AN EXPLORATORY FACTOR ANALYSIS OF THE WOODCOCK-JOHNSON IV TESTS OF COGNITIVE ABILITIES AND TESTS OF ORAL LANGUAGE FOR THE 14- TO 19- YEAR OLD AGE RANGE

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

DEPARTMENT OF PSYCHOLOGY AND PHILOSOPHY

COLLEGE OF ARTS AND SCIENCES

BY

ANGELIA R. SPURGIN, B.A., M.A.T.

DENTON, TEXAS

AUGUST 2018

DEDICATION

This dissertation is dedicated to my life partner, Holly Fields. Without your consistent encouragement, support, and guidance, this enormous achievement would not have been possible. These past six years have profoundly challenged our relationship and our family; however, you were continuously there to cheer me on and pick me up when needed. You, and you alone, know the immense sacrifices I have been forced to make over these past several years as well as the emotional toll involved in this process. Because of your unwavering support, moving forward I know I can accomplish anything. You will always be my always.

ACKNOWLEDGMENTS

I would like to acknowledge the many people who have been pivotal in supporting me throughout my graduate school career. First, I would like to thank my parents, Danny Henson and Nathena Squyres. Dad, you always encouraged me to push myself beyond my capabilities and to never give up. You possess this unbelievable ability to ground me when I feel like everything is spiraling out of control. Mom, while I embarked on this journey after you left this Earth, you instilled within me a work ethic beyond reproach. You taught me the importance of not procrastinating and always putting forth my absolute best work. I miss you every single day, but I know you are proud of me. I am the person I am today because of both of you.

Next, I want to acknowledge the sacrifices my daughter, Caleigh Spurgin, has made throughout this process. There were many games, ceremonies, and celebrations I was unable to attend due to my unshakeable devotion to my graduate studies. Despite missing so many things over these past several years, you never once felt slighted or made me feel guilty. I pray that I have demonstrated for you what a strong, educated woman can accomplish. You are my daughter and you possess within you the same drive that pushed me all these years. I cannot wait to see what you accomplish.

Finally, I want to thank all the professors, supervisors, and mentors who have guided me along this journey. I have learned something valuable from each of you, and

this has contributed to molding me into the person I am today. Dr. Maricle, you have been my most challenging professor and always demanded my absolute best. The lesson you have taught me will forever push me toward what is just and what is right. Dr. DeOrnellas, I would never have returned to graduate school in school psychology without your profound influence on me over a decade ago. You rooted within me the importance of looking beyond the numbers and focusing on the whole child. Dr. Johnson, you taught me the importance of looking deeper. The problem-solving strategies I learned under you have assisted me in developing the resources needed to be an effective practitioner. Dr. Miller, you introduced me to the world of neuropsychology. Your unyielding focus and energy often served to motivate me to push myself further and dig a little deeper. Finally, Dr. McGill, I never would have embarked on this journey of factor analysis without your influence. While you are no longer at TWU, your impact on my understanding of testing and testing instruments will continue to be a strong pillar in my professional life.

ABSTRACT

ANGELIA R. SPURGIN

AN EXPLORATORY FACTOR ANALYSIS OF THE WOODCOCK-JOHNSON IV TESTS OF COGNTIVE ABILITIES AND TESTS OF ORAL LANGUAGE FOR THE 14- TO 19- YEAR OLD AGE RANGE

AUGUST 2018

Not only has intelligence been an elusive construct, but methods for measuring it continue to be hotly debated in present day. A multitude of theories currently exist that attempt to objectively explain the mechanisms of intelligence, but the fact remains that any discussion regarding intelligence is theoretical in nature. In an attempt to understand the concept of general intelligence, numerous psychologists and researchers have attempted to quantitatively define this intangible paradigm through various forms of assessment. Accurate test interpretation centers around how well a test measures the construct it contends to measure. The Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG) and Tests of Oral Language (WJ IV OL) are two testing batteries in the Woodcock-Johnson IV (WJ IV) that purportedly measure general intelligence as well as seven broad cognitive factors. The publishers of the WJ IV denote that this most recent iteration of the test is based on modern Cattell-Horn-Carroll (CHC) theory as well as advances in neuropsychological research. Of concern, the WJ IV Technical Manual is exceedingly complex and the results from the presented studies supporting the factor structure of the WJ IV are profoundly obscured for the general practitioner. The primary purpose of this study is to provide an objective analysis of the WJ IV COG and WJ IV OL for the 14- to 19- year old age group to determine the factor structure of the assessment battery. Data analyses included an exploratory factor analysis utilizing the correlation matrix provided in the technical manual by the test publishers. Results from this study indicate the WJ IV COG and WJ IV OL measure five broad CHC factors in the identified age range: comprehension knowledge (G*c*), short-term working memory (G*wm*), auditory processing (G*a*), processing speed (G*s*), and a final factor that incorporates fluid reasoning (G*f*), long-term storage and retrieval (G*lr*), and visual processing (G*v*) together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The evolution of cognitive assessment is relatively complex; yet, the consistent utilization of various testing instruments is customary practice throughout the world (Boake, 2002). Presently, there are numerous assessment batteries available to practitioners to measure intelligence. The Woodcock-Johnson (WJ) is one such instrument that has been in the field since the late 1970s and continues to be utilized in the education system (McGrew, LaForte, & Schrank, 2014).

