LSU Doctoral Dissertations   Graduate School

2008

The construct validity of the clinical assessment of working memory ability

Benjamin David Hill

Louisiana State University and Agricultural and Mechanical College, bdhill@lsu.edu

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Part of the Psychology Commons

Recommended Citation

THE CONSTRUCT VALIDITY OF THE CLINICAL ASSESSMENT OF
WORKING MEMORY ABILITY

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
In partial fulfillment of the
Requirements for the degree of
Doctor of Philosophy

In

The Department of Psychology

by
Benjamin David Hill
B.S., Coastal Carolina University, 1999
M.A., Wake Forest University, 2003
August 2008

Acknowledgements

I would like to thank the large number of individuals who have aided me with the completion of this project. In particular, I would like to single out Dr. Drew Gouvier, Dr. Emily Elliott, Jill Shelton, Russ Pella, and Matt Calamia for helping me with the design, implementation, and/or collection of data for this dissertation. I would also like to thank my dissertation committee for providing many helpful suggestions that resulted in an improved final work. Finally, I am indebted to my family, friends, and loved ones for their support during these many long years that have led to the achievement of my doctorate. In a way, the completion of this project is a very sad moment for me as it brings to an end my time at Louisiana State University. I will always look back fondly on my doctoral training and will miss the many friends I have made during my stay at LSU. Geaux Tigers.

Table of Contents

iv

List of Tables

v

List of Figures

Abstract

Working memory is the cognitive ability to hold a discrete amount of information in mind in an accessible state for utilization in mental tasks. This cognitive ability is impaired in many clinical populations. There have been a number of theoretical shifts in the way that working memory is conceptualized and assessed in the experimental literature. The purpose of this study was to determine whether the Working Memory Indices (and their component subtests) from the WAIS-III and WMS-III accurately assess the working memory construct as it is currently defined in the experimental cognitive literature. Results generally supported the construct validity of the Arithmetic, Digit Span, Letter-Number Sequencing, and Spatial Span subtests as measures of working memory ability. Confirmatory factor analyses revealed that the Wechsler subtests do appear to be measuring the same working memory construct as the experimental cognitive measures used in this study (listening span task, operation span task, and modified lag task), though Arithmetic and Spatial Span had low factor loadings. Additionally, multiple regression analyses revealed that the best predictor models of Wechsler subtests for assessing working memory were composed of the Digit Span, Letter-Number Sequencing, Matrix Reasoning, Vocabulary, Symbol Search, Logical Memory I, and Spatial Span subtests.

Introduction

The field of clinical neuropsychology is based on reliably and validly measuring cognitive constructs related to clinical conditions. This allows clinicians to accurately assess cognitive strengths and weaknesses for the purpose of diagnosis and treatment planning. Accurate and valid assessment allows for the application of experimental research findings to better understand the condition of the individual patient. Particularly in the clinical neurosciences, the process of applying broad scientific research to a single individual patient involves perusing findings from many distinct but related scientific disciplines. Applying extant research is particularly difficult, however, when the same construct is labeled differently across fields or fields share the same name for different constructs. For example, inhibition is a popular term utilized in neuroscience, but cognitive inhibition is very different conceptually from neuronal inhibition. When this occurs, apparently contradictory findings may arise and the enterprise of research is hampered as studies are built upon inconsistent, or worse yet, irrelevant data. From a translational research standpoint, such conflicting definitions and operationalizations may have a deleterious effect on patient care if neuropsychologists base their diagnoses and recommendations on research findings that do not truly relate to the patient's condition.

A particularly important construct in this regard is the concept of working memory. Working memory refers to the cognitive ability to hold a discrete amount of information in mind in an accessible state for utilization in mental tasks. Working memory, and related executive functions, are impaired in a wide variety of neuropsychiatric conditions, including dementia (Collette, Van der Linden, & Salmon, 1999), attention-deficit hyperactivity disorder (ADHD; Pasini, Paloscia, Alessandrelli, Porfirio, & Curatolo, 2007), and schizophrenia (Fleming, Goldberg, Gold, & Weinberger, 1995; Goldman-Rakic, 1994). As efficacy for therapeutic interventions is typically demonstrated by measured improvements in executive functions, it is paramount that clinicians

1

validly measure the constructs they are reporting to measure. The objective of this dissertation is to examine whether clinical measures of working memory are validly assessing the same working memory construct that is commonly reported in the experimental cognitive and neuroscience literature. This issue is important as clinical neuropsychology uses this literature to inform its clinical interpretation of cognitive functioning, such as in the area of dementia diagnosis. A brief review of executive functions and working memory will now be provided. Next, a description of clinical measures used to assess working memory will be presented along with a synopsis of methodology used in the experimental cognitive literature to assess working memory ability. Lastly, the specific hypotheses of this study will be offered followed by the methodology, results, and discussion of research findings.

Executive Functions

Executive functions are typically defined in clinical usage as higher level cognitive abilities that modulate initiation of behavior, self-regulation, planning, and organization (Lezak, Howieson, Loring, & Hannay, 2004). These functions prioritize, regulate, and integrate other cognitive functions and are commonly associated with the frontal cortex. Due to this association with executive functions, the frontal lobes are reported to be the neural substrate for theory of mind (Stuss, Gallup, Jr., & Alexander, 2001), demonstrating how important executive functions are to the experience of being human. Deficits in these areas commonly arise from traumatic brain injury, frontal lobe lesions, and any number of neuropathological conditions that result in dementia or cognitive dysfunction. Executive dysfunction has also been reported to commonly occur in the course of common chronic diseases, such as diabetes mellitus and hypertension (Schillerstrom, Horton, & Royall, 2005).

Executive functions have wide ranging regulatory effects on other cognitive abilities that have neural underpinnings far removed from the frontal lobes. This is due to the organizational and

inhibitory properties of the frontal lobes. These properties allow for integration of neural systems via subcortical connections and neural feedback loops between various brain regions. Executive functions can be viewed as mechanisms that allow temporal discrimination of stimuli. They permit the individual to form an accurate perception of reality by processing information in real time as it is presented, comparing it to past events, and using this comparison to determine the probability of future events which can then be planned for accordingly (Constantinidis, Williams, & Goldman-Rakic, 2002). As such, executive functions represent an essential core cognitive system that, when impaired, can deleteriously impact a number of clinical factors such as functional outcome (Boyle et al., 2003), medication compliance (Hinkin et al., 2002), and capacity to give informed consent (Marson, Chatterjee, Ingram, & Harrell, 1996).

Working Memory

Working memory is perhaps the most important executive function and, therefore, one of the most studied. Working memory refers to the cognitive system that stores information in an accessible state for utilization in complex mental tasks. This system also allows for updating and manipulation of relevant information (Baddeley, 2001). It is composed of both passive storage and dynamic control processes in order to hold information in an active form. Working memory plays a key role in cognition in both *Homo sapiens* and infrahuman species (Brozoski, Brown, Rosvold, & Goldman, 1979; Müller, von Cramon, & Pollman, 1998). Humans utilize it in tasks such as reading to integrate early parts of sentences with ending segments and in arithmetic to hold the results of earlier steps in mind while completing multistep problems.

While the concept of working memory was popularized by Alan Baddeley (Baddeley & Hitch, 1974; Baddeley, 1986), others were using the term more than a decade prior to him (Miller, Galanter, & Pribram, 1960; Newell & Simon 1972). The current concept of working memory grew out of early memory models, with the Atkinson and Shiffrin (1968) model being the exemplar, that

3

typically conceptualized memory as being formed by the following three modules: 1) an initial

sensory store, 2) an intermediate short-term memory store (usually considered to hold information

up to 30 seconds; Peterson & Peterson, 1959), and 3) a final long-term memory store (see Figure 1).

Baddeley and Hitch (1974) pointed out in their seminal paper that the modal conceptualization of a

unitary short-term memory system was not sufficiently complex to explain research findings. They

instead proposed a three component model (see Figure 2) that relied on a central executive to

manipulate data and two systems that served the central executive by being able to either passively

hold or actively rehearse information. These two "slave" systems were termed the phonological

loop and visuospatial sketchpad for their abilities to hold auditory information and visuospatial

information respectively. These two systems also communicate bidirectionally with the long-term

memory system. The main difference between the earlier conceptualization of short-term memory

and the newer conceptualization of working memory is that the former was exclusively a passive

storage buffer while the latter is primarily characterized as a very active processing system that

utilizes some passive storage components.

Baddeley's model has stood as the gold standard in working memory research since its

inception, largely due to its ability to explain a majority of research data and its relative simplicity.

However, it has been faulted for not fully explaining factors such as the effect of unattended

information or where sensory information in modalities other than audition or vision is stored.

Cowan (1999), in particular, has attempted to address this last flaw by pushing the model in a more

generic direction that downplays the need to explicitly label the sensory domains of the buffers.

While Baddeley conceptualized working memory as a component of short-term memory that

is systematically distinct from long-term memory, Ericsson and Kintsch (1995) have suggested that

this division is illusory. They instead proposed that memory is more of a unitary phenomenon and

that data processing actually takes place in one system. In this system, currently relevant data is
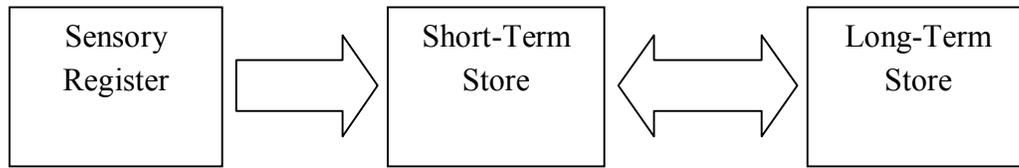
Figure 1

Atkinson and Shiffrin (1968) Model of Memory System
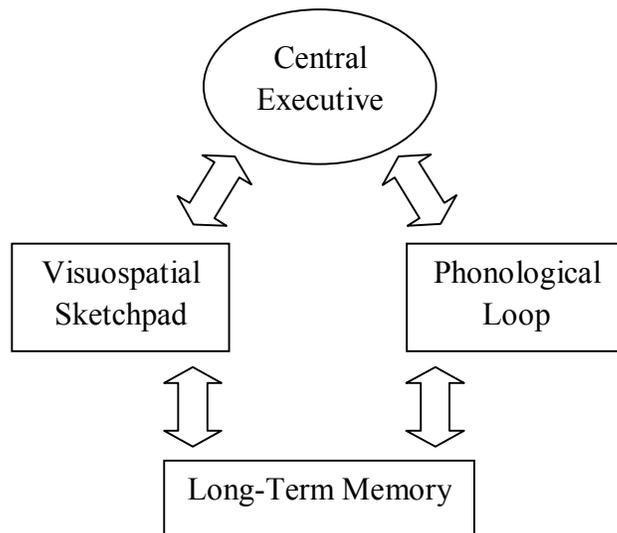
Figure 2

Baddeley and Hitch (1974) Model of Working Memory

marked for processing to distinguish it from other encoded information that is being held long-term.

Therefore, this can be predominantly characterized as a long-term memory system where

information can also be actively transformed and manipulated. This distinction is important from

the standpoint of exactly delineating cognitive architecture, but practically speaking, it is merely a

semantic argument when the focus is actually placed on working memory assessment. The

functional outcome should not differ regardless of whether working memory ability is housed in a

separate short-term memory system or in a monolithic long-term memory system.

In a nod to the evidence supporting some involvement of long-term memory in the working memory system, Baddeley (2000a) has amended his original three component model by adding an episodic buffer. This storage system communicates bidirectionally with the working memory subsystems and the long-term memory system (see Figure 3). It is called the episodic buffer since part of long-term memory is considered episodic memory. The episodic buffer appears to integrate spatial and phonological information (Baddeley & Andrade, 2000), which were previously held discretely, and links this new information to semantic categories nested in long-term memory as a form of information chunking. Consequently, it can be conceptualized as a short-term version of episodic memory that acts as an interface between working memory and long-term memory (for a review see Baddeley, 2001).
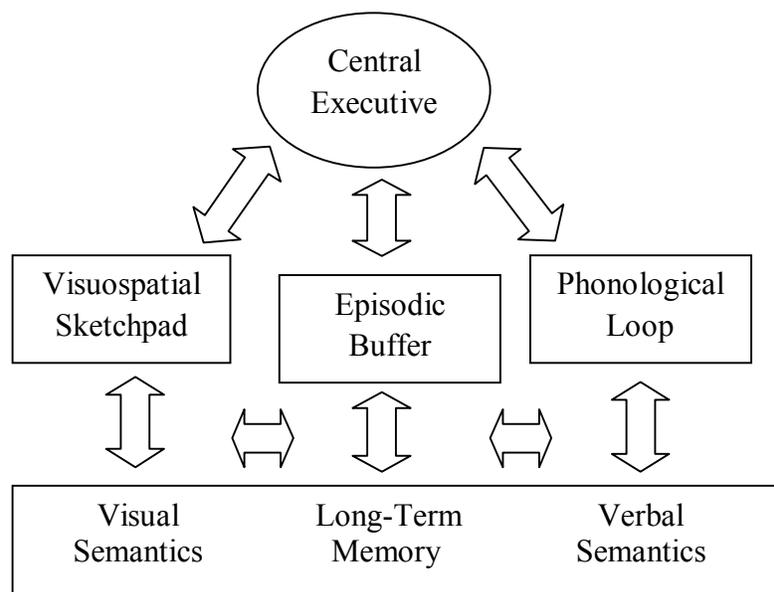
Figure 3

Baddeley's Current (2000a) Model of Working Memory

Working memory for the purpose of clinical assessment is sometimes defined as the number of items that can be recalled in an immediate-memory task such as the Digit Span subtest of the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III; Wechsler, 1997a). However, this is

probably more of a broad test of attention and short-term memory system capacity as there is no actual manipulation of data involved as is typically seen in experimental working memory tasks. A narrower definition of working memory capacity proposed by Cowan (1995, 2005) is known as the focus of attention (FoA). The FoA is the amount of information that can be maintained on-line without rehearsal or other strategic processes. This differs from the amount of information that can be held in mind in a specific component of the working memory system, also known as working memory span, as individuals typically employ strategic rehearsal processes when completing working memory span tasks like those used in clinical assessment. FoA can be defined as the information that can be maintained entirely within the individual's immediate awareness. Cowan's theory of working memory that incorporates the FoA, called the embedded process model (Cowan, 2001), builds on earlier work by Broadbent and extends the long-term working memory model (Ericsson & Kintsch, 1995). What truly distinguishes the embedded process model from previous conceptualizations is the assertion that the attentional mechanism itself has a storage capacity.

