nonwords and words. All nonwords were selected using the ARC Nonword Database (Rastle, Harrington, & Coltheart, 2002). Twenty nonwords were selected to be monosyllable and be pronounceable nonwords such as "yumph." All words were selected using the MRC Psycholinguistic Database (Coltheart, 1981). Twenty monosyllable, high-frequency English words such as "staff" were selected randomly. Words and nonwords are presented in random order on a computer screen. The participant's task is to indicate if the target stimulus is a real word or a nonword by pressing one of two different keys on a keyboard. Rather than being a measure of the participants' understanding of the meaning of the words, which is the purpose of the Extended Range Vocabulary Test, the Lexical Decision Test measures the speed of a participants' access to verbal information to recognize a word based on its orthography (Halpern & Wai, 2007).

Procedure

Pretest session.

Participants were given the tests of *Gf* and the cognitive tests one week prior to the first training session. Participants were seated in a well lit room that was free of distractions. Participants were randomly assigned to complete either the tests of *Gf* or the cognitive tests first. To avoid any practice effects or other systematic confounds, the order of the tests of *Gf* and the cognitive tests were also randomized for each participant. Participants were allowed 10 minutes to complete the Raven's Advanced Progressive Matrices, approximately 13 minutes to complete Cattell's Culture Fair Test, 5 minutes to complete the WASI Matrix Reasoning subtest, and 5 minutes to complete the BETA III Matrix Reasoning subtest. The total time to take the tests of *Gf* was approximately 55 to 60 minutes, including time to give instructions, answer any questions the participant may have, and actual testing time. The cognitive tests did not have any time restrictions and took approximately 15 to 30 minutes to complete. The total time to complete all tests ranged from 70 to 90 minutes.

Training sessions.

Participants were seated in front of a computer monitor, keyboard, and mouse in a room that was well lit and free of distractions. Participants who were assigned the dual *n*-back task or the auditory *n*-back task wore headphones to hear the consonants being spoken. All training sessions were on a computer program called BrainTwister 1.0.2 designed by Buschkuehl, Jaeggi, Kobel, and Perrig (2008). BrainTwister is designed specifically to act as a cognitive training program that saves trainees' demographic information as well as their progress. Thus, an identification code was used for each participant, which was entered into the program each day by the participant to begin the training session. Once the identification code was entered, a set of instructions appeared on the screen. For the first two days of training, the experimenter provided verbal instructions for the participants in addition to the written and visual instructions to make sure the participant understood the task. The actual training stimuli began after the participant pressed the space bar on the keyboard. Each participant completed 18 to 20 training sessions and each session lasted approximately 20 to 22 minutes. The computer program indicated the completion of the daily session to the participant. Attrition and missed training sessions was a concern for this study; therefore, strict criteria were set in case a person missed a training session. First, if the participant did not show up for the first two training sessions, then they were not allowed to continue the study. Second, if a participant missed a training session, then they were allowed to make up the training session on the following day by doing two training sessions. However, participants were not allowed to make up more than two training sessions. Only 5% of the participants who completed training did not have to make up a training session.

Posttest session.

The same tests of G*f* and cognitive tests used in the pretest session were used in the posttests. The same testing conditions and random ordering procedures were also used. The posttest was given no later than four days after the last training interval. Table 2 provides a detailed timeline for the testing and training schedule.

Table 2

Procedure Shown in Chronological Order with Hours and Number of Sessions per Week

Week	1	2	3	4	5	6
Event	Pretest	Training	Training	Training	Training	Posttest
Number of Sessions	1	5	5	5	5	1
Time Taken	1½ hours	1 hour 40 min	1 hour 40 min	1 hour 40 min	1 hour 40 min	1½ hours

Chapter Three

Results

The results of the current study are divided into five sections. The first section provides information about the training sessions to describe how many participants completed each type of training, their daily average performance and daily maximum performance, and whether there were any differences in training performance between females and males and among training groups. The second section provides descriptive information about the pretest session, which includes scores on each test of G*f*, the percent correct on the cognitive tasks, and the RTs for correct answers on the cognitive tasks. The third section replicates the second section for the posttests. The fourth section provides the correlations among the tests of G*f*, among the tests of G*f* and percent correct on cognitive tests, among tests of G*f* and RTs for correct answers on cognitive tests, and among percent correct on cognitive tests and RTs for correct answers on cognitive tests. The fifth section provides the analyses that tested the hypotheses directly for differences between pretest and posttest scores in each test of G*f*, percent correct on each cognitive task, and RTs for each cognitive task. Results of specific contrasts are also provided in the fifth section.

It is also important to note that all findings are reported with an unadjusted alpha level and a significance level of .05. Unadjusted values are used because a number of a priori comparisons are made and some important patterns in the data may fail to attain statistical significance with a highly conservative adjusted alpha level, when in fact, those differences are important and need to be recognized.

Training Performance

Analyses of the training sessions are divided into three sections. The first section describes the number of sessions completed by each training group. The second section describes the progression of the daily average performance for females and males per training group. The third section replicates the second section using the daily maximum performance levels.

Number of Sessions Completed.

Means and standard errors for the number of completed training sessions and are presented in Table 3. A mixed one-way ANOVA was used to calculate any differences between females and males and among training groups for the number of training sessions completed. There were no differences for overall average training sessions that were completed between females and males, $F\,(1, 102) = 1.08, p = .37, \eta^2 = .26$, or among training groups, $F\,(3, 102) = 2.61, p = .23, \eta^2 = .72$. There was also no interaction between sex and training, $F\,(3, 102) = .17, p = .92, \eta^2 = .01$.

Table 3

Means and Standard Errors for Training Sessions Completed by Males and Females per Training Group

	Females	Males	Total Average
Dual *N*-Back	19.29 (.22)	19.14 (.22)	19.21 (.16)
Visual N-Back	19.53 (.21)	19.36 (.22)	19.45 (.15)
Auditory N-Back	19.30 (.23)	19.25 (.24)	19.28 (.17)
Spatial Matrix Span	19.20 (.21)	19.31 (.23)	19.25 (.16)
Total Average	19.33 (.11)	19.26 (.11)	19.30 (.08)

Note: $n = 110$

The minimum number of training sessions required for participants to complete to take the posttest was 18 sessions. Thus, participants may have missed one or two sessions in the process. If a person missed a session in the middle of the training process, then the next session would count as the missed session. For example, if a person missed session 10, then session 11 became session 10; thus, there are missing data only for sessions 19 and 20. There are missing data from 21% of training session 19 and 47% missing from training session 20. A MANOVA showed no statistical differences for missing data in training sessions 19 and 20 between females and males, $F\,(2, 101) = .04$, $p = .96$, $\eta^2 = .00$, or among training groups, $F\,(6, 202) = .36$, $p = .90$, $\eta^2 = .01$. There was also no significant interaction between sex and training for missed sessions, $F\,(6, 202) = 1.83$, $p = .10$, $\eta^2 = .05$. Table 4 shows the number of participants who did not complete training sessions 19 and 20. Jaeggi et al. (2008) found no difference between participants who completed 17 sessions or 19 sessions when measuring gains on the Raven's Advanced Progressive Matrices. Because Jaeggi et al. found no difference between participants who completed 17 vs. 19 sessions and there are missing data in the current study for sessions 19 and 20, only sessions 1 through 18 will be analyzed for the remainder of the results.

Table 4

Number of Females and Males Who Missed 1 or 2 Training Sessions per Training Group

	Missed 1 Session		Missed 2 Sessions	
	Females	Males	Females	Males
Dual *N*-Back	3 (11)	4 (10)	7 (7)	7 (7)
Visual N-Back	1 (14)	4 (10)	6 (9)	5 (9)
Auditory N-Back	3 (10)	3 (9)	6 (7)	5 (7)
Spatial Matrix Span	5 (10)	1 (12)	7 (8)	8 (5)

Note: $n = 110$. Numbers in parentheses indicate the total number of participants who completed 19 training sessions as indicated under "Missed 1 Session" or 20 training sessions as indicated under "Missed 2 Sessions."

Daily Average Performance.

From this point forward, daily average performance refers to the average *n*-back that was reached for participants in the WMC training conditions or the number of sequences participants were able remember in the STM training condition. The overall daily average performance for each session was calculated for each sex per training group and is provided in Appendix A. A graphic display of the linear progression for the daily average performance for females and males per training group is provided in Figure 14.

There was a change in daily average performance over the 18 training sessions for all training groups, $F\,(17, 86) = 18.42, p < .001, \eta^2 = .79$. There was no significant two-way interaction between sex and change in the daily average performance over the 18 sessions, $F\,(17, 86) = 1.21, p = .28, \eta^2 = .19$. However, there was a significant two-way interaction between training groups and change in the daily average performance, $F\,(51, 257) = 1.56, p < .009, \eta^2 = .24$. To determine *where* the differences existed among training groups, the daily average performance for training session 1 was subtracted from training session 18 to calculate the total improvement made in each group. Between-subjects LSD unadjusted post-hoc analyses were conducted to determine differences among training groups. The post-hoc analyses revealed that no statistically significant differences existed among the three *n*-back training groups, but that differences existed between each *n*-back training group and the spatial matrix span group (see Table 5).

Table 5

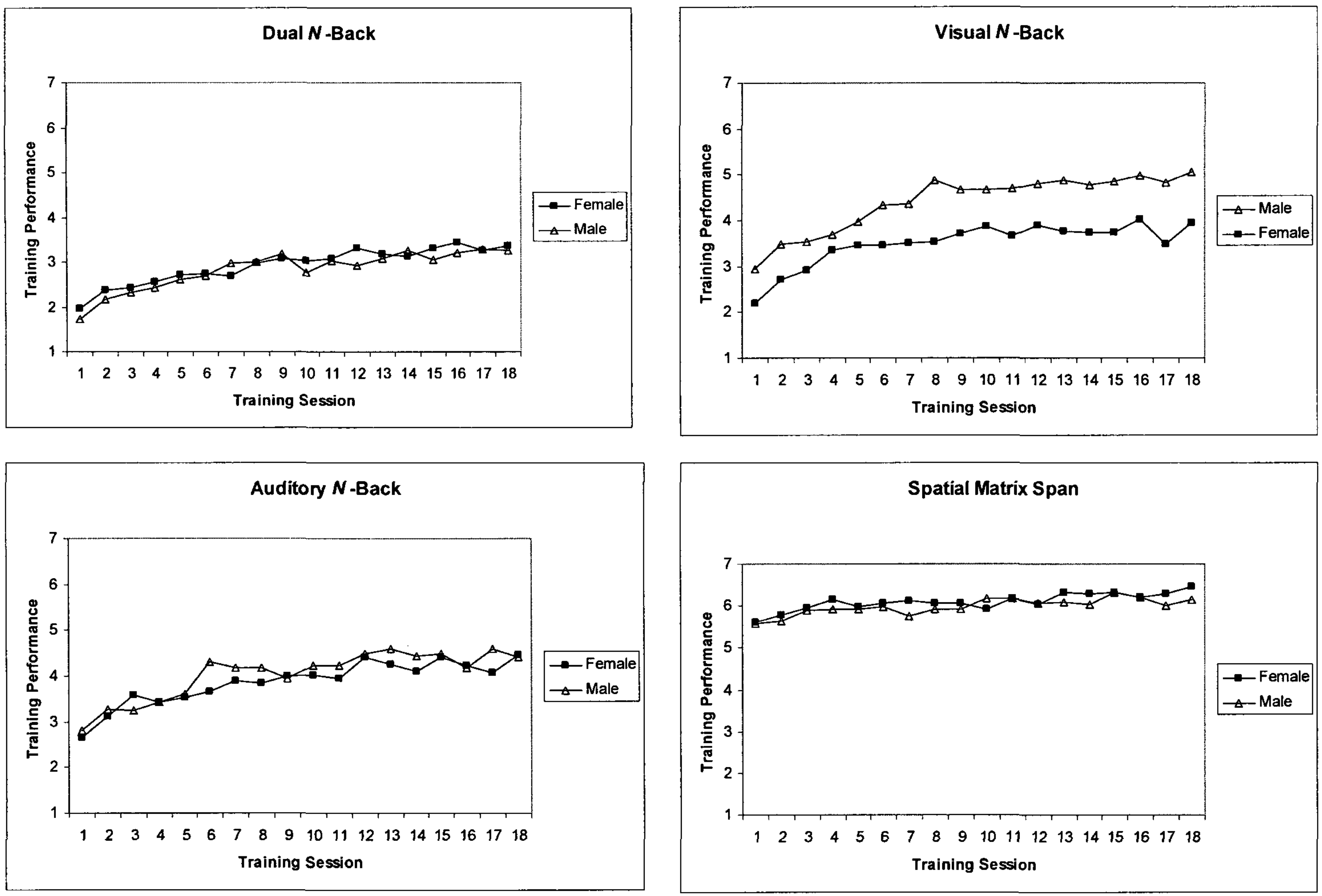
Post-Hoc Comparisons of Differences in Gains between Training Session 1 and 18

Comparison	*Mean Difference*	*SE*	*p*	95% CI *LL*	95% CI *UL*
Dual vs. Visual	-.47	.27	.085	-1.01	.07
Dual vs. Auditory	-.25	.28	.377	-.81	.31
Dual vs. Spatial*	.74	.27	.008	.20	1.29
Visual vs. Auditory	.22	.28	.428	-.33	.78
Visual vs. Spatial*	1.22	.27	.001	.68	1.75
Auditory vs. Spatial*	-.99	.28	.001	.43	-.43

Note. A positive Contrast indicates that the gain for the first group in the comparison was larger. Asterisk indicates significant ($p < .05$) difference between training group pair. SE = Standard Error. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Figure 14

Time Series for Daily Average Performance per Training Group for Females and Males



Finally, the potential 3-way interaction that includes training sessions, sex, and training groups was not significant, $F\,(51, 257) = .99, p = .51, \eta^2 = .16$. Independent *t*-tests were also conducted for each training session per training group to test for differences in daily average performance between females and males. The visual *n*-back training group was the only training group that had a consistent statistical significant difference between females and males in daily average performance level such that males performed significantly higher 14 out of 18 training sessions.

A repeated measures analysis showed that the overall trend for all four groups combined was linear, $F\,(1, 102) = 165.50, p < .001, \eta^2 = .62$, with a significant quadratic trend, $F\,(1, 102) = 105.50, p < .001, \eta^2 = .51$, and a statistically significant but much smaller cubic trend, $F\,(1, 102) = 17.29, p < .001, \eta^2 = .16$. A repeated measures analysis was also conducted for each training group to test for linearity over the 18 training sessions (see Table 6).

Table 6

Significant Trends in Daily Average Performance per Training Group

	F	*df*	*p*	η^2
Dual *N*-Back				
Linear	65.74	1, 26	.01	.72
Quadratic	43.22	1, 26	.01	.62
Visual *N*-Back				
Linear	79.68	1, 27	.01	.75
Quadratic	49.87	1, 27	.01	.65
Auditory *N*-Back				
Linear	30.72	1, 23	.01	.57
Quadratic	33.58	1, 23	.01	.59
Spatial Matrix Span				
Linear	13.00	1, 26	.01	.33
Quadratic	1.60	1, 26	.22	.06

All training groups with an *n*-back task had a statistically significant linear trend and a statistically significant quadratic trend. To determine the nature of the quadratic trend, the 18 training sessions were split in half and the change in performance was calculated for each half. Thus, session 1 was subtracted from session 9 to determine the change in the first half of the training sessions and session 10 was subtracted from session 18 to determine the change in the second half of the training sessions.

