Testing the Predictive Validity of Working Memory Capacity for Job Performance

By

Brenna Armfield

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts in Psychology

Middle Tennessee State University

August 2022

Thesis Committee:

Michael Hein, Ph.D., Chair

Judith Van Hein, Ph.D.

James Houston, Ph.D.

ACKNOWLEDGEMENTS

I would like to thank my thesis committee, Dr. Hein, Dr. Van Hein, and Dr. Houston for their time, energy, and support. Dr. Hein, I greatly appreciated your easy-going, yet meticulous style and commitment to our research. It has been an absolute joy working with you over the last two years. Dr. Van Hein, thank you for always providing feedback that is straightforward, relevant, abundantly helpful, and sometimes even humorous. Dr. Houston, thank you for lending your expertise and for your willingness to ensure cognitive psychology was properly represented in this study. I greatly appreciate your patience and flexibility. I would also like to thank Dr. Frame for making this project possible and for always being so generous with your time, energy, and words of encouragement.

I would also like to thank my family and friends for their love and support. To my parents, thank you for everything you have done for me and for always being there for me. To my siblings, Margaux and Connor, thank you for always providing me with advice and amusement, and Margaux, thank you for lending your coding skills that allowed my data collection to be smooth and painless (or nearly so anyway). Thank you to my cohort for being so kind and supportive and for being genuine friends to me over the last two years. I feel so thankful to have been part of this group. Lastly, thank you to Anna White for being a wonderful roommate and friend. You have been my rock throughout the program, and I can’t wait to see what you accomplish in the future.

ABSTRACT

This study tests whether working memory, short-term memory, and attentional control predict job performance and compares these measures to traditional intelligence and general mental ability (*g*) testing used for selection purposes. Researchers sought to present a viable alternative to tests of intelligence and *g*, which are often used for selection proposes despite evidence for differential validity and mean score differences in racial subgroups. The current paper seeks to address these issues by exploring viable alternatives to *g:* working memory, short-term memory, and attentional control. Results indicated that general mental ability, working memory, and attentional control were not predictive of performance, but short-term memory was found to have a significant relationship with structured interview performance. Short-term memory also contained no significant subgroup score differences between White and Non-White applicants, suggesting it may be a more culture-fair method of cognitive ability assessment than traditional measures of *g*.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1: LITERATURE REVIEW

Introduction

Selecting appropriate job candidates is a task of critical importance within human resources departments. Without valid and reliable selection tools, organizations may make personnel decisions based on unscientific, inaccurate, unfair processes such as relying on 'gut instincts' or judging a candidate's appearance or demeanor. Selection processes impact who is hired into an organization and who is internally promoted. Some potential consequences of poor selection choices include reduced individual and organizational performance, higher turnover, lawsuits over practices perceived as unfair, stifled organizational culture, lower customer satisfaction, and loss of revenue (Sutherland & Wöcke, 2011). Therefore, it is important that selection tools are as accurate as possible and are grounded in scientific research.

Currently, measures of intelligence are some of the most powerful predictors of job performance available (Schmidt, 2002; Murphy et al., 2003). General mental ability (*g*) has been shown to be highly predictive of performance across a range of jobs, and some researchers have identified the correlation between *g* and performance to be as high as .75 (Hunter, 1986). Consequently, many organizations utilize measures of *g* for selection purposes. However, researchers have identified a critical problem with using measures of *g* for organizational decision-making: significant differences exist between racial groups in average scores and in *g*'s predictive validity for performance measures (Berry et al., 2011; Garnder & Deadrich, 2011).

These subgroup differences suggest the possibility of a racial bias within the measures used to assess *g*. Consequently, utilizing these measures could lead to a host of detrimental outcomes for both individuals and organizations. These tests may create an adverse impact on minority groups and make selection errors that result in reduced employee and organizational performance (Aquinis & Smith, 2007). Businesses also expose themselves to legal action by utilizing unfair or inaccurate measures. These errors may also result in detrimental societal consequences by perpetuating and compounding existing racial inequalities (Murphy et al., 2003). These problems have prompted researchers to search for a better alternative to *g* and to try to identify measures that predict job performance but limit adverse impacts and race-based differential validity.

The current paper seeks to address these issues by proposing a viable alternative to *g:* working memory capacity. Previous researchers have found that working memory capacity and similar constructs (executive attention, attentional control, and short-term memory) are highly related to measures of general cognitive ability (Bosco et al., 2015) and fluid intelligence (Broadway & Engle, 2010) and that they can be used to predict job performance (Bosco et al., 2015; Hicks et al., 2016; Verive & McDaniel, 1996; Wang et al., 2018). Most importantly, measures of working memory and executive attention do not seem to have the same problems of substantial subgroup score differences or race-based differential validity that traditional measures of general mental ability often do (Bosco et al., 2015; Hicks et al., 2016; Larson, 2019). Therefore, this paper seeks to replicate these findings and expand on previous research investigating the relationships between intelligence (*g*), working memory, executive attention, and performance.

Measuring Intelligence

Intelligence is a broad yet complex construct, and the way it is defined and operationalized varies in research (Sternberg & Determan, 1987; Larson, 2019).

Intelligence can be conceptualized in two distinct ways: 1) the knowledge that an individual possesses (also known as crystallized intelligence), or 2) an individual's ability to process and manipulate information (also known as fluid intelligence; Cattell, 1943; Ackerman, 1996). Knowledge-based intelligence is influenced by any factor that affects a person's access to information, such as educational background, life experiences, socio-economic status, cultural norms, technology, and other factors (Cattell, 1943). Information-processing intelligence is determined by an individual's cognitive functioning and is typically less influenced by lifestyle or culture (Cattell, 1943).

Spearman (1927, as cited by Larson, 2019) proposed a foundational theory of intelligence that suggests there is a *general* mental ability (*g*) that can be assessed, and it underlines all other mental abilities (e.g., verbal ability, spatial ability, numerical ability, mechanical skills). Spearman's theory of general intelligence suggests that the usefulness of an intelligence measure is dependent upon the degree to which it assesses an individual's general intelligence, and thus, the specific content within the test is not important as long as it measures *g*. For this reason, many measures of intelligence utilized for selection purposes prioritize the measurement of *g* and are not sensitive to the degree to which their items measure intelligence as previously acquired knowledge or measure intelligence as information-processing abilities (Larson, 2019). Consequently, measures of intelligence often incorporate items that gauge knowledge-based intelligence.