Research into the FoA is difficult and has been hampered by the fact that the system components are neither easily dissociable nor readily observed. As such, this idea has gained little traction in the neuropsychology research community because the literature has not demonstrated its clinical utility. However, Cowan's idea of FoA is conceptually related to Engle's controlled attention component of working memory (Engle & Kane, 2004), which has resulted in a great deal of research in the cognitive realm, and Unsworth and Engle's (2006, 2007) more recent dual-component model of working memory. In Engle's original framework (Engle & Oransky, 1999), working memory was seen as being composed of two modules: immediate memory and controlled attention (also called executive attention). These modules fulfilled maintenance and retrieval functions, respectively. He also argued that working memory capacity was essentially the same as executive attention abilities and was, therefore, a proxy for general fluid intelligence (Engle, Kane,

& Tuholski, 1999), though Kyllonen and Christal (1990) should be credited with originating this idea. Kane and Engle (2002) defined executive attention as "a capability whereby memory representations are maintained in a highly active state in the presence of interference (p. 638)." Accordingly, a primary responsibility of executive attention was to increase processing efficiency by blocking intrusive stimuli from utilizing mental resources.

However, Unsworth and Engle (2006, 2007) have recently proposed a heavily modified version of Engle's original model. In this new conceptualization, working memory is proposed to be an interaction between two modules: 1) a primary memory storage system that maintains up to four chunks of information for a limited amount of time (though it can focus resources on fewer chunks) and 2) a cue-dependent search system that retrieves data from longer-term memory which they term secondary memory. This primary memory system is conceptually analogous to Cowan's FoA (1995, 2005) and Baddeley's episodic buffer (Baddeley, 2000). Unsworth and Engle postulate that all information is initially stored in primary memory but, whenever this buffer reaches capacity, information is automatically moved to secondary memory. Individual differences in working memory ability, then, are directly related to the efficiency with which the person can search the secondary memory system via cues, the most prominent of which are temporal-contextual cues. Unsworth and Engle (2005, 2007) still speculate that attentional control mechanisms play a role in this system by increasing search efficiency in secondary memory, largely by decreasing interference effects. However, they now believe that the connection between working memory and fluid intelligence is more related to the capability to efficiently search secondary memory, an ability that is important to performance on both constructs (Unsworth & Engle, 2006).

If they are correct and a significant component of working memory is shared with fluid intelligence, a function that is believed to be housed in the frontal lobes, then working memory ability would be a sensitive index of frontal functioning. This has obvious clinical applications.

However, other research appears to demonstrate that working memory and general fluid intelligence may be dissociable on specific tasks of executive functioning (Zook, Davalos, DeLosh, & Davis, 2004). Doubt has also been raised as to whether the neural substrate for fluid intelligence is found in prefrontal cortex (Tranel, Manzel, & Anderson, in press). Clearly, more research is required in this area to elucidate the exact cognitive mechanisms shared between working memory and fluid intelligence as well as what performance related to these constructs reveals about cortical functioning.

Clinical Assessment

Currently, the Wechsler Adult Intelligence Scale, 3rd edition (WAIS-III) and the Wechsler Memory Scale, 3rd edition (WMS-III) Working Memory Indices are the most commonly used clinical instruments to assess working memory ability. The WAIS-III (Wechsler, 1997a) is one of the most commonly utilized assessment instruments in the armamentarium of clinical neuropsychologists (Camara et al., 2000; Rabin, Barr, & Burton 2005). Among intellectual functioning assessment instruments, it is considered the gold standard (Ivnik et al., 1992) and has been utilized in some form for almost 70 years (Wechsler, 1939). For a single instrument, it provides an exceedingly broad view of an individual's cognitive functioning. The test's normative data covers an age range of 16-89 years old (Psychological Corporation, 1997) and scores are compared to the average of scores attained by the assessed individual's same age group. It contains 14 subtests, 3 of which were not found on previous versions of the WAIS. These include: Matrix Reasoning, Letter-Number Sequencing, and Symbol Search.

Typically, the main use of the WAIS-III is to provide the clinician with a Full Scale Intelligence Quotient (IQ) Index score. This represents a combination of the Performance IQ (PIQ) Index score and the Verbal IQ (VIQ) Index score, which are also provided. Further, specific subtests of the WAIS-III are used to derive a Processing Speed Index (PSI) score and a Working

Memory Index (WMI) score. These two measurement domains are also new additions to the WAIS-III that are not found on earlier versions of the WAIS (Strauss, Sherman, & Spring, 2006). Due to the influence of processing speed and working memory on the assessment of intellectual functioning, two other factor derived indexes of performance and verbal functioning are provided that remove the contribution of these abilities. These are the Perceptual Organization Index (POI) and the Verbal Comprehension Index (VCI), which reflect "pure" perceptual and verbal skills independent of the contributions of working memory and processing speed (Tulsky, Saklofske, & Zhu, 2003). The test developer has already announced that PIQ and VIQ will be phased out on the upcoming fourth edition of the WAIS. The VCI and a renamed POI (to be called the Perceptual Reasoning Index) will be retained along with the PSI and WMI (WAIS-IV, 2008).

The WAIS-III WMI is derived from the performance on the Digit Span, Arithmetic, and Letter-Number Sequencing subtests. These particular subtests were chosen to comprise the WMI based on both theoretical considerations and factor analytic studies. The first of these, the Digit Span subtest, has been used in various forms in the experimental literature since the late 1800's (Galton, 1887; Jacobs, 1887; Bolton, 1892). In the Digit Span subtest, there is both a digits forward and a digits backward condition. In the digits forward part of the subtest, the individual is presented with a string of numbers, presented at a rate of one number per second, and asked to repeat back to the examiner the number string in order immediately after stimuli presentation. In the digits backward condition, the individual is instructed to repeat back the presented string of numbers in reverse order. A maximum of nine digits is presented in the forward condition and eight in the digits backward condition. The Digit Span score combines the total number of digit strings correctly repeated in both the forward and back conditions. The practice of combining these forward and backward span scores originated with David Wechsler, the creator of the WAIS and WMS, because

he did not want two separate memory span scores to overly influence the final outcome of his tests (Wechsler, 1944).

Unfortunately, for a task whose primary purpose is the assessment of working memory, this subtest seems to have a heavy immediate memory component inherent in its design. For example, the Digit Span forward condition, which contributes half of the Digit Span total score, is essentially an immediate memory task with no manipulation of data required. It is a classic example of a short-term memory span task (Baddeley, 2000b). This was also recognized by Golden in *Clinical Interpretation of Objective Psychological Tests* (1990), where he introduces the Digit Span subtest as "a measure of immediate auditory memory (p. 23)." Reynolds (1997) noted that the forward and backward components of digit span tasks loaded on different factors and recommended that they not be interpreted together as a total score. It should be noted, however, that Reynolds' analysis only included children and adolescents and was not performed on WAIS-III Digit Span data. Additionally, differences in forward span compared to backward span have been reported to have clinical significance (Golden, Osmon, Moses, & Berg, 1981) which is obscured by obtaining a combined Digit Span score. Also, for a subtest that is supposed to measure a raw cognitive ability that should be relatively culture-free, Digit Span has been reported to be significantly affected by variables such as education and culture (Ostrosky-Solis & Lozano, 2006). Overall, this subtest (and Digit Span forward in particular) appears to have low face validity as a measure of working memory due to its lack of task complexity. However, recent research has found that immediate memory and working memory may not be discrete constructs as both appear to share a number of cognitive processes (Unsworth & Engle, 2006).

In the Arithmetic subtest, the individual is verbally presented with arithmetic word problems. It is also one of the subtests that was used in the original Wechsler-Bellevue Intelligence Scale (Wechsler, 1939) from which the WAIS evolved. This subtest has a high intercorrelation with other

11

subtests on the WAIS-III and is, therefore, believed to be a measure of general intelligence (Tulsky et al., 2003). Subjects have to respond with the correct answer in a set amount of time (between 15-120 seconds depending on the difficulty of the problem) and are not allowed to write down the problem in any form (such as using their finger to trace the problem on a table). The problems increase sequentially in difficulty. Also, each problem can only be repeated to the individual once for a maximum of two presentations. The testing protocol is constructed so that the problem must be solved using only purely cognitive abilities, including holding the details of the problem "on-line" while the problem is solved. While this task does appear to manipulate working memory load as part of the assessment procedure and is not an assessment of mathematic ability per se, it is likely influenced by mathematical abilities to some degree making it a less than ideal paradigm for assessing working memory performance. Additionally, anxiety related to mathematics has been demonstrated to negatively affect working memory performance (Ashcraft & Kirk, 2001) and would be a possible confound for this subtest as well.

The Letter-Number Sequencing subtest may be, perhaps, the best representative of a working memory measure among the Wechsler subtests. It is also one of the newest WAIS subtests and was originally developed by James Gold to study working memory deficits in schizophrenia (Gold, Carpenter, Randolph, Goldberg, & Weinberger, 1997). In this task, the individual is presented with a series of numbers and letters in random order. They are then instructed to repeat back the numbers in order first followed by the letters in alphabetical order. For example, if the subject were presented with the string "h-a-8-m-3" the correct response would be "3-8-a-h-m." Lists of two numbers or letters are presented first and list length is increased up to eight numbers or letters. Lists are presented until the individual responds incorrectly to the same list length twice. Unlike the Digit Span subtest, this subtest has no components that appear to be merely an immediate memory task

12

and, unlike the Arithmetic subtest, there is no minimum academic skill prerequisite other than knowing the numbers 1 through 9 and having a functional knowledge of the alphabet.

The second of the current Wechsler scales, the WMS-III (Weschler, 1997b), is broadly considered to be the first line choice in comprehensive clinical assessment of memory abilities (Rabin et al., 2005). It was co-normed with the WAIS-III allowing for accurate comparisons of standardized scores between the two instruments. Additionally, like the WAIS-III, it also provides a WMI score. However, despite the fact that both Wechsler scales produce a WMI score, the subtests utilized to obtain the index scores differ between the WAIS-III and the WMS-III. While the WAIS-III incorporates the data from three subtests into its WMI, the WMS-III uses data from only two subtests. Like the WAIS-III, the WMS-III includes the Letter-Number Sequencing subtest. However, the other subtest utilized is a task that is not found in the WAIS-III, the Spatial Span subtest.

Similar to the Digit Span subtest, Spatial Span has both a forward and backward component that are combined for a total score. Also, like the Digit Span subtest, the presentation protocol involves presenting a string of stimuli to be repeated in the order of presentation by the individual in either a forward order or a backward order. However, a major difference between Digit Span and Spatial Span is the sensory modality of stimulus presentation. While Digit Span is an auditory task (numbers are spoken by the examiner and repeated orally by the test subject), Spatial Span is a purely visuospatial task. In Spatial Span, the individual is presented with a board on which a series of identical 3-dimensional squares are attached in a way that makes the positions of the stimuli difficult to label verbally. The presentation protocol involves the examiner pointing in a preordained series to specific squares located in exact spatial locations on the board. The participant is asked to point to the squares presented by the examiner in the same order for the forward component and in reverse order for the separate backward component of the subtest. Again, like Digit Span, this

13

makes the forward component of the task essentially a nonverbal short-term memory span task (Baddeley, 2000b).

Spatial Span is basically a standardized and normed version of the Corsi Block Tapping Task (Corsi, 1972). However, an even earlier simplified form of the task, the Knox Cube, was used to screen immigrants at Ellis Island (Knox, 1914). Like Digit Span, Spatial Span stimuli are presented in ascending order of difficulty, starting with 2 blocks being touched in sequence and increasing up to a maximum of 9 blocks being presented in sequence. Also like Digit Span, testing is discontinued when incorrect responses are given for two consecutive series of the same length. The practice of combining both the forward and backward components scores is continued here as well, and, similar to Reynolds' (1997) recommendations against using a total score for Digit Span, experts in neuropsychological assessment have warned against doing this (Kaplan, Fein, Kramer, Delis, & Morris, 1999).

In essence, Spatial Span appears to be the visual counterpart to the auditory Digit Span task. However, research has not borne this out. Wilde, Strauss, and Tulsky (2004) reported that Spatial Span differed from Digit Span in patterns of performance obtained between forward and backward conditions and in age effects on the two tasks. Also, there is considerable evidence that visual memory is not the same as spatial memory as they demonstrate differential interference effects (Baddeley, 1994; Della Sala, Gray, Baddeley, Allamano, & Wilson, 1999). Therefore, it is probably safer to conceptualize Spatial Span as a spatial working memory task and not as a visual working memory task. However, it should be noted that the WMS-III technical manual (Psychological Corporation, 1997) does not make a visual versus spatial working memory distinction and refers to Spatial Span as a visual working memory task.

The most obvious potential problem in the way the WMS-III WMI is calculated is that both auditory and spatial tasks are combined to form a single index of working memory functioning. As

14

was discussed previously, the dominant model of working memory in the field (Baddeley, 1986) postulates two separate systems for auditory versus visuospatial information. Also, neuroimaging research has shown that visuospatial and verbal working memory systems rely on different neural substrates (Smith, Jonides, & Koeppe, 1996). It would seem conceptually inconsistent to combine measures of two separate systems into a single score as it would not provide useful information regarding whether one system is impaired or one is an area of strength.

For example, a WMS-III WMI standard score in the normal range could mean that both the auditory working memory system and spatial working memory system are functioning normally. However, it could just as well signify that the auditory system is below average but is balanced out by an above average spatial working memory system. The technical manual for the Wechsler scales specifically addresses this combining of scores from different sensory modalities, stating that the tests are "deemphasizing the distinction between verbal and visual material (p. 6)" and cites Carlson, Khoo, Yaure, and Schneider (1990) as evidence of a "single-workspace model" of working memory that integrates both auditory and visual abilities. However, this is not necessarily what the Wechlser scales are actually doing. Instead, the WAIS-III is in effect ignoring any visual component of working memory while the WMS-III is essentially pronouncing auditory and visuospatial working memory systems to be equal. Additionally, the Carlson et al. (1990) article does not explicitly support a single-workspace model and instead states that, while their obtained results are ambiguous, they do not believe a single-workspace model of working memory is correct.