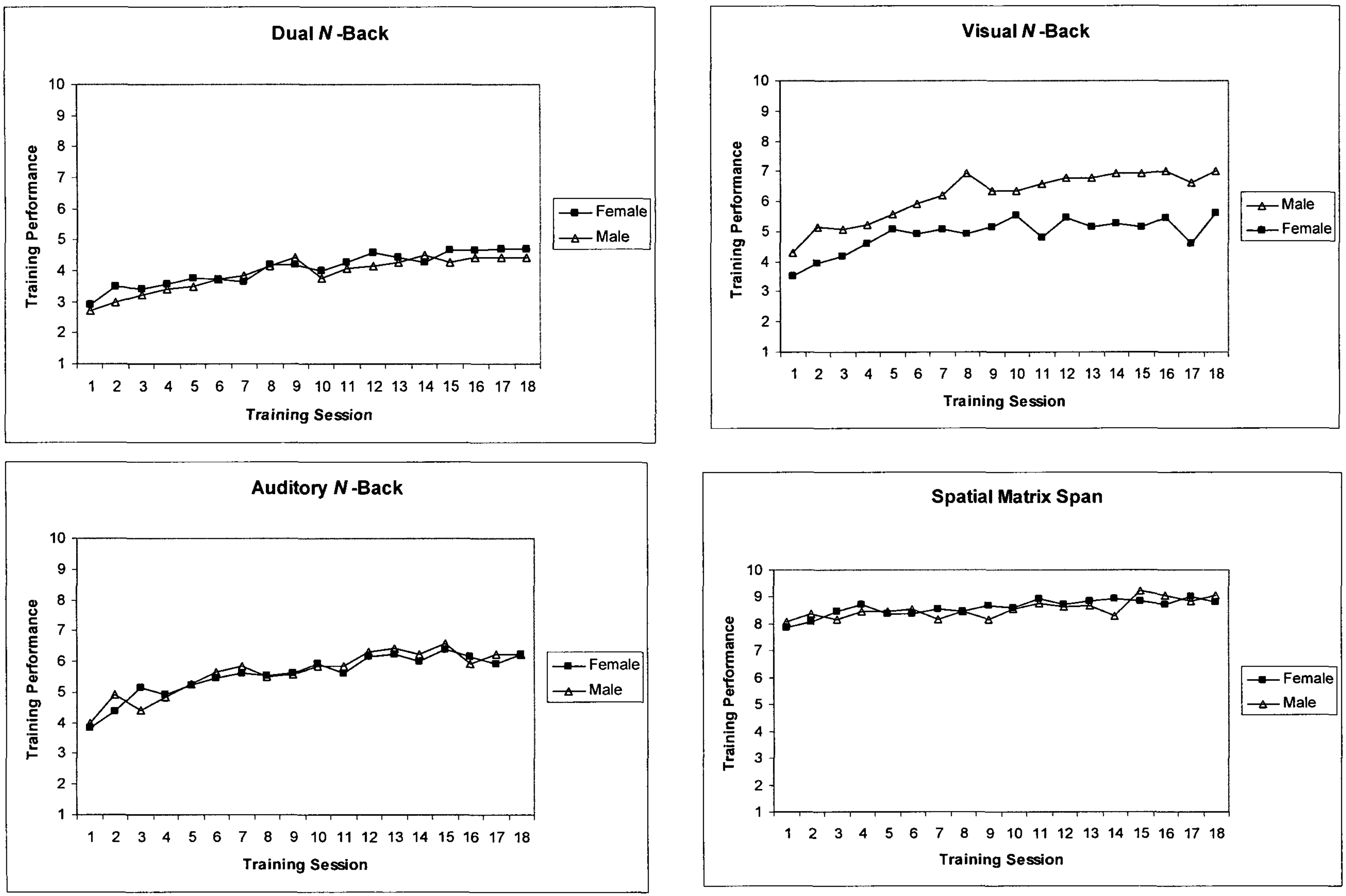
A paired *t*-test was used to determine if there was a difference in change between the first half and second half of the training sessions for each training group. There was a greater change in performance for the first half of the training sessions than the second half for the dual *n*-back training group, $t\ (27) = 6.90, p < .001, d = 2.66$, the visual *n*-back training group, $t\ (28) = 6.27, p < .001, d = 2.37$, and the auditory *n*-back training group, $t\ (24) = 4.00, p < .001, d = 1.63$. There was no difference between the first and second half of the training sessions for the spatial matrix span training group, $t\ (27) = .71, p = .48, d = .27$. The findings of a greater change in performance for the first nine training sessions by the groups with an *n*-back task suggests there was an asymptote in performance for the last nine training sessions whereas the change in performance for the spatial matrix span training group was more consistent throughout the 18 sessions.

Daily Maximum Performance.

The daily maximum performance refers to the maximum *n*-back reached by the WMC training groups or the maximum number of sequences remembered by the STM training group. The daily maximum performance for each training session was calculated for each sex per training session and is provided in Appendix B. A graphic display of the linear progression of the daily maximum performance per training group for females and males is shown in Figure 15.

Figure 15

Time Series for Daily Maximum Training Performance per Training Group for Females and Males



The same tests of statistical significance used for the daily average performance were used to test for differences in the daily maximum performance. There was a change in the daily maximum performance throughout the 18 training sessions, $F\,(17, 86) = 13.71, p < .001, \eta^2 = .73$. There was no significant two-way interaction between sex and change in the daily maximum performance over the 18 training sessions, $F\,(17, 86) = .83, p = .65, \eta^2 = .14$. Independent *t*-tests found a significant difference between females and males in the visual *n*-back training group for differences in daily maximum performance levels for 14 out of 18 training sessions.

As before with daily average performance, there was a significant difference among training groups for the change in the daily maximum performance, $F\,(51, 257) = 1.66, p < .006, \eta^2 = .25$. Table 7 provides post-hoc analyses that shows where the differences existed, which followed the same logic as calculating the difference in daily averages. The potential three-way interaction between training sessions, sex, and training groups was not significant, $F\,(51, 275) = 1.00, p = .51, \eta^2 = .16$

Table 7

Post-Hoc Results for Differences in Gains on Average Maximum between Training Sessions 1 and 18

Comparison	*Mean Difference*	*SE*	*p*	95% CI *LL*	95% CI *UL*
Dual vs. Visual	-.63	.41	.13	-1.44	.18
Dual vs. Auditory	-.57	.43	.18	-1.42	.28
Dual vs. Spatial	.79	.41	.06	-.04	1.61
Visual vs. Auditory	.06	.42	.89	-.18	.90
Visual vs. Spatial*	1.42	.41	.01	.60	2.23
Auditory vs. Spatial*	1.36	.43	.01	.51	2.20

Note. A positive Contrast indicates that the gain for the first group in the comparison was larger. Asterisk indicates significant ($p < .05$) difference between training group pair. SE = Standard Error. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

The linear trend for all training groups combined was significant, $F\,(1, 102) = 131.96, p < .001, \eta^2 = .56$, as well as the quadratic component, $F\,(1, 102) = 66.61, p <$

.001, $\eta^2 = .40$, and the cubic component, $F(1, 102) = 6.48, p < .01, \eta^2 = .06$. A repeated measures analysis was conducted for each training group to test for the linearity in the daily maximum performance across the 18 training sessions (see Table 8).

Table 8

Significant Trends in Daily Average Performance per Training Group

	F	*df*	*p*	η^2
Dual *N*-Back				
Linear	74.92	1, 26	.01	.74
Quadratic	16.37	1, 26	.01	.39
Visual *N*-Back				
Linear	41.19	1, 27	.01	.60
Quadratic	34.63	1, 27	.01	.56
Auditory *N*-Back				
Linear	27.61	1, 23	.01	.55
Quadratic	27.22	1, 23	.01	.54
Spatial Matrix Span				
Linear	13.94	1, 26	.01	.35
Quadratic	.09	1, 26	.69	.01

Similar procedures were used to determine if there were greater changes in the daily maximum performance for the first half of the training sessions compared to the second half. Paired *t*-tests showed there was a greater change in performance for the first half of the training sessions than the second half for the dual *n*-back training group, $t(27) = 3.76, p < .001, d = 1.45$, the visual *n*-back training group, $t(28) = 4.38, p < .001, d = 1.65$, and the auditory *n*-back training group, $t(24) = 3.83, p < .001, d = 1.56$. There was no difference between the first and second half of the training sessions for the spatial matrix span training group, $t(27) = .36, p = .73, d = .14$. A similar asymptote pattern existed in the daily maximum performance as the daily average performance such that training groups with the *n*-back task experienced less change in the second half of the training sessions and the spatial matrix training group experienced a more consistent change across the 18 training sessions.

Summary.

Participants in each training group improved their daily average performance and daily maximums performance over the span of 18 training sessions. There were no significant sex differences for the improvements made, but there was a statistically significant difference between females and males in the visual *n*-back training group such that the males outperformed females in 14 out of the 18 training sessions. There were some differences between the spatial matrix span training group and each *n*-back training group for change in performance throughout the training sessions. Finally, although all training groups had a strong linear trend showing improvement over the 18 sessions, the improvement was most rapid during the first half of the *n*-back training whereas the spatial matrix span training showed small but continuous improvement over 18 sessions.

Pretest Performance

The current section provides a descriptive summary of the pretest data. The tests of G*f* are summarized first, followed by the percent correct on cognitive tests, and then RTs for correct answers on cognitive tests.

Pretests of Gf.

Overall mean scores for tests of G*f* were calculated based on the number correct. A MANOVA was conducted to determine any differences among training groups and between females and males in scores on pretests of G*f* (see Table 9).

Table 9

MANOVA Results Testing for Differences among Training Groups and between Females and Males in Scores on Pretests of Gf

Test of G*f*	*F*	df	*p*	η^2
<u>Raven</u>				
Sex	1.62	1, 126	.21	.01
Training	.82	4, 126	.51	.03
Sex X Training	1.97	4, 126	.10	.06
<u>Cattell</u>				
Sex	2.10	1, 126	.15	.02
Training	.63	4, 126	.64	.02
Sex X Training	1.68	4, 126	.16	.05
<u>WASI</u>				
Sex	.45	1, 126	.51	.00
Training	1.59	4, 126	.18	.05
Sex X Training	1.26	4, 126	.29	.04
<u>BETA-III</u>				
Sex	1.05	1, 126	.31	.01
Training	.38	4, 126	.83	.01
Sex X Training	2.05	4, 126	.09	.06

Overall, there were no statistically significant differences between females and males or among training groups on pretest scores of G*f* or differences between females and males within training groups. Post hoc analyses showed marginal differences in pretest scores on the WASI subtest between the visual *n*-back training group and the dual *n*-back training group, $M_{Diff.} = 1.30, p = .06$; the auditory *n*-back training group, $M_{Diff.} = 1.34, p = .06$; and the control group, $M_{Diff.} = 1.35, p < .05$, such that the visual *n*-back training group had lower scores. The difference of the visual *n*-back training group is not a concern because these are marginal differences and did not differ with the other groups on the other tests of G*f*; rather, it is simply a fact in the data to be recognized.

Independent *t*-tests were also conducted to test if there were sex differences within training groups that may have been missed in the MANOVA for scores on pretests of G*f*. There was a difference between females and males who were in the visual *n*-back training group on scores for the Raven's and Cattell pretests (see Table 10 for statistical significance values). The differences observed are not a concern because the primary question of the current study is if improvements in scores are made as a result of training regardless of the pretest scores. However, though participants were assigned randomly to training conditions, there was a difference between females and males in the visual *n*-back training group in performance that may reflect in part a pretest difference.

Table 10

Means, Standard Errors, and Differences between Females and Males for Number Correct on Pretests of Gf

	Females		Males		Total Average					95% CI		
	M	*SE*	*M*	*SE*	*M*	*SE*	*t*	*df*	*p*	*LL*	*UL*	Cohen's *d*
Raven's												
Dual	15.29	.85	14.71	1.13	15.00	.69	.41	26	.69	-2.33	3.47	.16
Visual	11.20	1.02	15.57	1.40	13.31	.94	2.55	27	.02	-.85	7.89	.98
Auditory	13.54	1.00	14.25	1.12	13.88	.74	.48	23	.64	-2.38	3.81	.20
STM	13.07	.86	13.62	1.33	13.32	.76	.35	26	.73	-2.63	3.73	.14
Control	14.43	.82	13.67	.99	14.08	.63	.60	24	.56	-1.87	3.39	.24
Overall Average	13.46	.43	14.40	.54	13.91	.34	1.37	134	.17	-.45	2.29	.27
Cattell												
Dual	27.57	.98	26.64	1.30	27.11	.80	.57	26	.57	-2.42	4.28	.22
Visual	24.13	1.04	27.79	1.37	25.90	.91	2.14	27	.04	-.15	7.16	.82
Auditory	23.85	1.10	27.17	1.70	25.44	1.03	1.67	23	.11	-.81	7.45	.69
STM	26.13	.62	25.77	1.27	25.96	.67	.27	26	.79	-3.16	2.43	.11
Control	25.50	1.34	25.33	1.19	25.42	.89	.09	24	.93	-3.59	3.92	.04
Overall Average	25.45	.48	26.57	.61	25.99	.38	1.46	134	.15	-.39	2.63	.25
WASI												
Dual	20.79	.74	20.50	.71	20.64	.50	.28	26	.78	-1.81	2.39	.11
Visual	18.60	.75	20.14	.88	19.34	.58	1.34	27	.19	-.82	3.91	.52
Auditory	20.08	.57	21.33	.57	20.68	.42	1.56	23	.13	-.41	2.93	.65
STM	20.40	.58	19.38	.68	19.93	.45	1.14	26	.27	-.82	2.85	.45
Control	20.71	.44	20.67	.78	20.36	.42	.06	24	.93	-1.73	1.83	.02
Overall Average	20.10	.29	20.38	.33	20.24	.22	.65	134	.52	-.58	1.58	.11
BETA-III												
Dual	20.79	.57	19.50	.68	20.14	.45	1.46	26	.16	-.53	3.10	.56
Visual	19.20	.55	20.71	.85	19.93	.51	1.52	27	.14	-.54	3.56	.59
Auditory	20.46	.46	20.58	.54	20.52	.35	.17	23	.13	-1.34	1.59	.07
STM	19.67	.55	21.15	.49	20.69	.39	1.99	26	.06	-.05	3.02	.78
Control	19.93	.53	20.00	.46	19.96	.35	.10	24	.92	-1.40	1.54	.04
Overall Average	20.38	.24	20.38	.29	20.18	.19	1.07	134	.29	-.34	1.14	.18

Note: Total and overall averages are based on the designated group or the entire sample and not an average of group averages.

Cognitive pretests (Accuracy).

Performance on the cognitive tests was measured as the number correct and converted to a percentage. For the Lexical Decision Test, the percent correct is reported only for words and does not include nonwords. A MANOVA was conducted to determine any differences among training groups and between females and males in the percent correct on the cognitive pretests. The Levene's Test of Homogeneity indicated there was an issue with the homogeneity of variance for the Mental Rotation Pretest, *Levene Statistic* = 2.23, p = .02. The departure from homogeneity was due to a smaller variance in the control group for the Mental Rotation Pretest. Although the departure is statistically significant, the departure should not be a concern because the sample sizes are close to being equal, the test of equality of covariance matrices was not significant, and the spread vs. level plots did not show any alarming outliers.

The overall MANOVA results indicated a significant sex difference in the percent correct on the cognitive pretests, $F\,(4, 123) = 3.22, p < .02, \eta^2 = .10$, but no significant difference among training groups, $F\,(16, 376) = .39, p = .98, \eta^2 = .01$. A significant sex by training interaction was also found, $F\,(16, 376) = 1.82, p < .02, \eta^2 = .06$. The between-subjects results for each cognitive test are displayed in Table 11.

Table 11

MANOVA Results Testing for Differences among Training Groups and between Females and Males in Percent Correct on Cognitive Pretests

Test of G*f*	*F*	df	*p*	η^2
ERVT				
Sex	**6.31**	**1, 126**	**.01**	**.05**
Training	.41	4, 126	.80	.01
Sex X Training	.78	4, 126	.54	.02
LDT				
Sex	.23	1, 126	.63	.00
Training	.07	4, 126	.99	.00
Sex X Training	.55	4, 126	.70	.02
MRT				
Sex	**11.14**	**1, 126**	**.01**	**.08**
Training	.42	4, 126	.79	.01
Sex X Training	1.98	4, 126	.10	.06
PFT				
Sex	3.01	1, 126	.09	.02
Training	.65	4, 126	.63	.02
Sex X Training	**4.87**	**4, 126**	**.01**	**.13**

Note: ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.

There were two unexpected findings in the between-subjects effects relating to the sex difference in vocabulary. Typically, there is no sex difference in vocabulary, but in the current findings, males had a higher percent correct (44%) than females (37%). The other unexpected finding was an interaction effect between sex and training for the Paper Folding Test. Independent *t*-tests were conducted to identify which groups the sex differences existed and any other statistical sex differences that may have been missed by the MANOVA (see Table 12). Again, none of the statistical tests for sex differences are intended to infer cause and effect; rather, they are presented to describe how participants performed on the pretests and to establish a baseline to make adequate comparisons to the posttests.

Table 12

Means, Standard Errors, and Differences between Females and Males for Percent Correct on Cognitive Pretests

	Females		Males		Total Average					95% CI		
	M	*SE*	*M*	*SE*	*M*	*SE*	*t*	*df*	*p*	*LL*	*UL*	Cohen's *d*
Vocabulary												
Dual	37.05	3.66	49.11	3.46	43.08	2.73	2.39	26	.02	-1.70	22.41	.94
Visual	33.33	3.98	43.30	4.55	38.15	3.10	1.65	27	.11	-2.40	22.34	.64
Auditory	36.06	3.58	43.23	5.48	39.50	3.27	1.10	23	.28	-6.29	20.64	.46
STM	41.91	5.39	40.06	4.53	41.05	3.51	-.26	26	.80	-16.59	12.89	-.10
Control	35.27	3.56	42.19	4.18	38.46	2.75	1.27	24	.22	-4.33	18.17	.52
Overall Average	36.76	1.85	43.69	1.96	40.07	1.37	2.57	134	.01	-1.06	12.25	.44
Lexical Decision												
Dual	99.93	.57	98.57	.63	98.75	.42	-.42	26	.68	-2.10	1.38	-.17
Visual	98.67	.59	98.57	.63	98.62	.42	-.11	27	.91	-1.86	1.67	-.04
Auditory	98.85	.61	98.75	.65	98.80	.44	-.11	23	.92	-1.92	1.75	-.05
STM	98.67	.59	98.85	.61	98.75	.42	.21	26	.84	-1.57	1.93	.08
Control	97.86	.69	99.17	.56	98.46	.46	1.45	24	.16	-.56	3.18	.59
Overall Average	98.59	.27	98.77	.27	98.68	.19	.47	134	.64	-.58	.93	.08
Mental Rotation												
Dual	56.86	6.65	61.71	5.30	59.29	4.20	.57	26	.57	-12.61	22.33	.22
Visual	42.93	4.03	70.86	5.33	56.41	4.18	4.22	27	.00	-14.38	41.51	1.62
Auditory	48.92	5.91	65.67	6.52	56.96	4.62	1.91	23	.07	-1.42	34.90	.80
STM	53.07	5.27	55.38	6.98	54.14	4.22	.27	26	.79	-15.40	20.03	.11
Control	50.00	3.37	56.00	4.67	52.76	2.82	1.06	24	.30	-5.66	17.66	.43
Overall Average	50.31	2.30	62.09	2.62	55.94	1.80	3.39	134	.01	-4.90	18.67	.59
Paper Folding												
Dual	65.41	4.30	53.76	5.35	59.59	3.55	-1.70	26	.10	-25.77	2.46	-.67
Visual	44.91	5.21	72.18	3.55	58.08	4.06	4.26	27	.00	14.15	40.39	1.64
Auditory	56.28	4.86	67.54	3.82	61.68	3.27	1.80	23	.09	-1.66	24.20	.75
STM	56.49	4.31	51.82	6.69	54.32	3.82	-.60	26	.55	-20.60	11.26	-.24
Control	55.64	4.89	60.53	5.59	57.89	3.65	.66	24	.52	-10.38	20.16	.27
Overall Average	55.60	2.20	63.13	2.43	58.24	1.65	1.69	134	.09	-.94	12.01	.29

Note: Total and overall averages are based on the designated group or the entire sample and not an average of group averages.