Currently, some common terms used in psychological assessments to describe intelligence include general mental ability and cognitive ability. While these terms appear to focus on an information-processing conceptualization of intelligence, they still often incorporate measures that are influenced by knowledge-based intelligence (e.g., solving math problems, understanding vocabulary, interpreting situations that may or may not be familiar depending on one's background and culture). Measuring intelligence in a way that gauges an individual's background and existing knowledge structures may help explain why many intelligence measures have been found to have subgroup score differences and differential validity based on race.

Testing Intelligence and Cognitive Ability

General intelligence (*g*), general mental ability (GMA), and cognitive ability are common terms used in cognitive psychology when describing aspects of intelligence. While different researchers utilize varying definitions of *g*, GMA, and cognitive ability, this paper lumps all of these constructs together under the broad term general cognitive ability (GCA). GCA can be defined as an individual's relatively stable ability to learn (Schmidt, 2015).

GCA measures have widely been identified as powerful predictors of performance across a range of jobs (Schmidt, 2002; Murphy et al., 2003). This relationship is moderated by factors such as job complexity and performance criterion (Hunter, 1986). GCA measures more strongly predict performance in high complexity jobs than they do low complexity jobs (Schmidt, 2015; 2002). Conceptually, this may be because more complex jobs require job-holders to utilize cognitive abilities more effectively than less complex jobs do.

Due to their consistent validity in predicting performance, many organizations use measures of mental ability as selection tools for hiring employees. Some examples of commonly used assessment tools that measure intelligence and cognitive ability include the Wonderlic Personnel Survey, Wesman personnel classification test, Watson-Glazer

Critical Thinking Appraisal, and many others. Researchers argue that cognitive ability plays a central role in job performance because it is inextricably connected to an employee's ability to learn on the job, job knowledge, and cognitive tasks such as utilizing planning, judgment, and memory (Hunter, 1986). For these reasons, some psychologists believe "there can be no debate" about the role of mental abilities in predicting performance (Schmidt, 2002). However, utilizing cognitive ability testing for selection purposes is not without controversy (Murphy et al, 2003).

Controversy in GCA Testing

In the past few decades, an increasing number of researchers have identified racial biases in these measures, and the predictive validity of GCA measures in predicting performance has been found to be significantly moderated by race (Berry et al., 2011; Garnder & Deadrich, 2011). Researchers have identified patterns of racial groups displaying different normative data (e.g., differing average scores, standard deviations, etc.) on these assessments (Berry et al., 2011; Gardner & Deadrick, 2012), as well as different levels of predictive validity for different racial groups. Some potential causes of these findings include a range restriction in the data, the psychometric characteristics of tests or criteria (i.e., measurement error or bias), contextual influences (e.g., stereotype threat, socio-economic status disparities), or true differences between subgroups in the role cognitive ability plays in determining performance (Berry et al., 2011; Letang et al., 2021). Regardless of the reasons behind differential mean scores and predictive validity, it is clear that attempting to use and interpret GCA assessments in the same way for both White and non-White job candidates could result in racially-biased decision-making.

When employers utilize biased selection tools, they may create an adverse impact on minority groups and make problematic selection errors that harm both individuals and the organization. If organizations utilize the same cut-off score for all groups, despite differences in how the test predicts performance for certain groups or the existence of mean score differences between subgroups, the test will not produce fair, accurate predictions. One group's performance may be overestimated while another's performance may be underestimated. Consequently, some strong candidates may be overlooked while weaker candidates may be inappropriately selected. Reduced employee and organizational performance will occur if employers fail to hire applicants that are actually qualified or hire applicants that are actually unqualified (Aquinis & Smith, 2007). An adverse impact is created if the selection tool unfairly eliminates minority applicants at a disproportional rate. If this occurs, organizations are left with less diversity in their staff and are vulnerable to legal repercussions. Additionally, these errors can result in detrimental societal consequences by perpetuating and compounding racial inequalities (Murphy et al., 2003).

Because many of the assessment procedures used to detect bias have low statistical power, decision-makers may fail to realize that they are using selection tools that are racially biased and create adverse impact (Aquinis & Smith, 2007). Human resources managers have found themselves in a difficult situation whereby cognitive ability measures are currently some of the strongest predictors of performance (Hunter, 1986; Schmit, 2002), but using these measures may result in adverse impact, propagate unfair hiring decisions, mitigate diversity within their organization, and leave their organizations open to legal actions. Unless a viable solution to this problem is identified, practitioners face a tradeoff between utilizing cognitive ability measures (one of the most powerful selection tools currently available) and creating adverse impact, unfair selection practices, and potentially selecting less talented candidates (due to differential validity of measures).

Potential Solutions

Several suggestions have been presented to address this issue. Perhaps the simplest solution is to utilize different cut-off scores for differing racial groups to account for mean differences in subgroup test scores. However, this strategy is not ideal given that there may be an inadequate amount of data to perfectly align cut-off scores for every subgroup or fully understand how being a member of multiple subgroups may affect the predictive validity of the test (i.e. being bi-racial). Furthermore, the practice of using different cut-off scores based on race is illegal in the United States (The Civil Rights Act, 1991).

Another way to address the problem of selection tools with differential validity is to attempt to measure a construct related to GCA that accurately predicts job performance but does not demonstrate the same levels of differential validity, subgroup differences, or racial bias. Based on Akerman's theory of intelligence (1996) describing intelligence as either previously acquired knowledge structures or the ability to process information, cognitive ability measures that more exclusively focus on information-processing abilities may present a viable alternative to the traditionally-used intelligence and CGA selection assessments. One such potential construct is working memory capacity, and this paper seeks to validate it as a useful and less biased predictor of job performance than traditional cognitive ability tests and intelligence tests.