Also, there is considerable evidence that working memory is not a "single-workspace model" but is instead separated into separate verbal and spatial working memory components (Farmer, Berman, & Fletcher, 1986; Shah & Miyake, 1996). Additionally, children with ADHD have demonstrated deficits in verbal working memory without accompanying spatial working memory impairment (Jonsdottir, Bouma, Sergeant, & Scherder, 2005) while individuals with Alzheimer's

15

disease have been reported to have a specific impairment of visual working memory that does not extend to verbal working memory (Spaan, Raaijmakers, & Jonker, 2003). Even the Wechsler scales themselves have been used to show a dissociation as the backward span components of the Digit Span and Spatial Span subtests are not differentially affected compared to the forward span components by factors that should affect working memory performance (Wilde et al., 2004). Therefore, there appears to be evidence for a dissociation of working memory systems and, for the Wechsler scales to function as effective clinical measures, they should not obscure splits in functional abilities that may be diagnostically informative. However, it must be pointed out that Tulsky and Price (2003) did find using structural equation modeling (SEM) that specifying Digit Span, Arithmetic, Letter-Number Sequencing, and Spatial Span as a single factor in a larger six factor model provided a good fit for the WAIS-III/WMS-III co-norming data, though model fit was increased by allowing Arithmetic and Spatial Span to load on multiple factors. Consequently, these subtests, especially Digit Span and Letter-Number Sequencing, do appear to be measuring a unified construct. Whether that construct is truly working memory and, if so, how well they are measuring it remains to be seen.

In sum, the WMIs derived from both the WAIS-III and WMS-III may be poor approximations of actual working memory ability as it is conceptualized by current theoretical models. This is mainly due to the fact that there appears to be a considerable amount of variance associated with immediate memory or non-working memory factors in these measures. This situation could potentially be significantly rectified by removing the forward span conditions from the calculation of the WMIs, as these would seem to be inclusions of passive immediate memory and not working memory per se. However, Unsworth and Engle's (2006, 2007) dual-component conceptualization of working memory functioning would not support this line of reasoning.

Recently, Lamar and colleagues (2007) reported evidence that the working memory deficits in dementia are not due to impaired immediate recall, but an impaired ability to disengage and temporally sequence data. White matter degeneration was postulated to be the causal mechanism. Therefore, the Wechsler measures may miss some impaired individuals in this population if immediate memory contributes significantly to the derived WMI. This is not to say that the Wechsler WMIs do not have clinical utility, as they obviously do (DeLuca, Chelune, Tulsky, Lengenfelder, & Chiaravalloti, 2004), but they may not be optimally measuring working memory as they are currently used.

Consistently presenting stimuli solely in the auditory domain might also increase the ecological validity of these measurement tasks as phonological working memory has a much greater impact in daily life than visuospatial working memory. Also, this modality is easier to reliably assess (Psychological Corporation, 1997). Theoretically, this is also justified as there is evidence that various verbal memory tasks share cognitive resources (Fedorenko, Gibson, & Rohde, 2007). However, even with these changes, the Wechsler WMIs would still lack face validity, as viewed through the lens of established working memory theory, compared to cognitive experimental assessment techniques. The most glaring difference is that there is no ongoing secondary task to tax the attentional component of working memory that is commonly found in working memory tasks utilized by cognitive psychologists. These tasks will now be discussed in more detail.

Experimental Assessment

In experimental psychology and cognitive neuroscience, working memory is typically assessed via tasks that combine both storage and processing tasks in order to fully measure working memory. One of the first tasks developed for this purpose was the sentence span task (Daneman & Carpenter, 1980). In this task, the participant must indicate that they comprehend a presented sentence and also retain the last word of the sentence. Two to six sentences are presented in this

17

manner. Sentence comprehension represents the secondary ongoing attentional component of the task while recall of the last word in each sentence represents the storage function of the task. The sentence span is defined as the number of sentences that are both correctly comprehended and their last word correctly recalled in order. Since sentences can be presented either in print or orally, the type of sentence span task has been further categorized by the mode of administration. If the sentence is read by the participant, the task is called the reading span task and if the sentence is presented orally to the participant, the task is called the listening span task (l-span).

Turner and Engle (1989) built on this paradigm, developing the operation span task (o-span). In the o-span task, instead of a sentence comprehension component like in the sentence span task, an arithmetic problem is presented followed by a word to be remembered. Again, like the sentence span task, two to six of these arithmetic problems and accompanying words are presented in series. The number of words recalled in the correct order represents the working memory span. Also, unlike the clinical Wechsler subtests, there is a secondary ongoing attentional task in the solving of arithmetic problems.

These tasks are particularly good for studying modulation of working memory load because a necessary criterion level can be set for the processing component to ensure that the participant is fully engaged in the task. For example, data from participants who correctly answer less than 85% of the arithmetic problems can be excluded from analysis (and has been in past literature; Unsworth & Engle, 2005, 2006) as these subjects appear to have concentrated more on the storage component of the task (the words to be remembered) and less on the processing component (correctly answering the arithmetic problems). This ensures that this task is tapping working memory and not simply measuring immediate memory. This criterion level approach can also be considered a way of identifying those individuals giving suboptimal effort. Performance on these tasks is typically both reliable and highly correlated across tasks (Klein & Fiss, 1999; Daneman & Merickle, 1996;

18

Conway et al., 2005). Additionally, varying arithmetic difficulty of the o-span task to equalize performance across participants has been shown to not affect the relationship between o-span and reading span performance indicating that a common process is being assessed across tasks (Conway & Engle, 1996).

The *n*-back task, also known as the lag task, has also been commonly employed in cognitive psychology as a measure of working memory ability, to the point that it has been referred to as the gold standard working memory assessment technique in cognitive neuroscience (Kane & Engle, 2002). Because of this and the fact that its protocol is readily adaptable to the requirements of the magnetic imaging environment, it is the working memory assessment methodology of choice in the neuroimaging literature (Owen, McMillan, Laird, & Bullmore, 2005). In the *n*-back task, individuals are presented with stimulus sequences and required to recall a stimulus presented a specified number back in the sequence (*n* represents how far back in the sequence the person is to go). Stimuli can be numbers, letters, or even pictures. For example, at an *n* of one, the correct response would be whatever was presented immediately prior to the last stimulus. At an *n* of two, the correct response is what was presented two prior to the last stimulus. In experimental situations, *n* typically varies between one and four back, though in the neuroimaging literature the *n* requirement may be as low as two back. This task has high face validity as a measure of working memory due to the attentional requirement to continuously update the stimuli being held on-line.

Construct validity assessments of the *n*-back task have generally been supportive. The majority appear to show convergent validity with constructs such as fluid intelligence, attentional performance, and inhibitory abilities (Aronen, Vuontela, Steenari, Salmi, & Carlson, 2005; Hockey & Geffen, 2004; Kwong See, & Ryan, 1995). However, recently, Kane and colleagues reported negative findings regarding construct validity (Kane, Conway, Miura, & Coflesh, 2007). Also, some argue that the processing component of the task is not continuous, as in most other experimental

19

working memory tasks, and is utilized only when retrieval of the correct stimulus is required (i.e. when the command is given to the state the stimulus that was *n*-back; Conway *et al.,* 2005). This lack of a continuous processing component may allow for more rehearsal strategies than other experimental working memory tasks. However, this measure is widely utilized and generally accepted in the field of cognitive neuroscience as an index of working memory. Also, the majority of working memory research in cognitive neuroscience has been conducted using this task. Additionally, while the processing component may not be uniformly consistent across the task, it is present and necessary for successful completion of the task. Therefore, the *n*-back task has much more face validity as a working memory measure than the previously presented clinical indexes and deserves to be considered as a criterion measure.

After reviewing the differences between the methods used to measure working memory capacity in clinical versus experimental settings, it becomes clear that there is significant variability between these two paradigms. Cognitive psychologists defined the term working memory and have studied the reliability and validity of their tasks in depth to ensure that they are accurately measuring the correct construct (Conway et al., 2005). Thus, their experimental tasks of working memory should be held as the criterion standard to which the clinical measures are compared.

Such paradigm shifts in assessment are nothing new to the field of clinical neuropsychology. Something very similar happened in the 1970's when neuropsychologists changed their methods of assessment to explicitly include recognition memory, intrusion errors, and susceptibility to interference in reaction to the findings of cognitive psychologists studying amnesia (Butters, 1992). It has also been noted in the previous review that working memory is a critical cognitive faculty whose impairment is related to cognitive dysfunction in an array of cognitive disorders. Therefore, the accurate measurement of this construct is of critical importance to clinical neuropsychology as the field defines itself in terms of researching and assessing neurocognitive abilities.

Purpose of Study

This study has two main purposes. First, this study seeks to determine whether the WMIs produced by the Wechsler scales, the most commonly utilized instruments by clinical neuropsychologists, are validly measuring working memory as it is currently defined in the cognitive experimental literature. Second, this study seeks to improve the clinical measurement of working memory using the existing Wechsler scales. To this end, the construct validity of the WMIs from the WAIS-III and WMS-III will be evaluated with the experimental measures employed in cognitive psychology utilized as the comparison criterion and regression equations will be developed from the Wechsler scales that predict performance on the experimental working memory measures.

Research Questions and Hypotheses

Question 1

Do the WMIs from the WAIS-III and WMS-III validly measure the construct of working memory?

Hypothesis 1: It is predicted that the subtests that compose the WMIs will not display convergent validity with the experimental working memory measures. It is specifically hypothesized that the experimental cognitive measures of working memory will all highly positively correlate together but none of the Wechsler scale indexes or subtests will correlate as highly with the experimental measures as they do amongst themselves.

Hypothesis 2: It is predicted that the subtests that compose the WMIs will display discriminant validity with the cognitive experimental measures in that the clinical measures will be more related to a measure of immediate memory than the experimental cognitive measures. Further, it is predicted that the forward

21

components of the Digit Span and Spatial Span subtests will be more related to the immediate memory measure than the backward components of these subtests. It is specifically hypothesized that the cognitive experimental measures of working memory will either demonstrate a small positive relationship with performance on the Immediate Memory Index (IMI) while the subtests and WMIs from the Wechsler scales will demonstrate a medium to large positive relationship with WMS-III IMI performance. This is hypothesized to occur due to the inclusion of a separate immediate memory component in the Wechsler subtests that is not found to the same extent in the cognitive experimental working memory measures.

Hypothesis 3: It is predicted that a confirmatory factor analysis (CFA) will not support a single factor model for the clinical and experimental working memory measures. It is further predicted that a two factor model with one factor for the WMIs subtests and a separate factor for the experimental cognitive working memory tasks will provide a better fit for the data.

Hypothesis 4: It is predicted that an exploratory factor analysis (EFA) conducted with the forward and backward components of the Digit Span and Spatial Span subtests will result in a two factor model where the forward components load on a separate factor from the backward components.

Question 2

Can an improved WMI be derived from the subtests that comprise the WAIS-III and WMS-III?

Hypothesis 5: It is predicted that a multiple regression equation that includes additional subtests from the WAIS-III and WMS-III beyond those currently used to

form the WMIs will account for a greater percent of variance in the criterion

experimental working memory measures than the current clinical WMIs.

Further, it is predicted that subtests from the WAIS-III will account for the majority

of the variance in the derived regression equation as the WAIS-III subtests assess a

wide variety of attentional and concentration functions while the majority of WMS-

III subtests almost exclusively assess aspects of immediate and delayed memory.

Methods

<u>Participants</u>

A database assembled as part of a joint effort between the laboratories of Dr. Wm. Drew Gouvier and Dr. Emily Elliott was utilized in this study. These data were collected between May 2005 and December 2006 at a state university in the southern United States. The following factors excluded an individual from inclusion in this study: a visual and/or hearing impairment at the time of testing, a relevant psychiatric diagnosis that would result in cognitive impairment, identified English as their secondary language, or failing to attend all experimental sessions. A total of 188 participants met inclusion criteria and agreed to participate in this study. Participants were compensated for their donated time by receiving extra credit in a course offered in the Department of Psychology. No participants were excluded based on gender, race, ethnicity, or academic status.

Similar to other studies of this genre, the data from 8 participants were not utilized in this study since they performed at a less than 80% correct responding rate (more than 15 errors) on the mathematics section of the ao-span task. This performance indicated a less than acceptable level of engagement in the testing procedure and reflects that they may have focused on rehearsal strategies during the task. This left a sample of 180 participants (mean age in years 20.6, SD 3.6; mean WAIS-III full scale IQ 109.9, SD 11.4; 72.8% female, 27.2% male). While there is a significantly higher proportion of females than males in this sample (χ^2 =40.62, p<.001), the literature states that there is no significant gender difference in performance on any of the measures used in this study (Psychological Corporation, 1997; Robert & Savoie, 2006). Additionally, while 7 participants had missing data (either a subtest that had not been administered by the research assistant or a corrupted data file), the missing values for these individuals were imputed. The process used to impute these missing values will be explained further in the Results section.

The sample size in this study (n=180) was sufficient for all correlational and factor analyses (Tabachnick & Fidell, 2001) that were conducted. However, a power analysis was performed to determine the number of participants needed for power = .80 and α = .05 in the utilized multiple regression analyses. A small to moderate effect size was proposed based on both the literature and what would be relevant clinically. G*Power 3.0.3 software (Faul, Erdfelder, Lang, & Buchner, 2007) was utilized and it was estimated that only 160 participants would be needed to meet the proposed power and alpha requirements for 15 predictor variables.

Measures

The author (B.D.H.) was thoroughly involved in both the collection and scoring of data as well as the decision of what measures to include in this study. Data from the following measures was utilized in this study. For the clinical Wechsler scales, as the Letter-Number Sequencing and Digit Span subtests were present in both the WAIS-III and WMS-III, only one subtest was given and the score was used for both batteries.

WAIS-III

The following subtests of the WAIS-III (Wechsler, 1997a) were administered to all subjects according to the protocol described in the test manual: Picture Completion, Vocabulary, Digit Symbol Coding, Similarities, Block Design, Arithmetic, Matrix Reasoning, Digit Span, Information, Picture Arrangement, Comprehension, Symbol Search, and Letter-Number Sequencing. Optional subtests were not administered. From these subtests, both Full Scale IQ and the Working Memory Index (WMI) were calculated. While index scores were standardized in respect to the subject's age reference group using normative data found in the test manual, raw score data was utilized for analyses of individual WAIS-III subtests.

