



## A comparison of laboratory and clinical working memory tests and their prediction of fluid intelligence

Jill T. Shelton<sup>\*1</sup>, Emily M. Elliott<sup>2</sup>, B.D. Hill<sup>3</sup>, Matthew R. Calamia<sup>4</sup>, Wm. Drew Gouvier<sup>2</sup>

Louisiana State University, United States

### ARTICLE INFO

#### Article history:

Received 31 March 2008

Received in revised form 5 November 2008

Accepted 7 November 2008

Available online 21 December 2008

#### Keywords:

Working memory

Memory assessment

Intelligence

Fluid abilities

Translational research

### ABSTRACT

The working memory (WM) construct is conceptualized similarly across domains of psychology, yet the methods used to measure WM function vary widely. The present study examined the relationship between WM measures used in the laboratory and those used in applied settings. A large sample of undergraduates completed three laboratory-based WM measures (operation span, listening span, and n-back), as well as the WM subtests from the Wechsler Adult Intelligence Scale-III and the Wechsler Memory Scale-III. Performance on all of the WM subtests of the clinical batteries shared positive correlations with the lab measures; however, the Arithmetic and Spatial Span subtests shared lower correlations than the other WM tests. Factor analyses revealed that a factor comprising scores from the three lab WM measures and the clinical subtest, Letter-Number Sequencing (LNS), provided the best measurement of WM. Additionally, a latent variable approach was taken using fluid intelligence as a criterion construct to further discriminate between the WM tests. The results revealed that the lab measures, along with the LNS task, were the best predictors of fluid abilities.

© 2008 Elsevier Inc. All rights reserved.

Working memory (WM) was recently defined as “a temporary storage system under attentional control that underpins our capacity for complex thought” (Baddeley, 2007, p. 1). The strong relationship between WM and complex cognition underscores the key importance of this construct in many aspects of human behavior. WM has been heavily investigated by researchers, and has been shown to play a key role in complex behaviors such as reading comprehension (Daneman & Carpenter, 1980), the acquisition of language (Baddeley, Gathercole, & Papagno, 1998), and fluid abilities

(Gray, Chabris, & Braver, 2003; Salthouse & Pink, 2008). Additionally, the importance of the construct has been noted across areas of psychology, as researchers in clinical psychology have evaluated the relationship of WM to deficits in schizophrenia (Barch, 2003) and depression (Harvey et al., 2004), social psychologists have assessed the role of WM in stereotype threat (Bonnot & Croizet, 2007), and neuropsychologists have assessed WM ability as a way to identify the early onset of Alzheimer’s disease (Rosen, Bergeson, Putnam, Harwell, & Sunderland, 2002).

The prominence of the WM construct has resulted in the use of different measurement techniques that often vary substantially in their methodology. For example, cognitive and clinical psychologists typically define WM similarly, but use different methods to assess WM function. Cognitive psychologists use laboratory tasks that have been extensively analyzed for their reliability and construct validity (for review see Conway et al., 2005). Clinical psychologists often use psychometric indices, such as subscales of the Wechsler Adult Intelligence Scale (WAIS-III) and the Wechsler Memory Scale (WMS-III), to

\* Corresponding author. Campus Box #1125, Department of Psychology, Washington University, St. Louis, MO 63130, United States. Tel.: +1 314 935 7610; fax: +1 314 935 7588.

E-mail address: jshelton@arts.wustl.edu (J.T. Shelton).

<sup>1</sup> Now at Department of Psychology at Washington University in St. Louis, United States.

<sup>2</sup> Now at Department of Psychology, Louisiana State University, United States.

<sup>3</sup> Now at Department of Psychology at Brown University, United States.

<sup>4</sup> Now at Department of Psychology at the University of Iowa, United States.

measure WM function. The assumption is that the psychometric instruments used in the clinical setting accurately depict the WM construct discussed by cognitive psychologists. However, this assumption has not been fully tested. Ackerman, Beier, and Boyle (2005) commented on this issue in more general terms by saying that, “many intelligence measures have been developed with substantially greater criterion-related validity, as opposed to construct validity” (p. 31). The present study addressed these concerns by providing a systematic evaluation of the construct and criterion-related validity of various WM measures.

### 1. Laboratory assessment of WM

Laboratory studies often utilize tasks, such as the operation span task (Osplan; Turner & Engle, 1989), reading span task (Daneman & Carpenter, 1980), or the n-back task (Dobbs & Rule, 1989) to assess WM function. Complex span tasks, such as Osplan and reading span, require participants to retain a list of items (storage component), while simultaneously engaging in a secondary activity, such as solving math problems (processing component). It is assumed that the central executive distributes attentional resources to the memory system to enable an individual to meet the complex demands of the task. Storage and processing tasks were specifically designed to support the theoretical assumptions held about how the WM system operates. Furthermore, these tasks have been repeatedly shown to be reliable measures of WM that demonstrate excellent construct and criterion-related validity (for a list of the many higher order cognitive tasks that correlate with WM, see Conway et al., 2005, p.777).

The lag, or n-back, task is less commonly used to measure WM function in the laboratory setting; however, the demands inherent within this task make it a potentially good measurement tool. It was originally developed by Kirschner (1958) to examine general retrieval processes. More recently, researchers have used this task to examine WM function in neurological settings with brain injury patients (Cohen et al., 1994), in the aging population (Dobbs & Rule, 1989; Kwong See & Ryan, 1995), and to study focal attention (McElree, 2001). In the n-back task, participants are presented and asked to recall or recognize an item that fell in a particular serial position in a presented list. This task does not contain some of the features present in traditional WM measures (e.g., secondary processing task); however, several of its features do reflect important aspects of the WM construct. The items being presented must be actively maintained for later recall, while controlled attention is used to guide the retrieval process in meeting task demands (identifying the item located in a particular position in the list). The absence of a secondary processing component intermittent throughout the test trials is an important difference that likely leads to different strategies being used to perform these tasks. More specifically, the n-back task may not provide a clear indication of the capacity of WM, rather a person's ability to efficiently update the contents of WM to better maintain current task goals.

Recent studies have offered conflicting results on the utility of the n-back task as a valid measure of WM (Kane, Conway, Miura, & Colflesh, 2007; Shelton, Metzger, & Elliott, 2007). One key difference between these studies is the version of the task used; that is, participants were either

asked to perform recall or recognition. For example, Shelton et al. (2007) used a recall version of the task, and observed strong relationships between n-back and Osplan performance. Kane et al. (2007), on the other hand, used a recognition version of the task and did not find strong relationships between n-back performance and performance on traditional storage and processing measures. It is clear that further evaluation of this task is needed to support its usefulness as a measure of WM. This is one sub-goal of the present research.

### 2. Clinical assessment of WM

WM is important clinically as it is 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, Goldman, & Weinberger, 1995; Goldman-Rakic, 1994). Efficacy for therapeutic interventions in these groups is typically demonstrated by measured improvements in executive functions such as WM. Additionally, executive dysfunction has been shown to 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). Currently, it is unclear whether clinical tests of WM are assessing the same WM construct discussed in the experimental cognitive literature.