Working Memory Capacity as an Alternative to GCA

Though working memory has been conceptualized in a multitude of ways (Cowan, 2017), most definitions of working memory identify it as a temporary retrieval space that has a limited capacity, stores information, and processes demands (Baddeley & Hitch, 1974). Information from various sources is integrated within working memory via attentional control (Baddeley, 2000), and this information is kept in a temporary state of heightened accessibility (Cowan, 2017). Working memory is distinct from short-term memory because it incorporates the processing and manipulation of information rather than simply storing it, but short-term memory is an important component of working memory (Baddeley & Hitch, 1974; Cowan, 2017; Engle, 2002).

Measures of working memory capacity (WMC) are highly predictive of higher-order cognitive functioning (Engle, 2002), and some researchers believe WMC is effectively the same construct as executive attention (Engle, 2002). The validity of that assertion depends on the researcher's conceptualization and operationalization of WMC and executive attention. Whether or not they are truly equivalent, they seem to at least be related to one another, as many definitions of WMC identify the central importance of controlling attentional resources (e.g., Cowan, 2017; Hicks et al., 2016, Engle, 2002; Baddeley, 2000). WMC is related to general fluid intelligence (Kulikowski & Orzechowski, 2019), and measures of WMC and fluid intelligence often share much of the same variance (Broadway & Engle, 2010). Some researchers even consider WMC to be one of the components that makes up general mental ability (*g*; Kulikowski & Orzechowski, 2019). Due to its relationship with fluid intelligence and *g*, it follows theoretically that WMC should predict job performance as well. Conceptually, the ability to store and manipulate information seems like an important skill to effectively perform a variety of work-related tasks.

Importantly, measures of WMC do not seem to have the same degree of problems of mean score differences and race-based differential validity that traditional GCA measures often do (Bosco et al., 2015; Hicks et al., 2016; Larson, 2019). It has been demonstrated that measures of cognitive ability that focus on an individual's ability to process information (such as measures of working memory capacity and executive attention) produce significantly smaller subgroup differences than measures of cognitive ability that capture a participant's baseline knowledge (Larson, 2019). Thus, any GCA test that measures respondents' previously acquired knowledge is more likely to produce subgroup differences than cognitive ability measures of one's ability to process information, like WMC. Accordingly, WMC has been shown to be more culture-fair and have less adverse impact than tests of intelligence (Hicks et al., 2016 Bosco et al. (2015) found that subgroup differences between Black and White respondents were significantly smaller for executive attention and working memory measures than other mental ability measures.

Additionally, Larson (2019) conducted a meta-analysis and found much smaller group score differences between Black and White test-takers when using information processing measures of intelligence (d=.41) than for other common measures of intelligence with d's ranging from .61 to 1.14. Conceptually, it makes sense that compared to other measures of cognitive ability, measures of working memory would be less affected by respondents' educational background and baseline knowledge and would be less likely to contain assessment elements that are culturally biased—two factors that are thought to possibly account for subgroup scores differences.

WMC measures may also be less likely to produce stereotype threat in minority respondents. Stereotype threat is defined as a "socially premised psychological threat that arises when one is in a situation or doing something for which a negative stereotype about one's group applies" (Steele & Aronson, 1995). Intelligence measures are more saliently related to racially discriminatory beliefs that have been spouted historically than memory tests are. Thus, it seems probable that WMC measures would produce a smaller stereotype threat than measures of *g* or cognitive ability. These reasons may help explain why WMC measures have been found to have smaller subgroup differences than traditional GCA measures (Bosco et al., 2015; Hicks et al., 2016; Larson, 2019). These findings form the basis for hypotheses 1 and 2.

Hypothesis 1: Working memory capacity will be positively related to performance.

Hypothesis 2: Measures of working memory capacity (Backward Digit Span and Backwards Corsi task) will display smaller ethnic and racial group score differences than measures of general mental ability.

Though job performance is hypothetically the main criterion of interest, this study will instead examine predictor variable relationships with structured situational interview performance. Interview performance is used instead of job performance because the sample consisted of job applicants, so no job performance data was available. However, structured situational interview performance has been shown to be predictive of job performance with criterion validity coefficients ranging from .41 to .47 (Huffcutt &

Arthur, 1994; Kepes et al., 2012; Latham & Sue-Chan, 1999; McDaniel et al., 1994; Taylor & Small, 2002).

This paper seeks to validate WMC measures as a selection tool and to test whether WMC is a powerful predictor of performance that is stable across subgroups. Previous researchers have found that WMC and similar constructs (executive attention, attentional control, and short-term memory) are highly related to measures of *g* (Bosco et al., 2015) and fluid intelligence (Broadway & Engle, 2010) and that they are predictive of job performance (Bosco et al., 2015; Hicks et al., 2016; Verive & McDaniel, 1996; Wang et al., 2018). Most importantly, measures of WMC and executive attention have smaller mean score differences and predictive validity differences between racial subgroups than traditional general mental ability measures (Bosco et al., 2015; Hicks et al., 2016; Larson, 2019). This research seeks to replicate these findings and also examine the unique roles played in predicting performance by the related constructs: short-term memory and attentional control. Therefore, two research questions will also be investigated:

Will short-term memory measures predict performance?

Will attentional control measures predict performance?

CHAPTER 2: METHODOLOGY

Overview

This study analyzed archival data collected by an organizational consulting firm partnering with a mid-sized governmental law enforcement agency located in the southeastern United States. The organizational consulting group was contracted by the agency to manage and administer the agency's selection process. The data used in this paper was collected during the agency's yearly selection process in the Fall of 2021.

Participants

Participants were recruited from a group of job applicants applying for a law enforcement position at a mid-sized governmental agency. Participation in this research was voluntary, and participants were informed that choosing to participate in this research had no impact on the evaluation of their job application. Participants' individualized scores on the study measures were not shared with the participating governmental agency, nor were they shown to individuals evaluating their application materials. A total of 180 applicants completed at least one of the study measures. The majority of applicants identified as male (66.7% male, 7.2% female, 26.1% no data), White (55.6% White, 11.1% Black, 2.2% Hispanic, 0.6% Asian, 26.1% no data), and had completed a high school degree or higher (68.3% had, 5.0% had not, 26.7% no data). Tables 1-3 summarize demographic frequency data. The mean age of participants was 29.55 years old (*SD*= 7.21, ranging from 20 to 54).