<u>WMS-III</u>

The following subtests of the WMS-III (Wechsler, 1997b) were administered to all subjects according to the protocol described in the test manual: Logical Memory I (both stories), Faces I, Verbal Paired Associates (VPA) I, Family Pictures I, Letter-Number Sequencing, Spatial Span, Digit Span, Logical Memory II (recall and recognition), Faces II, VPA II (recall and recognition), and Family Pictures II. Again, optional subtests were not administered. The administered subtests allowed for the calculation of the Immediate Memory Index (IMI) and the WMI. The IMI is reported to be the best measure of global immediate memory ability derived from the WMS-III (Psychological Corporation, 1997) and is composed of Logical Memory I, Faces I, VPA I, and Family Pictures I. While index scores were standardized in respect to the subject's age reference group using normative data found in the test manual, raw score data was utilized for analyses of an individual WMS-III subtest.

<u>Automatic O-Span Task</u>

For this study, a modification of the original o-span task was utilized, referred to as the automatic o-span task (ao-span; Unsworth, Heitz, Schrock, & Engle, 2005). In the ao-span, a letter and a simple, two-step mathematics problem were simultaneously presented to the subject on a computer screen. The subject was instructed to both retain the presented letter and correctly respond in a true or false format to a number presented as the answer to the mathematic problem. The correct mathematical answer was presented half the time randomly. Randomly, between two to seven trials, the participant was presented on the computer screen with a list of all letters used during the experiment and instructed to select the letters that were previously presented in the order of presentation. List length varied from 3 to 7 letters with three trials at each length. Subjects were reminded periodically throughout the task to maintain a high level of correct responding to the mathematics problem. This ensured that they were actively taxing working memory by participating

in the secondary processing component and not simply rehearsing the list of presented letters. The ao-span differs from the original o-span task (Turner & Engle, 1989) in that words, not letters, are presented and recalled in the original task. The modifications to the task eased computer administration. The ao-span score that was used in this study is an index that reflects the number of letters recalled on errorless trials with more weight given to longer and more difficult trials.

L-Span Task

The l-span task (Cowan et al., 2005) utilized in this study was based on the paradigm utilized by Kail and Hall (1999). This was a modification of the reading span task (Daneman & Carpenter, 1980) where the stimuli were presented via headphones rather than being read by the subject. This task was chosen over the reading span to have consistency in stimuli presentation across tasks as most of the subtests from the Wechsler scales that comprise the WMI are presented orally. Presentation in the same modality reduced method variance that could affect the outcome of later analyses. In the l-span, participants were presented with an auditory sentence and had to then press a button characterizing the sentence's content as either true or false. They were also instructed to retain the last word of each presented sentence. Randomly between two to nine sentence presentations, the subject was instructed to type the last word of each sentence in sequence. The subject completed three trials at each span length (two to nine sentences), resulting in a trial block. The task was discontinued when three incorrect responses were given at any trial block. The l-span score was calculated by summing the number of words recalled on errorless trials.

Modified Lag Task

A version of the *n*-back task, known as the modified lag task (MLT), was used (Dobbs & Rule, 1989; Shelton, Metzger, & Elliott, 2007). The MLT has an advantage over other forms of the *n*-back task in that it has demonstrated convergent validity with other measures of working memory (Shelton et al., 2007). Additionally, it has a standardized length of stimulus presentation to control

27

for variability in rehearsal time that affects other versions of the *n*-back task. In the task, subjects were shown lists of words one word at a time. List length was either four or six words. All words used in the MLT were matched for familiarity and frequency of usage. Immediately after viewing the list, participants were asked to type either the last word from each list, the word that was one back in each list, two back in each list, or three back in each list. Each subject was given five trials at each of the four *n*-back conditions (0-back, 1-back, 2-back, and 3-back) for both the list length of four and the list length of six conditions with the end result being that each subject completed 40 trials total. The MLT score that was used is an index that reflects performance on errorless trials with more weight given to more difficult *n*-back conditions (e.g. 0-back performance x 1, 1-back performance x 2, 2-back performance x 3, and 3-back performance x 4).

Procedure

Data collection for this study was approved by the Louisiana State University Institutional Review Board. All data were collected in two sessions that took place approximately one week apart. This was necessary due to the amount of time required for data collection, typically four hours. Informed consent was always collected at the first session. The clinical measures (WAIS-III and WMS-III) were typically given together at one session and took approximately two and a half hours to complete. Both the WAIS-III and the WMS-III were administered by either B.D.H. or an undergraduate research assistant who was trained in the proper administration procedure by B.D.H. The experimental measures (ao-span, l-span, and MLT) were typically given as a separate session which typically required one and a half hours of time. While counterbalancing of presentation was done within each session, it was not feasible to counterbalance across sessions as specifically trained research assistants were required for both the clinical and experimental tasks. Also, preliminary analyses did not reveal a significant administration order effect. All experimental

working memory tasks were presented via desktop computer and were administered by a trained

research assistant.

Results

<u>Treatment of Missing Data</u>

As previously stated in the Participants section, maximum likelihood estimation with an expectation maximization algorithm (Jamshidian & Bentler, 1999) was used to impute missing data for these analyses. This process was done using the EQS 6.1 program (Bentler, 1995). Expectation maximization methods are superior to other imputation methods, such as regression, because they do not overfit the final solution and produce more realistic estimates of variance (Ullman, 2001). Also, maximum likelihood estimation has more lenient statistical assumptions. A total of 7 cases contained some missing data with the majority of cases having only one variable missing and none having more than two variables missing.

<u>Preliminary Data Analysis</u>

Descriptive statistics are presented for all of the measures that were given in this study (see Table 1, 2, and 3). All of the measures were normally distributed and had reasonable skewness and kurtosis coeffecients (>-1 and <1). The data were assessed for outliers by converting all utilized variables to z-scores. For the data converted to z-scores, scores in excess of 3.29 in either direction were considered potential outliers (Tabachnick & Fidell, 2001). Eight such cases were found and were examined further. Six of these cases were considered true outliers and their scores were replaced by the next closest score for that variable. New z-scores were then computed for those variables. No cases with z-scores in excess of 3.29 were observed in this second round of analysis.

Additionally, the WAIS-III and WMS-III data were checked for reliability by having the data rescored by a different individual than the person who initially scored the data and comparing the two sets of scores. High scoring reliability was observed across subtests (all in excess of $r =$ .90). Such a scoring reliability check was not necessary for the ao-span, l-span, and MLT data as these tasks were computer administered.

Table 1

Descriptive Statistics of WAIS-III Variables

WAIS-III Index Score or Subtest Raw Score	Mean (SD)
Full Scale IQ	110.41 (11.11)
Working Memory Index	107.16 (12.51)
Vocabulary	47.79 (7.56)
Similarities	24.58 (3.96)
Arithmetic	13.92 (3.12)
Digit Span Total	19.49 (3.75)
Digit Span Forward	11.47 (1.94)
Digit Span Backward	8.03 (2.34)
Information	18.82 (4.31)
Comprehension	22.12 (4.29)
Letter-Number Sequencing	12.53 (2.65)
Picture Completion	20.64 (3.03)
Digit Symbol Coding	88.17 (13.31)
Block Design	45.85 (10.89)
Matrix Reasoning	19.93 (2.88)
Picture Arrangement	15.57 (3.37)
Symbol Search	41.04 (6.87)

Note. Only Full Scale IQ and Working Memory Index are standardized scores. All others are raw scores.

Table 2

Descriptive Statistics of WMS-III Variables

WMS-III Index Score or Subtest Raw Score	Mean (SD)
Immediate Memory Index	105.52 (11.77)
Working Memory Index	107.27 (13.44)
Logical Memory I	46.25 (9.25)
Faces I	39.23 (4.25)
VPA I	23.06 (5.91)
Family Pictures I	51.36 (7.35)
Spatial Span Total	17.82 (3.12)
Spatial Span Forward	9.56 (1.83)
Spatial Span Backward	8.26 (1.74)
Logical Memory II	29.95 (6.87)
Faces II	39.82 (3.80)
VPA II Recall	7.36 (1.39)
Family Pictures II	51.62 (7.41)

Note. Only Immediate Memory Index and Working Memory Index are standardized scores. All others are raw scores.

Working Memory Criterion Construct Score

As it had been decided that the cognitive experimental measures would be utilized as the working memory criterion construct for this study, the ao-span, l-span and MLT data was subjected to an exploratory factor analysis (EFA). Principal axis factoring was used to assess the underlying construct. A single factor emerged (see Table 4) that explained 49.75% of shared variance in the three variables. The following assumptions were met: independent sampling,

Table 3

Descriptive Statistics of Experimental Cognitive Variables

Cognitive Tasks	Mean (SD)
Listening Span Score	29.70 (12.28)
Operation Span Score	44.13 (15.78)
Operation Span Errors	5.58 (2.94)
Modified Lag Task Score	54.41 (16.50)

Note. All scores are raw scores.

Table 4

Working Memory Criterion Construct Factor Loadings

Experiment Cognitive Task	Factor Loading	Initial Communality
Listening span	.81	.39
Operation span	.70	.35
Modified lag task	.59	.26
Eigenvalue	1.49	
Percent of variance explained	49.75	

normality, linear relationships between variables, and variables were correlated at a moderate level. Additionally, the Kaiser-Meyer-Olkin measure of sampling adequacy was sufficient (.67) and Bartlett's test of sphericity was significant [χ^2 (3, N=180) = 121.47, $p<.001$]. This factor was saved as a regression-derived factor score for each participant using SPSS 15.0 (2006). It is this factor score that is utilized in further analyses as the working memory criterion construct (WMCC) score. Such a data reduction approach reduces both method variance and error variance found in each of

the three individual experimental cognitive measures and provides a single relatively "pure" measure of their common variance which we assume to be working memory.

Convergent Validity

To address Hypothesis 1, two separate sets of Pearson product-moment correlations were used. The first compared the WMI scores from both the WAIS-III and the WMS-III to the scores from the experimental tasks as well as to the WMCC score (see Table 5). High positive correlation coefficients were found, demonstrating convergent validity. The second compared the WMCC to the Wechsler subtests that form the WMIs: Digit Span total score, Arithmetic, Letter-Number Sequencing, and Spatial Span total score. Forward and backward raw scores from the Digit Span and Spatial Span subtests were also examined. Hypothesis 1 was not supported as convergent validity of all measures was generally observed. However, there was some variability in the correlation coefficients for the Arithmetic and Spatial Span subtests from the Wechsler scales, calling into question the convergent validity of these two specific subtests (see Table 6).

Table 5

Pearson's Product-Moment Correlations of Working Memory Measures

Working Memory Measures	2	3	4	5	6
1. WAIS-III WMI	.76	.52	.50	.55	.62
2. WMS-III WMI	-	.47	.44	.43	.54
3. Listening Span		-	.57	.48	.92
4. Operation Span			-	.41	.80
5. Modified Lag Task				-	.67
6. Working Memory Criterion Construct					-

Note. All correlations are $p<.001$

Table 6

Pearson's Product-Moment Correlations of Subtests that Contribute to the WAIS-III and WMS-III Working Memory Indexes Compared to the Working Memory Criterion Construct

Subtests	2	3	4	5	6	7	8	9
1. Arithmetic	.46**	.32**	.26**	.31**	.20**	.20**	.15	.38**
2. Letter-Number Sequencing	-	.60**	.44**	.60**	.42**	.41**	.33**	.53**
3. Digit Span Total		-	.84**	.90**	.32**	.31**	.26**	.58**
4. Digit Span Forward			-	.52**	.22**	.23**	.16*	.51**
5. Digit Span Backward				-	.33**	.30**	.27**	.50**
6. Spatial Span Total					-	.88**	.87**	.41**
7. Spatial Span Forward						-	.52**	.38**
8. Spatial Span Backward							-	.33**
9. Working Memory Criterion Construct								-

*$p<.05$, **$p<.01$

Discriminant Validity

To address Hypothesis 2, two separate sets of Pearson's product-moment correlations were used. The standard score from the Immediate Memory Index (IMI) on the WMS-III was used as the memory criterion variable to which the other variables were compared. The first set of correlations compared the WMIs from both the WAIS-III and the WMS-III to the ao-span score, l-span score, MLT score, WMCC score, and the WMS-III IMI score. These all had a small to medium relationship with the immediate memory variable (see Table 7) with the exception of ao-span ($r =$ .08). The other correlation set examined the relationship between the IMI and the following Wechsler subtests that form the WMIs: Digit Span total score, Arithmetic, Letter-Number Sequencing, and Spatial Span total score. Forward and backward raw scores from the Digit Span and Spatial Span subtests were also examined. These also all had a small to medium relationship with the immediate memory variable (see Table 8) with the exception of Spatial Span forward

Table 7

Pearson's Product-Moment Correlations of Working Memory Measures Compared to WMS-III
Immediate Memory Index

Working Memory Measures	Immediate Memory Index
WAIS-III WMI	.21**
WMS-III WMI	.17*
Listening Span	.23**
Operation Span	.08
Modified Lag Task	.25***
Working Memory Criterion Construct	.22**

$*p<.05; **p<.01; ***p=.001$

Table 8

Pearson's Product-Moment Correlations of Subtests that Contribute to the WAIS-III and WMS-III
Working Memory Indexes Compared to WMS-III Immediate Memory Index

Working Memory Measures	Immediate Memory Index
Arithmetic	.17*
Letter-Number Sequencing	.19**
Digit Span Total	.18*
Digit Span Forward	.13
Digit Span Backward	.17*
Spatial Span Total	.10
Spatial Span Forward	.03
Spatial Span Backward	.15*

$*p<.05; **p<.01$

($r=.03$). It is particularly interesting that the forward components of both Digit Span and Spatial

Span had some of the lowest overall correlations with the Immediate Memory Index of all the

measures. This finding is directly contradictory to the second hypothesis of this study. Overall, the

hypothesis that the clinical measures assess immediate memory abilities more than the cognitive

experimental measures was not supported (though ao-span appears to be an exception).