Cognitive pretests (RTs).

Mean RTs for correct answers in each cognitive pretest were calculated. Mean RTs for the Lexical Decision Test are provided only for correct answers given for words. A MANOVA showed no overall statistical sex differences, F (4, 123) = .93, p =.45, η^2 = .03, or among training groups, F (16, 376) = .89, p = .59, η^2 = .03. The interaction between sex and training was also nonsignificant, F (16, 376) = 1.01, p = .44, η^2 = .03. There was one significant between-subjects sex by training interaction for RTs in the Paper Folding Test, F (4, 126) = 2.45, $p < .05$, η^2 = .07, such that males had significantly slower RTs than females in the visual *n*-back training group (see Table 13).

Independent *t*-tests were calculated to determine differences between females and males in each training group for RTs that may not have been detected in the MANOVA (see Table 13). Overall, there were no significant differences between females and males for any of the cognitive pretests with the exception of the sex difference in the visual *n*-back training group for the Paper Folding Test.

Table 13

Means and Standard Errors for RTs for Correct Answers on Cognitive Pretests

	Females		Males		Total Average					95% CI		
	M	*SE*	*M*	*SE*	*M*	*SE*	*t*	*df*	*p*	*LL*	*UL*	Cohen's *d*
Vocabulary												
Dual	8250	712	7860	630	8055	468	-.41	26	.69	-2345	1564	-.16
Visual	10439	1018	10324	1039	10380	714	-.08	27	.94	-3101	2870	-.03
Auditory	10899	1385	8844	580	9913	785	-1.33	23	.20	-5256	1145	-.55
STM	8991	925	9056	758	9021	597	.05	26	.96	-2442	2572	.02
Control	8596	837	9142	756	8848	561	.48	24	.64	-1814	2907	.20
Overall Average	9423	447	9048	354	9243	288	-.65	134	.52	-1515	766	-.11
Lexical Decision												
Dual	768	41	821	33	795	26	1.01	26	.32	-55	162	.40
Visual	806	30	818	32	812	22	.28	27	.78	-78	103	.11
Auditory	782	32	803	31	792	22	.48	23	.64	-72	115	.20
STM	800	28	813	37	807	22	.29	26	.77	-80	106	.11
Control	780	30	803	32	792	22	.52	24	.61	-68	113	.21
Overall Average	788	14	812	15	799	10	1.21	134	.23	-15	65	.21
Mental Rotation												
Dual	17192	2374	19185	2117	18188	1573	-.63	26	.54	-8531	4545	-.25
Visual	14005	1791	18455	1645	16153	1270	-1.82	27	.08	-9463	562	-.70
Auditory	16688	2072	17561	2296	17506	1511	-.28	23	.98	-7254	5509	-.01
STM	17364	2158	12613	2033	15300	1534	1.59	26	.16	-1402	10904	.57
Control	13894	1580	16201	1824	14959	1195	-.96	24	.35	-7260	2646	-.40
Overall Average	15812	894	16863	910	16314	637	-.82	134	.41	-3577	1476	-.11
Paper Folding												
Dual	17368	1143	18356	2363	17862	1292	-.38	26	.71	-6383	4408	-.15
Visual	14881	1302	20059	1561	17381	1106	-2.56	27	.02	-9328	1029	-.99
Auditory	19885	1588	16360	1483	18193	1126	1.62	23	.12	-990	8040	.68
STM	16944	1424	15259	1898	16162	1154	.72	26	.48	-3116	6485	.28
Control	14166	1289	17755	1824	15882	1127	-1.64	24	.11	-8103	924	-.68
Overall Average	16582	633	17624	840	17080	520	-1.00	134	.32	-1017	3100	-.17

Note: Total and overall averages are based on the designated group or the entire sample and not an average of group averages.

Posttest Performance

The posttests summary will follow the same formant as the pretests summary by using the same analyses and the same ordering of sections: tests of G*f*, percent correct on cognitive tests, and RTs for correct answers on cognitive tests.

Posttests of G*f*.

A MANOVA was used to determine if there were sex differences or differences among training groups in scores on the posttests of G*f*. There was no overall significant sex difference on the G*f* posttest scores, $F\,(4, 123) = 1.14, p = .34, \eta^2 = .04$. There were, however, significant differences in posttest scores among training groups, $F\,(16, 376) = 2.02, p < .01, \eta^2 = .06$. The between-subjects tests indicate that there was only one significant difference among training groups for the BETA-III posttest (as shown in Table 14). The interaction between sex and training was not significant, $F\,(16, 376) = .97, p = .49, \eta^2 = .03$.

Table 14

MANOVA Results Testing for Differences among Training Groups and between Females and Males in Scores on Posttests of Gf

Test of G*f*	*F*	df	*p*	η^2
Raven				
Sex	1.30	1, 126	.26	.01
Training	2.11	4, 126	.08	.06
Sex X Training	1.23	4, 126	.30	.04
Cattell				
Sex	.45	1, 126	.50	.00
Training	1.33	4, 126	.26	.04
Sex X Training	1.88	4, 126	.19	.06
WASI				
Sex	.32	1, 126	.57	.00
Training	1.16	4, 126	.33	.04
Sex X Training	.82	4, 126	.51	.03
BETA-III				
Sex	3.92	1, 126	.05	.03
Training	**5.42**	**4, 126**	**.01**	**.15**
Sex X Training	1.41	4, 126	.26	.04

Post hoc analyses were conducted to determine if there were any differences between paired training groups. No significant differences existed between pairwise groups for scores on the Cattell posttest and one significant difference existed in the WASI posttest scores such that the dual *n*-back training group performed better than the visual *n*-back training group, $M_{Diff.} = 1.24$, $SE = .62$, $p < .05$. There were more significant differences between pairwise training groups for the Raven's and BETA-III posttests and they are summarized in Table 15.

Table 15

Post-Hoc Results for Differences among Training Groups in Posttest Scores on the Raven's and BETA-III

Comparison	*Mean Difference*	*SE*	*p*	95% CI *LL*	95% CI *UL*
Raven's					
Dual vs. Visual	2.19	1.16	.06	-.11	4.49
Dual vs. Auditory	2.22	1.21	.07	-.17	4.60
Dual vs. STM	2.04	1.17	.09	-.28	4.36
Dual vs. Control*	3.38	1.19	.01	1.02	5.75
Visual vs. Auditory	.02	1.20	.98	-2.34	2.39
Visual vs. STM	-.16	1.16	.89	-2.45	2.14
Visual vs. Control	1.19	1.18	.32	-1.15	3.54
Auditory vs. STM	-.18	1.21	.88	-2.57	2.21
Auditory vs. Control	1.17	1.23	.34	-1.02	3.60
STM vs. Control	1.34	1.19	.26	-1.02	3.71
BETA-III					
Dual vs. Visual	1.11	.58	.06	-.03	2.25
Dual vs. Auditory*	1.47	.60	.02	.29	2.65
Dual vs. STM*	1.14	.58	.05	-.01	2.29
Dual vs. Control*	2.68	.59	.01	1.51	3.85
Visual vs. Auditory	.36	.59	.56	-.81	1.53
Visual vs. STM	.04	.58	.95	-1.10	1.17
Visual vs. Control*	1.58	.59	.01	.42	2.74
Auditory vs. STM	-.32	.60	.59	-1.51	.86
Auditory vs. Control*	1.22	.61	.05	.01	2.42
STM vs. Control*	1.54	.59	.01	.37	2.71

Note. A positive Contrast indicates that the gain for the first group in the comparison was larger. Asterisk indicates significant ($p < .05$) difference between training group pair. SE = Standard Error. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

Independent *t*-tests were conducted to test for differences between females and males in posttest scores on tests of G*f* that may not have been detected by the MANOVA. There was a significant sex difference only for the BETA-III posttest scores such that males had higher scores. As with the pretests, the differences between females and males existed in the visual *n*-back training group for the Raven's and Cattell posttests; however, there was also a sex difference for BETA-III posttest, which was not present in the pretest scores (see Table 16).

Table 16

Means, Standard Errors, and Differences between Females and Males for Number Correct on Posttests of Gf

	Females		Males		Total Average					95% CI		
	M	*SE*	*M*	*SE*	*M*	*SE*	*t*	*df*	*p*	*LL*	*UL*	Cohen's *d*
Raven's												
Dual	18.00	1.11	17.07	1.07	17.54	.76	-.60	26	.55	-4.41	2.25	-.24
Visual	13.47	1.06	17.36	1.30	15.34	.90	2.33	27	.03	-.46	7.32	.90
Auditory	15.23	1.27	15.42	1.10	15.32	.83	.11	23	.91	-3.32	3.70	.05
STM	15.13	1.06	15.92	1.79	15.50	.99	.39	26	.70	-3.35	4.92	.15
Control	13.86	.85	14.33	1.04	14.08	.65	.35	24	.73	-2.31	3.26	.14
Overall Average	15.14	.50	16.09	.58	15.60	.38	1.25	134	.22	-.53	2.50	.21
Cattell												
Dual	29.57	1.16	27.93	1.53	28.75	.95	-.86	26	.40	-5.58	2.30	-.34
Visual	25.67	1.06	29.64	1.27	27.59	.89	2.41	27	.02	-.60	7.36	.93
Auditory	25.54	1.21	27.75	1.37	26.60	.92	1.22	23	.24	-1.55	5.98	.51
STM	26.60	1.14	26.38	1.57	26.50	.93	-.11	26	.91	-4.13	3.70	-.04
Control	27.00	1.36	25.42	1.15	26.27	.90	-.87	24	.39	-5.33	2.16	-.36
Overall Average	26.87	.54	27.49	.63	27.17	.41	.75	134	.46	-1.03	2.27	.13
WASI												
Dual	22.36	.54	21.71	.71	22.04	.44	-.72	26	.48	-2.47	1.19	-.28
Visual	20.33	.48	21.29	.71	20.79	.43	1.11	27	.27	-.80	2.70	.43
Auditory	21.15	.65	22.17	.61	21.64	.45	1.13	23	.27	-.84	2.87	.47
STM	21.67	.67	21.08	.68	21.39	.47	-.62	26	.54	-2.56	1.38	-.24
Control	20.92	.59	21.54	.62	21.12	.42	.47	24	.64	-1.36	2.17	.19
Overall Average	21.28	.27	21.51	.30	21.39	.20	.57	134	.57	-1.02	.56	.10
BETA-III												
Dual	22.36	.46	21.86	.61	22.12	.38	-.65	26	.52	-2.08	1.08	-.25
Visual	20.00	.38	22.07	.51	21.00	.36	3.30	27	.01	-.78	3.36	1.27
Auditory	20.31	.54	21.00	.62	20.64	.40	.85	23	.40	-.99	2.37	.35
STM	20.47	.67	21.54	.56	20.96	.45	1.21	26	.24	-.75	2.90	.47
Control	19.29	.59	19.58	.82	19.42	.48	.30	24	.77	-1.74	2.34	.12
Overall Average	20.48	.26	21.26	.29	20.85	.20	2.00	134	.05	-.01	1.57	.35

Note: Total and overall averages are based on the designated group or the entire sample and not an average of group averages.

Cognitive posttests (Accuracy).

A MANOVA showed a significant sex difference in the percent correct in the cognitive posttests, $F\,(4,\,123) = 5.34, p < .001, \eta^2 = .12$, and no significant differences among training groups, $F\,(16,\,376) = 1.09, p = .36, \eta^2 = .03$. The interaction between sex and training was not significant, $F\,(16,\,376) = 1.55, p = .08, \eta^2 = .05$. Table 17 shows the between-subjects significance analysis for each cognitive posttest.

Table 17

MANOVA Results Testing for Differences among Training Groups and between Females and Males in Percent Correct on Cognitive Posttests

Test of G*f*	*F*	df	*p*	η^2
ERVT				
Sex	**6.69**	**1, 126**	**.01**	**.05**
Training	.78	4, 126	.54	.02
Sex X Training	.68	4, 126	.61	.02
LDT				
Sex	.00	1, 126	.99	.00
Training	.83	4, 126	.51	.03
Sex X Training	.76	4, 126	.55	.02
MRT				
Sex	**20.06**	**1, 126**	**.01**	**.14**
Training	.53	4, 126	.71	.02
Sex X Training	2.04	4, 126	.27	.04
PFT				
Sex	**4.55**	**1, 126**	**.04**	**.04**
Training	2.04	4, 126	.09	.06
Sex X Training	**3.70**	**4, 126**	**.01**	**.11**

Note: ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.

Independent *t*-tests were conducted to determine which training groups experienced the significant sex differences and are presented in Table 18. Overall, more significant differences between females and males are present in the percent correct on cognitive posttests than the cognitive pretests and may be a result of training, which will be determined in a later section.

Table 18

Means, Standard Errors, and Differences between Females and Males for Percent Correct on Cognitive Posttests

	Females		Males		Total Average					95% CI		
	M	*SE*	*M*	*SE*	*M*	*SE*	*t*	*df*	*p*	*LL*	*UL*	Cohen's *d*
Vocabulary												
Dual	41.37	3.38	51.19	3.44	46.27	2.55	2.04	26	.05	-.08	19.72	.80
Visual	35.42	3.83	45.39	4.14	40.23	2.92	1.77	27	.09	-1.58	21.52	.68
Auditory	38.14	3.60	43.92	5.30	40.92	3.15	.92	23	.37	-7.28	18.85	.38
STM	44.03	4.78	42.95	4.33	43.53	3.20	.17	26	.87	-14.51	12.35	.07
Control	36.76	3.04	44.96	3.60	40.55	2.43	1.76	24	.09	-1.45	17.87	.72
Overall Average	39.17	1.70	45.80	1.85	42.34	1.28	2.64	134	.01	-1.67	11.59	.47
Lexical Decision												
Dual	99.29	.49	98.57	.63	98.23	.39	.90	26	.38	-2.34	.91	-.29
Visual	99.33	.45	99.29	.49	99.31	.33	.07	27	.94	-1.41	1.31	.00
Auditory	98.08	.70	99.17	.56	98.60	.46	1.20	23	.24	-.79	2.97	-.55
STM	98.67	.59	98.08	.70	98.39	.45	.65	26	.52	-2.46	1.28	-.52
Control	98.93	.57	99.17	.56	99.04	.40	.30	24	.77	-1.42	1.90	-.05
Overall Average	98.87	.25	98.85	.26	98.86	.18	.08	134	.94	-.74	.69	-.04
Mental Rotation												
Dual	60.86	6.05	66.00	4.66	63.43	3.78	.67	26	.51	-10.56	20.85	.26
Visual	46.13	6.17	73.14	5.24	59.17	4.75	3.31	27	.01	-10.28	43.74	1.27
Auditory	55.54	6.52	77.33	6.76	64.96	5.20	2.53	23	.02	-4.36	43.23	1.06
STM	55.47	5.50	66.15	5.98	60.43	4.11	1.32	26	.20	-6.01	27.39	.52
Control	50.86	4.35	65.00	5.33	57.38	3.61	2.09	24	.05	-.09	28.19	.85
Overall Average	53.30	2.57	69.48	2.49	61.03	1.92	4.51	134	.01	-9.08	23.28	.78
Paper Folding												
Dual	72.18	3.46	66.92	5.19	69.55	3.10	-.84	26	.41	-18.09	7.56	-.33
Visual	57.19	4.65	77.07	2.73	66.79	3.28	3.62	27	.01	-8.60	31.15	1.39
Auditory	55.06	5.67	75.44	3.26	64.84	3.88	1.80	23	.09	-6.54	34.22	.75
STM	62.46	5.23	59.11	5.13	60.90	3.63	3.05	26	.01	-18.52	11.82	1.20
Control	57.89	4.48	57.89	5.90	57.89	3.56	.00	24	.99	-15.06	15.06	.00
Overall Average	61.01	2.19	67.45	2.22	64.09	1.58	2.07	134	.04	-.28	12.61	.36

Note: Total and overall averages are based on the designated group or the entire sample and not an average of group averages.

Cognitive posttests (RTs).

A MANOVA indicated that there was no overall significant sex difference in RTs for correct answers on the cognitive posttests, $F(4, 123) = 2.06, p = .09, \eta^2 = .06$. There were, however, differences among the training groups, $F(16, 376) = 1.88, p = .02, \eta^2 = .06$. No significant interaction between sex and training was present for RTs on the cognitive posttests, $F(16, 376) = 1.27, p = .22, \eta^2 = .04$. Table 19 shows the between-subjects analyses for RTs in each cognitive posttest.

Table 19

MANOVA Results Testing for Differences among Training Groups and between Females and Males for RTs on Cognitive Posttests

Test of G*f*	*F*	df	*p*	η^2
ERVT				
Sex	.57	1, 126	.45	.00
Training	**2.83**	**4, 126**	**.03**	**.08**
Sex X Training	1.03	4, 126	.39	.03
LDT				
Sex	3.61	1, 126	.06	.03
Training	**2.68**	**4, 126**	**.04**	**.08**
Sex X Training	1.19	4, 126	.32	.04
MRT				
Sex	.90	1, 126	.35	.01
Training	.50	4, 126	.74	.02
Sex X Training	1.11	4, 126	.36	.03
PFT				
Sex	2.06	1, 126	.15	.02
Training	2.18	4, 126	.08	.07
Sex X Training	1.06	4, 126	.38	.03

Note: ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.