Table 1

Frequency Data for Gender

	N	Percentage
Male	120	66.7%
Female	13	7.2%
No data	47	26.7%

Table 2

Frequency Data for Race/Ethnicity

	N	Percentage
White	100	55.6%
Black	20	11.1%
Hispanic	8	4.4%
Asian	1	0.6%
Other	4	2.2%
No data	47	26.1%

Table 3

Frequency Data for High School Degree

	N	Percentage
Graduated high school	123	68.3%
Did not graduate high school	9	5.0%
No data	48	26.7%

Procedures

Participants were sampled from a pool of job candidates applying to work for a governmental agency focusing on law enforcement and highway safety. Applicants completed the first stage of the agency's standard job application process, and they were then offered the chance to complete some additional measures of cognitive ability and general mental ability. They were made aware that the measures were purely for research purposes and that their completion (or non-completion) of these measures would have no impact on the evaluation of their job application. They were also informed that they were not required to complete the measures in order to advance to the next phase of the selection process. It was emphasized that completing these measures was entirely voluntary.

The cognitive ability measures and the GMA measure served as predictor variables, while interview performance serves as the criterion variable. The structured situational interview is a standard component of the agency's current selection process. Thus, it is not being introduced for this study; it is simply functioning as the criterion variable.

All measures were taken online by applicants during the Fall of 2021. The memory tasks took approximately 20-30 minutes to complete, and the GMA measure took approximately 15 minutes to complete. However, these measures were administrated online without a time limit or proctor, so the time participants spent completing the measures varied. It is also impossible to know whether a participant's entire time was spent working on the tasks or if they took any breaks during the process.

Measures

Participants completed an online questionnaire featuring four components: short-term memory measures, attentional control measures, working memory capacity measures, and a measure of general mental ability. Short-term memory and attentional control are components of working memory capacity. These constructs were tested independently to better understand the relative importance of each in predicting job performance. Participants also completed the HEXACO assessment for a separate research project. The present study did not make any predictions relating to HEXACO performance, and researchers did not include HEXACO scores in their analysis.

Short-term Memory. Participants completed the Digit Span and the Corsi task (Corsi, 1972; Wechsler, 2008). Both tasks were administered online and un-proctored via PsyToolKit.org (Stoet, 2010; Stoet, 2017).

Digit Span. In this task, participants are asked to remember a series of digits that increases in length in consecutive trials. For each series, a digit appears centrally on the screen (1000 ms) and is replaced by another digit after a short interstimulus interval (500 ms). After the final digit in each series, participants are asked to type the presented series of digits into a textbox in the same serial order as the digits were presented. Scoring is determined by the span length reached by the participant. In other words, the minimum score possible is a 3, since the task begins with a series of 3 digits and increases in length if participants are able to correctly recall the series of digits. A respondent's score is the largest span of digits that the respondent is able to correctly recall on at least two trials.

Corsi task. This task is similar to the digit span, but it requires participants to remember the order in which a set of blocks have been illuminated on a screen. A set of blocks is present on the screen and then they light up in a pattern. Respondents are required to click the box in the same serial order as they light up on the screen. As with the digit span task, the sequence begins with three and increases in length if the respondent correctly identifies the correct blocks. The longer sequence of blocks that a participant can correctly recall, the longer their Corsi span. Their final score represents the maximum span correctly identified on at least two trials.

Working Memory Capacity Battery. Participants completed two simple working memory tasks: the *Backward Digit Span* (BDS; Wechsler, 2008) and the *Backward Corsi Task* (BCT; Corsi, 1972). These measures were also administered online via PsyToolKit.org (Stoet, 2010; Stoet, 2017). Simple working memory tasks are shorter and less complex to administer than other working memory tasks such as complex span tasks (e.g., *Operation span*, *Reading span*, *Counting span*), so if they are able to provide the same level of predictive validity, they would be preferable in practical use. Additionally, the *Operation span* and *Reading Span* tasks require basic math skills and logical judgments of sentence appropriateness, respectively. Therefore, a participant's score may be somewhat swayed by educational and cultural background. This is much less likely to be the case for the simple WM tasks, which do not require numerical or verbal understanding. Therefore, this research will only utilize simple measures of working memory capacity: *Backward Digit Span* and *Backward Corsi Task*. These tasks are identical to the forward versions except that participants are asked to type the digits or click the blocks in reverse order (opposite to the order the stimuli were presented in).

Attentional Control. Participants completed the Flanker task (Erikson & Erikson, 1974) and the Stroop task (MacLeod, 1991; Stroop, 1935). These are two commonly used attentional control tasks that are fairly simple to administer online. These tasks were administered online via PysToolKit.org (Stoet, 2010; Stoet, 2017). Both tasks test attentional control by requiring a participant to ignore irrelevant stimuli and focus on relevant information (also known as inhibitory control).

Flanker task. The Flanker task involves responding to a stimulus that is flanked by irrelevant information that must be ignored. The version used in this study differed slightly from the original Flanker task designed by Erikson and Erikson (1974). In the PsyToolKit implementation of this task, respondents are presented with 5 letters but are instructed to respond to the middle letter. Depending on the letter presented, participants must click either the A button on the keyboard (if an X or C is the middle letter) or the L button (if a V or B is the middle letter). Respondents are scored based on the number of correct responses given and their reaction time.

Stroop task. In the Stroop task, participants are presented with the name of a color (e.g., “red” or “yellow”), and a word is displayed in colored font. Sometimes the color of the font matches the word presented (such as “red” in red font) and this is considered a “compatible” presentation of the stimulus. Other times, the word presented is written in a font color that does not match the word (such as “red” written in blue font) and this is considered an “incompatible” presentation of the stimulus. Participants are instructed to type the first letter of the name of the color of the presented stimulus (e.g., if the word “green” is presented in yellow font, the correct answer would be “y”). Respondents are scored based on their response time for incompatible trials, their response time for compatible trials, and the accuracy of their responses.