Factor Analyses

All of the following CFA analyses that are reported below were performed using EQS 6.1

(Bentler, 1995). In CFA, the factor structure is based on theoretical considerations and specified a

priori. The obtained data is then compared with the specified model. The χ^2 statistic is one way to

assess the "goodness of fit" of the model to the actual data but this approach is sensitive to sample

size. The fit of each model to the data can also be examined by interpreting fit indices. The

following fit indices were used in this study to determine data to model fit: the Bentler-Bonett

Normed Fit Index (NFI), the Comparison Fit Index (CFI), and the Standardized Root-Mean Square

Residual (SRMR). The values of the NFI and CFI, which are incremental fit indices (though both

use different approaches), range from 0 to 1.0. Higher values are indicative of greater model fit with

0 indicating that the data does not fit the proposed model and 1.0 indicative of a perfect fit. For the

NFI, values over .90 are considered a good-fitting model while for the CFI, values over .95 are

indicative of exceptional fit of the data to the model (Ullman, 2001). For the SRMR, an absolute fit

index, values closer to zero are considered to indicate a better fit of the data to the proposed model.

SRMR values less than .08 are considered indicative of acceptable model fit. Within these classes of

fit indices, the ones selected were those considered sensitive to model misspecification (i.e., models

that lack necessary parameters or cluster the variables inappropriately) while at the same time being

relatively insensitive to small sample sizes (i.e., n<170 ; Hu & Bentler, 1998).

To address Hypothesis 3, a CFA was performed on the following variables: Arithmetic,

Digit Span, Letter-Number Sequencing, Spatial Span, ao-span, l-span, and MLT. Two different

models were tested: a single factor model representing all the variables measuring a single latent

construct and a two factor model where all of the experimental measures load on one factor and the

clinical Wechsler subtests load on a separate factor. A confirmed single factor model would support that the clinical measures are validly measuring the cognitive experimental construct of working memory and would also lend support to the view espoused in the Wechsler scales technical manual that there is single working memory work-space (Psychological Corporation, 1997). It was hypothesized that a two factor model would best explain the data.

Single Factor Model with All Clinical Measures

The first single factor CFA that was performed specified a single latent factor composed of Arithmetic, Digit Span, Letter-Number Sequencing, Spatial Span, ao-span, l-span, and MLT. All assumptions were met for CFA (multivariate kurtosis, no multivariate outliers, normality, adequate sample size, and moderate intercorrelations without multicollinearity). For this single factor model, the χ^2 (14, N=180) = 41.80, p <.001 and the following fit indices values were reported: NFI .90, CFI .93, and SRMSR .05. These are all indicative of acceptable, though not excellent, model fit. Factor loadings ranged from .50 for Arithmetic to .74 for Digit Span (see Table 9).

Table 9

Standardized Solution for Single Factor CFA of All Working Memory Measures

Variable	Factor Loading	Error	R^2
Arithmetic	.50	.87	.24
Digit Span	.74	.67	.55
Letter-Number Sequencing	.73	.68	.53
Spatial Span	.50	.87	.25
Automatic Operation Span	.67	.74	.45
Listening Span	.68	.73	.47
Modified Lag Task	.65	.73	.42

Two Factor Model

The second CFA that was performed specified two latent factors. The first factor was composed of

Arithmetic, Digit Span, Letter-Number Sequencing, and Spatial Span and can be considered a

clinical working memory assessment factor. The second factor was composed of ao-span, l-span,

and MLT and can be conceptualized as an experimental cognitive working memory tasks factor.

These latent factors were modeled as being correlated since it was theoretically implausible that

they were orthogonal constructs. For this two factor model, the χ^2 (13, N=180) = 33.16, p <.01 and

the following fit indices values were reported: NFI .92, CFI .95, and SRMSR .05. These fit indices

are again indicative of good model fit and appear slightly better than the single factor model. Factor

loadings ranged from .50 for Spatial Span to .78 for Letter-Number Sequencing (see Table 10).

Table 10

Standardized Solution for Two Factor CFA with Separate Factors for Clinical Working Memory
and Cognitive Experimental Working Memory Measures

| Variable | Factor Loadings | | Error | R^2 |
	Factor 1	Factor 2		
Arithmetic	.51		.86	.26
Digit Span	.76		.65	.57
Letter-Number Sequencing	.78		.63	.61
Spatial Span	.50		.87	.25
Automatic Operation Span		.71	.70	.51
Listening Span		.74	.67	.55
Modified Lag Task		.65	.76	.43

Single Factor Model with Arithmetic and Spatial Span Removed

As both the previous single factor and the two factor model fit the data adequately but not exceptionally, a third CFA was performed that specified a single latent factor composed of all of the previous measures with the exceptions of Arithmetic and Spatial Span. This model was proposed based on the hypothesis that model fit in the first CFA could be improved by removing variables with factor loadings that were obviously lower than the other variables. For this alternative single factor model, the χ^2 (5, N=180) = 22.55, $p <.001$ and the following fit indices values were reported: NFI .93, CFI .94, and SRMSR .04. Again, these indexes indicate good model fit. Factor loadings ranged from .63 for MLT to .77 for Digit Span (see Table 11). A χ^2 difference test (Ullman, 2001) did not find a significant difference between the first and third models, both of which specified a single factor [χ^2 (9, N=180) = 19.25, $p <.05$]. In summary, the third hypothesis of this study, that a two factor model would fit the data better than a single factor model, was not fully supported, though the fit indices were slightly better for the two factor model and the single factor model removing Arithmetic and Spatial Span.

Table 11

Standardized Solution for Single Factor CFA of All Working Memory Measures Except Arithmetic and Spatial Span

Variable	Factor Loading	Error	R^2
Digit Span	.77	.64	.59
Letter-Number Sequencing	.70	.72	.48
Automatic Operation Span	.70	.72	.49
Listening Span	.69	.73	.47
Modified Lag Task	.63	.78	.40

<u>EFA of Digit Span and Spatial Span</u>

As a primary hypothesis of this study was that the addition of forward spans in the Digit Span and Spatial Span subtests includes a significant amount of variance that is not accounted for by working memory ability, an exploratory factor analysis was also run on the forward and backward components of the Digit and Spatial Span subtests from the Wechsler scales. This addressed Hypothesis 4 and was an attempted partial replication of Reynolds' (1997) findings which found that forward span loaded separately from backward span. However, Reynolds' analysis included only adolescents whereas the dataset that was utilized in this study was composed predominantly of young adults. Also, Reynolds only looked at verbal span tasks (digit and letter), and did not include a spatial working memory task, such as Spatial Span task, as was done in the present analysis.

Principal axis factoring was utilized to assess whether forward and backward spans of the Digit Span and Spatial Span subtests are assessing the same construct. All assumptions previously listed for exploratory factor analysis were met. Two analyses where performed, one with Varimax rotation and the other with Promax rotation. Varimax rotation clarifies factor structure but assumes orthogonality of the underlying structure (which is unlikely with cognitive functions) while Promax assumes that the factors are oblique (a more likely assumption). In the analysis with Varimax rotation, two factors emerged (see Table 12). The first factor was defined by Digit Span forward and Digit Span backward and accounted for 26.36% of the variance. The second factor was defined by Spatial Span Forward and Spatial Span backward and accounted for 26.25% of the variance. The analysis with Promax rotation yielded almost identical results. Again, a two factor structure was found (see Table 13). The first factor was defined by Digit Span forward and Digit Span backward and accounted for 38.36% of the variance. The second factor was defined by Spatial Span Forward and Spatial Span backward and accounted for 14.25% of the variance. In summary, the fourth hypothesis of this study was not supported and Reynolds' (1997) previous findings were not

Table 12

Exploratory Factor Analysis with Varimax Rotation of Forward and Backward Components of Digit Span and Spatial Span Subtests

| | Rotated Factor Loadings | | |
Span Tasks	Factor 1	Factor 2	Initial Communalities
Digit Span Forward	.72	.11	.53
Digit Span Backward	.69	.25	.53
Spatial Span Forward	.22	.67	.30
Spatial Span Backward	.12	.73	.28
Eigenvalues	1.06	1.05	
Percent of variance explained	26.36	26.25	

Table 13

Exploratory Factor Analysis with Promax Rotation of Forward and Backward Components of Digit Span and Spatial Span Subtests.

| | Rotated Factor Loadings | | |
Span Tasks	Factor 1	Factor 2	Initial Communalities
Digit Span Forward	.76	-.07	.30
Digit Span Backward	.68	.09	.28
Spatial Span Forward	.06	.67	.28
Spatial Span Backward	-.06	.76	.32
Eigenvalues	1.53	.57	
Percent of variance explained	38.36	14.25	

replicated as forward and backward span tasks loaded on a single factor. However, Digit Span and Spatial Span components loaded on separate factors indicating that they are likely assessing different constructs. This does support other findings of this study that Spatial Span appears to be a measure somewhat distinct from the other clinical working memory subtests.

Multiple Regression

To address the fifth hypothesis of this study that additional Wechsler subtests would account for a greater percent of variance in the WMCC than the current subtests that compose the WMIs, multiple regressions were conducted to determine the best linear combination of subtests from the WAIS-III and WMS-III for predicting the WMCC score. These multiple regressions compared the following predictor models: 1) all WAIS-III subtests with Digit Span total score used, 2) all WAIS-III subtests with Digit Span forward and backward scores used instead of Digit Span total score, 3) all WMS-III subtests with Spatial Span total score used, 4) all WMS-III subtests with Spatial Span forward and backward scores used instead of Spatial Span total score, 5) all of the significant predictor subtests from the previous four analyses, 6) only the subtests that make up the WAIS-III WMI, and 7) only the subtests that make up the WMS-III WMI. Including models with Digit and Spatial Span total scores and separate forward and backward scores helps to resolve whether forward and backward spans account for unique variance in working memory performance. The first five multiple regressions used backward elimination which initially includes all predictor variables before eliminating the least significant predictor. It then refits the model and begins the process again until only the best predictors are left. All assumptions for multiple regression (linearity, normality of errors, and uncorrelated errors) were met.

WAIS-III Predictor Models

The first multiple regression included all of the subtests from the WAIS-III with the WMCC score being the predicted variable. The following combination of subtests significantly predicted [$F(5, 174) = 27.78$, $p<.001$] WMCC score: Letter-Number Sequencing, Digit Span, Matrix Reasoning, Symbol Search, and Vocabulary. This model accounted for 43% of the variance in WMCC score (adjusted R^2 value = .43). The standardized beta weights (see Table 14) suggested that Digit Span and Letter-Number Sequencing contributed most to predicting the WMCC score.

43

Table 14

Backward Elimination Multiple Regression Analysis Final Model for All WAIS-III Subtests Predicting WMCC Score

WAIS-III Subtest	B	SEB	β
Constant	-4.91	.52	-
Digit Span total	.08	.02	.36***
Letter-Number Sequencing	.07	.02	.22**
Matrix Reasoning	.05	.02	.15**
Vocabulary	.02	.01	.15*
Symbol Search	.01	.01	.11*

Note. $R^2 = .44$; $F(5,174) = 27.78$, $p < .001$
*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

The second multiple regression was identical to the first but included Digit Span forward and backward raw scores as separate variables instead of as a combined Digit Span total raw score in the first multiple regression. The following combination of subtests significantly predicted [$F(6, 173) = 23.38$, $p<.001$] WMCC score: Letter-Number Sequencing, Digit Span forward, Digit Span backward, Matrix Reasoning, Symbol Search, and Vocabulary. Both forward and backward components of Digit Span were retained in the model. This model accounted for 43% of the variance in WMCC score (adjusted R^2 value = .43). While it may appear at first glance that separating Digit Span into its forward and backward components had no effect, the standardized beta weights (see Table 15) suggested that Digit Span forward and Letter-Number Sequencing contributed the most to predicting the WMCC score. Digit Span backward only contributed slightly more than Matrix Reasoning and Vocabulary to the final model.

WMS-III Predictor Models

The third multiple regression included all of the subtests from the WMS-III with the WMCC score as the predicted variable. The following combination of subtests significantly predicted [$F(5,$

Table 15

Backward Elimination Multiple Regression Analysis Final Model for WAIS-III Subtests (with Digit Span Broken into Forward and Backward Components) Predicting WMCC Score

WAIS-III Subtest	B	SEB	β
Constant	-5.02	.53	-
Digit Span forward	.08	.03	.24***
Letter-Number Sequencing	.08	.03	.23**
Digit Span backward	.07	.03	.18*
Matrix Reasoning	.05	.02	.15**
Vocabulary	.02	.01	.15*
Symbol Search	.01	.01	.11

Note. $R^2 = .45$; $F(6,173) = 23.38$, $p < .001$
*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

174) = 27.78, p<.001] WMCC score: Letter-Number Sequencing, Spatial Span, and Logical Memory I. This model accounted for 36% of the variance in WMCC score (adjusted R^2 value = .36). The standardized beta weights (see Table 16) suggested that Letter-Number Sequencing and Logical Memory I contributed the most to predicting the WMCC score.

Table 16

Backward Elimination Multiple Regression Analysis Final Model for All WMS-III Subtests Predicting WMCC Score

WMS-III Subtest	B	SEB	β
Constant	-3.46	.37	-
Letter-Number Sequencing	.13	.02	.39***
Logical Memory I	.02	.01	.23***
Spatial Span total	.05	.02	.18**

Note. $R^2 = .37$; $F(3,176) = 33.79$, $p < .001$
*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

The fourth was identical to the third multiple regression but included the Spatial Span forward and backward raw scores as separate variables instead of as a combined Spatial Span total raw score in the third multiple regression. Similar to Digit Span, this was done to ascertain whether the forward component of the subtest accounted for any unique variance in the WMCC score. The following combination of subtests significantly predicted [$F(3, 176) = 33.39$, $p<.001$] WMCC score: Letter-Number Sequencing, Spatial Span forward, and Logical Memory I. Please note that Spatial Span backward was not retained in the model as a significant predictor variable. Similar to the previous WMS-III predictor model, this model accounted for 35% of the variance in WMCC score (adjusted R^2 value = .35). The standardized beta weights (see Table 17) again suggested that Letter-Number Sequencing and Logical Memory I contributed the most to predicting the WMCC score.