The WAIS-III (Wechsler, 1997a) is the most commonly used measure of intelligence in clinical settings (Camara, Nathan, & Puente, 2000; Heijden & Donders, 2003; Rabin, Barr, & Burton, 2005). This test generates four summary indices, one of which is the Working Memory Index. The WM Index is derived from the following WAIS-III subtests: Digit Span (Dspan), Arithmetic, and Letter/Number Sequencing (LNS). Performance on the Dspan subtest reflects a combined measure of accuracy in the forward and backward conditions. Although there have been some questions raised regarding this practice (see Reynolds, 1997), the combined score was used in the current research as this is common clinical methodology. It should also be pointed out that the Dspan task is considered a simple span test because it only contains a storage component, in contrast to complex span measures that also contain a secondary processing requirement.

The Arithmetic portion of the WM index consists of word problems read aloud to participants with increasing levels of difficulty with each new problem. Difficulty level is determined by the amount of information that has to be held in memory in order to successfully complete the problem. This task does place demands on the WM system; however, other factors likely contribute to accuracy on this task, such as mathematical efficiency (Stearns, Dunham, McIntosh, & Dean, 2004). Math anxiety has also been linked to impaired performance on WM tasks that contain a mathematical element, such as computation span (Ashcraft & Kirk, 2001).

In the LNS task, the experimenter reads mixed lists of digits and letters aloud to the participants and they are asked to recall this list in correct numeric and alphabetic order. This task involves additional processing requirements similar to that of traditional WM tasks. Haut, Kuwabara, Leach, and Arias (2000) examined LNS performance in a PET study and

concluded that the brain regions activated by the task were consistent with known patterns of activation associated with other WM measures. This neuroimaging evidence lends support to the potential benefits of using the LNS task as a measure of WM, but additional empirical support is needed.

The WMS-III (Wechsler, 1997b) is one of the most commonly used memory assessment tools in the clinical setting (Rabin et al., 2005). It also has a WM Index, and it includes the Spatial Span (Sspan) and the LNS subtests. The Sspan was designed to be a visual analogue to the WAIS-III Dspan subtest. In the Sspan task, the experimenter points to a series of raised 3-dimensional squares, and the participant has to repeat this pointing sequence. Similar to Dspan, this occurs in both a forward and backward order and the Sspan score reflects performance in the forward and backward conditions combined.

Some clinicians have raised concerns regarding the construct validity of the subtests that comprise the WM indices. For example, Stearns et al. (2004) investigated the relationship between WM and ADHD in adults using the WM Indices from the WAIS-III and the WMS-III. The results did not reveal significant correlations between WM and self-reported ADHD symptoms despite the fact that WM impairment is commonly observed in the disorder. The authors questioned the validity of the clinical WM tasks from the test batteries, stating “it is difficult to defend the inclusion of tasks that require the examinee to simply repeat verbal or visual stimuli (such as Digit Span or Spatial Span forward) into the calculation of the WAIS-III and WMS-III Working Memory Indexes.” (p.283).

### 3. Present study

The primary goal of the present research was to provide a bridge between experimental methods of WM assessment used in the laboratory and the psychometric tools used in applied settings. Laboratory psychologists have successfully demonstrated the validity and reliability of several key tasks (Conway et al., 2005), and this knowledge could be utilized by clinicians to promote more efficient measurement of WM function. It is becoming increasingly clear that an effective way to better understand the cognitive functioning associated with disorders, such as schizophrenia, is to implement basic experimental methods into the clinical setting (Carter, 2005; Nuechterlein, Pashler, & Subotnik, 2006). The present research examined how the laboratory and clinical tests of WM related to one another and which subset of these tests provided the best measurement of the WM construct.

In general, we predicted that the laboratory and clinical WM tests would be highly correlated with one another; however, certain clinical subtests were believed to be more similar to the laboratory tests. For example, the demands present within the LNS task appeared to be most comparable to those present within the laboratory tests. The processing requirements of re-ordering the sequence of letters and numbers were similar to the processing demands present in many laboratory tests.

There were particular concerns about the use of the Arithmetic subtest as a measure of WM. The previously discussed concerns regarding the mathematical efficiency component of the task supported the prediction that performance on

this subtest would share relatively low correlations with the laboratory WM measures. Furthermore, in contrast to all of the laboratory measures, there was less reliance on serial order information in the Arithmetic subtest.

The expectation for how Dspan and Sspan performance would relate to the laboratory WM tests was less straightforward. Research has demonstrated that in some instances simple span tests (e.g., Dspan and Sspan) were quite comparable to traditional laboratory measures of WM (Colom, Rebollo, Abad, & Shih, 2006), while other studies suggested that simple span tests do not provide the clearest measurement of WM capacity (Engle, Tuholski, Laughlin, & Conway, 1999; Kane et al., 2004). Additionally, some evidence suggests that the Sspan task performs differently than the Dspan task in brain injured populations (Wilde & Strauss, 2002), leaving uncertainty surrounding its construct validity. One of the arguments against using simple span tests as measures of WM is that performance relies heavily on a storage component, and does not demand the level of attentional control required by the inclusion of a processing component. Given the inconsistent findings associated with simple span tests, we predicted that Dspan and Sspan performance would be highly related to the laboratory tests, but to a lesser degree than the LNS test.

General fluid intelligence was used as a criterion construct in the present study to further discriminate between the various WM tasks under consideration. General fluid intelligence (gF) is considered a form of intelligence that allows people to think and reason abstractly in novel situations (Cattell, 1987). Performance on cognitive tests can be used to help clinicians make predictions about a person's ability to function in other areas; thus, a criterion construct constituting higher-order cognitive abilities was a useful tool for examining the relative predictive power of various WM tests. Further support for the use of gF as a criterion construct stems from the considerable body of evidence that demonstrates a sizeable relationship between WM and gF (Ackerman, Beier, & Boyle, 2002; Conway, Cowan, Bunting, Theriault, & Minkoff, 2002; Engle et al., 1999; Kane et al., 2004), and WM and reasoning ability (Kyllonen & Christal, 1990).

Taken together, it was predicted that the three laboratory tests, along with the LNS subtest of the WAIS-III and WMS-III would provide the best measurement of the WM construct. Specifically, it was predicted that a model depicting a hybrid latent WM construct comprising these four tests would provide the best fit for the data. Furthermore, we predicted that the hybrid WM construct would be the best predictor of fluid abilities.

## 4. Method

### 4.1. Participants

One hundred and seventy-four participants (age  $M = 20.55$ ,  $SD = 3.74$ ; 43 males) from undergraduate psychology classes at Louisiana State University were retained in the final sample and were given course credit for their participation. Through extensive neuropsychological research with this population over the last 25 years, their psychometric properties have been well established. Based on the outcome of this research with the LSU student body it is anticipated that no

more than 5% of this population would include clinically-relevant individuals (Gouvier, Cubic, Jones, Brantley, & Cutlip, 1992; Gouvier, Uddo-Crane, & Brown, 1988). We also examined the full scale IQ of the sample ( $M = 110.38$ ,  $SD = 11.07$ , range 87–142) to determine if using undergraduate students or having a majority female sample could have resulted in a limited range. The current sample of undergraduate participants produced a range of FSIQ that was similar to other published work using versions of the WAIS (Johnson & Bouchard, 2007). Additionally, research has shown that gender differences are not statistically significant in either clinical or laboratory measures of WM (Robert & Savoie, 2006; Wechsler, 1997a). Thus, the range of this sample of undergraduates was considered appropriate.