General Mental Ability/Verbal Ability. To test general mental ability, participants completed a 45-question vocabulary IQ test found on the Open-Source Psychometrics Project website (https://openpsychometrics.org). This test measures verbal ability/verbal IQ which is a component of many cognitive tests used in personnel selection (e.g., Wesman personnel classification test)., Vocabulary measures have been shown to be highly correlated with verbal IQ ($r = .80$; Smith et al., 2005), which is a significant factor in many general mental ability measures. Indeed, several researchers have argued that vocabulary IQ is more closely linked with general mental ability than any other cognitive measure (Crawford et al., 1989; Jensen, 2001; Schipolowski, Wilhelm, & Schroeders, 2014). Thus, verbal ability serves as a representation of general mental ability in this research.

Job Performance. Job performance was not directly measured because this sample is job applicants. Job performance was represented by structured situational interview performance. The interview used consisted of 10 questions: 9 questions asking applicants to describe how they would deal with specific behavioral situations and one question asking interviewers to rate the applicant's communication skills. Applicants' responses to the interview questions were scored using behaviorally-anchored rating scales on a scale of 0 (Did not address the question) to 5 (Excellent). For the purposes of this study, performance was measured in three ways: 1) the sum of scores on all 10 interview items, 2) the sum of scores on the 9 situational interview questions (excluding the overall communication item), and 3) the score on the overall communication item.

Bilingualism. Participants also completed a one-item question to determine if they are bi- or multi-lingual. This was asked in order to investigate whether being bi- or multi-lingual impacts cognitive performance on these tasks. This question was asked within the demographics section of the application completed for job applicants.

CHAPTER 3: RESULTS

Duplicates and participants that did not complete any of the predictive measures were deleted from the data set. The remaining sample (N = 180) was analyzed using Microsoft Excel and SPSS. Participant age ranged from 20 to 54, with an average age of 29. Descriptive statistics are presented in Table 4.

Table 4

Descriptive Statistics

	N	Mean	Standard Deviation	Min	Max
Digit Span	179	7.20	2.40	3	13
Backward Digit Span	166	6.58	2.37	3	13
Corsi	157	5.21	1.76	0	9
Backward Corsi	153	4.27	1.71	0	8
Flanker	141	-5.71	131.58	-680.95	291.21
Stroop	147	123.38	391.64	-579.00	4367.73
Verbal Ability	137	17.98	8.99	-15.75	42.30
Interview Performance	110	33.85	7.00	13	49
Interview Performance (excluding overall communication item)	110	29.91	6.28	12	44
Overall Communication	110	3.94	.97	1	5

Relationships between measures were analyzed using correlations. Digit span was significantly positively correlated with backward digit span, Corsi, verbal ability, Flanker, interview performance including and excluding the overall communication, and overall communication. Backward digit span was significantly positively correlated with digit span and backward Corsi. Corsi was significantly positively correlated with digit span and backward Corsi. All of the performance measures (interview performance, interview performance excluding overall communication, and overall communication) were significantly correlated with digit span, as well as each other. Flanker was not significantly correlated with any measure except digit span. Stroop was not significantly correlated with any other measure. Table 5 contains a full correlation matrix.

Table 5

Pearson Correlation Coefficients

	1	2	3	4	5	6	7	8	9
1. Digit Span	---								
2. Backward Digit Span	.226** (n=165)	---							
3. Corsi Task	.233** (n=156)	-.092 (n=157)	---						
4. Backward Corsi Task	.154 (n=152)	-.184* (n=152)	.339** (n=157)	---					
5. Flanker	.184* (n=141)	.062 (n=140)	.137 (n=140)	.111 (n=140)	---				
6. Stroop	-.046 (n=147)	-.020 (n=146)	.057 (n=146)	-.002 (n=146)	.025 (n=140)	---			
7. Verbal Ability	.198* (n=137)	.149 (n=135)	.154 (n=135)	.132 (n=135)	.075 (n=131)	.039 (n=134)	---		
8. Interview Performance	.204* (n=110)	.004 (n=103)	.181 (n=97)	.022 (n=96)	.132 (n=86)	-.114 (n=92)	.111 (n=85)	---	
9. Interview Performance excluding Overall Communication item	.188* (n=110)	-.007 (n=103)	.191 (n=97)	.017 (n=96)	.142 (n=86)	-.108 (n=92)	.113 (n=85)	.995** (n=110)	---
10. Overall Communication	.255** (n=110)	.070 (n=103)	.059 (n=97)	.054 (n=96)	.026 (n=86)	-.125 (n=92)	.072 (n=85)	.773** (n=110)	.707** (n=110)

* Correlation is significant at the .05 level.

** Correlation is significant at the .01 level.

The predictive validity of digit span for the interview performance scores was analyzed using regression. Digit span score explained a significant amount of variance in interview performance, $F(1,108) = 4.687$, $p = .033$, $R^2 = .042$, $R^2_{adjusted} = .033$. The regression coefficient ($B = .594$) indicated that an increase in one point on the digit span on average corresponded to an increase in .594 in interview performance. Digit span score also explained a significant amount of variance interview performance excluding overall communication, F(1,108) = 3.952, p = .049, $R^2 = .035$, $R^2_{adjusted} = .026$. The regression coefficient ($B = .491$) indicated that an increase in one point on the digit span on average corresponded to an increase in .491 in interview performance. Lastly, digit span explained a significant amount of variance in overall communication score, F(1,108) = 7.527, p = .007, $R^2 = .065$, $R^2_{adjusted} = .056$. The regression coefficient ($B = .103$) indicated that an increase in one point on the digit span on average corresponded to an increase in .103 on interview performance.

Comparisons by Racial/Ethnic Group

Next, scores were compared based on race/ethnicity. Sample sizes were insufficient to compare each racial/ethnic group sampled, so minority candidates were grouped together and compared to White candidates. Because none of the predictor variables were significant except for digit span, the other cognitive variables were excluded from the analysis. Descriptive Statistics for each condition are shown in Table 6.