Table 17

Backward Elimination Multiple Regression Analysis Final Model for WMS-III Subtests (with Spatial Span Broken into Forward and Backward Components) Predicting WMCC Score

WMS-III Subtest	B	SEB	β
Constant	-3.37	.36	-
Letter-Number Sequencing	.13	.02	.40***
Logical Memory I	.02	.01	.23***
Spatial Span forward	.08	.03	.17**

Note. $R^2 = .36$; $F(3,176) = 33.39$, $p < .001$
*$p \le .05$, **$p \le .01$, ***$p \le .001$

Combined WAIS-III and WMS-III Predictor Model

The fifth multiple regression included the subtests from the WAIS-III and WMS-III that were identified as being significant predictors of the WMCC score from the previous multiple regression models. These were Letter-Number Sequencing, Digit Span forward, Digit Span backward, Matrix

46

Reasoning, Vocabulary, Symbol Search, Logical Memory I, Spatial Span forward, and Spatial Span backward. The WMCC score was again used as the predicted variable. Both Digit Span and Spatial Span were entered as separate forward and backward variables as the previously discussed models had demonstrated that these components contribute different amounts of predictive ability. The following combination of subtests significantly predicted [$F(6, 173) = 24.28$, $p<.001$] WMCC score: Letter-Number Sequencing, Digit Span forward, Digit Span backward, Matrix Reasoning, Spatial Span forward, and Logical Memory I. Please note that the Vocabulary and Spatial Span backward subtests were the only predictor variables removed from the model in this analysis. This combined WAIS-III/WMS-III subtests model accounted for 44% of the variance in WMCC score (adjusted R^2 value = .44), nearly identical to the WAIS-III alone predictor models (see Tables 14 and 15). While the standardized beta weights (see Table 18) suggested that Digit Span forward and Letter-Number Sequencing contributed the most to predicting the WMCC score, the majority of beta weights were relatively close to one another.

Table 18

Backward Elimination Multiple Regression Analysis Final Model for All Previous Significant Predictors of WMCC Score

WAIS-III/WMS-III Subtests	B	SEB	β
Constant	-4.67	.45	-
Digit Span forward	.12	.03	.26***
Letter-Number Sequencing	.06	.03	.19**
Logical Memory I	.02	.01	.16**
Digit Span backward	.06	.03	.16*
Matrix Reasoning	.04	.02	.13*
Spatial Span forward	.07	.03	.14*

Note. $R^2 = .46$; $F(6,173) = 24.28$, $p < .001$
*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

<u>WAIS-III and WMS-III Working Memory Index Models</u>

As the previously discussed best-derived predictors models were composed of more WAIS-III and WMS-III subtests that just those currently used to formulate the WMIs (Digit Span, Arithmetic, Letter-Number Sequencing, and Spatial Span), it was deemed relevant to determine the amount of variance explained by just those subtests used to compute the current WMIs. This was important in order to determine whether the additional variables included in the previous best-derived predictor models explained a clinically significant degree of additional variance above that which is currently being assessed using only the WMIs subtests. Two multiple regressions were performed in which only the subtests that comprise the WMIs were entered: Arithmetic, Digit Span, Letter-Number Sequencing, and Spatial Span. Two models were assessed: 1) WAIS-III WMI model composed of Arithmetic, Digit Span, and Letter-Number Sequencing and 2) WMS-III WMI model composed of Letter-Number Sequencing and Spatial Span. Separate forward and backward components of Digit Span and Spatial Span were not entered as only the total scores of these subtests are used to calculate the WMI scores. The WMCC score was again used as the predicted variable for this final analysis. The first WAIS-III WMI subtests model significantly predicted WMCC score, $F(3, 176) = 39.34$, $p<.001$. This model accounted for 39% of the variance in WMCC score (adjusted R^2 value = .39). The beta weights ranged from .15 to .39 (see Table 19) and suggested that the Digit Span subtest contributed the most to predicting the WMCC score. The second model composed of the WMS-III WMI subtests also significantly predicted WMCC score, $F(2, 177) = 41.63$, $p<.001$. This model accounted for 31% of the variance in WMCC score (adjusted R^2 value = .31). For this model, the beta weights ranged from .22 to .44 (see Table 19) and suggested that the Letter-Number Sequencing subtest contributed the most to predicting the WMCC score.

In summary, the fifth hypothesis of this study was fully supported by these results. Additional subtests from the Wechsler batteries beyond those currently used to the form the WMIs

Table 19

Multiple Regression Analysis of Working Memory Indices Subtests as Predictors of WMCC Score

Working Memory Indices Subtests	B	SEB	β
WAIS-III			
Constant	-3.32	.32	-
Digit Span	.09	.02	.39***
Letter-Number Sequencing	.07	.03	.22**
Arithmetic	.04	.02	.15*
WMS-III			
Constant	-2.92	.35	-
Letter-Number Sequencing	.15	.02	.44***
Spatial Span	.06	.02	.22***

Note. WAIS-III model $R^2 = .40$; $F(3,176) = 39.34$, $p < .001$
WMS-III model $R^2 = .32$; $F(2,177) = 41.63$, $p < .001$
*$p \leq .05$, **$p \leq .01$, ***$p \leq .001$

were shown to account for a greater percent of the variance in the WMCC. Also, the WAIS-III

subtests were found to account for more variance in the criterion construct than the WMS-III

subtests. For the WMIs in particular, the WAIS-III WMI explained substantially more variance

(8%) in the WMCC score than the WMS-III WMI. However, the best fit model that was derived

from the combined WAIS-III/WMS-III subtests explained 5% more variance than the WAIS-III

WMI.

Discussion

At the onset of this project, it was hoped that the findings from this study would confirm that past literature based on the WMIs from the WAIS-III and WMS-III is easily translatable to the current concept of working memory and that both clinical neuropsychology and cognitive neuroscience are examining the same working memory construct. Conversely, it was also possible that the results of this study would discount the current methodology for assessing working memory clinically and suggest that new clinical assessment methods need to be developed based on current experimental cognitive methodologies and theoretical conceptualizations. Somewhat unexpectedly, both of these outcomes have occurred to an extent, though the current clinical methodology was generally supported.

It was anticipated that this study would find that the working memory assessment tests used clinically are measuring a construct different than the cognitive experimental working memory tasks. This assumption was based on the fact that the experimental assessment methods appear to be much more demanding than the current clinical subtests that are used to form the Wechsler WMIs. Additionally, none of the clinical assessment methods appear to require the secondary continuous processing component that is a hallmark of the experimental cognitive methodology for measuring working memory ability. However, the results revealed that all of the clinical subtests used to assess working memory are doing just that, though some are better than others and there is demonstrable room for improvement. This supports the premise that clinicians are assessing the same working memory construct that is being studied in the experimental cognitive literature. However, the results of this study also demonstrated that the Arithmetic and Spatial Span subtests are accounting for only about a quarter of the variance in working memory and that other subtests from the Wechsler clinical batteries can be utilized to more fully assess working memory ability.

This study addressed two primary questions. First, do the WMIs from the WAIS-III and WMS-III validly measure the construct of working memory? Second, can an improved WMI be derived from the subtests that comprise the WAIS-III and WMS-III? The four hypotheses that were proposed to answer the first question will now be addressed to fully explain the previously discussed findings, followed by the fifth hypothesis which addresses the second question.

Hypothesis 1

The first hypothesis predicted that the subtests that compose the WMIs would not display convergent validity with the experimental cognitive working memory measures. This was not the case. In fact, the results of this study demonstrated good convergent validity for the two clinical WMIs and the subtests that comprise them compared to the WMCC (see Tables 5 and 6). These results are also compatible with Unsworth and Engle's (2006, 2007) contention that all memory tasks that require the encoding of more than four chunks of information in primary memory will effectively function as working memory tasks. However, it was also apparent that the WMS-III WMI is not equivalent to the WAIS-III WMI as it accounted for approximately 8% less variance in the WMCC. Also, while the Arithmetic and Spatial Span subtests do appear to be tapping into working memory, they were found to be the weakest of the clinical subtests for assessing working memory as it was defined in this study. This strongly suggests that, though these measures do appear to tap into the working memory construct, they may have a high degree of error variance in regards to working memory assessment and should be reconsidered as tests of working memory ability. This possibility will be discussed in more detail later.

Hypothesis 2

The second hypothesis stated that the subtests that compose the WMIs would display discriminant validity with the experimental cognitive measures when compared to a measure of immediate memory. This hypothesis was based on the belief that the clinical measures rely more on

the immediate storage component of working memory while the experimental measures place a greater demand on the controlled attention aspect of the construct and retrieving data from secondary memory (Engle, Kane, & Tuholski, 1999; Unsworth & Engle, 2006). This hypothesis was not supported (see Table 6). A significant mild to moderate relationship was found between immediate memory and the following measures: WAIS-III WMI, WMS-III WMI, l-span, MLT, and the WMCC. The only measure not to have a significant relationship to immediate memory was the ao-span task. It appears that this task may not rely on the immediate storage of information to the same degree as other working memory measures and perhaps is assessing more cue-dependent search processes for retrieval from secondary memory (see Unsworth & Engle 2007 for a review).

Related to the second hypothesis of this study, it was also specifically hypothesized that the forward components of the Digit Span and Spatial Span subtests would be more related to the immediate memory measure than the backward components of these subtests. Again, this was not supported (see Table 8). In fact, the forward components of both Digit Span and Spatial Span forward along with the Spatial Span total score were the only individual subtests that did not demonstrate a significant relationship with immediate memory. This finding is somewhat confusing and counterintuitive as the forward components of these span tasks have the highest face validity for immediate memory of all the measures utilized in this study. Also, the WMS-III IMI that was used as the measure of immediate memory is composed of subtests that assess both auditory and visual memory, which should exclude sensory modality as a possible confound.

Attempting to explain this finding for the Spatial Span subtest, it is possible that Spatial Span requires much more abstract spatial immediate memory than the visual memory subtests that are used to form the IMI (Faces I and Family Pictures). This could explain why Spatial Span forward and total score was not significantly related to immediate memory. However, this assumes

that abstract spatial immediate memory is a separate construct from more global immediate memory.

Another possible explanation is that others have noted that factor analytic studies do not support an immediate memory versus general memory split for the WMS-III memory indexes and that the memory subtests that make up the WMS-III break down better as auditory memory versus visual memory subtests (Tulsky, Ivnik, Price, & Wilkins, 2003). Still, it would be expected that an index composed of four memory subtests would demonstrate some significant relationship to Digit Span and Spatial Span forward and Spatial Span total score. It should be noted that Digit Span backward was significantly related to IMI ($r = .17$) and that Digit Span forward's relation was relatively similar ($r = .13$), but this still does not account for Spatial Span forward and total scores' lack of relationship to the IMI, particularly when all other subtests used to form the WMIs (except Digit Span forward) were significantly related (including Spatial Span backward). However, this finding does support the possibility that Spatial Span is potentially assessing additional constructs distinct from the other WMIs subtests.

Hypothesis 3

The third hypothesis of this study was that a single factor working memory model that included all of the clinical and experimental working memory measures would not be supported. It was anticipated that a two factor model with separate factors for the WMIs subtests and experimental cognitive working memory tasks would provide a better fit for the data. The results of this study, however, found that both single and two factor models provided good, though not excellent, fits for the data (see Table 9, 10, and 11). Also, while the χ^2 statistic did not support model fit for any of the models (a common problem with smaller sample sizes; see Ullman, 2001), the results of the fit indices (which are less affected by sample size) all indicated adequate to good model fit and were consistent across all models. The best fitting models based on fit indices appear

to be the two factor model and the single factor model with Arithmetic and Spatial Span removed. However, it should be noted that a χ^2 difference test (Ullman, 2001) did not find a significant difference between the two single factor models (though this test is based on a statistic that, as previously discussed, is more greatly affected by sample size). Based on the results of this study, it appears that the WMIs can be validly conceptualized as measuring the same construct as the experimental cognitive measures (see Tables 9 and 11). Conversely, they could also be measuring a construct that is highly correlated to, but possibly separate from, the construct measured by experimental cognitive measures (second CFA model; see Table 10), though this is a less parsimonious explanation.

In the end, the results of this study support all of the clinical measures as being valid indicators of working memory ability. The results also lend some credence to the WMS-III WMI practice of combining Letter-Number Sequencing and Spatial Span as both subtests do appear to be tapping into the same construct. However, the percentage of variance accounted for by Arithmetic and Spatial Span in all of the models was approximately half of what was accounted for by other variables. This strongly suggests that while the Arithmetic and Spatial Span subtests are measuring working memory (or a related construct) to a degree, the other clinical and experimental measures are better indicators of working memory performance. Also, as these two clinical tasks only account for a quarter of the variance in working memory across models, it is likely that they account for a significant amount of variance in other constructs. Differences in ability on these unaccounted for sources of error variance may deleteriously impact assessed working memory performance when Arithmetic and Spatial Span subtests are utilized. As an example, it is likely that performance on the Arithmetic subtest is at least partially affected by level of mathematic achievement. Therefore, two individuals with similar working memory abilities may score very differently on this subtest if one has inferior mathematic ability compared to the other.

The fourth hypothesis in this study veered away from examining all of the subtests that comprise the WMIs and focused on only the Digit Span and Spatial Span subtests. This hypothesis predicted that an EFA would reveal that the forward spans of these tasks would load on a separate factor from the backward spans. This hypothesis was based on two things. First, Reynolds (1997) had previously found that forward spans loaded on a separate factor from backward spans for a digit span and a letter span task in children and adolescents. Second, clinical lore postulates that forward span tasks measure more immediate memory functions while backward span tasks assess working memory abilities to a greater degree. The hypothesis was not supported. Instead, this study found that forward and backward spans loaded together and appear to be measuring the same construct. Separate factors only emerged for different modalities of presentation. Specifically, this study found that Digit Span forward and backward loaded together on one factor and Spatial Span forward and backward loaded together on a separate factor. Also, though not included in the results section due to repetition, this combined forward and backward span factor loading was replicated when EFAs were performed using only Digit Span or only Spatial Span data. Therefore, these findings do not appear to be an artifact of factor analyzing auditory and spatial tasks together.

There are a number of possible reasons why the present findings differ from Reynolds' (1997) results. First, Reynolds employed a much larger sample that was collected as part of a nationally representative normative group. Though our study had adequate power, perhaps our smaller and more geographically specific sample allowed a degree of sampling error to distort our findings. Alternatively, as the primary way the present study differs from Reynolds (1997) is in the age of subjects utilized, it is possible that there may be a developmental aspect to our divergent findings. Perhaps forward and backward spans do represent separate factors in adolescents; that the working memory system matures over time has been well documented in the literature (Cowan,

1998). It is possible that ceiling effects in adolescents differentially affect forward compared to backward span tasks. Such an effect could lead to forward span tasks measuring an ability very distinct from backward span task in this age group. As the working memory system matures and ceiling effects are pushed upward, perhaps forward and backward span tasks become more commiserate with one another. Whether such a shift is gradual or is more of an all-or-nothing event would be an interesting area for future research to explore. Regardless of the mechanism, this study's findings support continuing the practice of combining forward and backward spans within tasks but demonstrate that the sensory modality of the task is a consideration and potential confound.