Post hoc analysis indicated a number of significant pairwise differences among the training groups for RTs in the Extended Range Vocabulary Posttest, the Lexical Decision Posttest, and the Paper Folding Posttest (see Table 20).

Table 20

Post-Hoc Results for Differences in RTs for Correct Answers on Posttests for Vocabulary, Lexical Decision, and Paper Folding

Comparison	*Mean Difference*	*SE*	*p*	95% CI *LL*	95% CI *UL*
Vocabulary					
Dual vs. Visual*	-1681	734	.02	-3133	-230
Dual vs. Auditory*	-1550	762	.04	-3058	-42
Dual vs. STM	-229	740	.76	-1694	1235
Dual vs. Control	287	754	.70	-1205	1779
Visual vs. Auditory	132	756	.86	-1363	1627
Visual vs. STM *	1452	734	.05	0	2904
Visual vs. Control*	1968	748	.01	488	3448
Auditory vs. STM	1321	762	.09	-187	2828
Auditory vs. Control*	1837	776	.02	302	3372
STM vs. Control	516	754	.50	-976	2009
Lexical Decision					
Dual vs. Visual*	-73	30	.01	-132	-15
Dual vs. Auditory	-24	31	.43	-85	36
Dual vs. STM	16	30	.58	-43	75
Dual vs. Control	-9	30	.78	-69	51
Visual vs. Auditory	49	30	.11	-11	109
Visual vs. STM *	90	30	.01	31	148
Visual vs. Control*	65	30	.03	5	124
Auditory vs. STM	41	31	.19	-20	101
Auditory vs. Control	16	31	.61	-46	78
STM vs. Control	-25	30	.41	-85	35
Paper Folding Test					
Dual vs. Visual	744	1619	.65	-2458	3946
Dual vs. Auditory	1082	1667	.52	-2244	4407
Dual vs. STM	1077	1619	.51	-2153	4307
Dual vs. Control*	4440	1650	.01	1148	7731
Visual vs. Auditory	338	1667	.84	-2960	3636
Visual vs. STM	333	1681	84	-2869	3535
Visual vs. Control*	3696	1650	.03	432	6960
Auditory vs. STM	-5	1681	.99	-3330	3320
Auditory vs. Control*	3358	1711	.05	-27	6743
STM vs. Control*	3363	1664	.05	71	6654

Note. A positive Contrast indicates that the gain for the first group in the comparison was larger. Asterisk indicates significant ($p < .05$) difference between training group pair. SE = Standard Error. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

In general, differences existed between the dual and visual *n*-back training groups for the Extended Range Vocabulary Posttest with the dual *n*-back training group generally having faster RTs and the visual *n*-back training group having slower RTs. There were no specific patterns in differences for the Lexical Decision Posttest. Surprisingly, the control group had significantly faster RTs for correct answers on the Paper Folding Posttest than any of the training groups. No pairwise differences were present in the Mental Rotation Posttest.

Independent *t*-tests were also conducted within each training group to test for sex differences that may not have been detected by the MANOVA. No overall sex differences were present in RTs for correct answers on cognitive posttests when the training groups were pooled together (see Table 21). There was an unexpected sex difference exists in the auditory *n*-back training group for the Lexical Decision Test where females' RTs were significantly faster than males. In general, there were no specific patterns when comparing females and males in RTs for cognitive posttests.

Table 21

Means, Standard Errors, and Differences between Females and Males for RTs for Correct Answers on Cognitive Posttests

	Females		Males		Total Average					95% CI		
	M	*SE*	*M*	*SE*	*M*	*SE*	*t*	*df*	*p*	*LL*	*UL*	Cohen's *d*
Vocabulary												
Dual	6774	719	7001	571	6887	451	.25	26	.81	-1661	2114	.10
Visual	9206	978	7885	670	8569	603	1.10	27	.28	-1146	3788	.42
Auditory	9203	1206	7607	773	8437	732	1.09	23	.29	-1423	4616	.45
STM	6658	598	7645	665	7116	447	1.11	26	.28	-848	2821	.44
Control	6642	495	6551	540	6600	358	.12	24	.90	-1420	1601	.05
Overall Average	7682	388	7349	287	7523	244	.68	134	.49	-636	1302	.12
Lexical Decision												
Dual	725	35	763	36	863	187	.76	26	.46	-65	141	.30
Visual	823	33	811	32	979	174	.26	27	.80	-82	106	.10
Auditory	721	22	821	39	937	193	2.26	23	.03	8	191	.12
STM	731	21	724	25	835	115	.21	26	.84	-59	72	.94
Control	724	15	787	38	887	147	1.62	24	.12	-17	142	.66
Overall Average	746	12	781	15	900	171	1.76	134	.08	-4	74	.30
Mental Rotation												
Dual	13481	1890	15179	1997	14330	1359	-.62	26	.54	-7349	3953	-.24
Visual	9759	1673	15982	2351	12720	1516	-2.15	27	.04	-11992	-273	-.83
Auditory	15740	3401	14266	1930	15033	1962	.37	23	.72	-6801	9749	.15
STM	15085	2345	13599	1735	14395	1473	.50	26	.62	-4669	7642	.27
Control	11737	1940	13264	1183	12442	1167	-.65	24	.53	-6414	3359	-.27
Overall Average	13103	1026	14495	841	13768	670	-1.04	134	.30	-4042	1259	-.18
Paper Folding												
Dual	13904	1095	18755	1929	16329	1184	-2.19	26	.04	-9410	-290	-.86
Visual	14137	1893	17137	1709	15585	1289	-1.17	27	.25	-8256	2257	-.45
Auditory	15138	1514	15366	2357	15247	1349	-.08	23	.94	-5931	5474	-.03
STM	14987	1816	15558	1474	15252	1169	- .24	26	.81	-5477	4335	-.09
Control	12426	1150	11264	894	11890	739	.78	24	.44	-1921	4245	.32
Overall Average	14117	684	15758	818	14901	532	-1.55	134	.12	-3738	454	-.27

Note: Total and overall averages are based on the designated group or the entire sample and not an average of group averages.

Correlations

Because tests of *Gf* theoretically predict other cognitive abilities, correlation matrices were created to identify the relationships between tests of *Gf*, between tests of *Gf* and percent correct for cognitive tests, and between tests of *Gf* and RTs for correct answers on the cognitive tests. The current section includes all 136 participants and is only for descriptive purposes. Correlations within each training group were also calculated and are presented in Appendices C through G. Correlation matrices are set up so that pretest correlations are below the diagonal and the posttest correlations are above the diagonal on the matrix. Correlations between each test of *Gf* and between the tests of *Gf* and percent correct on cognitive tests are described first. The second part of the current section covers correlations between tests of *Gf* and RTs for correct answers on cognitive tests to determine if higher scores on *Gf* would correlate with faster RTs. The third section correlates the percent correct on cognitive tests with RTs to determine if there was a speed-accuracy tradeoff in the cognitive tests.

Correlations (*Gf* and percent correct).

The first correlation matrix identifies the relationships between each test of *Gf* and between each percent correct on the cognitive tests (see Table 22). Overall, the correlations are what would be expected such that all correlations are in the positive direction. Two correlations that are important to note involve the nonsignificant correlations between the Lexical Decision Making Test and the Mental Rotation Test and between the Lexical Decision Making Test and the Paper Folding Test. What is important about the two correlations is that the Lexical Decision Making Test does not correlate with the two visual tasks, but does correlate with the Extended Range

Vocabulary. Moreover, Extended Range Vocabulary Test correlates with both visual tests. This could suggest the Lexical Decision Making Test is not a test of visuospatial abilities, but is actually a test of verbal abilities, which is supported by Ratcliff et al. (2004).

Table 22

Intercorrelations between Tests of Gf and Percent Correct on Cognitive Tests

Measure	1	2	3	4	5	6	7	8
1. Raven	—	.64**	.55**	.54**	.44**	.18*	.42**	.58**
2. Cattell	.69**	—	.59**	.57**	.42**	.23**	.46**	.58**
3. WASI	.61**	.61**	—	.52**	.34**	.24**	.46**	.44**
4. BETA-III	.54**	.46**	.51**	—	.40**	.24**	.48**	.55**
5. ERVT	.31**	.50**	.42**	.20*	—	.41**	.43**	.37**
6. LDT	.29**	.37**	.28**	.21*	.33**	—	.19*	.08
7. MRT	.43**	.43**	.38**	.31**	.40**	.17	—	.56**
8. PFT	.55**	.52**	.48**	.45**	.40**	.15	.47**	—

Note: $N = 136$. Correlations below the diagonal represent pretests and correlations above the diagonal represent posttests. ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.

** indicates *p*-values at or < .01

* indicates *p*-values at or < .05

Correlations (G*f* and RTs).

Correlations between tests of G*f* and RTs for correct answers were also calculated and are provided in Table 23. Overall, there were significant negative correlations between the tests of G*f* and the Extended Range Vocabulary Test such that longer reactions time for correct answers were correlated with lower scores on tests of G*f*. There were also significant correlations between the tests of G*f* and the Lexical Decision Test that followed the same pattern as the correlations with the Extended Range Vocabulary Test. The significant correlation between the Lexical Decision Test and the Extended Range Vocabulary Test was positive such that longer RTs on one test meant longer RTs on the other, which is what would be expected. A similar positive correlation also existed for the RTs between the Paper Folding Test and the Mental Rotation Test. In general, there were few, if any, correlations between tests of *Gf* and RTs for correct answers on the Mental Rotation Test and the Paper Folding Test.

Table 23

Intercorrelations between Tests of Gf and RTs for Correct Answers on Cognitive Pretests and Posttests

Measure	1	2	3	4	5	6	7	8
1. Raven	—	.64**	.55**	.54**	-.24**	-.27**	-.02	-.06
2. Cattell	.69**	—	.59**	.57**	-.22**	-.25**	.04	-.03
3. WASI	.61**	.61**	—	.52**	-.20*	-.26**	.05	-.08
4. BETA-III	.54**	.46**	.51**	—	-.11	-.25**	.03	-.05
5. ERVT (RT)	-.26**	-.25**	-.22*	-.17*	—	.43**	.23**	.41**
6. LDT (RT)	-.25**	-.29**	-.24**	-.19*	.40**	—	.08	.22**
7. MRT (RT)	-.01	.04	.01	.00	.11	.00	—	.53**
8. PFT (RT)	.04	.06	.02	.02	.11	.10	.43**	—

Note: $N = 136$. Correlations below the diagonal represent pretests and correlations above the diagonal represent posttests. ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test; RT = reaction times.
** indicates p-values at or $< .01$
* indicates p-values at or $< .05$

Correlations (Accuracy and RTs).

Correlations were calculated to identify the relationship between percent correct on cognitive tests and RTs for correct answers on cognitive tests. The correlations between these two measures may provide information about the amount of time it requires processing a problem that leads to a correct answer and how that processing time may be related to providing a correct answer on other tests. Table 24 shows the correlations for pretests and posttests.

Table 24

Intercorrelations between Percent Correct on Cognitive Tests and RTs for Correct Answers on Cognitive Tests

Measure	1	2	3	4	5	6	7	8
1. ERVT	—	.41**	.43**	.37**	-.32**	-.31**	.12	-.04
2. LDT	.33**	—	.19*	.08	-.21*	-.22**	-.04	-.08
3. MRT	.40**	.17	—	.56**	-.09	-.16	.46**	.21
4. PFT	.40**	.15	.47**	—	-.05	-.13	.18*	.20
5. ERVT (RT)	-.37**	-.22*	-.10	-.23**	—	.43**	.23**	.41**
6. LDT (RT)	-.23**	-.40**	-.15	-.13	.40**	—	.08	.22**
7. MRT (RT)	.19	.20*	.51**	.20*	.11	.00	—	.53**
8. PFT (RT)	.12	.00	.33**	.33**	.11	.10	.43**	—

Note: $N = 136$. Correlations below the diagonal represent pretests and correlations above the diagonal represent posttests. ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test; RT = reaction times.
** indicates p-values at or < .01
* indicates p-values at or < .05

Correlations between RTs and percent correct were significant such that the verbal tests correlated significantly with each other and the visuospatial tests correlated significantly with each other. In general, as RTs increased on verbal tests, the percent correct decreased, which suggests that participants were not engaging in speed-accuracy tradeoffs. The opposite was true for the visuospatial tasks, in that the slower the RTs, the higher percent correct.

Comparisons between Pretests and Posttests

The section to follow provides analyses that test the specific hypothesis set forth in the current study. A number of tests were used to determine differences between training groups and between females and males for gains on the posttests. Gains on tests of G*f*, percent correct on cognitive tests, and RTs were calculated by subtracting the pretest value from the posttest value. The first section of the current section analyzes the tests of G*f*, the second section analyzes differences in the cognitive tests, and the third section analyzes differences in RTs for correct answers on cognitive tests.

Tests of G*f*.

The first research question in the current study was, "Do scores on tests of G*f* increase after participants complete a cognitive training program using the dual *n*-back task?" Similar to Jaeggi et al. (2008), paired *t*-tests comparing pretest and posttest scores were conducted only for the participants in the dual *n*-back training group. Participants who completed the dual *n*-back training program experienced improved scores on the Raven's, t (27) = 6.19, p < .001, d = 1.17; Cattell's test, t (27) = 4.18, p < .001, d = .56; the WASI subtest, t (27) = 2.99, p = .006, d = .75; and the BETA-III subtest, t (27) = 3.98, p < .001, d = .79. Based on these analyses, Jaeggi et al.'s results were replicated (i.e., Jaeggi et al. found an effect size of .65) to the extent that there was improvement in scores on Raven's Advanced Progressive Matrices. Jaeggi et al.'s results were also expanded in that there were improvements on other tests of G*f* as well. The analyses to follow answers the other research questions posed in the current study regarding differences in improvements based on sex and training.

A repeated measures analysis was used to determine whether posttest scores were higher than pretest scores and if there were any simple effects, main effects, and interactions. There was no significant sex difference in the four tests, F (4, 123) = .90, p = .47, η^2 = .03; thus females and males did not differ significantly in their scores on tests of G*f*. There was a marginally significant difference between training groups on scores of tests of G*f*, F (16, 504) = 1.64, p = .06, η^2 = .05. There was no significant interaction between sex and training, F (16, 504) = 1.52, p = .09, η^2 = .05. Overall, there were no differences among groups on the tests of G*f*. Differences among training groups were not predicted because participants were assigned randomly to training conditions. However, it was expected that there would be differences between females and males such that males would have higher scores (Lynn & Irwing, 2004); thus, the current study does not support the literature that there are sex differences on tests of G*f*.

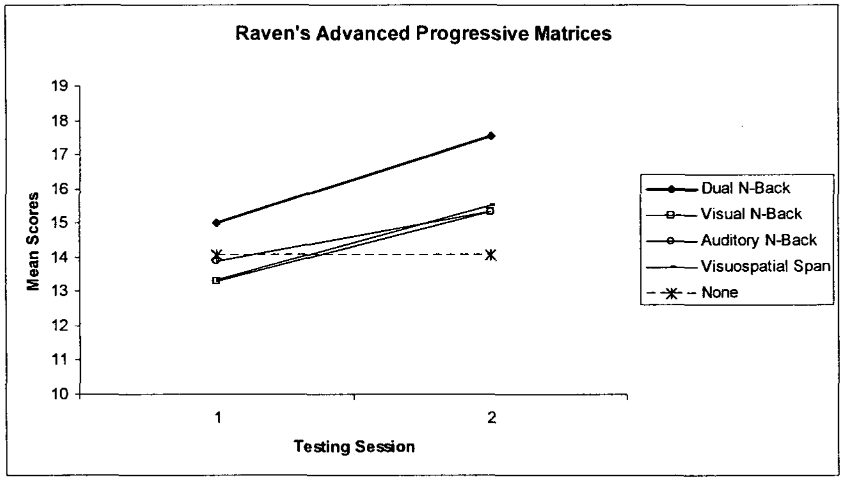
There was a significant difference between pretest scores and posttest scores, F (4, 123) = 28.73, $p < .001$, $\eta^2 = .48$, such that posttest scores were generally higher than pretest scores. Sex did not have an effect on the difference in pretest and posttest scores, F (4, 123) = .57, $p = .69$, $\eta^2 = .02$. Training, however, did have a significant effect on the differences between pretest and posttest scores, F (16, 504) = 2.26, $p = .004$, $\eta^2 = .07$. There was no interaction effect of sex and training on the difference in scores on pretests and posttests, F (16, 504) = .46, $p = .96$, $\eta^2 = .01$. Based on these findings, the hypothesis that females would experience greater gains on tests of G*f* as a result of training was not supported. Furthermore, there is no evidence to support the hypothesis that females would experience greater gains if they were in the dual *n*-back or visual *n*-back training groups. Because of the nonsignificant main effect for sex or interaction between sex and training group (see Appendix I), there is no need for further analyses for sex differences in gains on tests of G*f*.