A conundrum currently surrounds the methodology involved in test interpretation and disability identification. For example, evaluators may qualify a child for special education using the strengths and weaknesses model via subtest profile analysis, the discrepancy model, or data collected through Response to Intervention (RtI; McGrew & Wendling, 2010; Watkins, Glutting, & Youngstrom, 2005). Moreover, the hallmark of test interpretation revolves around the testing instruments themselves. Considering the high stakes associated with regular education as well as special education identification, it remains imperative that practitioners understand the potential limitations surrounding testing instruments.

This chapter provides a brief review of the literature pertaining to the history of intellectual assessment, intelligence theory, and test interpretation. Additionally, the Woodcock-Johnson assessment family is presented and a brief discussion regarding each instrument is provided. Finally, the chapter concludes with an overview of the rationale and significance for the present study.

Brief Literature Review

To appreciate the current theory behind modern intelligence assessment, it is essential to review how this area came into existence. Understanding the history of test development, intelligence theory, and interpretative procedures provides a conceptual foundation for evaluating the evolution of the Woodcock-Johnson test family. The following discussion addresses the historical context of intelligence testing through an empirical lens as opposed to a philosophical approach.

Intellectual Assessment

Prior to discussing the theory behind intelligence, it is critical to understand how researchers began to assess this abstract construct. Originally, the middle of the 19th century witnessed intelligence testing that was centrally focused on discerning between bodily dimensions and sensations. Eventually, this transitioned to examining physiological and psychological characteristics in men and women. The areas of interest were not directly identified as intelligence, rather research targeted how well one's body performed under a variety of different situations. This performance then provided information regarding the future potential of that individual (Wasserman, 2012).

The beginning of the 20th century marked the inception of modern-day intelligence testing. The father of intellectual assessment, Alfred Binet, was a pioneer in the evaluation of intelligence and paved the way for testing in the school setting. Binet and Theodore Simon jointly developed an instrument used to identify intellectually deficient children within the school system in France. The Binet-Simon Intelligence Scale included measures of language, auditory and visual processing, memory, and problem solving (Boake, 2002; Wolf, 1961). Eventually, Henry Goddard and Elizabeth Kite Goddard translated Binet and Simon's intelligence test into English. Once translated, Lewis Terman conducted extensive research in the United States to revise the intelligence scale. In 1916, Terman published the Stanford-Binet Tests of Intelligence (Terman, 1916) which became one of the most well-known intelligence tests used in the 20th century. The Stanford-Binet founded the use of a single score that represented a general intelligence factor (Boake, 2002; Terman & Merrill, 1937).

The commencement of large scale assessment for individuals became prominent during World War I with the Army mental tests: Group Examinations Alpha and Beta. These tests determined the potentiality of recruits and examined a variety of basic functions (Wasserman, 2012). Additionally, these two batteries allowed for the evaluation of recruits who were fluent English speakers as well as those who were illiterate or non-English speakers. The Army Alpha and Beta tests were the first large scale evaluations conducted and had a profound impact on testing in the education system.

Prior to the war, a majority of the research conducted on intelligence testing centered on children. However, David Wechsler acknowledged and tackled the profound shortage of adult research and assessment tools available. His first adult assessment battery, the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939) provided a Full Scale, Verbal, and Performance Intelligence Quotient. In addition to the Stanford-Binet, the Wechsler intelligence tests continue to dominate the field of assessment (Boake, 2002; Wasserman, 2012).