The four previously discussed hypotheses attempted to determine whether the WMIs from the WAIS-III and WMS-III, and the subtests that comprise them, validly measure the construct of working memory. With a few caveats, the results generally affirmed that the WMIs from the Wechsler batteries are measuring working memory as it is defined in the experimental cognitive literature. However, while all of the clinical measures do appear to account for some variance in working memory, the Digit Span and the Letter-Number Sequencing subtests appear to be superior working memory measures compared to the Arithmetic and Spatial Span subtests. Also, the results of this study do not support clinical lore that forward and backward components of Digit Span and Spatial Span measure different memory constructs. However, the results do suggest that the Digit Span and Spatial Span subtests may each be looking at separate, though possibly related, constructs.

Hypothesis 5

The second question which this study originally sought to answer was whether other subtests from the WAIS-III and WMS-III than the ones currently utilized could be used to better assess working memory. The specific hypothesis for this question predicted that a multiple regression equation that includes additional subtests from the WAIS-III and WMS-III beyond those currently

used to form the WMIs would account for a greater percent of variance in the WMCC than the current clinical WMIs. Even more specifically, it was hypothesized that subtests from the WAIS-III, compared to the WMS-III, would account for more variance in the WMCC as the WAIS-III subtests assess a wide variety of attentional and concentration functions and do not focus on aspects of "pure" memory. The results of this study supported this hypothesis in its entirety and will now be discussed.

The multiple regressions that focused on the WAIS-III found that the best combination of subtests for assessing working memory were Digit Span, Letter-Number Sequencing, Matrix Reasoning, Vocabulary, and Symbol Search. This combination of subtests was able to account for a greater amount of the variance in the WMCC than the WAIS-III WMI (see Table 15 and 19). The fact that Digit Span and Letter-Number Sequencing were the strongest predictors is not unexpected. However, it was not anticipated that Digit Span forward would account for the most variance explained (see Table 15) as Letter-Number Sequencing has much more face validity as a measure of working memory ability. This supports previous findings that simple span tasks are roughly equivalent to much more complex working memory tasks (Unsworth & Engle, 2006) as retaining anything more than four chunks of information is believed to require retrieval from secondary memory similar to more complex working memory tasks.

More surprising was that Arithmetic was not included in this model and that Matrix Reasoning and Vocabulary were included. The exclusion of Arithmetic is consistent with earlier findings in this study that this subtest appears to weakly assess the working memory construct. This result is also consistent with factor analytic studies that have found that Arithmetic tends to load across WAIS-III factors and has high intercorrelations with other Wechsler subtests (Tulsky et al., 2003). Therefore, it is likely Arithmetic is more an index of g than a specific working memory measure, and the current results would support this supposition.

Symbol Search being included in the best predictor model (though it was the weakest predictor) is not completely unexpected as it is a part of the Processing Speed Index. Working memory is known to be affected by processing speed variables (Barrouillet, Bernardin, & Camos, 2004) and factor analytic models have shown that processing speed is highly correlated with working memory (Kyllonen & Christal, 1990). Presumably, rate of processing speed affects how quickly information can be refreshed in the working memory system to prevent decay (Miller, 1956; Li, 1999). Also, as Symbol Search performance is increased if the individual is able to hold the symbols with their matching numbers "on-line" in memory versus visually scanning back and forth between the symbol-number key and the response portion of the test, it makes intuitive sense that there may be a working memory component to this subtest. More intriguing theoretically is the inclusion in the prediction model of Matrix Reasoning and Vocabulary.

The inclusion of Matrix Reasoning fits well with Engle's theory that specific aspects of working memory are shared with fluid intelligence (Engle, Tuholski, Laughlin, & Conway, 1999; Kane & Engle, 2002; Unsworth & Engle, 2005, 2006). The Matrix Reasoning subtest from the WAIS-III is generally considered the best index of fluid intelligence in the battery (Tulsky et al., 2003). Its inclusion in the best fit model of WAIS-III subtests for predicting working memory ability suggests that it is accounting for some unique variance (15%; see Table 14) that is not accounted for by Digit Span or Letter-Number Sequencing. Theoretically, it can be deduced that Matrix Reasoning is accounting for either the controlled attention component of working memory (Engle et al., 1999; Unsworth & Engle, 2005) or retrieval from secondary memory processes (Unsworth & Engle, 2006) in a way that cannot be done by either Digit Span or Letter-Number Sequencing. This result strongly suggests that future versions of the WAIS should include more measures of fluid intelligence and derive ways to include such subtests in future WMIs. While the

58

upcoming WAIS revision is reported to include more measures of fluid intelligence, it does not appear these will be utilized in any way to measure working memory ability (WAIS-IV, 2008).

The inclusion of the Vocabulary subtest in the WAIS-III best predictor model of working memory ability was the single variable that initially made the least intuitive sense. However, this subtest is more than just a measure of verbal ability as it has the strongest relationship to full scale IQ of all WAIS-III subtests (Psychological Corporation, 1997) and is generally considered the best index of g among the Wechsler subtests (Tulsky et al., 2003). Additionally, it has been reported that other tests that assess vocabulary abilities are highly related to an updating function in working memory (Friedman, Miyake, Corley, Young, DeFries, & Hewitt, 2006), allowing for the addition or removal of information from the working memory system. If, similar to other measures of vocabulary ability, the WAIS-III Vocabulary subtest is accounting for an information updating aspect in working memory that is not found in other WAIS-III subtests, then its inclusion as a predictor of working memory ability becomes conceptually understandable. However, it is also possible that Vocabulary acts as a predictor via its relation with g and is not causally associated with working memory ability.

To focus on the WMS-III subtests, another multiple regression was performed. Letter-Number Sequencing was included as a possible predictor variable in this model as this subtest is used to form the WMS-III WMI along with Spatial Span. The best predictor model that emerged for the WMS-III subtests included both of these subtests along with Logical Memory I and was able to account for 5% more variance than the WMS-III WMI (see Tables 16 and 19). However, despite the fact that both subtests that comprise the WMS-III WMI were included in this model, overall these three subtests could only account for 36% of the variance in the WMCC. This is in comparison to the best WAIS-III predictors that were able to account for 43% of the variance in the WMCC. This result also supports the more specific aspect of the fifth hypothesis of this study that the WAIS-III

59

subtests would be able to account for more of the variance in the WMCC due to the fact that the battery contains a broader range of tasks that assess various aspect of attention and concentration compared to the memory-focused tasks that comprise the WMS-III. This makes the WAIS-III subtests more likely to be able to account for variance in working memory that is attributable to controlled attention functions.

Another interesting observation from this analysis was the amount of variance accounted for by Logical Memory I (30%). Logical Memory I is a contextual memory task in that the information to be remembered is presented to the individual in the form of a short story. It is easy to see how working memory ability comes into play on this task when one considers the process required for understanding an unfamiliar story. In general, the individual has to hold onto the earliest presented parts of the story until the last elements of the story are given. While this information is being maintained, the complete story can then be integrated to derive the gist of what was presented. Such a process requires a substantial amount of working memory. In Unsworth and Engle's (2006, 2007) framework, it can be assumed that recall of such a story also requires substantial retrieval from secondary memory and a story format provides a significant amount of contextual cues to aid such a process. The fact that other subtests of immediate memory that provide less context (such as Verbal Paired Associates) where not included in the model supports this interpretation.

Concerning the Spatial Span subtest in this model, it and Arithmetic had previously been identified as likely being the poorest clinical measures of working memory. Therefore, it was interesting to note that Spatial Span accounted for less than half the variance that Letter-Number Sequencing accounted for in working memory (see Tables 16 and 17). Further analysis revealed that Spatial Span backward was not a significant predictor of working memory ability. Only Spatial Span forward actually predicted a significant amount of the variance in the WMCC. This may be due to a practice effect that has been postulated to confound the Spatial Span backward component

due to the fact that the same stimuli is used for the forward component, which is always assessed first (Wilde et al., 2004). It is also plausible that Spatial Span backward relies more on primary memory and requires less retrieval from secondary memory (making it less of a working memory task) as it had the lowest mean raw score of all the clinical span tasks (see Tables 1 and 2).

Finally, as the WAIS-III and WMS-III are co-normed as a combined battery, it was deemed clinically relevant to see which subtests between the two batteries provided the best assessment of working memory functioning. Interestingly, when all of the previously significant subtest predictors from both the WAIS-III and WMS-III were entered together into a multiple regression model, several subtests were not retained in the final best predictor model of working memory ability. Specifically, Vocabulary and Spatial Span backward were not included. Instead, a model composed of Digit Span forward, Digit Span backward, Letter-Number Sequencing, Matrix Reasoning, Spatial Span forward, and Logical Memory I emerged and accounted for 44% of the variance in the WMCC. While this combined WAIS-III/WMS-III model accounted for the largest amount of variance of any of the models, it was only marginally better (1%) than the amount of variance that was able to be accounted for by the WAIS-III alone best predictor model and still left more than half the variance in working memory unexplained.

In summary, several conclusions can be made from the results of this study. First, the WMIs derived from both the WAIS-III and WMS-III do appear to be effectively measuring the same working memory construct that is reported in the experimental cognitive literature. Second, the subtests that form the WMIs vary in the degree with which they assess this construct with Letter-Number Sequencing and Digit Span being superior to Arithmetic and Spatial Span for measuring working memory. Third, while forward and backward components of Digit Span and Spatial Span appear to be assessing a single construct (though their factor structures separate by sensory modality of presentation), they may still be indexing different aspects of the same construct as they accounted

61

for different amounts of variance in the multiple regression models predicting working memory ability. Fourth, the clinical assessment of working memory using the WAIS-III can be improved by including additional subtests that are not currently included in the WMI, particularly subtests potentially related to fluid intelligence and updating aspects of working memory. However, these additional subtests would still account for less than half the variance in working memory ability.

The eminent clinician and researcher Nelson Butters (1992) stated during his Distinguished Clinical Neuropsychologists Award address that he believed that clinical memory assessment typically trailed the current cognitive literature by a decade. Others have offered a similar view (Spaan et al., 2003). This verdict likely applies to other cognitive domains, such as working memory, as well. The WAIS-III (Wechsler, 1997a) and WMS-III (Wechsler, 1997b) WMIs have both been in use for 10 years. While they will soon be replaced by new versions of these well-respected clinical assessment batteries, history has demonstrated that older test batteries continue to be used by many clinicians well into the future. The results of this study have demonstrated that many of the current clinical assessment methodologies for working memory ability appear sound. However, this study also confirmed that the clinical assessment of cognitive constructs can be improved and the use of existing batteries such as the WAIS-III can be enhanced with only slight modifications if researchers decide to make improving them an area of focus. As such, this study has direct relevance to longitudinal studies that may be using the WAIS-III and WMS-III for a prolonged period of time. The regression algorithms presented in this paper give such studies an additional way to calculate working memory ability using these batteries that is superior to the current WMIs.

Limitations

There are a number of limitations to this study which mainly focus on the sample that was utilized. This research utilized a young adult sample that had a high-average group full scale IQ (see

Table 1). Therefore, the current findings may not generalize to older individuals, particularly geriatric populations, and individuals with lower levels of intellectual functioning.

Also, the majority of these individuals demonstrated average to high average working memory performances. Previous research has demonstrated that many working memory tasks, particularly the experimental cognitive tasks used in this study, operate differently in individuals with high working memory spans compared to those with low working memory spans (Conway & Engle, 1994). As such, the findings of this study may not apply to individuals with lower levels of working memory ability. Specifically, the regression models that were derived for the Wechsler scales to predict the WMCC score may not accurately predict working memory ability in individuals with low working memory spans.

Future Research

There are several areas of research that are suggested by the results of the present study. The most obvious is that the field of clinical neuropsychology could potentially benefit from modifying the current experimental cognitive methodologies that are used to measure working memory to make it more applicable to the clinical environment. Such an endeavor would also require the collection of normative data for these assessment methods. This would give clinicians a different and potentially superior assessment paradigm to study pathological working memory impairment in clinical populations.

Related to this, the ability to parse out the controlled attention and retrieval from secondary memory processes in working memory would yield another cognitive modality for neuropsychologists to investigate that has not previously been accessed. Developing measures that specifically index these aspects of working memory would open new avenues of research and potentially explain phenomenon like why so many of our tests of executive functioning fail to

correlate highly together. Perhaps it is because they vary in the amount of functions such as controlled attention?

Finally, while this study needs to be replicated in a more diverse sample, it can be seen as providing a template for assessing the construct validity of other areas of neuropsychological assessment. The field of clinical neuropsychology shares many terms with cognitive psychology. Constructs such as memory, attention, and inhibition are discussed in both fields but sometimes measured very differently in experimental versus clinical contexts. In order to assure that clinicians can validly and reliably apply the experimental literature to the data provided by their clinical assessment methods, studies such as this are necessary to demonstrate that we are not all only using the same terms but also assessing the same constructs.

References

Aronen, E.T., Vuontela, V., Steenari, M.R., Salmi, J., & Carlson, S. (2005). Working memory, psychiatric symptoms, and academic performance. *Neurobiology of Learning and Memory, 83,* 33-42.

Ashcraft, M.H., & Kirk, E.P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General, 130(2),* 224-237.

Atkinson, R.C., & Shiffrin, R.M. (1968). Human memory: A proposed system and its control processes. In K.W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89-195). New York: Academic Press.

Baddeley, A.D. (1994). Working memory: The interface between memory and cognition. In D.L. Schacter & E. Tulving (Eds.), *Memory systems* (pp. 351-367). Cambridge, MA: MIT Press.

Baddeley, A. D. (1986). *Working memory.* Oxford, England: Clarendon Press.

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

Baddeley, A.D. (2000b). Short-term and working memory. In E. Tulving and F.I.M. Craik (Eds.), *The Oxford handbook of memory*. New York: Oxford University Press.

Baddeley, A.D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General, 129*, 126–145.

Baddeley, A.D. (2001). Is working memory still working? *American Psychologist, 56,* 849-864.

Baddeley, A.D., & Hitch, G.J. (1974). Working memory. In G. Bower (Ed.), *Recent advances in learning and motivation* (Vol. VIII). New York: Academic Press.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General, 133*, 83-100.

Bentler, P.M. (1995). *EQS Structural Equations Program Manual.* Encino, CA: Multivariate Software, Inc.