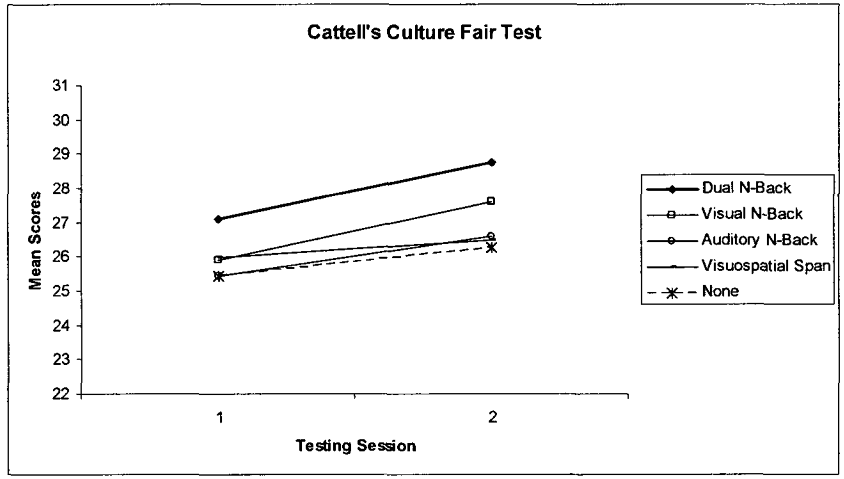
Results of the repeated measures analysis indicated that training had a significant impact on the difference in scores between pretests and posttests; thus, a closer look at the univariate results of the repeated measures analysis was necessary. Surprisingly, type of training had a significant effect on the difference in pretest and posttest scores for only the Raven's, F (4, 126) = 3.49, $p = .01$, $\eta^2 = .10$, and the BETA-III scores, F (4, 126) = 4.86, $p = .001$, $\eta^2 = .13$. Training type did not have a statistically significant effect on the difference for pretest and posttest scores on the Cattell test, F (4, 126) = .91, $p = .46$, $\eta^2 =$

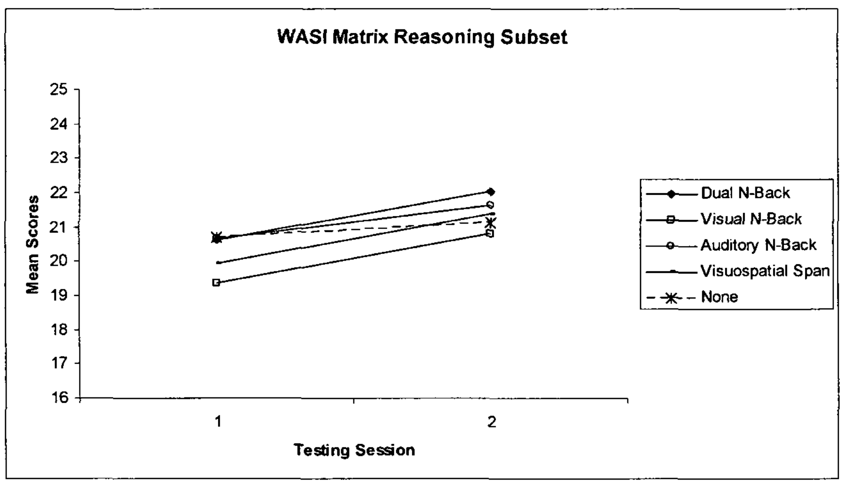
.03, or the WASI, F (4, 126) = 1.33, p = .26, η^2 = .04. The fact that all tests of G*f* were highly correlated, but participants experienced significant gains on only two of the four brings up many questions that will be discussed further in the discussion section. Visual representations depicting the changes in scores on tests of G*f* are shown in Figure 16.

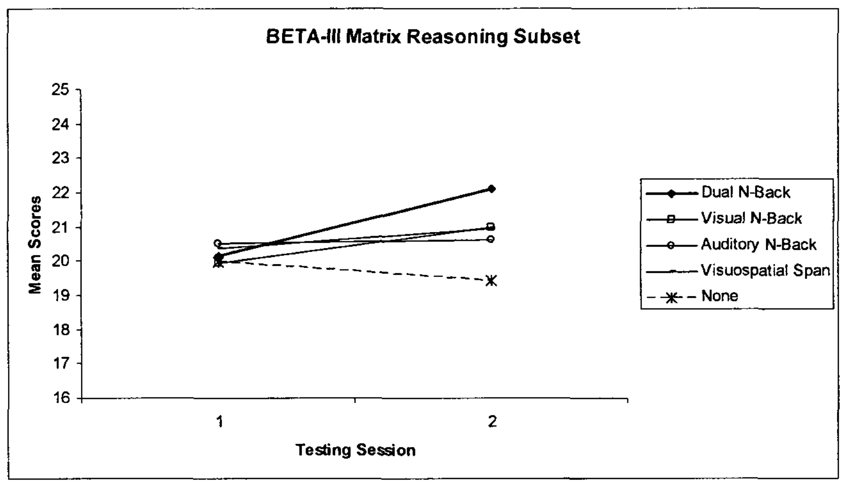
Figure 16

Improvements between Pretest and Posttest Scores on Tests of Gf









Although training types may not have differed from one another in the improvement of scores on the Cattell and WASI tests, there was generally significant improvement as shown by the one-sample *t*-tests to compare the mean gains experienced by each training group per test of G*f* to the null hypothesis of zero gain (see Table 25).

Table 25

Mean Gains, Standard Errors, and Differences between Mean Gains and Null Hypothesis (Gain = Zero) on Tests of Gf

	Total Average					95% CI		
	M	*SE*	*t*	*df*	*p*	*LL*	*UL*	Cohen's *d*
Raven's								
Dual	2.53	.41	6.19	27	.01	1.69	3.38	2.38
Visual	2.03	.47	4.29	28	.01	1.06	3.01	1.62
Auditory	1.44	.49	2.92	24	.01	-.42	2.46	1.19
STM	2.18	.63	3.44	27	.01	-.88	3.48	1.32
Control	.08	.42	.18	25	.86	-.79	.94	.07
Cattell								
Dual	1.64	.55	2.99	27	.01	-1.57	3.00	1.15
Visual	1.69	.51	3.31	28	.01	-2.46	1.81	1.25
Auditory	1.16	.55	2.12	24	.04	-1.15	3.37	.87
STM	.54	.60	.90	27	.38	-2.65	2.36	.35
Control	.85	.47	1.81	25	.08	-.46	3.29	.72
WASI								
Dual	1.39	.35	3.98	27	.01	-1.10	1.82	1.53
Visual	1.44	.40	3.63	28	.01	-1.06	2.24	1.37
Auditory	.96	.35	2.72	24	.01	-1.25	1.73	1.11
STM	1.46	.38	3.86	27	.01	-2.01	1.16	1.49
Control	.42	.39	1.08	25	.29	-2.10	1.20	.43
BETA-III								
Dual	1.96	.47	4.18	27	.01	-2.73	1.16	1.61
Visual	1.07	.41	2.59	28	.02	-2.27	1.15	.98
Auditory	.12	.38	.31	24	.76	-2.18	1.04	.13
STM	.61	.36	1.66	27	.11	-1.11	1.94	.64
Control	-.54	.46	-1.16	25	.26	-2.19	1.74	-.46

Based on the current statistical evidence, Jaeggi et al.'s (2008) study was, in general, replicated. However, there are qualifications that need to be made for the claim that training on the dual *n*-back task improves G*f*. The current evidences suggests that specific types of training have different effects on the improvement in scores on the

Raven's and BETA-III tests; thus, specific a priori contrasts were conducted to test predicted group differences in improvement on the Raven's and BETA-III scores.

A priori contrasts were conducted to determine if each training group experienced higher gains than the control group on the tests of G*f*. Even though there were no overall statistically significant differences between pretest and posttest scores on the Cattell and WASI tests, contrasts were conducted on Cattell and WASI first to determine if there were any specific differences between groups. Again, post hoc analyses were not used because there were a priori hypotheses for all tests of G*f*, and no adjustments have been made to the computed p-values. Contrasts for Cattell's test revealed no differences among any of the training groups (see Appendix H). However, there was a marginal difference between the control group and the visual n-back training group and a significant difference between the control group and the spatial matrix span training group on the WASI subtest (see Table 26).

Table 26

Contrasts of Training Groups for Gains on the WASI Subtest

Comparisons	Contrast	*SE*	*t*	*df*	*p*	*d*
Dual vs. Control	.97	.54	1.81	131	.07	.32
Visual vs. Control	1.03	.53	1.93	131	.06	.34
Auditory vs. Control	.54	.55	.98	131	.33	.17
STM vs. Control*	1.04	.54	1.95	131	.05	.34
Dual vs. Visual	-.05	.52	-.11	131	.92	-.02
Dual vs. Auditory	.43	.54	.80	131	.43	.14
Dual vs. STM	-.07	.53	-.14	131	.89	-.02
Visual vs. Auditory	.48	.54	.91	131	.36	.16
Visual vs. STM	-.01	.52	-.03	131	.98	-.01
Auditory vs. STM	-.50	.54	-.93	131	.35	-.16

Note: A positive Contrast indicates that the gain for the first group in the comparison was larger. Asterisk indicates significant ($p < .05$) difference between training group pair.

Results of the a priori contrasts for the Raven's are provided in Table 27. Based on the contrasts, the first hypothesis was supported and Jaeggi et al.'s (2008) results were replicated, once again, in that participants in the dual *n*-back training group experienced significant gains compared to the control group on the Raven's, $t\ (131) = 3.49, p < .001$, Cohen's $d = .61$.

Table 27

Contrasts of Training Groups for Gains on the Raven's

Comparisons	Contrast	*SE*	*t*	*df*	*p*	*d*
Dual vs. Control*	2.46	.70	3.49	131	.01	.61
Visual vs. Control*	1.96	.70	2.80	131	.01	.49
Auditory vs. Control	1.36	.72	1.88	131	.06	.33
STM vs. Control*	2.10	.70	2.98	131	.03	.52
Dual vs. Visual	.50	.69	.73	131	.47	.13
Dual vs. Auditory	1.10	.71	1.54	131	.13	.27
Dual vs. STM	.36	.69	.52	131	.61	.09
Visual vs. Auditory	.59	.71	.84	131	.40	.15
Visual vs. STM	-.14	.69	-.21	131	.83	-.04
Auditory vs. STM	-.74	.71	-1.00	131	.30	-.17

Note: A positive Contrast indicates that the gain for the first group in the comparison was larger. Asterisk indicates significant ($p < .05$) difference between training group pair.

However, the current evidence shows that participants do not have to train WMC using the dual *n*-back task as training. Participants who went through the visual *n*-back training and, surprisingly, the spatial matrix span training also experienced significant improvement on the Raven's compared to the control group. There was a marginal effect of improvement on the Raven's for participants who completed the auditory *n*-back training compared to the control group. The first two contrasts are evidence that partially supports the hypothesis that participants who completed the dual and visual *n*-back training programs would experience significant gains on the Raven's when compared to participants in the control group. In a serendipitous finding, participants in the spatial matrix span training also experienced greater gains on the Raven's than the control group and the auditory *n*-back training group also experienced close to significant gains compared to the control group. The hypothesis that the dual *n*-back training group would experience greater gains than the other training groups was not supported at all in that there were no differences between the training groups when it came to improvements on the Raven's.

The repeated measures analysis also showed an overall improvement in the BETA-III subtest and the same contrast used for the other tests of G*f* were conducted (see Table 28).

Table 28

Contrasts of Training Groups for Gains on the Beta-III Subtest

Comparisons	Contrast	SE	*t*	*df*	*p*	*d*
Dual vs. Control*	2.50	.60	4.18	131	.01	.73
Visual vs. Control*	1.61	.59	2.70	131	.01	.47
Auditory vs. Control	.66	.62	1.07	131	.29	.19
STM vs. Control	1.15	.60	1.91	131	.06	.33
Dual vs. Visual	.89	.58	1.54	131	.13	.27
Dual vs. Auditory*	1.84	.61	3.05	131	.01	.53
Dual vs. STM*	1.36	.59	2.31	131	.02	.40
Visual vs. Auditory	.95	.60	1.58	131	.12	.28
Visual vs. STM	.46	.58	.79	131	.43	.14
Auditory vs. STM	-.48	.61	-.80	131	.42	-.14

Note: A positive Contrast indicates that the gain for the first group in the comparison was larger. Asterisk indicates significant ($p < .05$) difference between training group pair.

The dual and visual *n*-back training groups experienced greater improvements than the control group. However, the auditory *n*-back training group did not differ from the control group in the improvements and the spatial matrix span training group only experienced a marginal improvement compared to the control group. Furthermore, the auditory *n*-back training group and the spatial matrix span training group experienced significantly less improvement than the dual *n*-back training group, but not the visual *n*-

back training group. The evidence from the contrasts for the BETA-III subtest suggests that the dual *n*-back training was superior for making gains, followed by the visual *n*-back training. The auditory *n*-back training did little in improving scores on the BETA-III subtest and the spatial matrix span may or may not improve scores on the BETA-III subtest.

Cognitive tests (Accuracy).

Analyses that were similar to those used with the tests of G*f* were conducted on the percent correct answers for cognitive tests. According to a repeated measures analysis, the between-subjects analyses showed a sex difference in the percent correct on the four cognitive tests, $F\,(4, 123) = 4.81, p < .001, \eta^2 = .14$, no difference among training groups, $F\,(16, 504) = .76, p = .74, \eta^2 = .02$, and there was a marginal significant interaction between sex and training, $F\,(16, 376) = 1.62, p = .06, \eta^2 = .05$. Because there was a sex difference, an examination of the tests of between-subjects effects was necessary. Sex differences existed for the Extended Range Vocabulary Test, $F\,(1, 126) = 6.88, p < .01, \eta^2 = .05$, the Mental Rotation Test, $F\,(1, 126) = 17.15, p < .001, \eta^2 = .12$, and the Paper Folding Test, $F\,(1, 126) = 4.37, p < .05, \eta^2 = .03$. Males had a higher percent correct than females for all three tests (refer to Tables 12 and 19 for means and standard deviations). There was no sex difference for the Lexical Decision Test, $F\,(1, 126) = .10, p = .76, \eta^2 = .00$.

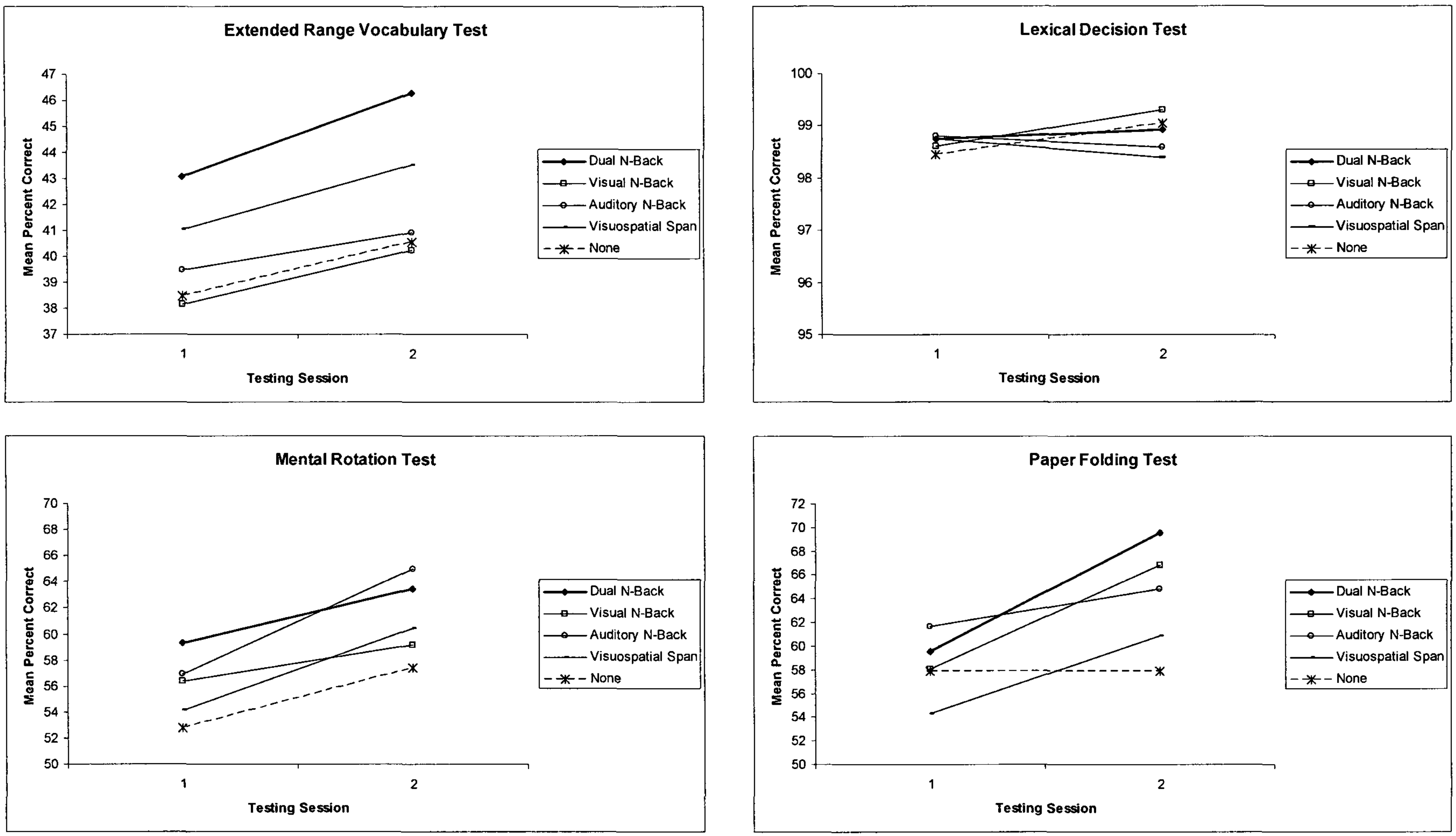
Although the between-subjects interaction between sex and training was not significant, the marginally significant value justified a closer look at each test to determine if there was an interaction in one of the cognitive tests. There was an interaction between sex and training for the Paper Folding Test, $F\,(4, 126) = 4.75, p =$

.001, $\eta^2 = .13$. The differences existed such that females in the dual *n*-back training group and the spatial matrix span training group had a higher percent correct than males, whereas males had a higher percent correct than females in the visual and auditory *n*-back training groups, as well as the control group (refer to Tables 12 and 19 for means and standard errors). The presence of this interaction does not impact the analysis of whether training improved performance on the cognitive tests.