Intelligence Theory

Throughout history, the conceptualization of intelligence has been cogitated by numerous philosophers; however, the current study examines intelligence theory as it pertains to cognitive assessment. To date, there is not one all-encompassing definition regarding intelligence or the components of intelligence (Boake, 2002). Intelligence theory will continue to transform as neuropsychological research advances and the factors of intelligence become more tangible.

Early research into intelligence theory was marked with strongly competing views, yet they all played a significant role in the evolution of modern-day theory. Charles Spearman is the researcher most closely associated with general intelligence or *g*. He originally hypothesized an empirical, two-factor model of intelligence: a general factor known as *g* and specific factors identified as *s*. He posited that *g* represented the shared variance and *s* represented the unique variance within a testing instrument (Spearman, 1904, 1923). His seminal studies provided the initial groundwork for the quantitative approach to studying intelligence through assessment procedures (Guilford, 1967). Louis Thurstone challenged Spearman's *g* and proposed intelligence was best described as a vast collection of cognitive abilities. Thurstone pioneered the use of multivariate statistics and factor analysis in his study of the psychometrics of intelligence assessment (Wasserman, 2012). Additionally, Thurstone advocated that intelligence was better conceptualized as a collection of primary mental abilities (PMA); thus, laying the foundation for the concept of broad and narrow abilities (Thurstone, 1938, 1948). Philip Vernon was the first researcher to hypothesize that intelligence was hierarchical in nature. He acknowledged the existence of a general factor, *g*, in addition to two lower-order factors: verbal and spatial abilities (Vernon, 1950, 1969). Vernon's work played a pivotal role in more current conceptualizations of intelligence.

The early psychometric work of Spearman, Thurstone, and Vernon was critical in providing the basis for the hierarchical model of intelligence. This tiered approach postulates that intelligence is composed of many related factors that maintain their independence from each other (Carroll, 1993; Fabrigar & Wegener, 2012; Horn & Cattell, 1966). While there are other theories of intelligence, the current study examines the evolution of the hierarchical model.

Raymond Cattell extended the work of Thurstone and introduced a dichotomous theory of intelligence known as the fluid-crystallized (G*f*-G*c*) theory of intelligence. The hallmark of his theory focused on two highly correlated factors of intelligence, fluid and crystalized ability, rather than a single intelligence factor (Cattell, 1943, 1944). Roughly two decades after Cattell's original proposition of intelligence, John Horn further advanced the G*f*-G*c* theory of intelligence. While Horn's research supported fluid and crystallized ability, he also identified additional factors that make up overall intelligence.

Throughout their work together, Cattell and Horn expanded G*f*-G*c* theory to include a total of nine broad, second-order abilities (Cattell & Horn, 1978; Horn, 1968; Horn & Blankson, 2012).

Building upon preceding investigations into cognitive abilities, John Carroll postulated a hierarchical model of intelligence consisting of three tiers. The top tier represented a general intelligence factor, *g*, with the second order broad factors or abilities in the middle, and finally the first order narrow abilities on the bottom (Carroll, 1993). Carroll acknowledged that while each of the second order factors maintained their own uniqueness, they also shared some variance resulting from *g* (Bickley, Keith, & Wolfe, 1995). Carroll's three stratum theory of intelligence was seminal to creating a method of identifying and classifying cognitive abilities.

Through extensive research and debate, Cattell and Horn's G*f*-G*c* theory and Carroll's three stratum theory eventually amalgamated together to create a modern theory of intelligence, Cattell-Horn-Carroll (CHC) theory (McGrew, 1997). Today, CHC theory continues to evolve through new research; however, it allows for a common taxonomy to identify the proposed components of intelligence. While this is not the only current theory used to describe intelligence and its subsequent constructs, CHC theory is a prominent and respected theory in the field of psychology and test development. Furthermore, for nearly two decades, many cognitive assessment instruments have utilized CHC theory in some capacity during test conceptualization and interpretation (Flanagan & Dixon, 2013).

Test Interpretation

In addition to assessment and intelligence theory, test interpretation has also witnessed dramatic revisions over the past century. Initially interpretation was solely based on qualitative evaluation from the examiner, but has progressed into more sophisticated processes. Test interpretation has been identified as progressing through four separate and distinct stages (Kamphaus, Winsor, Rowe, & Kim, 2012).