Participants were excluded from the final sample for the following reasons: failing to attend both experimental sessions, having a hearing impairment, speaking a native language other than English, performing below 80% on the math portion of the Ospan task, being in a session in which the experimenter inadvertently did not administer one required test or not completing all of the tasks administered within the two experimental sessions. The purpose of a processing accuracy criterion in the Ospan task was to identify participants who may have focused too narrowly on the recall portion of the task, at the expense of the processing component (e.g., doing poorly on the math portion to enhance the memory component).

## 4.2. Laboratory tasks

All of the laboratory tasks were completed on individual computers.

### 4.2.1. Operation span (Ospan)

The task used in the present study was an automated version of the Ospan task, developed by Unsworth, Heitz, Schrock, and Engle (2005). Participants were instructed to mentally solve basic incomplete math equations (e.g.,  $6 * 3 = ?$ ). After clicking the mouse to indicate they had solved the problem, a number appeared on the screen and participants had to decide if it was the correct answer to the equation. Following this choice, a letter appeared on the screen, which they were told to remember. At the conclusion of each trial, participants were instructed to recall the letters from that trial in the correct order by selecting them from a screen containing all the letters used in the experiment (F, H, J, K, L, N, P, Q, R, S, T, and Y). The list lengths ranged from 3 to 7 with three trials presented at each list length, and the order of set presentation was random for each participant.

Three practice sessions were completed in this task. The first session provided training on the letter recall portion, followed by training on the math completion. The final practice session mirrored the test trials by providing training on the combined storage and processing segments. Mean Ospan Score was used as the performance index for this task in order to provide a comparable index between the laboratory-based tasks. This index was a weighted summary (more letters led to more points being given for a trial) of performance that reflected the number of letters recalled in perfectly recalled trials.

### 4.2.2. Listening span (Lspan)

The version of the Lspan used in the present study was taken from Cowan et al. (2005). Previous research suggests that Lspan

is highly related to other laboratory WM tasks, such as counting span (Cowan et al., 2003). It was similar to the reading span task, except that the sentences were read aloud to the participants. The presence of an auditory component in Lspan is similar to the auditory administration of the clinical WM measures, which is one of the primary reasons it was chosen. In the task, participants heard sentences through headphones and used the keyboard to make true/false judgments. For example, one practice sentence was, "In winter it is very hot." Following each test block, participants used the keyboard to recall the last word from each sentence in the correct serial order. The full task included seven blocks with three lists in each. List length was 2 in block 1 and increased by 1 for each successive block. Administration was terminated if participants missed all three sets within one block. A weighted performance index was also used for this task, referred to as Lspan Score, and it reflected the number of words recalled in perfectly recalled trials.

### 4.2.3. Modified lag task (Nback)

The task used in the present study was developed by Shelton et al. (2007). In this task, participants saw lists of 4 or 6 words and were asked to recall either the last, next to last (1-back), second to last (2-back), or third to last word (3-back) presented in the list. Participants typed their responses. The list lengths were presented in a mixed order to prevent participants from predicting the end of the list. The serial position requested also varied randomly from trial to trial. Words were either four or five letters long and all had high levels of familiarity, meaning, and frequency according to the MRC Psycholinguistic Database. Participants completed 5 trials at each list length for the four N-back positions, making a total of 40 trials completed in this task. The performance index used, LagScore, was a weighted summary of correctly recalled trials, with the number of points assigned to trials increasing as the lag conditions increased with difficulty. The formula used for this index was as follows:

$$\text{LagScore} = (\text{Lag } 0\# \text{ correct} * 1) + (\text{Lag } 1\# \text{ correct} * 2) \\ + (\text{Lag } 2\# \text{ correct} * 3) + (\text{Lag } 3\# \text{ correct} * 4)$$

## 4.3. Clinical WM tasks

All clinical tests were administered following the instructions present in the WAIS-III and WMS-III manuals. An experimenter recorded answers on the designated answer sheets.

### 4.3.1. Digit span (Dspan)

For Digits Forward, the experimenter read strings of digits aloud to the participants at the approximate rate of one number a second. The participants were asked to repeat them back in the correct order. The Dspan Forward condition consisted of 8 blocks of 2 trials at each list length. The number of digits in the initial block was 2 and increased by 1 in each successive block. Administration was terminated if participants incorrectly recalled the digits for both trials within the same block. For Dspan Backward, the same procedure was used, except that participants had to repeat the digits in reverse order. In addition, 7 as opposed to 8 blocks were presented. Following all trials, responses were recorded by the experimenter as correct (if all digits were recalled in the correct serial order) or

incorrect. Scores reflected a sum of the total number of correct trials from the Forward and Backward conditions.

#### 4.3.2. Arithmetic

Participants were read aloud arithmetic problems and asked to respond with the correct answer. They were not allowed to write any information. If requested, the experimenter could repeat the entire problem once. Both the level of difficulty and time allotted to solve each problem increased through the task. Answers were only considered correct if they were given within the time limits allotted, and administration was terminated if four consecutive incorrect answers were given. The problems required a variety of mathematical operations (e.g. addition) and concepts (e.g. percentages). Accuracy on this task reflected the total number of correct responses out of the possible 20.

#### 4.3.3. Letter number sequencing (LNS)

In this task, the experimenter read a series of numbers and letters aloud to the participants at the approximate rate of one item per second. Participants were asked to recall each list with the numbers in numerical order followed by the letters in alphabetical order. The complete test consisted of eight blocks with three trials in each. List length was three for the first block and increased by one for each successive block. Administration was terminated if participants missed all three trials within a block. The total number of correct trials was summed to create a LNS score.

#### 4.4. WMS-III WM tasks

The LNS is a component of both the WAIS and WMS WM Indices. Dspan is an optional subtest presented in the WMS, but scores from this subtest are not incorporated into its WM Index. Dspan and LNS were typically administered in the WAIS-III in the present study; thus, the Sspan subtest was the only WM task completed during the WMS administration.

##### 4.4.1. Spatial span (*Sspan*)

This task is a variation of Corsi's block-tapping test (Lezak, Howieson, & Loring, 2004). Participants were shown a white board with raised, equally-sized blue blocks. In the Forward condition, the experimenter tapped the blue blocks in a sequence and participants attempted to repeat the correct order. In the Backwards condition, the test procedure was the same, except that participants had to repeat the tapped sequence in the reverse order. Both conditions were composed of eight sets of items with two trials in each set. The sequence length was two for trial 1 and increased by one for each successive set. Administration was terminated if participants missed both trials of any set. Scores reflected a sum of the total number of correct trials from the Forward and Backward conditions.

#### 4.5. Fluid intelligence tasks

##### 4.5.1. Raven's advanced progressive matrices (RAPM)

This version of the task included three blocks of 12 items each (Raven, Raven, & Court, 1998). For each item, a portion of a geometric pattern was missing and participants were instructed to choose the response that correctly completed the pattern. Six response options were given for items in Set 1, and 8 response options were given for items in Set 2. The items

increased in difficulty across each block and 5 min were allotted to solve each block. This task was computer administered and participants advanced by making responses with the mouse. Individual scores on RAPM represented the total number of items they responded to correctly across the three blocks.