Table 6

Descriptive Statistics for Digit Span, General Mental Ability, Performance, and Race

Race/Ethnicity		Digit Span	General Mental Ability	Interview Performance (IP)	IP excluding overall communication	Overall Communication
White	*M*	7.26	18.33	34.54	30.48	4.06
	N	99	83	83	83	83
	SD	2.36	9.03	6.67	6.06	.89
Non-White	*M*	6.39	12.36	31.70	28.15	3.56
	N	33	21	27	27	27
	SD	2.28	7.17	7.66	6.74	1.12
Total	*M*	7.05	17.26	33.85	29.91	3.95
	N	132	104	110	110	110
	SD	2.36	8.99	7.00	6.28	.97

Relationships between measures for both White and Non-White applicants were analyzed using correlations. For Non-White applicants, digit span was significantly positively related to interview performance (with and without the overall communication item) and overall communication. Verbal ability was not significantly related to any other measure. Interview performance was significantly positively related to interview performance excluding overall communication, as well as overall communication. Interview Performance excluding the overall communication item was significantly positively correlated to overall communication. Table 7 contains a full correlation matrix for Non-White applicants.

For White applicants, digit span was not correlated with any other measures, nor was verbal ability. Interview performance was significantly positively related to interview performance excluding overall communication, as well as overall communication. Interview Performance excluding the overall communication item was significantly positively correlated to overall communication. See Table 8 for a full correlation matrix for White applicants.

Table 7

Pearson Correlation Coefficients for Non-White Applicants

	1	2	3	4	5
Non-White Applicants					
1. Digit Span	--				
2. Verbal Ability	-.048 (n= 21)	--			
3. Interview Performance	.445* (n= 27)	.170 (n= 15)	--		
4. Interview Performance excluding overall communication item	.442* (n= 27)	.173 (n= 15)	.996** (n= 27)	--	
5. Overall Communication	.385* (n= 27)	.106 (n= 15)	.844** (n= 27)	.793** (n= 27)	--

* Correlation is significant at the .05 level.

** Correlation is significant at the .01 level.

Table 8
Pearson Correlation Coefficients for White Applicants

	1	2	3	4	5
White Applicants					
1. Digit Span	--				
2. Verbal Ability	.103 (n= 83)	--			
3. Interview Performance	.089 (n= 83)	.046 (n= 70)	--		
4. Interview Performance excluding overall communication item	.074 (n= 83)	-.050 (n= 70)	.995** (n= 83)	--	
5. Overall Communication	.162 (n= 83)	.006 (n= 70)	.729** (n= 83)	.657** (n= 83)	--

* Correlation is significant at the .05 level.
** Correlation is significant at the .01 level.

A one-way ANOVA was used to compare performance scores for the between-subject factor of race/ethnicity (White, Non-White). Performance scores significantly differed based on race for Verbal Ability, $F(1, 102) = 7.898$, $p < .006$, $\eta = .268$, $\eta^2 = .072$ and Overall Communication, $F(1, 108) = 5.756$, $p < .018$, $\eta = .225$, $\eta^2 = .051$. Performance scores were not significantly different based on race for Digit span [$F(1,130) = 3.405$, $p < 0.067$, $\eta = .160$, $\eta^2 = .026$], Interview Performance [$F(1,108) =3.425$, $p < .067$, $\eta = .175$, $\eta^2 = .031$], or Interview Performance excluding overall communication [$F(1,108) = 2.861$, $p < .094$, $\eta = .161$, $\eta^2 = .026$]. See Table 9 for a summary of the ANOVA analysis.

Table 9

Predictor	Sum of Squares	*df*	Mean Square	*F*	η^2
Digit Span	713.051	130	5.485	3.405	.026
Verbal Ability	7720.461	102	75.691	7.898**	.072
Interview Performance	5176.232	108	47.928	3.425	.031
Interview Performance excluding overall communication item	4188.130	108	38.779	2.861	.026
Overall Communication	97.365	108	. 902	5.756*	.051

* Significant at the .05 level.

** Significant at the .01 level.

Researchers also sought to test the impact of bilingualism on cognitive ability performance. However, performance scores were not analyzed by bilingual status because the sample of bi- or multi-lingual participants was too small (n= 16).

DISCUSSION

This study sought to test the predictive validity of memory and executive attention measures for interview performance and compared their performance to a more traditional intelligence/cognitive ability measure. Verbal ability was used to represent general mental ability, while digit span, Corsi, backward digit span, backward Corsi, Flanker, and Stroop were used to represent alternative measures of cognitive ability that captured short-term memory, working memory, and attentional control. Overall, hypotheses were not met for working memory capacity or executive attention measures, but the research questions investigating short-term memory produced significant results.

Hypotheses and Research Questions

Hypotheses were expected to be met such that the working memory capacity was found to be a valid predictor of performance. Additionally, it was expected that the working memory measures would display smaller subgroup differences in scores and validity than the general mental ability measure. Results did not support these hypotheses because the working memory measures (Backward digit span and Backward Corsi) were not significantly related to any of the performance measures. Neither were the attentional control measures (Flanker and Stroop).

However, investigating the first research question relating to short-term memory did provide significant findings. Digit span was found to predict interview performance and to contain no significant score differences between White and Non-White candidates. Verbal ability, however, did not predict interview performance, and it contained significant score differences between White and Non-White candidates.

Another interesting finding was that digit span was strongly related to performance for minority applicants, but it was not significantly related to performance for White applicants. There are many explanations for this finding, one being that short-term memory plays a stronger role in interview performance for Minority candidates than for White candidates. Another explanation is that the interview performance measure contained a degree of bias while the short-term memory measure is objective and not influenced by bias. Therefore, it is possible that having a higher short-term memory capacity helped Non-White applicants overcome bias in interview scores but did not impact White applicant scores.

Implications

These findings suggest that memory measures may indeed be better tools for predicting performance than general mental ability measures, which have been shown to produce significant differences in scores and predictive validity between racial groups (Berry et al., 2011; Garnder & Deadrich, 2011). Though interview performance and overall communication were measured rather than performance ratings of job-holders, these findings may extend to job performance as well. Structured situational interview performance is known to be a robust predictor of job performance with meta-analytic evidence indicating criterion-related validity ranging from .41 to .47 in mean-corrected correlations (Huffcutt & Arthur, 1994; Kepes et al., 2012; Latham & Sue-Chan, 1999; McDaniel et al., 1994; Taylor & Small, 2002). Therefore, short-term memory measures may be useful predictors of job performance as well as interview performance. Organizations may wish to use digit span or similar measures as screening tools for selection purposes.