The first wave of interpretation witnessed the development of criterion referenced testing, norming procedures, and identification based on scores. The Simon-Binet Intelligence Scale initiated the inception of these interpretative procedures and tended to focus on a single score to represent intelligence (Wasserman, 2012). A more multidimensional approach marked the second era of interpretation. Administrators began to compare subtests as well as individual testing scores. Also, specific diagnoses began to emerge during this time and clinicians began to compare intra-individual cognitive strengths and weaknesses (Kamphaus et al., 2012). More statistical analyses regarding interpretation ensued during the third wave. Factor analysis and psychometric profile analysis began to be utilized. This allowed for a more quantitatively sound interpretative process for intelligence testing (Fruchter, 1954). The fourth wave of interpretation witnessed a materialization of theory guiding test development. This contributed to more selective assessment procedures and a more evidenced-based interpretative process for practitioners (Horn & Blankson, 2012). Furthermore, there has been an extension to the previous generations of test interpretation to include a much broader, holistic conceptualization of the individual. Best practice in the field encourages a test administrator to look beyond test scores when completing an evaluation and to conduct an interpretation through a more developed lens (Kamphaus, 2001; McDermott, Watkins, & Rhoad, 2014; Watkins, 2009).

The Woodcock-Johnson Tests

The inception of the Woodcock-Johnson (WJ) testing family began in the 1970s and the series continues to be a well-respected assessment tool used by clinicians. The WJ has undergone three revisions, with the most recently released fourth edition of the test in 2014. The current study examines the factor structure of the most current iteration of the WJ cognitive and oral language battery; therefore, it is important to trace the evolution of the WJ assessment family beginning with the first testing battery.

The first generation of the WJ family was pioneered by Dr. Richard Woodcock and is known as the Woodcock-Johnson Psycho-Educational Battery (WJPEB; Woodcock & Johnson, 1977). This instrument identified no theoretical foundation for its development; rather, the test stemmed from many years of experimentation and data collection by Dr. Woodcock regarding learning and the components of learning (Schrank, Decker, & Garruto, 2016). The Woodcock-Johnson Psycho-Educational Battery – Revised (WJ-R; Woodcock & Johnson, 1989) was the next edition in the testing family and included a cognitive measure (WJ-R COG) and an academic measure (WJ-R ACH). The test creator, Dr. Woodcock, asserted the WJ-R was a functional test that accurately conceptualized Cattell and Horn’s fluid-crystallized (G*f*-G*c*) theory of intelligence (Horn

& Cattell, 1966; Woodcock, 1990). Specifically, through detailed analytic procedures, the WJ-R was modeled to accurately represent the G*f*-G*c* theory of intelligence (McGrew, Werder, & Woodcock, 1991).

In 2001, the Woodcock-Johnson III (WJ III; Woodcock, McGrew, & Mather, 2001, 2007) made its debut. This was the third generation in the WJ family and included a cognitive and academic measure. The WJ III was postulated to be the first intelligence test solely grounded in Cattell-Horn-Carroll (CHC) theory (McGrew & Woodcock, 2001; Schrank, Miller, Wendling, & Woodcock, 2010; Schrank & Wendling, 2012). The WJ III was proposed to be an assessment tool that operationally represented CHC theory. This version of the WJ was the first to include a general intelligence score known as General Intellectual Ability (GIA). The GIA score was described as representative of Spearman's *g* and was identified as the apex of the hierarchy of cognitive abilities (McGrew & Woodcock, 2001). Interpretation also included up to nine broad cognitive factors in the second tier and numerous narrow factors in the third tier. The WJ III was subjected to widespread critical analyses from the time it debuted. While there was extensive support presented for the WJ III through a variety of studies conducted during the norming procedures as well as after its release (Floyd, McGrew, Barry, Rafael, & Rogers, 2009; Keith & Reynolds, 2010; McGrew & Woodcock, 2001; Taub & McGrew, 2004), there was also critical examination of the methodology used in these studies. Furthermore, there were numerous articles hypothesizing that the factor structure proposed by the authors of the WJ III lacked support through additional, external analyses; thus, concluding that the WJ III was potentially over factored (Dombrowski, 2013, 2015; Dombrowski & Watkins, 2013; Strickland, Watkins, & Caterino, 2015).