Bolton, T.L. (1892). The growth of memory in school children. *American Journal of Psychology, 4,* 362-380.

Boyle, P.A., Malloy, P.F., Salloway, S., Cahn-Weiner, D.A., Cohen, R., & Cummings, J.L. (2003). Executive dysfunction and apathy predict functional impairment in Alzheimer's disease. *American Journal of Geriatric Psychiatry, 11,* 214-221.

Brozoski, T., Brown, R.M., Rosvold, H.E., & Goldman, P.S. (1979). Cognitive deficits caused by regional depletion of dopamine in prefrontal cortex of rhesus monkeys. *Science, 205*, 929-932.

Butters, N. (1992). Memory remembered: 1970-1991. *Archives of Clinical Neuropsychology, 7,* 285-295.

Camara, W.J., Nathan, J.S., & Puente, A.E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research and Practice, 31(2),* 141-154.

Carlson, R.A., Khoo, B.H., Yaure, R.G., & Schneider, W. (1990). Acquisition of a problem solving skill: Levels of organization and use of working memory. *Journal of Experimental Psychology: General*, *119*, 193-214.

Collette, F., Van der Linden, M., & Salmon, E. (1999). Phonological loop and central executive functioning in Alzheimer's disease. *Cortex, 35(1),* 57-72.

Constantinidis, C., Williams, G.V., & Goldman-Rakic, P.S. (2001). A role for inhibition in shaping the temporal flow of information in prefrontal cortex. *Nature Neuroscience, 5(2),* 175-180.

Conway, A.R.A., & Engle, R.W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General, 123,* 354-373.

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

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

Corsi, P.M. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34(2),* 819B. (University Microfilms No. AAI05-77717).

Cowan, N. (1998). The development of working memory. In N. Cowan (Ed.), *The development of Memory in Childhood* (pp. 163-200). New York: Psychology Press

Cowan, N. (1995). *Attention and memory: An integrated framework.* Oxford, UK: Oxford University Press.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). Cambridge, UK: Cambridge University Press.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavior and Brain Sciences, 24,* 87.

Cowan, N. (2005). *Working memory capacity.* New York: Psychology Press.

Della Sala, S., Gray, C., Baddeley, A.D., Allamano, N., & Wilson, L. (1999). Pattern span: A means of unwelding visuo-spatial memory. *Neuropsychologia, 37(10),* 1189-1199.

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

Daneman, M., & Merickle, P.M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422-433.

DeLuca, J., Chelune, G.J., Tulsky, D.S., Lengenfelder, J., & Chiaravalloti, N.D. (2004). Is speed of processing or working memory the primary information processing deficit in multiple sclerosis? *Journal of Clinical and Experimental Psychology, 26,* 550-562.

Dobbs, A.R., & Rule, B.G. (1989). Adult age differences in working memory. *Psychology and Aging, 4,* 500-503.

Engle, R.W., & Kane, M.J. (2004). Executive attention, working memory capacity, and a two-factor model of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation, Vol. 44,* 145-199. New York: Elsevier.

Engle, R.W., Kane, M.J. & Tuholski, S.W. (1999). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence and functions of the prefrontal cortex. In Miyake, A. & Shah, P. (Eds.) *Models of working memory: Mechanisms of active maintenance and executive control,* 102-134. London: Cambridge Press.

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

Engle, R.W. & Oransky, N. (1999). The evolution from short-term to working memory: Multi-store to dynamic models of temporary storage. In R. Sternberg (Ed.), *The nature of human cognition.* Cambridge, MA: MIT Press.

Ericsson, K., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211-245.

Farmer, E.W., Berman, J.V.F., & Fletcher, Y.L. (1986). Evidence for a visuo-spatial scratchpad in working memory. *Quarterly Journal of Experimental Psychology, 38,* 675-688.

Faul, F., Erdfelder, E., Lang, A.G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39(2),* 175–191.

Fedorenko, E., Gibson, E., & Rohde, D. (2007). The nature of working memory in linguistic, arithmetic, and spatial integration processes. *Journal of Memory and Language, 56(2),* 246-269.

Fleming, D., Goldberg, T.E., Goldman, J.M., & Weinberger, D.R. (1995). Verbal working memory dysfunction in schizophrenia: Use of a Brown-Peterson paradigm. *Psychiatry Research, 56,* 155-161.

Galton, F. (1887). Supplementary notes on 'prehension' in idiots. *Mind, 12,* 71-82.

Gold, J.M., Carpenter, C., Randolph, C., Goldberg, T.E., & Weinberger, D.R. (1997). Auditory working memory and Wisconsin card sorting test performance in schizophrenia. *Archives of General Psychiatry, 54,* 159-165.

Golden, C.J. (1990). *Clinical interpretation of objective psychological tests, 2nd edition.* Boston: Allyn and Bacon.

Golden, C.J., Osmon, D.C., Moses, J.A., & Berg, R.A. (1981). *Interpretation of the Halstead-Reitan neuropsychological test battery: A casebook approach.* New York: Grune & Stratton.

Goldman-Rakic, P.S. (1994). Working memory dysfunction in schizophrenia. *Journal of Neuropsychiatry, 6,* 348-357.

Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.G., & Curtiss, G. (1993). *Wisconsin Card Sorting Test Manual: Revised.* Odessa, FL: Psychological Assessment Resources.

Hinkin, C.H., Castellon, S.A., Durvasula, R.S., Hardy, D.J., Lam, M.N., Mason, K.I., Thrasher, D., Goetz, M.B., & Stefaniak, M. (2002). Medication adherence among HIV + adults: Effects of cognitive dysfunction and regimen complexity. *Neurology, 59,* 1944-1950.

Hockey, A., & Geffen, G. (2004). The concurrent validity and test-retest reliability of a visuospatial working memory task. *Intelligence, 32,* 591-605.

Hu, L.T., & Bentler, P.T. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424-453.

Jacobs, J. (1887). Experiments on "prehension." *Mind, 12,* 75-79.

Jamshidian, M., & Bentler, P.M. (1999). ML estimation of mean and covariance structures with missing data using complete data routines. *Journal of Educational and Behavioral Statistics, 24,* 21-41.

Jonsdottir, S., Bouma, A., Sergeant, J.A., and Scherder, E.J.A. (2005). The impact of specific language impairment on working memory in children with ADHD combined subtype. *Archives of Clinical Neuropsychology, 20,* 443-456.

Kail, R. & Hall, L.K. (1999). Sources of developmental change in children's word-problem performance. *Journal of Educational Psychology, 9,* 660-668.

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

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

Kaplan, E., Fein, D., Kramer, J., Delis, D.C., & Morris, R. (1999). *WISC-III as a process instrument.* San Antonio, TX: The Psychological Corportation.

Klein, K., & Fiss, W.H. (1999). The reliability and stability of the Turner and Engle working memory task. *Behavior Research Methods, Instruments, and Computers, 31,* 429-432.

Knox, H.A. (1914). A scale based on the work at Ellis Island for estimating mental defect. *Journal of the American Medical Association, 62,* 741-747.

Kwong See, S.T., & Ryan, E.B. (1995). Cognitive mediation of adult age differences in language performance. *Psychology and Aging, 10,* 458-468.

Kyllonen, P.C., & Christal, R.E. (1990). Reasoning ability is (little more than) working memory capacity. *Intelligence, 14,* 389-433.

Lamar, M., Swenson, R., Kaplan, E., & Libon, D.J. (2004). Characterizing alterations in executive functioning across distinct subtypes of cortical and subcortical dementia. *Clinical Neuropsychologist, 18(1),* 22-31.

Lezak, M.D., Howieson, D.B, Loring, D.W., & Hannay, H.J. (2004). *Neuropsychological Assessment, 4th Edition.* New York: Oxford University Press.

Li, K. Z. H. (1999). Selection from working memory: On the relationship between processing and storage components. *Aging, Neuropsychology, and Cognition, 6*, 99-116.

Marson, D.C., Chatterjee, A., Ingram, K.K., & Harrell, L.E. (1996). Toward a neurologic model of competency: Cognitive predictors of capacity to consent in Alzheimer's disease using three different legal standards. *Neurology, 46,* 666–672.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81-97.

Miller, G.A., Galanter, E., & Pribram, K.H. (1960). *Plans and the structure of behavior.* New York: Holt, Rinehart, and Winston.

Newell, A., & Simon, H.A. (1972). *Human problem solving.* Englewood Cliff, NJ: Prentice-Hall.

Ostrosky-Solis, F., & Lozano, A. (2006). Digit span: Effect of education and culture. *International Journal of Psychology, 41,* 333-341.

Owen, A.M., McMillan, K.M., Laird, A.R., & Bullmore, E. (2005). *N*-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25,* 46-59

Pasini, A., Paloscia, C., Alessandrelli, R., Porfirio, M.C., & Curatolo, P. (2007). Attention and executive function profile in drug naïve ADHD subtypes. *Brain & Development, 29,* 400-408.

Peterson, L.R., & Peterson, M.J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58,* 193-198.

Psychological Corporation (1997). *Weschler Adult Intelligence Scale-Third Edition and Wechsler Memory Scale-Third Edition technical manual.* San Antonio, TX: The Psychological Corporation.

Rabin, L.A., Barr, W.B., & Burton, L.A. (2005). Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. *Archives of Clinical Neuropsychology, 20,* 33-65.

Reynolds, C. (1997). Forward and backward memory span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology, 12,* 29-40.

Robert, M., & Savoie, N. (2006). Are there gender differences in verbal and visuospatial working-memory resources? *The European Journal of Cognitive Psychology, 18(3),* 378-397.

Schillerstrom, J.E., Horton, M.S., & Royal, D.R. (2005). The impact of medical illness on executive function. *Psychosomatics, 46(6),* 508-516.

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, 125,* 4-27.

Shelton, J.A., Metzger, R.L., & Elliott, E.M. (2007). Modified lag task as a measure of working memory. *Behavior Research Methods, 39(3),* 482-493.

Smith, E.E., Jonides, J., & Koeppe, R.A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 7,*11-20.

Spaan, P.E.J., Raaijmakers, J.G.W., & Jonker, C. (2003). Alzheimer's disease versus normal aging: A review of the efficiency of clinical and experimental memory measures. *Journal of Clinical and Experimental Neuropsychology, 25(2),* 216-233.

SPSS Inc. (2006). SPSS Base 15.0 for Windows User's Guide. Chicago: SPSS, Inc.

Strauss, E., Sherman, E.M.S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary.* New York: Oxford University Press.

Stuss, D.T., Gallup, Jr., G.G., & Alexander, M.P. (2001). The frontal lobes are necessary for 'theory of mind'. *Brain, 124(2),* 279-286.

Tabachnick, B.G., & Fidell, L.S. (2001). *Using Multivariate Statistics (4ᵗʰ Edition).* Boston: Allyn & Bacon.

Tranel, D., Manzel, K., & Anderson, S.W. (in press). Is the prefrontal cortex important for fluid intelligence? A neuropsychological study using Matrix Reasoning. *The Clinical Neuropsychologist.*

Tulsky, D.A., Ivnik, R.J., Price, L.R., & Wilkins, C. (2003). Assessment of cognitive functioning with the WAIS-III and WMS-III: Development of a six factor model. In D. Tulsky, G. Chelune, R. Ivnik, A. Prifitera, D. Saklofske, R. Heaton, R. Bornstein, and M. Ledbetter (Eds.), *Clinical Interpretation of the WAIS-III and WMS-III.* San Diego: Academic Press, pp. 147-179.

Tulsky, D.A., & Price, L.R. (2003). The joint WAIS-III and WMS-III factor structure: Development and cross-validation of a six-factor model of cognitive functioning. *Psychological Assessment, 15(2),* 149-162.

Tulsky, D.A., Saklofske, D.H., & Zhu, J. (2003). Revising a standard: An evaluation of the origin and development of the WAIS-III. In D. Tulsky, G. Chelune, R. Ivnik, A. Prifitera, D. Saklofske, R. Heaton, R. Bornstein, and M. Ledbetter (Eds.), *Clinical Interpretation of the WAIS-III and WMS-III.* San Diego: Academic Press, pp. 43-92.

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

Ullman, J. (2001). Structural equation modeling. In B. Tabachnick and L. Fidell (Eds.), *Using Multivariate Statistics (4th Edition).* Boston: Allyn & Bacon, pp. 653-771.

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

Unsworth, N., & Engle, R.W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language, 54,* 68-80.

Unsworth, N., & Engle, R.W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review, 114(1),* 104-132.

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

WAIS-IV (2008) WAIS-IV product information listed on the Harcourt Assessment website. Retrieved May 13, 2008, from http://www.WAIS-IV.com

Wechsler, D. (1939). *Wechsler-Bellevue Intelligence Scale.* New York: The Psychological Corporation.

Wechsler, D. (1944). *Measurement of adult intelligence-Third Edition.* Baltimore, MD: Williams & Wilkins.

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

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

Wilde, N.J., Strauss, E., & Tulsky, D.S. (2004). Memory span on the Wechsler scales. *Journal of Clinical and Experimental Neuropsychology, 26(4),* 539-549.

Zook, N.A., Navalos, D.B., DeLosh, E.L., & Davis, H.P. (2004). Working memory, inhibition, and fluid intelligence as predictors of performance on Tower of Hanoi and Tower of London tasks. *Brain and Cognition, 56,* 286-292.

Vita

Benjamin David Hill was born in Lexington County, South Carolina, on November 24, 1976. He grew up in the northern Sandhills region of South Carolina in a small town called Pageland. Ben attended New Covenant Christian School where he was valedictorian in 1995 and a National Merit Scholarship finalist. He graduated from Coastal Carolina University in 1999 *Magna Cum Laude* with a Bachelor of Science in psychology and was later accepted to the master's program in general/experimental psychology at Wake Forest University in Winston-Salem, North Carolina. He earned a Master of Arts degree from Wake Forest University in 2001. He is presently obtaining his doctoral degree in clinical psychology with an emphasis in neuropsychology under the mentorship of Wm. Drew Gouvier, Ph.D., at Louisiana State University in Baton Rouge, Louisiana. Ben is currently completing his clinical internship at the University of Mississippi Medical Center/G.V. (Sonny) Montgomery VAMC in Jackson, Mississippi. In August 2008, he will begin his postdoctoral fellowship in clinical neuropsychology at the Alpert Medical School of Brown University in Providence, Rhode Island.