It is important to note that the digit span is a factor used in some IQ tests, such as the Weschler Adult Intelligence Scale. Thus, using a short-term memory measure is not necessary a new idea in intelligence testing. However, this research indicates that organizations using traditional IQ measures may wish to place more emphasis on the memory components of these tests rather than the general mental ability components relying more on crystalized knowledge rather than fluid intelligence. Furthermore, organizations may be better off only considering short-term memory measures and avoiding mental ability subscales shown to have substantial subgroup score differences.

This study suggests short-term memory may be a more fair and accurate way of predicting performance than some commonly used general mental ability measures that can lead to an adverse impact on minority candidates. Digit span was shown to contain no significant score differences between White and minority candidates, indicating it may be a more culture-fair measure of performance. Organizations wishing to increase predictive validity, avoid measures containing racial bias, reduce adverse impact, and encourage more diversity within their organizations should consider replacing a general mental ability or IQ measure with a short-term memory measure.

Limitations

A key limitation in this research is that interview performance is measured instead of job performance. Interview performance is serving as a proxy measure of job performance, but a direct study of the relationship between job performance and cognitive measures would provide greater evidence of predictive validity for selection purposes. However, this study does provide information that could help organizations filter the candidate pool to only those candidates predicted to do well in an interview.

Administering these measures would likely be significantly less expensive than putting all applicants through the interview process, so these measures could be used to reduce the applicant pool in a multiple hurdle selection approach.

Another limitation of this study concerns the sample used. First, this research was conducted only on individuals applying to work in a law enforcement position at a governmental agency. Thus, the findings may not generalize to a diverse range of job applicants in different industries. Additionally, applicants self-selected whether or not they chose to complete the research measures once they finished the required portion of the job application materials. Random selection did not occur, so it is possible that candidates who chose to complete these measures differed from job candidates that did not in some significant way. For example, they may be more conscientious or more motivated to do well in the application process than the job applicants that chose not to complete the additional measures used for this research. Also, the sample was largely male, indicating that results may be different in a sample with a more even gender distribution.

Additionally, the online, un-proctored administration of these measures could be considered a limitation as well. Online administration can reduce the precision of the protocols provided as compared to a controlled laboratory environment due to a lack in standardization of the equipment used (e.g., trial response times may vary compared to an equivalent on-ground application that uses the same keyboard, same monitor, consistent server, etc.) Additionally, it is possible that some participants attempted to cheat, particularly on the Vocabulary IQ test. However, the fact that these measures (the working memory measures, short-term member measures, and attentional control measures) are easy to administer and available for free online is one of the key practical benefits associated with them. Therefore, this limitation primarily acts on the general mental ability measure rather than the measures proposed as alternative measures of cognitive ability (i.e., the working memory, short-term memory, and attentional control measures).

Lastly, this study presented all of the cognitive measures in the same order with the digit span presented first, followed by the backward digit span, Corsi, backward Corsi, Stroop, flanker, and vocabulary IQ. Attrition occurred as participants progressed through the tasks, meaning that more participants completed the digit span than any of the other tasks. Additionally, this task occurred first, when participants were likely paying the most attention. This may contribute to why it was the only variable found to be significantly related to the performance measures. It is possible participants were less focused when completing the tasks occurring later, such as the attentional control tasks and the general mental ability measure.

Additionally, a small sample size, attrition, and maturation lead to the sample of minority applicants who completed both the interview performance measures and the general mental ability measure (which was presented last) to be relatively small (n=15). Therefore, the small sample size may have contributed to this research failing to find a significant relationship between general mental ability and performance. Thus, it is possible a relationship may have been found in a larger sample size with greater power. However, the sample size of White applicants that completed the both general mental ability measure and the performance measures was significantly larger (n= 70) and still found no significant relationship between the two. Also, the correlation between minority candidates' verbal ability performance and interview performance ranged from .173 to .106. This suggests the relationship may still not have been practically significant even if a relationship had been found in a sample size with greater power.

Future Directions

Future studies should use a more diverse sample in both gender and race so that any potential subgroup differences in the relationships between cognitive ability measures and performance can be identified. Similarly, researchers should explore potential explanations if race or gender is found to be a moderator of these relationships. It is unclear why this study found differences in the predictive validity of short-term memory for interview performance between White and Non-White candidates. Hopefully, future researchers can shed light on this surprising finding.

Researchers should also explore other measures of memory, attentional control, and general mental ability and their relationships with performance, given that only a few of these were used in the present study. Also, more research is needed to understand the conditions and explanations surrounding whether short-term memory or working memory is a better predictor of performance, as this study failed to replicate previous findings demonstrating the predictive validity of working memory. Future researchers may wish to randomize the order in which participants are presented cognitive measures, in order to avoid maturation effects (which may have impacted this study). Additionally, these measures should be tested with other measures of performance such as task performance and direct measures of job performance (e.g., performance ratings by supervisors) to illuminate which cognitive measures are most related to varying dimensions of performance.

Lastly, other moderators of the relationship between cognitive ability and performance such as job complexity should be investigated. It seems probable the impact that memory, executive attention, intelligence, and other cognitive measure have on job performance varies based on the characteristics and demands of the job.

Conclusion

This study investigated the relationships between working memory, short-term memory, attentional control, general mental ability, and performance. Though it failed to replicate previous studies that found a predictive relationship between working memory and performance, short-term memory was found to be a significant predictor of performance. Additionally, short-term memory contained no significant score differences between racial subgroups, while general mental ability did. These findings suggest short-term memory may be a viable alternative to general mental ability as a predictor of performance that mitigates racial bias and promotes fairer selection practices. However, more research is needed to understand why the predictive validity of short-term memory was moderated by race such that it was predictive for Non-White applicants but not White applicants. Overall, this study demonstrates the importance of continuing to explore the predictive validity of cognitive ability for forecasting job performance in order to achieve fairer, more useful selection tools that will benefit both job-seekers and organizations.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*, 227-257.

Aquinis, H., & Smith, M. A. (2007). Understanding the impact of test validity and bias on selection errors and adverse impact in human resource selection. *Personnel Psychology*, *60*(1), 165–199. https://doi.org/10.1111/j.1744-6570.2007.00069.x

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

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, 47–89. https://doi.org/10.1016/s0079-7421(08)60452-1

Berry, C.M., Clark, M.A. & McClure, T.K. (2011), Racial/ethnic differences in the criterion-related validity of cognitive ability tests: a qualitative and quantitative review, *Journal of Applied Psychology, 96*(5), pp. 881-906.