##### 4.5.2. Matrix Reasoning

Matrix Reasoning is a WAIS-III subtest similar to RAPM. Participants were asked to choose which of five choices best completed a geometric sequence. The complete test contained 26 items. Administration was terminated if participants were incorrect on either four consecutive items or four out of five consecutive items. The number of correct responses was summed to create the Matrix Reasoning score.

##### 4.5.3. Block design

This was also a task administered from the WAIS-III. Participants were given red and white blocks and asked to make specific designs. In the first trial, the experimenter made a design and asked the participant to replicate it. If they were successful, the experimenter then modeled a design from the stimulus book and asked participants to do the same. On these trials, two points were received if they were successful on the first attempt, one point if they were successful on a second attempt that followed the experimenter again making the design, and no points if they twice failed to complete the design. For all of the 8 following items, the experimenter did not model the design. Participants attempted the designs found in the stimulus book on their own. If they made an error or failed to complete the design within the time limit, they received no points. If they successfully completed the design within the time limit, they received between 4 and 7 points, depending on how quickly they finished. Participants were given 60 s to complete the 4-block designs and 120 s to complete the 9-block designs. The total number of points earned determined the Block Design score.

#### 4.6. Procedure

Data for this experiment were collected during two sessions that took place approximately one week apart, in order to minimize fatigue effects. Two procedure orders were utilized. In the first procedure order, participants completed the entire WAIS-III and WMS-III during the first session, and completed the Ospan, RAPM, n-back and Lspan tasks in four randomly-assigned orderings in the second session. In the second procedure order, participants completed the WAIS-III, Ospan, and RAPM during the first session, and the WMS-III, n-back, and Lspan during the second session. Preliminary analyses were conducted with procedure order as a between-subjects variable, and no significant differences emerged for any of the tasks. The data were collapsed across this variable in all remaining analyses.

Informed consent always took place at the beginning of the first session and debriefing during the second session. The cognitive tests were always completed by the participants at individual computer stations and the clinical tasks were administered by a trained experimenter to each participant individually. All of the subtests present in the WAIS-III and WMS-III were completed by the participants, except for the tests that were designated as optional. It was not necessary to administer the optional psychometric tests given that all of

**Table 1**  
Descriptive statistics and bivariate correlations

	Mean	SD	1	2	3	4	5	6	7	8	9	10
<i>Clinical measures</i>												
1. Arithmetic	13.97	3.05	.75									
2. Spatial span	17.75	2.99	.22 **	.74								
3. Digit span	19.55	3.76	.31 **	.33 **	.81							
4. Letter number	12.56	2.67	.45 **	.43 **	.60 **	.74						
<i>Lab measures</i>												
5. Ospan score	44.15	15.54	.23 **	.35 **	.54 **	.41 **	.73					
6. Lspan score	29.06	10.93	.34 **	.29 **	.43 **	.45 **	.55 **	.74				
7. Lag score	54.23	16.36	.41 **	.33 **	.48 **	.44 **	.38 **	.45 **	.79			
<i>gF measures</i>												
8. RAPM	25.50	4.04	.34 **	.25 **	.16 *	.29 **	.29 **	.30 **	.40 **	.75		
9. Block design	45.79	10.94	.41 **	.43 **	.27 **	.37 **	.29 **	.32 **	.38 **	.41 **	.75	
10. Matrix Reasoning	20.17	2.51	.19 *	.30 **	.18 *	.21 **	.12	.23 **	.36 **	.37 **	.33 **	.65

Note:  $N = 174$ . Raw scores were used in the Clinical measures; weighted summary scores were used in the Lab measures.

\*\* indicates  $p < .01$ , \* indicates  $p < .05$ . Italicized scores on the diagonal represent Cronbach's Alpha measure of internal consistency.

the required tests were administered; therefore, the required tests could be used to calculate the separate index scores.

## 5. Results

All of the subtests of the WAIS-III and WMS-III were scored by project staff and then re-scored independently by an additional trained psychometrician to ensure that the data had been scored correctly. This was an important step, given the potential for human error that is present when hand-scoring techniques are used. In addition, the data from each subtest were screened for outliers using a criterion of 3 SD. This led to the identification of a total of 5 scores (four on Matrix Reasoning and one on the Span total) which were replaced with values equivalent to the mean  $\pm$  3 SD. As described above, a weighted summary of performance was used for each of the laboratory-based WM measures in order to establish a comparable measurement scale between tasks (i.e., OspanScore, LagScore, and LspanScore). The raw scores from the clinical measures were used to represent performance on all of the WM subtests of the WAIS-III and WMS-III. Given the narrow age range of the sample, the raw scores were used as opposed to the scaled scores. The use of raw scores also facilitated comparisons with the laboratory-based WM measures, for which scaled scores were not available. Performance on the gF measures taken from the WAIS-III (Block Design and Matrix Reasoning) reflected the raw scores as well, and the index used for RAPM reflected the total number of successfully completed matrices. The reliability estimates for all of the measures were based on item-level analyses using Cronbach's measure of internal consistency, and these values are presented in Table 1. All performance indices used in the primary analyses reached acceptable levels of reliability (.65–.81).

Bivariate correlations were calculated to examine the relationships among the various WM tasks, as well as their relationship with measures of gF (see Table 1). Confirmatory factor analysis (CFA) was used to address the primary question at hand: Are the WM tests used in the laboratory and applied settings measuring the same thing? Finally, structural equation modeling (SEM) was used to assess the best prediction of higher order cognition.

### 5.1. Preliminary analyses

We examined the individual variables for skewness and kurtosis, and these all fell within acceptable ranges (skewness  $< 2$ , kurtosis  $< 4$ ; Kline, 2005). As expected, the correlational analyses revealed that all of the WM measures shared a significant positive relationship. In general, the pattern of correlations between the laboratory and clinical task sets were similar to those observed within task sets; however, there was a trend for the Span and Arithmetic subtests of the WMS-III and WAIS-III, respectively, to share lower correlations ( $r_s = .23$ – $.34$ ) with the traditional laboratory WM measures (Ospan and Lspan).

### 5.2. CFAs and SEMs

Covariance matrices were fit using the maximum-likelihood procedure of AMOS 7.0. Several fit indices were used to evaluate these models including the chi-square goodness of fit test ( $\chi^2$ ; Bollen, 1989), comparative fit index (CFI; Bentler, 1990), root-mean square error of approximation (RMSEA; Brown & Cudeck, 1993) and standard root mean residual (SRMR; Hu & Bentler, 1998). To establish the existence of a “good-fitting” model it is most desired to have a non-significant  $\chi^2$ ; however,  $\chi^2$  is sensitive to sample size. The CFI and RMSEA are less sensitive to sampling characteristics and take degrees of freedom into account. Following published guidelines, a CFI value close to .95, a RMSEA value below .08, and an SRMR value close to .08 are considered indicative of good model fit (Hu & Bentler, 1999; Vandenberg & Lance, 2000). The SRMR is the standardized root mean residual and reflects the difference between the observed and predicted covariance matrices. In this index, lower values are indicative of a better fit (Hu & Bentler, 1998).