A multivariate test of the four cognitive tests as a group showed a significant difference between the pretests and posttests, $F(4, 123) = 13.18$, $p < .001$, $\eta^2 = .30$. An analysis of the univariate tests indicated a significant change in performance on the Extended Range Vocabulary Test, $F(1, 126) = 12.56$, $p = .001$, $\eta^2 = .09$; the Mental Rotation Test, $F(1, 126) = 21.52$, $p < .001$, $\eta^2 = .15$; and the Paper Folding Test, $F(1, 126) = 23.30$, $p < .001$, $\eta^2 = .16$. There was no significant difference in the percent correct between the Lexical Decision pretest and posttest, $F(1, 126) = .50$, $p = .48$, $\eta^2 = .00$. The three tests that were significant such that there was a higher percent correct on the posttest than the pretest (see Figure 17).

Figure 17

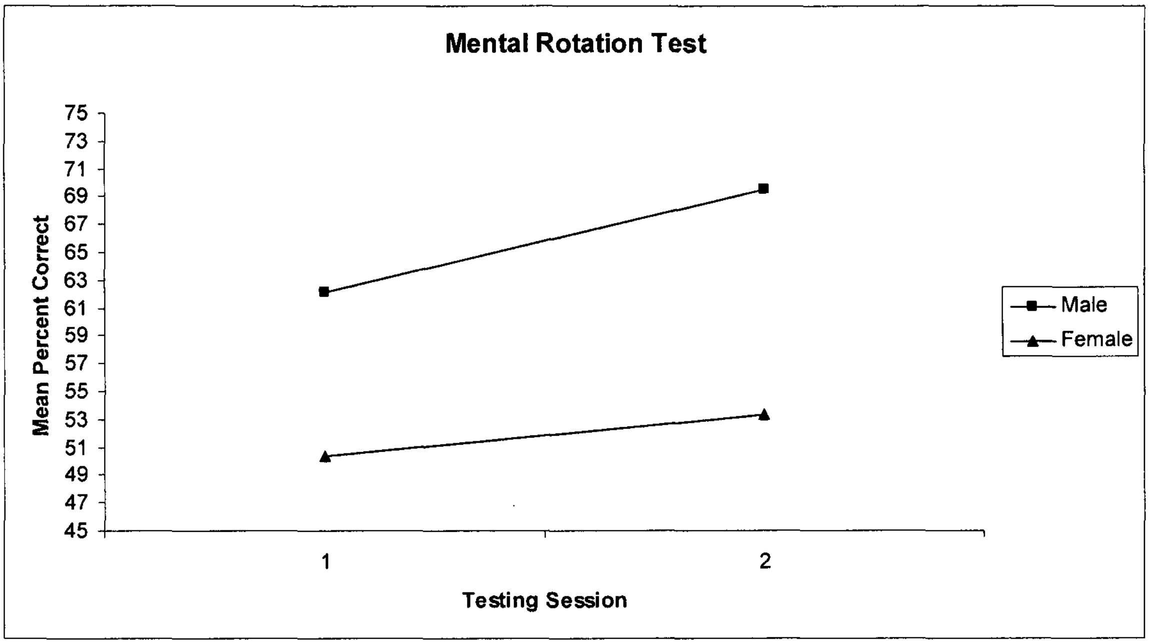
Changes between Pretest and Posttest for Percent Correct on Cognitive Tests



There was no effect of sex on gains in percent correct on cognitive tests with the exception of the Mental Rotation Test, $F(1, 126) = 4.02, p < .05, \eta^2 = .03$. Surprisingly, the sex difference was not a result of females improving their performance more than males; instead, males ($M_{Gain} = 7.60$, $SE = 1.65$) improved significantly more than females ($M_{Gain} = 3.02$, $SE = 1.58$, see Figure 18).

Figure 18

Difference in Gains on Mental Rotation Tests between Females and Males



Note: $N = 136$

Change in percent correct was significantly influenced by training only for the Paper Folding Test, $F(4, 126) = 2.42, p = .05, \eta^2 = .07$; training group effects were not significant for changes in performance on the other tests. Contrasts were conducted to identify specific differences among training groups for gains in terms of percent correct (see Table 29).

Table 29

Contrasts of Training Groups for Gains on the Paper Folding Test

Comparisons	Contrast	SE	*t*	*df*	*p*	*d*
Dual vs. Control*	9.96	3.75	2.66	131	.01	.46
Visual vs. Control*	8.71	3.72	2.34	131	.02	.41
Auditory vs. Control	3.16	3.86	.82	131	.41	.14
STM vs. Control	6.58	3.75	1.76	131	.08	.31
Dual vs. Visual	1.25	3.65	.34	131	.73	.06
Dual vs. Auditory	6.80	3.79	1.80	131	.08	.31
Dual vs. STM	3.38	3.68	.92	131	.36	.16
Visual vs. Auditory	5.55	3.76	1.48	131	.14	.26
Visual vs. STM	2.13	3.65	.59	131	.56	.10
Auditory vs. STM	-3.42	3.79	-.90	131	.37	-.16

Note: A positive Contrast indicates that the gain for the first group in the comparison was larger. Asterisk indicates significant ($p < .05$) difference between training group pair.

Based on the contrasts, participants in the dual and visual *n*-back training conditions experienced greater improvements compared to the control group. However, the improvements made by participants in the dual and visual *n*-back training groups did not differ significantly from the other training groups. Thus, it could be inferred that the dual and visual *n*-back training is better than doing nothing, but that the other two training tasks may provide similar results. A final test of the potential three-way interaction between changes in percent correct, sex, and training did not exist for any of the cognitive tests, $F(16, 504) = .86, p = 49, \eta^2 = .03$.

Based on the evidence presented thus far, the hypothesis that there would be improvements on the Mental Rotation Test and the Paper Folding Test and not on the Extended Range Vocabulary Test and the Lexical Decision Test has been partially confirmed. The hypothesis was correct in that there were significant improvements on the Mental Rotation Test and the Paper Folding Test. However, the hypothesis was not fully supported because there were also significant improvements in the Extended Range Vocabulary Test. The Lexical Decision Test was the only test that there was not significant improvement, which was hypothesized. The hypothesis also stated that the improvements in the tests would vary with types of training; however, this was not the case as there was not statistically significant evidence that type of training had an effect on the changes between the pretests and posttests with the exception of the Paper Folding Test.

The hypothesis stating that improvements on the Mental Rotation Test and the Paper Folding Test would be the result of the dual or visual *n*-back training was also partially supported in that participants in the dual and visual *n*-back training conditions improved only on the Paper Folding Test, but this could be an isolated finding. Finally, the hypothesis stating that females would experience greater improvement in visuospatial abilities than males was not supported by the current statistical evidence which, in fact, went in the opposite direction such that males experienced greater improvements on the Mental Rotation Test than the females.

Cognitive tests (RTs).

Training was hypothesized to improve RTs on the cognitive tests. A repeated measure analysis showed no between-subjects effect for sex, $F(4, 123) = 1.99, p = .10, \eta^2 = .06$, or training, $F(16, 376) = 1.32, p = .18, \eta^2 = .04$. The potential sex by training interaction was also nonsignificant, $F(16, 376) = .96, p = .50, \eta^2 = .03$. The analysis for within-subjects effect did show that there was a difference between pretests and posttests RTs, $F(4, 123) = 20.48, p < .001, \eta^2 = .40$, with the posttests having faster RTs. However, there was no sex difference in the change in RTs, $F(4, 123) = .14, p = .97, \eta^2 = .00$, or differences among training groups, $F(16, 376) = 1.11, p = .35, \eta^2 = .04$. The potential 3-way interaction between change in RTs, sex, and training was also nonsignificant, $F(16, 376) = 1.19, p = .28, \eta^2 = .04$.

Because of the change in RTs between cognitive pretests and posttests, an inspection of the univariate tests was necessary. The univariate tests indicated a change in RTs for the Extended Range Vocabulary Test, $F(1, 126) = 56.23, p < .001, \eta^2 = .31$; the Lexical Decision Test, $F(1, 126) = 10.65, p < .001, \eta^2 = .08$; the Mental Rotation Test, $F(1, 126) = 21.55, p < .001, \eta^2 = .15$; and the Paper Folding Test, $F(1, 126) = 15.82, p < .001, \eta^2 = .11$. Figure 19 provides a graphical representation of the change in RTs for correct answers in each cognitive test per training group.

Summary

The results are a number of significant results in the current study and are summarized in Table 30. The table shows each research question followed by a list of items that answers each question accordingly.

Figure 19

Improvements between Pretest and Posttest for RTs for Correct Answers on Cognitive Tests

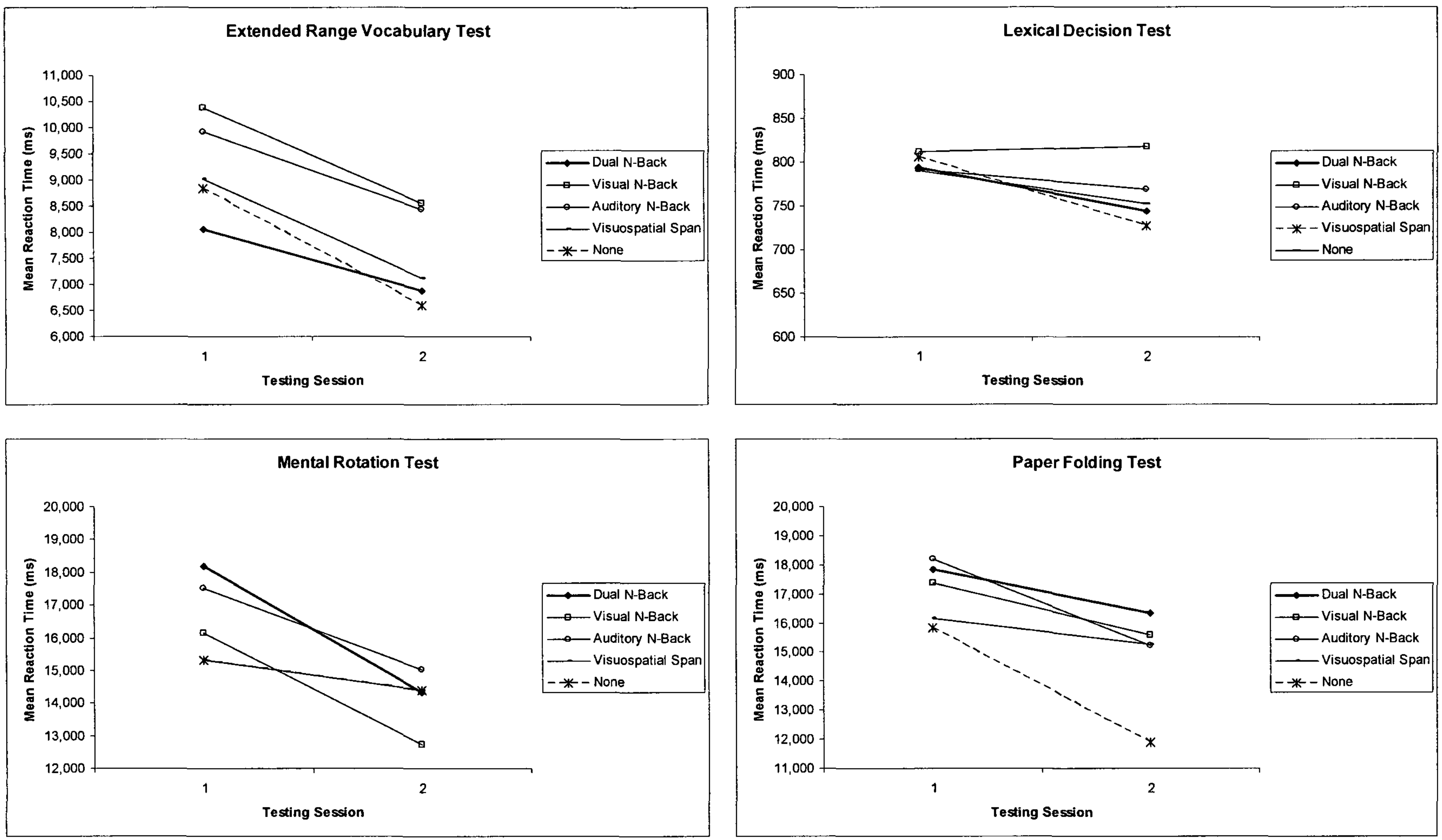


Table 30

Summary of Results for the Current Study

Research Question #1

Do scores on tests of Gf increase after participants complete a cognitive training program using the dual n-back task?

1) Yes, scores increase on all tests of *Gf* when participants complete a training program using the dual *n*-back task.

Research Question #2

Will participants in the dual, visual, and auditory n-back training groups and the STM training group experience greater gains on tests of Gf than the control group? Do participants who complete the dual n-back task training programs experience greater increases in scores on tests of Gf compared to groups completing the visual or auditory n-back training or the STM training?

1) Dual and visual *n*-back and STM training groups had greater gains than the control group on the Raven's.
    a. No differences among training groups in gains on scores on Raven's.
2) Dual and visual *n*-back training groups had greater gains than control on BETA-III.
    a. Dual *n*-back training group had greater gains than auditory *n*-back and STM training groups on BETA-III.
    b. No difference between control group and auditory *n*-back or STM training groups on BETA-III.
3) STM training group had greater gains than control group on WASI.
    a. No other differences between training groups in gains in scores for WASI.
4) No training effect on gains in scores on Cattell.

Research Question #3

Will participants' performance on the cognitive tests improve as a result of the training?

1) Overall improvement, but only Paper Folding Test greater than control.
2) Overall decrease in RTs but no difference as a result of training.

Research Question #4

If participants improve on the Mental Rotation Test and the Paper Folding Test, will there be differences in their improvement based on the type of training they received?

1) Dual and visual *n*-back training groups experienced greater gains than control group on Paper Folding Test.
    a. No differences among training groups existed.
2) No differences in RTs as a result of training.

Research Question #5

Will females and males experience an improvement in visuospatial abilities? Furthermore, will female's improvement be greater than males in visuospatial abilities and tests of Gf?

1) Yes, females and males experienced improvement in scores on the Mental Rotation Test and the Paper Folding Test.
    a. There was also improvement on the Extended Range Vocabulary Test, which was not predicted.
    b. Males experienced greater gains on Mental Rotation Test than females, which was not predicted.
2) No other differences were found in gains on cognitive tests.
3) No sex difference in gains on tests of *Gf*.

Chapter Four

Discussion

The results of the current study raise many questions about how researchers define *Gf* regarding plasticity in terms of improvement through training and the psychometric tests that are used to test *Gf*. *Gf* is conceptually defined as an ability to solve novel problems without the use of prior strategies, and *Gf* is assumed to function without relying on strategies derived from verbal and visuospatial abilities (Horn & Cattell, 1966; & Raven, 2000). Thus, psychometric tests such as the Raven's Progressive Matrices and Cattell's Culture Fair Test are assumed to test *Gf* without the test taker using other abilities (Cattell, 1963; Horn & Cattell, 1966; Jensen, 1998). However, recent research on tests of *Gf* has provided evidence that the tests may not be testing *Gf* exclusively and the tests may be multidimensional (van der Ven & Ellis, 2000; Vigneau & Bors, 2005). In addition, Kane, et al. (2004) found a strong relationship between visuospatial storage and *Gf* similar to the relationship between attentional control and *Gf*, which casts further doubt on the notion that tests of *Gf* measure only a single construct.

The current study lends support for multidimensionality in tests of *Gf*. The claim that the current study supports multidimensionality in tests of *Gf* is based on the evidence that the greatest gains in the tests of *Gf* were experienced by training groups that had a visuospatial component, including the WMC *and* the STM training. Furthermore, the Paper Folding Test had the strongest correlation with the tests of *Gf* and it was the only cognitive test in which participants experienced gains as a result of the dual and visual *n*-back training, which both have a substantial visuospatial component. However, did improvements on the Paper Folding Test result in an improvement in *Gf*, attentional control, or an improvement in visuospatial skills, or are these merely the same labels for the same construct? Tests of G*f* are assumed to predict other cognitive abilities (Sternberg, 2008). If there was an improvement in G*f*, then there should have also been an improvement in the Mental Rotation Test because G*f* is not domain or task specific. There were no improvements in the Mental Rotation Test as a result of training in the current study, but there were improvements in the Paper Folding Test. Therefore, attentional control and *some* visuospatial abilities were improved that led to improved scores on tests of G*f*; suggesting a multidimensional component to those tests.

The evidence from the current study also offers some support for the argument of unidimesionality in tests of G*f* based on the fact that attentional control may have been enhanced instead of visuospatial abilities because females did not experience an advantage over males in gains on tests of G*f* after the training. Males outperformed females on visuospatial tasks for pretests and posttests without the females making greater gains than males on the tests. However, the finding that males made greater gains than females is not necessarily counterintuitive. Ceci and Papierno (2003) argued that if training is provided for a group of people that already have an advantage in a cognitive ability and a group that does not have an advantage, then the group with the advantage will experience greater gains from the training, which is referred to as the Matthew Effect. Thus, the training that males received in the current study most likely sharpened or enhanced their mental rotation abilities that gave them an even greater advantage than the females. If tests of G*f* rely on visuospatial abilities *and* attentional control, then males should have made greater gains on tests of G*f* as a result of sharpening or enhancing their visuospatial abilities, which are already at an advantage over females' visuospatial abilities. However, males did not experience greater gains on the tests of G*f*, thus, the training most likely influenced participants' attentional control, which had an effect on G*f*.