Bosco, F., Allen, D., & Singh, K. (2015). Executive attention: An alternative perspective on general mental ability, performance, and subgroup differences. *Personnel Psychology*, *68,* 859–898. https://doi.org/10.1111/peps.12099.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40(3),* 153–193. https://doi.org/10.1037/h0059973.

Civil Rights Act of 1991 § 109, 42 U.S.C. § 2000e et seq (1991). Retrieved from Equal Employment Opportunity Commission website: https://www.eeoc.gov/civil-rights-act-1991-original-text.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain. McGill University.

Cowan, N. (2017). The many faces of working memory and short-term storage. 2016, 1158–1170. https://doi.org/10.3758/s13423-016-1191-6

Crawford, J. R., Stewart, L. E., Cochrane, R. H. B., Parker, D. M., & Besson, J. A. O. (1989). Construct validity of the National Adult Reading Test. *Personality and Individual Differences*, *10*(5), 585-587.

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*, 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

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

Eriksen, B. A. & Eriksen, C. W. (1974). Effects of noise letters upon identification of a target letter in a non-search task. *Perception and Psychophysics, 16*, 143-149.

Gardner, D. G., & Deadrick, D. L. (2012). Moderation of selection procedure validity by employee race. *Journal of Managerial Psychology*, *27*(4), 365–382. https://doi.org/10.1108/02683941211220180

Hicks, K. L., Foster, J. L., & Engle, R. W. (2016). Measuring working memory capacity on the web with the online working memory lab (the OWL). *Journal of Applied Research in Memory and Cognition*, *5*(4), 478–489. https://doi.org/10.1016/j.jarmac.2016.07.010.

Huffcutt, A.I. (2011). An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment*, *19*, 62–81.

Huffcutt A.I., Weekley, J.A, Wiesner, W.H, DeGroot, TG, & Jones C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. *Personnel Psychology*, *54*, 619–644

Huffcutt, A. I., & Arthur, W. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79*, 184-190. doi:10.1037/0021- 9010.79.2.184

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340–362.

Jensen, A. R. (2001). Vocabulary and general intelligence. *Behavioral and Brain Sciences, 24*, 1109-1110.

Kepes, S., Banks, G. C., McDaniel, M., & Whetzel, D. L. (2012). Publication bias in the organizational sciences. *Organizational Research Methods, 15,* 624-662. doi:10.1177/1094428112452760

Kulikowski, K., & Orzechowski, J. (2019). Working memory and fluid intelligence as predictors of work engagement—Testing preliminary models. *Applied Cognitive Psychology, 33*(4), 596–616.

Kulikowski, K., & Orzechowski, J. (2018). All employees need job resources—Testing the job demands-resources theory among employees with either high or low working memory and fluid intelligence. *Medycyna Pracy, 69*(5), 483–496. https://doi.org/10.13075/mp.5893.00709

Larson, E. C. (2019). A meta-analysis of information processing measures of intelligence, performance, and group score difference, *CUNY Academic Works, 80*(7).

Latham, G. P., & Sue-Chan, C. (1999). A meta-analysis of the situational interview: An enumerative review of reasons for its validity. *Canadian Psychology, 40*, 56-67. doi:10.1037/h0086826

Letang, S. K., Lin, S. S., Parmelee, P. A., & McDonough, I. M. (2021). Ethnoracial disparities in cognition are associated with multiple socioeconomic status-stress pathways. *Cognitive research: principles and implications*, *6*(1), 64. https://doi.org/10.1186/s41235-021-00329-7

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*, 163-203.

McDaniel, M. A., Whetzel, A. M., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology*, *79*, 599-616. doi:10.1037/0021-9010.79.4.599

Mojzisch, A., Krumm, S., & Schultze, T. (2014). Do high working memory groups perform better? A conceptual approach linking individual differences in working memory capacity to group performance. *Journal of Personnel Psychology*, *13*(3), 134–145. https://doi.org/10.1027/1866-5888/a000111

Murphy, K.R., Cronin, B.E. and Tam, A.P. (2003). Controversy and consensus regarding the use of cognitive ability testing in organizations. *Journal of Applied Psychology*, *88*, 660-71.

Schipolowski, S., Wilhelm, O., & Schroeders, U. (2014). On the nature of crystallized intelligence: the relationship between verbal ability and factual knowledge. *Intelligence*, *46*(1), 156–168. https://doi.org/10.1016/J.INTELL.2014.05.014

Schmidt, F.L. (2002). The role of general cognitive ability and job performance: why there cannot be a debate. *Human Performance, 15*, 187-210.

Schmidt, F. L. (2015). Select on intelligence. *Handbook of Principles of Organizational Behavior,* 1–17. https://doi.org/10.1002/9781119206422.ch1

Smith, B. L., Smith, T. D., Taylor, L., & Hobby, M. (2005). Relationship between intelligence and vocabulary. *Perceptual and motor skills*, *100*(1), 101–108. https://doi.org/10.2466/pms.100.1.101-108.

Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 1096-1104.

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology, 44(1)*, 24-31.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*, 643-662.

Taylor, P. J., & Small, B. (2002). Asking applicants what they would do versus what they did do: A meta-analytic comparison of situational and past behaviour employment interview questions. *Journal of Occupational and Organizational Psychology, 75*, 277-294. doi:10.1348/096317902320369712

Verive, J. M., & McDaniel, M. A. (1996). Short-term memory tests in personnel selection: Low adverse impact and high validity. *Intelligence*, *23*(1), 15–32. https://doi.org/10.1016/S0160-2896(96)80003-1

Wang, H., Su, Y., Shang, S., Pei, M., Wang, X., & Jin, F. (2018). Working memory: A criterion of potential practicality for pilot candidate selection. *International Journal of Aerospace Psychology*, *28*(3–4), 64–75. https://doi.org/10.1080/24721840.2018.1535832

Wechsler, D. (2008). Wechsler Adult Intelligence Scale-Fourth Edition: Administration and scoring manual. San Antonio, TXNCS Pearson.