### 5.3. Are WM tests used in the laboratory and applied settings measuring the same thing?

The primary goal of the present study was to test the assumption that the psychometric indices that have been developed to assess WM function are measuring the same cognitive processes assessed by laboratory-based WM tests.

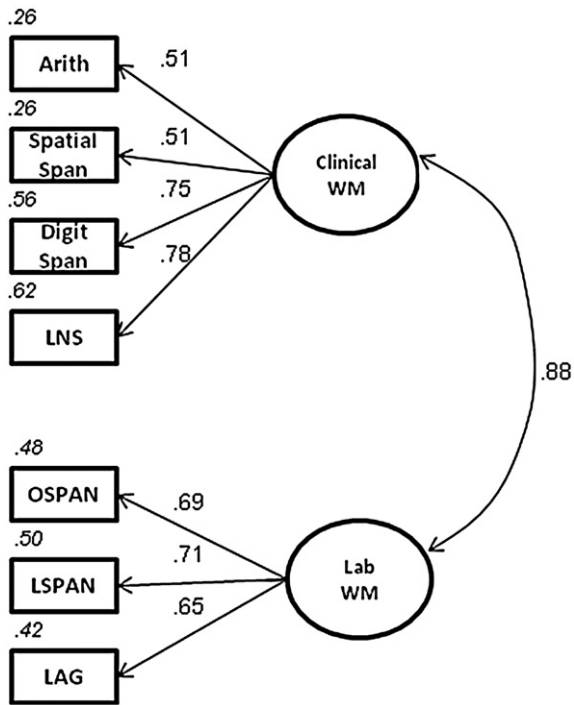
**Table 2**  
Confirmatory factor analysis of working memory constructs

Models	$\chi^2$	df	Fit indexes		
			CFI	RMSEA	SRMR
One factor model	39.43**	14	.93	.10	.05
Correlated two factor model	33.20**	13	.94	.10	.05
Hybrid one factor model	3.48	2	.99	.07	.03

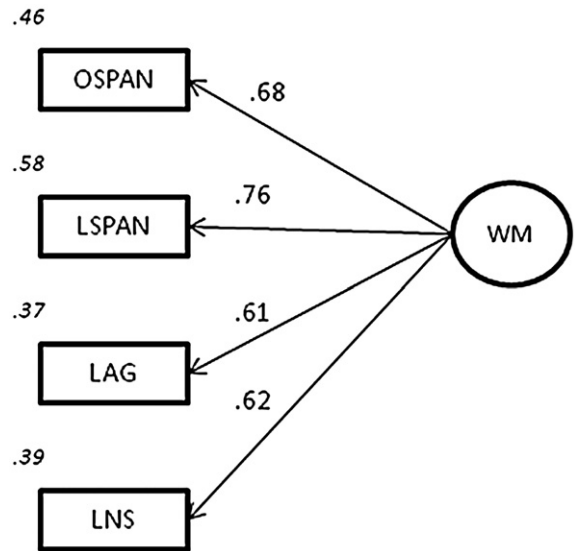
Note:CFI= comparative fit index; RMSEA= root mean square error of approximation, SRMR= squared root mean residual. \*\* $p < .01$ .

As discussed in the introduction, cognitive and clinical psychologists define the construct of WM in similar ways, and the WM subtests of the WAIS-III and WMS-III are believed to reflect this similar orientation. Two CFA's were run initially to assess this basic assumption:

- *One-factor model*: This model tested the fit of a single overall factor that contained all of the laboratory and clinical tests of WM. If a one factor model demonstrated good fit, it could be argued that the different measures



**Fig. 1.** Latent variable model representing the correlated twofactor model of working memory. In all figures, the circles represent latent constructs, and the squares represent the observed variables. The path coefficients appearing on the arrows from the latent construct to the observed variables are standardized values, comparable to standardized regression weights and reflect the factor loadings of the observed variables (Kline, 2005). All paths are significant at  $p < .05$ . Italicized numbers next to the observed variables represent the squared multiple correlations. The double-headed curved arrow represents the correlation between the two WM constructs. The three laboratory measures were Ospan = Operation Span Score; Lspan = Listening Span Score; Lag = N-back or Lag Score; The clinical measures from the WAIS-III and WMS-III were LNS = Letter Number Sequencing raw score; Arith = Arithmetic raw scores; Spatial Span = raw scores from the total of the forward and backward scales of the spatial span task; Digit span = raw scores from the total of the forward and backward scales of the digit span task.



**Fig. 2.** Latent variable model representing the hybrid model of the working memory construct. Numbers on the arrows between the latent variable and the observed variables reflect the loadings for each task; all paths are significant at  $p < .05$ . Italicized numbers next to the observed variables represent the squared multiple correlations. The three laboratory measures were Ospan = Operation Span Score; Lspan = Listening Span Score; Lag = N-back or Lag Score; The clinical measure from the WAIS-III was LNS = Letter Number Sequencing raw score.

were all related to the same underlying construct (i.e., WM).

- *Correlated two factors model*: This model tested whether the two WM latent factors (laboratory and clinical) were bidimensional, but related.

The results for these two models are reported in Table 2. The one factor model only moderately fit the data; only the SRMR achieved the desired level of fit. The two factor correlated model fit the data significantly better than the one factor model [ $\chi^2$ -difference(1) = 6.23,  $p < .05$ ]; however its overall fit to the data was still relatively poor. These results suggest that the laboratory and psychometric tests of WM were not measuring two separate, unrelated constructs; however, neither the unidimensional model nor the correlated two factor model provided a good fit to the data. Additionally, the correlation between the two WM constructs was very high (see Fig. 1). It should be noted that the Sspan and Arithmetic subtests had the lowest factor loadings for the clinical WM construct (both .51).

Given the theoretical stance outlined in the introduction, it was predicted *a priori* that the LNS subtest from the WAIS-III and WMS-III would be the best measure of WM from the clinical subtests. As indicated by the poor fit in the previous models, it was clear that the best possible assessment of WM had not yet been attained. Thus, the raw scores from the LNS subtest and the laboratory measures of WM were added to a hybrid one factor model, depicted in Fig. 2. As reported in Table 2, this model demonstrated acceptable fit, importantly including a non-significant  $\chi^2$ , which none of the other models included. This approach demonstrated that the laboratory measures (Ospan, Lspan, and n-back), coupled with the LNS task, was the best

combination of WM measures that were tested in this sample.<sup>5</sup> Furthermore, these tasks were the most highly related to one another.

#### 5.4. Do the WM measures used in the laboratory and applied setting differentially predict gF?

The present investigation used gF as a criterion construct, given the substantial body of research that suggests one way to establish the construct validity of a WM test is to demonstrate that performance on a given measure will share a significant relationship with fluid abilities (Conway et al., 2002; Engle et al., 1999; Kane et al., 2004). Another reason for establishing a form of criterion-related validity in the present investigation is the importance of prediction in clinical settings, and that assessing a clients' WM function could lead to predictions being made about their ability to perform other complex cognitive activities, such as performing financial responsibilities or weighing medical decisions.