Improving G*f*

The question, "Does training to enhance working memory capacity improve fluid intelligence," can be answered with a "yes," depending on which test of G*f* is being used, which in the current study includes the Raven's Advanced Progressive Matrices and the BETA-III subtest. The improvement in scores on the Raven's Advanced Progressive Matrices and BETA-III subtest was the result of participants completing training programs that enhanced their attentional control using visuospatial stimuli; as seen by the participants who completed the dual and visual *n*-back training. However, participants who completed the STM training also experienced greater gains than the control group and did not differ in gains compared to the WMC training groups. Why did participants in the STM training group experience gains similar to the WMC training groups? The most likely explanation is that the STM training enhanced participants' attentional control. According to Engle et al.'s (1999) structural equation model, STM and WMC have equal weight in the factor labeled as "COMMON," which refers to attentional control. Therefore, attentional control could be enhanced through training STM similar to training WMC and then have an effect on G*f*.

Although training STM or WMC improves G*f*, the training that contains a visuospatial component should experience the greatest gains. Participants who completed the auditory *n*-back training experienced marginal gains compared to the control group, but did not differ from the other training groups. Furthermore, the lack of difference in gains among training groups and a difference in gains between the control group and each training group suggests that some training programs are better than others. There were also differences among training groups depending on the test of G*f*. The current study suggests that the dual and visual *n*-back training programs are superior to the auditory *n*-back and spatial matrix span training programs because the dual and visual *n*-back tasks have a strong attentional control component, especially the dual *n*-back task, with visuospatial stimuli. The third "best" training program is the STM training. Although the STM training does not have as complicated of a task to train attentional control, the task still requires attention, which may lead to improvements in G*f*. The visuospatial aspect of the STM memory most likely has an influence as well, but needs to be compared to an auditory STM training test in future research. Finally, the auditory *n*-back training is not useless, but definitely does not provide the training necessary for participants to experience strong gains. The only difference between the auditory *n*-back training and the other WMC training programs is the type of stimuli used to train participants. Because the auditory *n*-back training lacks a visuospatial component and participants who completed the training did not differ from the control group, a visuospatial component most likely provides an advantage in obtaining gains in G*f* but is not a requirement.

Conceptualizing and Measuring G*f*

A primary issue regarding the psychometric tests of G*f* is whether the tests are measuring the same construct. Although the tests of G*f* were all significantly correlated with each other in the current study, there was improvement in only two of the four tests. If the tests of G*f* are truly measuring the same construct, then there should have been a consistent pattern of improvement across all tests of *Gf* as a result of training. Even though the Raven's Advanced Progressive Matrices had strong pretest and posttest correlation values with Cattell's Culture Fair Test, absolutely no gains on the Cattell's was made as a result of specific training. This could have been because Cattell's Culture Fair Test is supposed to, at least psychometrically, test *g* as defined by Spearman (1904) and not *Gf* exclusively. The directions given to test takers for Cattell's Culture Fair Test are also different than the Raven's Advanced Progressive Matrices in that the directions for Cattell's explicitly states the rules for each subtest whereas the directions for the Raven's Advanced Progressive Matrices do not identify the rules for each problem; thus, the test taker ends up needing to identify the rules in addition to producing an answer.

Pretest and posttest correlations were also strong between the Raven's Advanced Progressive Matrices and the WASI subtest and, yet, there was only a marginal impact on gains in scores for WASI subtest. A possible explanation for the marginal impact on the WASI subtest scores is that the subtest was only a small portion of a larger test of intelligence, and may not have as strong validity as the Raven's Advanced Progressive Matrices for measuring *Gf* exclusively. Although it seems probable that the lack of gains on the WASI subtest was a result of this test being only a part of a larger test and it may not be as valid as a stand alone test, the reasoning does not coincide with the impact that training had on the BETA-III subtest because it is also part of a larger test of intelligence, but there were improvements on the BETA-III subtest. Furthermore, the correlation between the Raven's Advanced Progressive Matrices and the BETA-III subtest were lower than the other tests' correlations with the Raven's Advanced Progressive Matrices.

Why would there be an improvement on two tests of *Gf* that had a lower correlation value than some of the other tests? When the Beta-III and the WASI were created, they were correlated with other tests using an aggregate score from all of the subtests (Kellogg & Morton, 1999; WASI Manual, 1999). In the current study, only one subtest was used and could have impacted the correlation between the BETA-III subtest and WASI subtest and the Raven's Advanced Progressive Matrices. In other words, the correlation value could have been affected by having a value from only a portion of the WASI or BETA-III tests instead of the entire test, which could have resulted in a less valid correlation value between the subtests (i.e., WASI and BETA-III) and the tests that were given in their entirety (i.e., Raven's and Cattell's). The other explanation is that the training is specific to improving *Gf* exclusively as it is measured by the Raven's Advanced Progressive Matrices, but not other tests or subtests of *Gf*. If the training is specific to one test of *Gf*, then it is apparent that the tests of *Gf* are not measuring the same construct.

Improving Cognitive Abilities

The current study found limited improvements on the two verbal tests and the two visuospatial tests. The Paper Folding Test was the only cognitive test that participants experienced greater improvements as a result of training when compared to the control group. If scores on the Mental Rotation Test were also improved as a result of training, then a conclusion could have been made that training improves visuospatial abilities. The finding that there was improvement only on Paper Folding Test scores does suggest, however, that the training was not domain specific and attentional control was most likely improved, which is most likely a critical requirement for completing the Paper Folding Test.

There was also an unpredicted improvement in the Extended Range Vocabulary Test that was not a result of training. The improvement in percent correct on the vocabulary test is most likely a practice effect. Furthermore, training did not have an effect on RTs for any of the cognitive tests, which could be a result of the training not necessarily having an effect on processing speed. If training does not help with processing speed as observed in the current study, then training probably has an effect on the allocation of attention resources to the appropriate information while ignoring irrelevant information, which could explain the increase in percent correct in the Paper Folding Test. If attention is directed more efficiently, then this would lead to more correct answers, but not necessarily more questions being answered.

The evidence from the lack of statistically significant differences among training groups in improvement on cognitive tests, with the exception of the Paper Folding Test, supports the conclusion that attentional control was most likely improved through training instead of visuospatial abilities. If attentional control was improved and no improvements were made on visuospatial or verbal tests specifically, then the improvements on the tests of G*f* were most likely a result of improving attentional control.

Limitations and Future Research

The limitations in the current study do not necessarily detract from the final results; rather, as with the limitations in Jaeggi et al.'s study (2008), the limitations lead to empirical questions that provide opportunities for future research. For example, time restrictions on the Raven's Advanced Progressive Matrices and the WASI subtest are not typically used. If a time restriction is used on the Raven's Advanced Progressive Matrices, then it is typically set at a 45 minute limitation (Raven et al., 1999). Moody (2009) pointed out that having participants complete items in the Raven's Advanced Progressive Matrices within a restrictive time limit does not allow them to attempt the more difficult items in the test, which are located toward the end of the test. Furthermore, Moody argued that if the participants do not attempt the difficult items, then the potential high score a participant may receive on the test is less predictable.

A restricted time was used on all four tests of G*f* in the current study to replicate Jaeggi et al.'s (2008) study as closely as possible. The fact that a time restriction was still used in the current study provides an opportunity for further research in which no time restrictions are used on tests of G*f* that do not typically use the restriction. Furthermore, a study replicating the current study and Jaeggi et al.'s can be done with the knowledge the dual *n*-back training program has been successful in two studies and that only one other training program (i.e., visual *n*-back training) has a similar success rate for improving scores on tests of G*f*.

In a similar vein to using time restrictions on tests of G*f*, the current study did not test if the training has an effect on a larger test of intelligence such as the Stanford-Binet Intelligence Scales, the complete Wechsler Scale of Intelligence, or the complete BETA-III test. The larger scale tests of intelligence do not test G*f* exclusively, but the increase in G*f* could lead to solving problems on the larger tests of intelligence with greater ease.

Another limitation of the current study that opens the doors for future research is that there has been no determination for whether the training transfers to everyday situations or habits such as taking notes, improving study habits, improving skills learning, improving test scores in college courses, or improved attention during class lectures. If the training is truly improving attentional control and G*f*, then a reasonable hypothesis would be that a person would be able to take notes, study for tests, and attend to information more efficiently. Tests of G*f* also predict job performance, and the improvement could transfer to those aspects as well. The training may transfer to academic performance or skills performance because exercising cognitive abilities may follow the old adage that, "Doing something is better than doing nothing."

There are also other nuanced areas that the training could be applied. For example, does training WMC increase abilities in children? There are specific training programs on BrainTwister for children (Buschkuehl, et al., 2008). There could be a stronger effect of the training on children because of their plasticity. Similarly, does the training improve scores on tests of G*f* for people who are elderly? Or, does the training help slow the cognitive decline in cognitive abilities? The training to improve WMC can be applied to a range of cognitive development studies now that there is confirmation that the training works for young adults.

Another question that arises from the current study and Jaeggi et al.'s (2008) study is whether the improvements experienced in the tests of G*f* last over an extended period of time once the training has stopped. Most likely, the training would have to be maintained for the effects to remain. However, the training tasks are not exciting and are quite tedious, which leads to other opportunities for future research that could determine if the dual *n*-back task could be incorporated into a video game or a task that is more enjoyable to play. Research has shown that playing first or third person video games that require a person to navigate through complex routes and maps improve visual attention (Green & Bavelier, 2003). If a dual or visual *n*-back task was incorporated into such a video game, then there could be the potential for a training program that is entertaining and potentially improve scores on tests of G*f*.

Another question that has been left unanswered concerns the training protocol. The current study and Jaeggi et al. (2008) have determined, so far, that improvements on scores of tests of G*f* result from training five days per week, 20 minutes per day, and for the duration of four weeks. Does the training have to be five days per week for 20 minutes per day? Could the training be reduced to two or three days per weeks for 30 to 40 minutes per training session? Participants in the current study reported reaching a threshold after about 11-12 minutes during each session and then "crashing" in training performance after reaching the threshold and then slowly reaching the threshold again during the remain 8-9 minutes. The report made by participants was not validated by empirical means, but is a phenomenon that is worth looking into in future research.

Participants also started at *2*-back on WMC training programs or remembering two placements in the spatial matrix span training program each time they started a new training session. Is it necessary to start at *2*-back each time? Why not start a new training session where the participant left off in the previous training session? A reasonable hypothesis would be that greater improvements might be made in the training if participants could start each new training session where they left off. The other possibility is that the progression from the *2*-back is necessary in making overall improvements, but this hypothesis needs to be tested empirically.

Overall, the current study leaves many important questions unanswered. Some questions are the result of the limitations in the current study. However, none of the limitations or unanswered questions diminishes the validity or the valuable results observed in the current study. The limitations and questions provide opportunities for unique and fruitful future research.

Conclusion

The results of the current study provide evidence that G*f* is more malleable than was previously thought. The current study was successful in replicating Jaeggi et al.'s (2008) study in that participants' scores improved on the Raven's Advanced Progressive Matrices and the BETA-III subtest after completing the dual *n*-back training program. However, participants who completed the visual *n*-back training program or the spatial matrix span training program also experienced gains in the tests of G*f*. The current study, in conjunction with Jaeggi et al.'s study has clearly demonstrated that scores on tests of G*f* can be improved after training to improve WMC and can be used as a foundation for future research investigating ways to improve intelligence.

References

Abad, F. J., Colom, R., Rebollo, I., & Escorial, S. (2004). Sex differential item functioning in the Raven's Advanced Progressive Matrices: Evidence for bias. *Personality and Individual Differences, 36*, 1459-1470. doi:10.1016/S0191-8869(03)00241-1

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs. *Psychological Bulletin, 131*, 30-60. doi:10.1037/0033-2909.131.1.30

Arthur, W., & Woehr, D. J. (1993). A confirmatory factor analytic study examining the dimensionality of the Raven's Advanced Progressive Matrices. *Educational and Psychological Measurement, 53*, 471-478. doi:10.1177/0013164493053002016

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417-423. doi:10.1016/S1364-6613(00)01538-2

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

Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., & Noll, D. C. (1997). A parametric study of prefrontal involvement in human working memory. *Neuroimage, 5*, 49-62. doi:10.1006/nimg.1996.0247

Buschkuehl, M., Jaeggi, S. M., Kobel, A., & Perrig, W. J. (2008). *BrainTwister: A collection of cognitive training tasks.* University of Bern, Switzerland: Department of Psychology.

Case, R., Kurland, D. M., & Goldberg, J. (1983). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology, 33*, 386-404. doi:10.1016/0022-0965(82)90054-6

Castel, A. D., Pratt, J., & Drummond, E. (2005). The effects of action video game experience on the time course of inhibition of return and efficiency of visual search. *Acta Psychologica, 119*, 217-230. doi:10.1016/j.actpsy.2005.02.004

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*, 1-22. doi:10.1037/h0046743

Ceci, S. J., & Papierno, P. B. (2003). The rhetoric and reality of gap closing: When the "have-nots" gain but the "haves" gain even more." *American Psychologist, 60*, 149-160. doi: 10.1037/0003-066X.60.2.149

Cherney, I. D. (2008). Mom, let me play more computer games: They improve my mental rotation skills. *Sex Roles, 59*, 776-786. doi:10.1007/s11199-008-9498-z

Colom, R., Escorial, S., & Rebollo, I. (2004). Sex differences on the Progressive Matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences, 37*, 1289-1293. doi:10.2307/1130467

Colom, R., & García-Lopez, O. (2002). Sex differences in fluid intelligence among high school graduates. *Personality and Individual Differences, 37*, 445-451. doi:10.1016/S0191-8869(01)00040-X

Coltheart, M. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology, 33A*, 497-505. doi: 10.1080/14640748108400805

Court, J. H. (1983). Sex differences in performance on Raven's Progressive Matrices: A review. *The Alberta Journal of Educational Review, 29*(1), 54-74.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* New York, NY: Oxford University Press.

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

Engle, R. W. (2001). What is working memory capacity? In H. L. Roediger, J. S. Nairne, I. Neath, & A. M. Suprenant (Eds.), *The nature of remembering: Essays in honor of Robert G. Crowder* (pp. 297-314). Washington, DC: American Psychological Association Press.

Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*, 19-23. doi:10.1111/1467-8721.00160

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. doi:10.1037/0096-3445.128.3.309

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science, 18*, 850-855. doi:10.1111/j.1467-9280.2007.01990.x

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature, 423*, 534-537. doi:10.1038/nature01647

Green, C. S. & Bavelier, D. (2006a). Effect of action video games on the spatial distribution of visuospatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 32*, 1465-1478. doi:10.1037/0096-1523.32.6.1465

Green, C. S. & Bavelier, D. (2006b). Enumeration versus multiple object tracking: The case of action video game players. *Cognition, 101*, 217-245. doi:10.1016/j.cognition.2005.10.004

Green, C. S. & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science, 18*, 88-94. doi:10.1111/j.1467-9280.2007.01853.x

Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Halpern, D. F. (2004). A cognitive-process taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science, 13*, 135-139. doi:10.1111/j.0963-7214.2004.00292.x

Halpern, D. F., & Wai, J. (2007). The world of competitive scrabble: Novice and expert differences in visuospatial and verbal abilities. *Journal of Experimental Psychology: Applied, 13*, 79-94. doi: 10.1037/1076-898X.13.2.79

Heitz, R. P., Unsworth, N., & Engle, R. W. (2004). Working memory capacity, attention control, and fluid intelligence. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 61-77). Thousand Oaks, CA: Sage Publications, Inc.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*, 258-270. doi:10.1037/h0023816

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581-592. doi:10.1037/0003-066X.60.6.581

Irwing, P., & Lynn, R. (2005). Sex differences in means and variability on the progressive matrices in university students: A meta-analysis. *British Journal of Psychology, 96*, 505-524. doi:10.1348/000712605X53542

Jaeggi, S. M., Buschkuehl, M., Etienne, A., Ozdoba, C., Perrig, W. J., & Nirkko, A. C. (2007). On how high performers keep cool brains in situations of cognitive overload. *Cognitive, Affective, & Behavioral Neuroscience, 7*, 75-89. doi:10.3758/CABN.7.2.75

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences, 105*, 6829-6833. doi:10.1073/pnas.0801268105

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

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, 66-71. doi:10.1037/0033-2909.131.1.66

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*, 189-217. doi:10.1037/0096-3445.133.2.189

Kellogg, C. E., & Morton, N. W. (1999). *BETA III manual.* San Antonio, TX: Harcourt Assessment.

Kline, P. (2000). *Handbook of psychological testing* (2nd ed.). New York: Routledge.

Kyllonen, P. C. (2002). *g*: Knowledge, speed, strategies, or working-memory capacity? In R. J. Sternberg, & E. L. Gigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 415-445). Mahwah, NJ: Erlbaum.

Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence, 14*, 389-433. doi:10.1016/S0160-2896(05)80012-1

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479-1498. doi:10.2307/1130467

Lynn, R., & Irwing, P. (2004). Sex differences on the advanced progressive matrices in college students. *Personality and Individual Differences, 37*, 219-223. doi:10.1016/j.paid.2003.08.028

Mackintosh, N. J., & Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence, 33*, 663-674. doi:10.1016/j.intell.2005.03.004

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive *Gf-Gc* framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 131-150). New York: Guilford.