Given the sizable relationship that was observed between the laboratory and clinical WM construct in the previous analyses, it was more parsimonious to form a unitary WM construct to predict our gF latent variable. As a reminder, our gF construct was composed of RAPM, Block Assembly, and Matrix Reasoning. Our theoretical views led to the prediction that the three laboratory tests, along with the LNS test, would provide the most coherent measurement of WM, and the results from the CFA have supported this prediction. Thus, we tested two models in the following SEM analyses.

The first model tested depicted the unitary WM latent construct as a predictor of a gF latent construct. Fit information is reported in Table 3. Only one of the three fit indices (SRMR) indicated that this model provided a good fit for the data. The unitary WM construct accounted for 53% of the variance in gF performance (i.e., squared multiple correlation = .53). When the four-variable hybrid model was tested (see Fig. 3), all four fit indices demonstrated adequate fit, including a non-significant  $\chi^2$  statistic. Furthermore, this model accounted for approximately the same amount of variance, explaining 54% of the variance in gF performance, with three fewer observed variables. These results demonstrate that a construct consisting of the laboratory WM measures along with the LNS test, provided optimal predictive utility for fluid abilities.

#### 5.5. Is the N-back task a valid measure of WM?

Another contribution of the present research was the direct comparison between a recall version of the n-back task and other tests of WM. Conway et al. (2005) conducted an extensive review of cognitive-experimental WM measures, and they stressed the need for direct comparisons to be made between the n-back task and other WM measures. Several

<sup>5</sup> Given the relatively high bivariate correlations between Digit Span and the laboratory measures of WM, an additional CFA was conducted in which the three laboratory measures, LNS, and DS were included. This analysis did not result in a good fit, as the  $\chi^2 = 18.74$  was significant. Furthermore, of the other fit statistics, the CFI = .95 was at a borderline level of acceptability, the RMSEA = .13 exceeded the recommended .08, and the SRMR = .04 was the only statistic in the acceptable range.

**Table 3**  
Predictive models of gF

Models	$\chi^2$	df	Fit indexes			Squared multiple correlation
			CFI	RMSEA	SRMR	
<i>Unitary WM → gF</i>						
Structural model	88.49**	34	.89	.10	.06	.53
<i>Hybrid WM → gF</i>						
Structural model	19.44	13	.98	.05	.04	.54

Note: CFI = comparative fit index; RMSEA = root mean square error of approximation, SRMR = squared root mean residual. \*\* $p < .01$ .

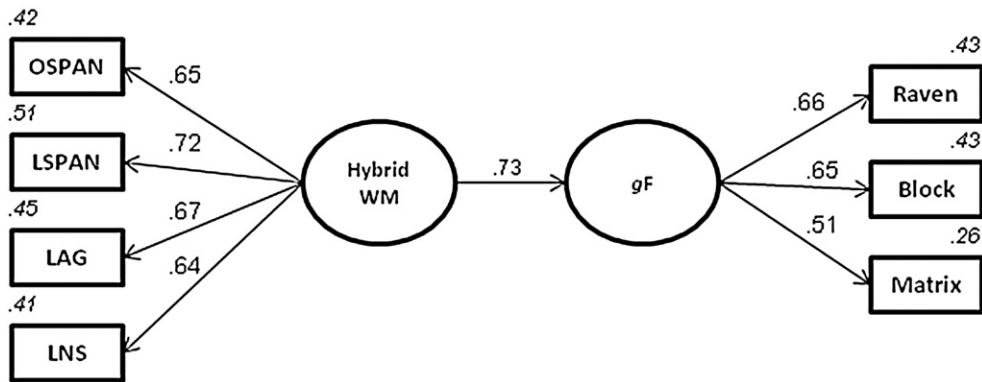
outcomes of the present research suggested that the version of the n-back task used in this study was a valid measure of WM. First, performance on the n-back task was moderately correlated with all other measures of WM ( $r$ s range of .33–.45). The lowest correlation was shared with the Sspan subtest of the WMS-III, which is consistent with the previously discussed problems associated with this test. Second, the results of the CFA revealed that scores from the n-back task loaded well on a factor that included scores from other WM tests that have been consistently shown to be valid measures of WM (Ospan and Lspan; see Fig. 1). Consistent with previous research (Friedman et al., 2006; Gray et al., 2003), n-back performance was also significantly correlated to measures of fluid intelligence ( $r$ s range from .37–.40), and tended to correlate more strongly across the gF measures relative to the other WM tests used in the study. Overall, the recall version of the n-back task used in the present study proved to be a valid measure of WM function.

## 6. General discussion

The findings from the present study offered insight into the relationship between tests used to assess working memory function in the laboratory and those used in an applied clinical setting. In general, the WM subtests of the WAIS-III and WMS-III clinical batteries were highly related to the laboratory WM measures, suggesting that performance on these tests was essentially tapping resources from the same psychological construct. There were, however, a few exceptions to this general pattern.

In line with our predictions, the WM tests that portrayed the lowest levels of construct validity were the Arithmetic and Sspan subtests of the WAIS-III and WMS-III WM indices, respectively. The LNS and Dspan tasks were the most highly related of the psychometric tests to laboratory measures of WM. However, inspection of all of the fit indices used in the present study indicated that a model containing the laboratory measures (n-back, Lspan, and Ospan) and the LNS task provided a good fit for these data, suggesting that this combination of tasks allowed for the best measurement of WM function. Furthermore, the best prediction of individual differences in fluid intelligence was accomplished using a hybrid model that depicted a latent construct comprising scores from the LNS and laboratory WM tests. Taken together, these findings suggested that while the laboratory and psychometric indices of WM may be measuring similar cognitive processes, there were subtle differences that should be considered, including their predictive utility. These differences and their practical implications will now be discussed.





**Fig. 3.** The path between the latent variables represents the strength of the relationship between the constructs. Italicized numbers next to the observed variables represent the squared multiple correlations. A) Prediction of gF from the variables representing the hybrid model of the WM construct. Ospan = Operation Span Score; Lspan = Listening Span Score; Lag = N-back or Lag Score; LNS = Letter Number Sequencing raw score; Raven = total score from the Raven's Advanced Progressive Matrices; Block = raw score from the Block Design task; Matrix = raw score from the Matrix Reasoning task.

### 6.1. Predictive utility of WM tests

As previously discussed, WM function has consistently been shown to be a good predictor of a variety of higher-order cognitive abilities. Uncertainty still exists regarding the nature of this relationship; however, strong theoretical arguments have been made favoring the role of executive attention processes in this relationship (Engle et al., 1999; Kane, Hambrick, & Conway, 2005; Kane et al., 2004; Unsworth & Engle, 2006, 2007; for alternative views see Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008; Oberauer, Süß, Wilhelm, & Wittman, 2008). Recent neuroimaging evidence supports this claim by demonstrating that the WM–gF relationship is mediated by individual differences in neural activity from brain regions associated with controlled attention processes (Gray et al., 2003).

The importance of a general attention factor in WM function, as well as its relationship with complex cognition, has recently been discussed in terms of retrieval from secondary memory (SM). Unsworth and Engle (2005) argued that the processing component present in complex span tasks forces the list items from the storage component to be displaced into SM. Controlled attention must be used to retrieve these items by aiding the search process. They further predicted that the relationship between WM and gF is driven by the need to retrieve task-relevant information, and the existence of a relationship between simple span tasks (e.g., digit span, spatial span) and gF will be present to the extent that these retrieval processes are emphasized (i.e., the longer list lengths of simple span tasks). Unsworth and Engle tested this prediction by creating estimates of WM (complex span performance), primary memory (short list lengths of simple span tests), and SM (longer list lengths of simple span tests) to examine their relative prediction of variation in fluid abilities. The results revealed that only the estimates of SM and WM predicted a significant amount of unique and shared variance in gF; furthermore, the complex span–gF relationship remained stable across list lengths, while the simple span–gF relationship strengthened as list length increased. These findings provided support for their claim that individual differences in WM will be good predictors of fluid abilities to the extent that the individual memory tests emphasize a controlled search of SM.

The present findings could have been influenced by the degree of emphasis placed on controlled attention processes in the various WM tests that were administered. Complex span tests, such as the Ospan and Lspan, emphasized these processes by asking respondents to maintain activation of the to-be-remembered items while simultaneously completing the processing task. Furthermore, engagement in the processing task leads to stored items being displaced to SM, facilitating a controlled search of SM at the time of retrieval. The LNS task also includes a processing component by requiring respondents to manipulate the order of presented items. The key difference between the LNS and other complex span measures is that the processing task takes place on a secondary set of items in the latter case, while the former involves manipulation of the actual to-be-remembered items. The n-back task also emphasizes the need for a controlled search process by forcing participants to retrieve an item that fell in a specific position in the list. The Dspan and Sspan tasks, on the other hand, involve the simple storage of information without an additional processing demand. This does not imply that performance on these tasks is void of attention, but the extent to which controlled attention is emphasized in the task may be different. Performance on the longer list lengths of these tests would presumably yield better predictive power, but the way in which the data from the WAIS-III and WMS-III were collected (absolute accuracy scores for each list rather than item-level accuracy) does not allow for a direct examination of this possibility.

The problems associated with the Arithmetic subtest could also reflect the fact that performance is less reliant on general attention factors and more reliant on a specific skill set. For example, performance on the Arithmetic subtest requires a certain level of math competency in addition to a WM load. Furthermore, research has demonstrated that math anxiety can influence performance on WM tests that include a considerable mathematical component (Ashcraft & Kirk, 2001). Thus, individual scores on the Arithmetic subtest of the WAIS-III likely reflect a combination of math competency, math-related anxiety, and memory function. This promotes difficulty in interpreting why an individual performs in a certain way on this task, and creates problems for its construct validity as a measure of WM function.

Although the present data cannot fully speak to the nature of the predictive relationship between WM and higher-order cognitive function, it still offers insight into the specific WM tests that hold the most predictive power. This is particularly important for clinical evaluation where performance on cognitive tests could be used to predict how a patient will function in other areas. It is clear that certain memory tests are more sensitive to variation in other cognitive abilities, and simple evaluation of the correlational relationship between these tests will not necessarily speak to these subtle, but important, differences.

### 6.2. Practical implications

Studies have linked WM function with a variety of clinical diagnoses, including depression (Harvey et al., 2004), schizophrenia (Barch, 2003), dementia (Collette et al., 1999), and attention-deficit hyperactivity disorder (Pasini et al., 2007). The concept is even more relevant in the domain of clinical neuropsychology, as attention and executive function are among the most vulnerable neurocognitive abilities and are often impacted by both developmental changes in brain function and acquired brain injuries (Lezak et al., 2004; Strauss, Sherman, & Spreen, 2006). This clearly demonstrates one example of the overlap between cognitive and clinical disciplines, and the need for better communication between the respective fields, and the present study sought to address one aspect of this issue.

These findings have diagnostic significance for clinicians given certain WM tests have proven to be more sensitive to clinical disorders. For example, one study demonstrated that a laboratory test of WM (Ospan) was successful in discriminating between healthy older adults who had genetic markers for Alzheimer's disease and those who did not have these markers; however, clinical tests of memory were not able to discriminate between these two groups (Rosen et al., 2002). Another study examined the updating function of WM in depressed patients versus controls using the n-back task, and found that depressed patients experienced significant disruptions in their ability to update contents of WM (Harvey et al., 2004). The Dspan and Sspan were also administered in this study but failed to effectively discriminate depressed patients from normal controls. Such results support the potentially greater sensitivity of laboratory tests for detecting neurocognitive deficits that have the potential to develop into clinical disorders. The results of the present research provided insight into the construct validity of widely used clinical WM indices in a normal college population. Future research is needed to examine the same principles in a clinical population.

### 6.3. Conclusions

The translation of basic experimental research techniques to the applied setting is becoming an increasingly popular scientific endeavor, and the outcome of this empirical work has the potential to be quite useful for both researchers and clinicians. The present research took the first step towards examining this issue in the assessment of working memory function. The findings suggest that the WM tests used in the laboratory and applied setting are measuring the same general set of cognitive processes. There is a subset of these tests, however, that appear to provide the most valid assessment of WM ability. Specifically, the three laboratory-based tests (operation span, listening span, and n-back), along with the

Letter/Number-sequencing subtest of the WAIS-III and WMS-III represent the purest battery of tests for the WM construct. Furthermore, these four tests offer the best predictive capability of higher-order cognitive abilities in this college-student sample. When making the decision to use a particular set of memory tests one should consider several factors, including how well these tests predict performance in other areas and the other skills or cognitive processes that are being represented in the tasks. Future research is needed to better understand how effective these tests will be at detecting specific cognitive abnormalities in other populations. It is possible that certain tests will be differentially sensitive to specific clinical disorders, and this information could provide a powerful tool for diagnosis and treatment planning. It is hoped that studies such as this constructive collaboration will encourage additional joint efforts between experimental and clinical psychology to the enrichment of the field of psychology as a whole.

### Acknowledgements

This research was supported in part by a pre-doctoral fellowship awarded to Jill T. Shelton by the National Institute of Health. Portions of these data were presented at the 47th Annual Conference of the Psychonomic Society held in Houston, Texas in November, 2006. We would like to acknowledge Amanda Exner, Amineh Abbas, and John Holmes for their help with data collection. We would also like to thank Russell Matthews for his expertise in the statistical analyses and Russ Pella for his help with scoring on the psychometric tests.

### References

- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2002). Individual differences in working memory within a monological network of cognitive and perceptual speed abilities. *Journal of Experimental Psychology: General*, *131*, 567–589.
- Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *31*, 30–60.
- Ashcraft, M. H., & Kirk, E. P. (2001). The relationship between working memory, math anxiety and performance. *Journal of Experimental Psychology: General*, *130*, 224–237.
- Baddeley, A. D. (2007). *Working memory thought and action*. Oxford, England: Oxford University Press.
- Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*, 158–173.
- Barch, D. W. (2003). Cognition in schizophrenia: Does working memory work? *Current Directions in Psychological Science*, *12*, 146–150.
- Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, *107*, 238–246.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bonnot, V., & Croizet, J. (2007). Stereotype internalization and women's math performance: The role of interference in working memory. *Journal of Experimental Social Psychology*, *43*, 857–886.
- 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.
- Brown, M. N., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J.S. Long (Eds.), *Testing Structural Equation Models* (pp. 136–162). Newbury Park, CA: Sage.
- 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.
- Carter, C. S. (2005). Applying new approaches from cognitive neuroscience to enhance drug development for the treatment of impaired cognition in Schizophrenia. *Schizophrenia Bulletin*, *31*, 810–815.
- Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action*. Amsterdam: North Holland.