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence, 37*, 327-328. doi: 10.1016/j.intell.2009.04.005

Norman, D. A., & Shallice, T. (1986). Attention to action: Willed and automatic control of behavior. In R. J. Davidson, G. E. Schwartz, & D. Shapiro, (Eds.), *Consciousness and self-regulation: Advances in research and theory, Vol. 4* (pp. 1-18). New York: Plenum Press.

Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology, 15*, 33-58. doi:10.1016/0193-3973(94)90005-1

Rastle, K., Harrington, J., & Coltheart, M. (2002). 358,534 nonwords: The ARC nonword database. *The Quarterly Journal of Experimental Psychology, 55A*, 1339-1362. doi: 10.1080/02724980244000099

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111*, 159-182. doi:10.1037/0033-295X.111.1.159

Raven, J. (2000). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*, 1-48. doi:10.1111/j.1745-3984.1989.tb00314.x

Raven, J., Raven, J. C., & Court, J. H. (1998). *Advanced Progressive Matrices: 1998 edition.* San Antonio, TX: Harcourt Assessment.

Schweizer, K., Goldhammer, F., Rauch, W., & Moosbrugger, H. (2007). On the validity of Raven's matrices test: Does spatial ability contribute to performance? *Personality and Individual Differences, 43*, 1998-2010. doi:10.1016/j.paid.2007.06.008

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125*, 4-27. doi:10.1037/0096-3445.125.1.4

Shepard, R., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703. doi:10.1126/science.171.3972.701

Spearman, C. (1904). "General intelligence," objectively determined and measured. *The American Journal of Psychology, 15*, 201-292. doi:10.2307/1412107

Spearman, C. (1914). The theory of two factors. *Psychological Review, 21*, 101-115. doi:10.1037/h0070799

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences, 105*, 6791-6792. doi:10.1073/pnas.0803396105

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22*, 996-1013. doi:10.1002/acp.1420

Thompson, P. M., Cannon, T. D., Narr, L. K., van Erp, T., Poutanen, V., Huttunen, M., Toga, A. W. (2001). Genetic influences on brain structures. *Nature Neuroscience, 4*, 1253-1258. doi:10.1038/nn758

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127-154. doi:10.1016/0749-596X(89)90040-5 van der Ven, A. H. G. S., & Ellis, J. L. (2000). A Rasch analysis of Raven's standard progressive matrices. *Personality and Individual Differences, 29*, 45-64. doi:10.1016/S0191-8869(99)00177-4

Vandenburg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599-601.

Vigneau, F., & Bors, D. A. (2008). The quest for item types based on information processing: An analysis of Raven's Advanced Progressive Matrices, with a consideration of gender differences. *Intelligence, 36*, 702-710. doi:10.1016/j.intell.2008.04.004

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250-270. doi:10.1037/0033-2909.117.2.250

WASI Manual. (1999). San Antonio, TX: Harcourt Assessment.

Watter, S., Geffen, G. M., & Geffen, L. B. (2001). The *n*-back as a dual-task: P300 morphology under divided attention. *Psychophysiology, 38*, 998-1003. doi:10.1111/1469-8986.3860998

Wilhoit, B. E., & McCallum, R. S. (2003). Cross-battery assessment of nonverbal cognitive ability. In R. S. McCallum (Ed.), *Handbook of nonverbal assessment* (pp.63-78). New York: Plenum.

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 15*, 763-771. doi:10.3758/PBR.15.4.7

Appendix A

Means and Standard Errors for Average Performance per Training Session

	Dual		Visual		Auditory		STM	
	Males	Females	Males	Females	Males	Females	Males	Females
Session 1	1.74 (.11)	1.97 (.10)	2.94 (.24)	2.18 (.13)	2.81 (.15)	2.66 (.18)	5.59 (.21)	5.62 (.11)
Session 2	2.17 (.15)	2.36 (.11)	3.49 (.31)	2.70 (.15)	3.27 (.16)	3.11 (.24)	5.64 (.22)	5.77 (.14)
Session 3	2.33 (.19)	2.43 (.15)	3.53 (.22)	2.90 (.20)	3.26 (.19)	3.58 (.27)	5.90 (.21)	5.95 (.15)
Session 4	2.43 (.19)	2.56 (.14)	3.70 (.30)	3.36 (.27)	3.42 (.22)	3.54 (.29)	5.93 (.19)	6.14 (.17)
Session 5	2.62 (.22)	2.71 (.17)	3.98 (.29)	3.45 (.30)	3.62 (.21)	3.46 (.19)	5.92 (.23)	5.99 (.15)
Session 6	2.70 (.19)	2.74 (.20)	4.35 (.23)	3.45 (.23)	4.30 (.26)	3.68 (.34)	5.99 (.24)	6.05 (.15)
Session 7	2.97 (.27)	2.68 (.18)	4.37 (.34)	3.50 (.30)	4.17 (.26)	3.89 (.29)	5.75 (.29)	6.11 (.17)
Session 8	3.00 (.21)	2.97 (.17)	4.87 (.37)	3.54 (.25)	4.17 (.30)	3.84 (.32)	5.92 (.20)	6.06 (.17)
Session 9	3.17 (.23)	3.08 (.16)	4.68 (.26)	3.70 (.27)	3.95 (.27)	3.99 (.31)	5.93 (.24)	6.06 (.20)
Session 10	2.76 (.21)	3.03 (.20)	4.67 (.32)	3.88 (.26)	4.22 (.33)	3.88 (.35)	6.17 (.17)	5.93 (.24)
Session 11	3.02 (.27)	3.09 (.21)	4.71 (.34)	3.66 (.22)	4.23 (.40)	3.66 (.29)	6.17 (.22)	6.19 (.20)
Session 12	2.91 (.25)	3.32 (.24)	4.80 (.36)	3.90 (.28)	4.48 (.40)	3.90 (.32)	6.06 (.22)	6.04 (.20)
Session 13	3.09 (.32)	3.17 (.21)	4.89 (.28)	3.76 (.33)	4.59 (.30)	3.76 (.38)	6.08 (.22)	6.33 (.23)
Session 14	3.25 (.28)	3.12 (.23)	4.79 (.36)	3.75 (.28)	4.44 (.33)	3.75 (.28)	6.05 (.21)	6.30 (.21)
Session 15	3.04 (.27)	3.31 (.25)	4.85 (.32)	3.73 (.24)	4.50 (.43)	3.73 (.40)	6.33 (.22)	6.31 (.20)
Session 16	3.21 (.30)	3.43 (.22)	4.97 (.38)	4.03 (.30)	4.17 (.40)	4.03 (.35)	6.20 (.23)	6.21 (.25)
Session 17	3.29 (.29)	3.27 (.22)	4.82 (.39)	3.48 (.21)	4.60 (.36)	3.48 (.38)	6.02 (.30)	6.28 (.22)
Session 18	3.27 (.27)	3.36 (.25)	5.06 (.31)	3.94 (.23)	4.40 (.44)	3.94 (.37)	6.16 (.20)	6.46 (.26)

Appendix B

Means and Standard Errors for Max Performance per Training Session

	Dual		Visual		Auditory		STM	
	Males	Females	Males	Females	Males	Females	Males	Females
Session 1	2.71 (.19)	2.93 (.13)	4.29 (.27)	3.53 (.22)	4.00 (.25)	3.84 (.25)	8.08 (.33)	7.87 (.22)
Session 2	3.00 (.18)	3.50 (.17)	5.14 (.42)	3.93 (.21)	4.92 (.26)	4.38 (.31)	8.38 (.35)	8.07 (.21)
Session 3	3.21 (.21)	3.43 (.20)	5.07 (.30)	4.20 (.26)	4.42 (.26)	5.15 (.42)	8.15 (.34)	8.47 (.17)
Session 4	3.43 (.29)	3.57 (.23)	5.21 (.46)	4.60 (.35)	4.83 (.21)	4.92 (.43)	8.46 (.31)	8.73 (.28)
Session 5	3.50 (.25)	3.79 (.21)	5.57 (.37)	5.07 (.34)	5.25 (.28)	5.23 (.30)	8.46 (.31)	8.40 (.21)
Session 6	3.71 (.29)	3.71 (.27)	5.93 (.35)	4.93 (.33)	5.67 (.33)	5.46 (.49)	8.54 (.27)	8.40 (.16)
Session 7	3.86 (.31)	3.64 (.25)	6.21 (.53)	5.07 (.38)	5.83 (.32)	5.62 (.37)	8.15 (.36)	8.53 (.24)
Session 8	4.14 (.27)	4.21 (.21)	6.93 (.64)	4.93 (.32)	5.50 (.34)	5.54 (.51)	8.46 (.27)	8.47 (.27)
Session 9	4.43 (.39)	4.21 (.24)	6.36 (.39)	5.13 (.34)	5.58 (.31)	5.62 (.45)	8.15 (.34)	8.67 (.27)
Session 10	3.79 (.24)	4.00 (.30)	6.36 (.50)	5.53 (.36)	5.58 (.44)	5.92 (.45)	8.54 (.24)	8.60 (.32)
Session 11	4.07 (.38)	4.29 (.27)	6.57 (.50)	4.80 (.22)	5.83 (.52)	5.62 (.42)	8.77 (.20)	8.93 (.27)
Session 12	4.14 (.29)	4.57 (.27)	6.79 (.54)	5.47 (.39)	6.33 (.58)	6.15 (.45)	8.62 (.21)	8.73 (.34)
Session 13	4.29 (.42)	4.42 (.23)	6.79 (.47)	5.13 (.51)	6.42 (.40)	6.23 (.58)	8.69 (.29)	8.87 (.27)
Session 14	4.50 (.33)	4.29 (.29)	6.93 (.62)	5.27 (.36)	6.25 (.41)	6.00 (.48)	8.31 (.24)	8.93 (.28)
Session 15	4.29 (.35)	4.64 (.32)	6.93 (.56)	5.13 (.34)	7.58 (.57)	6.38 (.60)	9.23 (.30)	8.87 (.24)
Session 16	4.43 (.31)	4.64 (.32)	7.00 (.51)	5.47 (.46)	5.92 (.67)	6.15 (.58)	9.08 (.24)	8.73 (.27)
Session 17	4.43 (.37)	4.71 (.29)	6.64 (.76)	4.60 (.29)	6.25 (.51)	5.92 (.59)	8.85 (.36)	9.00 (.34)
Session 18	4.43 (.33)	4.71 (.32)	7.00 (.56)	5.60 (.41)	6.25 (.63)	6.23 (.53)	9.08 (.24)	8.80 (.31)

Appendix C

Dual N-Back Training Group's Intercorrelations between Tests of Gf and Percent Correct on Cognitive Tests

Measure	1	2	3	4	5	6	7	8
1. Raven	—	.69**	.60**	.44*	.43*	.13	.28	.55**
2. Cattell	.72**	—	.71**	.57**	.30	.14	.30	.64**
3. WASI	.60**	.64**	—	.63**	.30	.33	.10	.50**
4. BETA	.32	.35	.34	—	.29	.02	.30	.64**
5. ERVT	.24	.36	.46*	.02	—	.36	.34	.15
6. LDT	.11	.14	.10	-.04	.15	—	-.08	.00
7. MRT	.20	.17	.05	.07	.12	.08	—	.28
8. Paper	.49	.33	.22	.29	.24	-.04	.16	—

Note: $n = 28$. Correlations below the diagonal represent pretests and correlations above the diagonal represent posttests. ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.
** indicates p-values at or < .01
* indicates p-values at or < .05

Appendix D

Visual N-Back Training Group's Intercorrelations between Tests of Gf and Percent Correct on Cognitive Tests

Measure	1	2	3	4	5	6	7	8
1. Raven	—	.61**	.63**	.69**	.25	.09	.44*	.63**
2. Cattell	.65**	—	.61**	.61**	.52**	.23	.66**	.61**
3. WASI	.59**	.67**	—	.47**	.36	.19	.62**	.47*
4. BETA	.74**	.55**	.67**	—	.44*	.25	.51**	.49**
5. ERVT	.21	.55**	.37*	.32	—	.55**	.70**	.45**
6. LDT	.15	.37*	.10	.29	.37*	—	.41*	.11
7. MRT	.48**	.67**	.50**	.44*	.65**	.37*	—	.66**
8. Paper	.65**	.70**	.61**	.57**	.48**	.07	.70**	—

Note: $n = 29$. Correlations below the diagonal represent pretests and correlations above the diagonal represent posttests. ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.
** indicates p-values at or < .01
* indicates p-values at or < .05

Appendix E

Auditory N-Back Training Group's Intercorrelations between Tests of Gf and Percent Correct on Cognitive Tests

Measure	1	2	3	4	5	6	7	8
1. Raven	—	.60**	.37	.21	.41*	.22	.33	.39
2. Cattell	.66**	—	.53**	.62**	.63**	.37	.67**	.70
3. WASI	.67**	.57**	—	.52**	.29	.08	.74**	.56**
4. BETA	.74**	.61**	.53**	—	.33	.18	.68**	.61**
5. ERVT	.44*	.65**	.62**	.33	—	.48*	.44*	.37
6. LDT	.10	.25	.19	.04	.24	—	.07	.14
7. MRT	.53**	.59**	.71**	.56**	.55**	-.10	—	.83**
8. Paper	.58**	.56**	.79**	.53**	.58**	.06	.65**	—

Note: $n = 25$. Correlations below the diagonal represent pretests and correlations above the diagonal represent posttests. ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.

** indicates p-values at or < .01

* indicates p-values at or < .05

Appendix F

Spatial Matrix Span Training Group's Intercorrelations between Tests of Gf and Percent Correct on Cognitive Tests

Measure	1	2	3	4	5	6	7	8
1. Raven	—	.64**	.63**	.56**	.59**	.08	.68**	.60**
2. Cattell	.72**	—	.64**	.48*	.40*	.10	.41*	.45*
3. WASI	.67**	.75**	—	.57**	.52**	.22	.25	.26
4. BETA	.50*	.49*	.55**	—	.53**	.05	.57**	.47*
5. ERVT	.28	.42*	.36	.18	—	.28	.48*	.47*
6. LDT	.58**	.53**	.47*	.29	.11	—	.00	-.08
7. MRT	.60**	.43*	.47*	.47*	.37	.29	—	.64**
8. Paper	.55**	.59**	.52**	.39*	.35	.23	.52**	—

Note: $n = 28$. Correlations below the diagonal represent pretests and correlations above the diagonal represent posttests. ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.
** indicates p-values at or < .01
* indicates p-values at or < .05

Appendix G

Control Group's Intercorrelations between Tests of Gf and Percent Correct on Cognitive Tests

Measure	1	2	3	4	5	6	7	8
1. Raven	—	.64**	.62**	.57**	.29**	.24**	.46**	.57**
2. Cattell	.81**	—	.63**	.48**	.50**	.32**	.48**	.54**
3. WASI	.49**	.54**	—	.54**	.43**	.22**	.41**	.52**
4. BETA	.35	.36	.43*	—	.21*	.17	.36**	.44**
5. ERVT	.45*	.52**	.40*	.10	—	.25**	.43**	.40**
6. LDT	.60**	.57**	.59**	.42*	.67**	—	.17	.08
7. MRT	.21	.13	.25	-.11	.14	.15	—	.51**
8. Paper	.48*	.47**	.30	.49*	.38	.43	.19	—

Note: $n = 26$. Correlations below the diagonal represent the control group ($n = 26$) and correlations above the diagonal represent training groups pooled together ($n = 110$). ERVT = Extended Range Vocabulary Test; LDT = Lexical Decision Test; MRT = Mental Rotation Test; PFT = Paper Folding Test.

** indicates p-values at or < .01

* indicates p-values at or < .05

Appendix H

Contrasts of Training Groups for Gains on Cattell's Culture Fair Test

Comparisons	Contrast	SE	*t*	*df*	*p*
Dual vs. Control	.80	.76	1.04	131	.30
Visual vs. Control	.84	.76	1.11	131	.27
Auditory vs. Control	.31	.79	.40	131	.69
STM vs. Control	-.31	.76	-.41	131	.69
Dual vs. Visual	-.05	.74	-.06	131	.95
Dual vs. Auditory	.48	.77	.63	131	.53
Dual vs. STM	1.11	.75	1.48	131	.14
Visual vs. Auditory	.53	.77	.69	131	.49
Visual vs. STM	1.15	.74	1.55	131	.12
Auditory vs. STM	.62	.77	.81	131	.42

Appendix I

Repeated Measures Results Testing for Interaction Effects between Differences in Each Gf Pretest and Posttest Scores, Sex, and Training

Test of G*f*	*F*	*df*	*p*	η^2
<u>Raven</u>				
Change	**53.03**	**1, 126**	**.01**	**.30**
Change X Sex	.00	1, 126	.96	.00
Change X Training	**3.68**	**4, 126**	**.01**	**.11**
Change X Sex X Training	.53	4, 126	.72	.02
<u>Cattell</u>				
Change	**22.87**	**1, 126**	**.01**	**.15**
Change X Sex	1.30	1, 126	.26	.01
Change X Training	.91	4, 126	.46	.03
Change X Sex X Training	.50	4, 126	.74	.02
<u>WASI</u>				
Change	**44.40**	**1, 126**	**.01**	**.26**
Change X Sex	.03	1, 126	.86	.00
Change X Training	1.33	4, 126	.26	.04
Change X Sex X Training	.39	4, 126	.82	.01
<u>BETA-III</u>				
Change	**11.43**	**1, 126**	**.01**	**.08**
Change X Sex	.81	1, 126	.37	.01
Change X Training	**4.86**	**4, 126**	**.01**	**.13**
Change X Sex X Training	.31	4, 126	.87	.01