- Cohen, J., Forman, S. D., Braver, T. S., Casey, B. J., Servan-Schreiber, D., & Noll, D. C. (1994). Activation of prefrontal cortex in the non-spatial working memory task with functional MRI. *Human Brain Mapping*, 1, 293–304.
- Collette, F., Van der Linden, M., & Salmon, E. (1999). Phonological loop and central executive functioning in Alzheimer's disease. *Cortex*, 35(1), 57–72.
- Colom, R., Abad, F. J., Quiroga, M. A., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, 36, 584–606.
- Colom, R., Rebollo, I., Abad, F. J., & Shih, P. C. (2006). Complex span tasks, simple span tasks, and cognitive abilities: A re-analysis of key studies. *Memory and Cognition*, 34, 158–171.
- Conway, A. R. A., Cowan, N., Bunting, M. F., Theriault, D. J., & Minkoff, S. R. B. (2002). A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. *Intelligence*, 30, 163–183.
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: a methodological review and user's guide. *Psychological Bulletin and Review*, 12, 769–786.
- Cowan, N., Elliott, E. M., Saults, J. S., Morey, C. C., Mattox, S., & Hismjatlullina, A. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology*, 51, 42–100.
- Cowan, N., Towse, J. N., Hamilton, Z., Saults, J. S., Elliott, E. M., Lacey, J. F., Moreno, M. V., & Hitch, G. J. (2003). Children's working memory processes: A response-timing analysis. *Journal of Experimental Psychology: General*, 132, 113–132.
- Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Behavior*, 19, 450–466.
- Dobbs, A. R., & Rule, B. G. (1989). Adult age differences in working memory. *Psychology and Aging*, 4, 500–503.
- 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.
- Fleming, D., Goldberg, T. E., Goldman, J. M., & Weinberger, D. R. (1995). Verbal working memory dysfunction in schizophrenia: Use of a Brown-Petersen paradigm. *Psychiatry Research*, 56, 155–161.
- Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science*, 17, 172–179.
- Goldman-Rakic, P. S. (1994). Working memory dysfunction in schizophrenia. *Journal of Neuropsychiatry and Clinical Neurosciences*, 6, 348–357.
- Gouvier, W. D., Cubic, B., Jones, G., Brantley, P., & Cutlip, Q. (1992). Postconcussion symptoms and daily stress in normal and head injured college populations. *Archives of Clinical Neuropsychology*, 7, 193–211.
- Gouvier, W. D., Uddo-Crane, M., & Brown, L. M. (1988). Base rates of post-concussional symptoms. *Archives of Clinical Neuropsychology*, 3, 273–278.
- Gray, J. R., Chabris, C. F., & Braver, T. S. (2003). Neural mechanisms of general fluid intelligence. *Neuroscience*, 6, 316–322.
- Harvey, P. O., Le Bastard, G., Pochon, J. B., Levy, R., Allilaire, J. F., Dubois, B., et al. (2004). Executive functions and updating of the contents of working memory in unipolar depression. *Journal of Psychiatric Research*, 38, 567–576.
- Haut, M. W., Kuwabara, H., Leach, S., & Arias, R. G. (2000). Neural activation during performance of number-letter sequencing. *Applied Neuropsychology*, 7, 237–242.
- Heijden, P., & Donders, J. (2003). A confirmatory factor analysis of the WAIS-III in patients with traumatic brain-injury. *Journal of Clinical and Experimental Neuropsychology*, 25, 59–65.
- Hinkin, C. H., Castellon, S. A., Durvasula, R. S., Hardy, D. J., Lam, M. N., Mason, K. I., et al. (2002). Medication adherence among HIV + adults: Effects of cognitive dysfunction and regimen complexity. *Neurology*, 59, 1944–1950.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424–453.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternative. *Structural Equation Modeling*, 6, 1–55.
- Johnson, W., & Bouchard, T. J., Jr. (2007). Sex dimensions in mental abilities: g masks the dimensions on which they lie. *Intelligence*, 35, 23–39.
- 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, 615–622.
- Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, & Boyle (2005). *Psychological Bulletin*, 131, 66–71.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T., & Engle, R. W. (2004). The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning ability. *Journal of Experimental Psychology: General*, 133, 189–217.
- Kirschner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, 33, 352–358.
- Kline, R. (2005). *Principles and practice of structural equation modeling*. New York, New York: The Guilford Press.
- Kwong See, S. T., & Ryan, E. B. (1995). Cognitive mediation of adult age differences in performance. *Psychology and Aging*, 10, 453–468.
- Kyllonen, P. C., & Christal, R. E. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence*, 14, 389–433.
- Lezak, M. D., Howieson, D. B., & Loring, D. W. (2004). *Neuropsychological assessment*, (4th Edition). New York, NY: Oxford University Press.
- 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.
- McElree, B. (2001). Working memory and focal attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27, 817–835.
- Nuechterlein, K. H., Pashler, H. E., & Subotnik, K. L. (2006). Translating basic attentional paradigms to schizophrenia research: Reconsidering the nature of the deficits. *Development and Psychopathology*, 18, 831–851.
- Oberauer, K., Süß, H. M., Wilhelm, O., & Wittman, W. W. (2008). Which working memory functions predict intelligence? *Intelligence*, 36, 641–652.
- 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.
- 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.
- Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford, England: Oxford Psychologists Press.
- 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? *European Journal of Cognitive Psychology*, 18, 378–397.
- Rosen, V. M., Bergeson, J. L., Putnam, K., Harwell, A., & Sunderland, T. (2002). Working memory and apolipoprotein E: What's the connection? *Neuropsychologia*, 40, 2226–2233.
- Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, 15, 364–371.
- Shelton, J. T., Metzger, R. L., & Elliott, E. M. (2007). Modified lag task as a measure of working memory. *Behavior Research Methods*, 39, 482–493.
- Stearns, C., Dunham, M., McIntosh, D., & Dean, R. S. (2004). Attention deficit/hyperactivity disorder and working memory in clinically referred adults. *International Journal of Neuroscience*, 114, 273–287.
- Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: administration, norms, and commentary*, 3rd Edition. Oxford, England: Oxford University Press.
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependant? *Journal of Memory and Language*, 28, 127154.
- Unsworth, N., & Engle, R. W. (2005). Working memory capacity and fluid abilities: Examining the correlation between Ospan and Raven. *Intelligence*, 33, 67–81.
- Unsworth, N., & Engle, R. W. (2006). Simple and complex memory span 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, 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.
- Vandenmerg, R., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70.
- Wechsler, D. (1997). *Wechsler adult intelligence scale-Third Edition: Technical Manual*. San Antonio, TX: The Psychological Corporation.
- Wechsler, D. (1997). *Wechsler memory scale-Third Edition: Administration and Scoring Manual*. San Antonio, TX: The Psychological Corporation.
- Wilde, N., & Strauss, E. (2002). Functional equivalence of WAIS-III/WMS-III digit and spatial span under forward and backward recall conditions. *The Clinical Neuropsychologist*, 16, 322–330.