The most recent revision, the Woodcock-Johnson IV (WJ IV; Schrank, McGrew, & Mather, 2014a) was released in 2014 and underwent numerous alterations prior to its release. In addition to the cognitive and academic measures, the publishers also included a battery that assesses oral language (WJ IV OL; Schrank, Mather, & McGrew, 2014b). According to the test developers, the WJ IV was developed based on the most contemporary understanding of CHC theory as well as current neuropsychological research. In addition to the GIA score, the authors indicate there are up to ten broad factors that can be measured dependent on the administered subtests, as well as other composite scores available for interpretation (McGrew et al., 2014).

Purpose and Significance of the Study

Children entering adolescence are faced with a multitude of changes in cognitive development (Watson & Gable, 2013). Comprehensive school evaluations are essential for successful identification of a disability; thus, allowing for meaningful and targeted evidenced-based interventions. While effective and reliable test interpretation is important, clinicians must understand the limitations of intelligence testing. Assessing latent, or unobservable, constructs of intelligence can only be guided by theory; therefore, there are inherent limitations resulting from working with theoretically grounded constructs.

The Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014b) and Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b) are presented as measuring seven broad CHC factors when utilized together: comprehension-knowledge (G*c*), fluid reasoning (G*f*), short-term working memory (G*wm*), long-term storage and retrieval (G*lr*), visual processing (G*v*), auditory processing (G*a*), and processing speed (G*s*). The information provided in the technical manual regarding the factor structure of the WJ IV is extremely dense and complicated for the traditional clinician. Furthermore, the discussion regarding the factor structure explicitly focuses on the full battery and not the individual batteries (e.g., cognitive, academic, or oral language). The data analyses conducted during the norming procedures for the WJ IV are provided in the technical manual, but are fairly limited in scope and appear to lack transparency.

For the current study, the data analyses incorporate an exploratory factor analysis (EFA) of the correlation matrix found in the Technical Manual (McGrew et al., 2014) for the adolescent standardization sample. Exploratory factor analysis (EFA) is a quantitative procedure used to examine the factor structure of an instrument without any preconceived hypothesis; essentially, the data drive the results (Fabrigar & Wegener, 2012). There are several varieties of EFA available to researchers; however, this study utilizes an iterated principal axis factoring with promax rotation.

The purpose of the present study was to examine the factor structure of the WJ IV COG and WJ IV OL for the 14- to 19- year old age range. The following research question and hypothesis were proposed:

Research Question:

> When utilizing exploratory factor analysis, what is the factor structure of the WJ IV COG and WJ IV OL in the 14- to 19- year old age group?

Hypothesis:

> The results from the exploratory factor analysis will *not* support the proposed seven CHC broad factors identified by the test publishers.

CHAPTER II

REVIEW OF THE LITERATURE

The purpose of this chapter is to provide a systematic and thorough review of the literature pertinent to this study. The Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014b) and the Woodcock-Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b) are based on the most current conceptualization of the Cattell-Horn-Carroll (CHC) theory of intelligence (McGrew et al., 2014). Modern CHC theory developed through many years of research into intelligence and intelligence assessment. The proposed study examines the factor structure of the WJ IV COG and WJ IV OL; thus, it is imperative to trace the evolution of intelligence theory to garner an understanding of the development of current CHC theory and its utilization in testing.

The areas discussed in this chapter include a historical review of seminal intelligence tests employed in the field, an examination of the evolution of intelligence theory, and current theoretical perspectives regarding intelligence. There is also a discussion regarding the interpretation of assessment tools, as well as a thorough analysis of each Woodcock-Johnson assessment battery up to the most recent iteration of the test. Finally, the chapter concludes with a discussion of the rationale for the study.

Development of Intellectual Assessments

According to Raymond B. Cattel (1998), the commencement of modern theoretical understanding of intelligence began in the early 20th century; however, the inception of quantitative study on intelligence began in the middle to late 1800s. For the purpose of this literature review, it is important to envision intelligence from an empirical framework; therefore, a discussion regarding the development of assessment tools and the role this played in understanding intelligence is needed prior to discussing intelligence theory.

Francis Galton

Francis Galton is credited with the beginning of the empirical assessment era with his Anthropometric Laboratory in England. While not explicitly identified as intelligence assessment, Galton conducted a series of tests on willing participants that measured a variety of bodily dimensions and basic sensory discrimination. Galton believed the results from his testing battery provided an understanding to the basis of one's intellectual capability. Additionally, his research provided the foundation for the normal distribution curve, which continues to be used for conceptualizing and describing human characteristics and performance. Galton was also a pioneer in scientific data collection through his use of test batteries, comprehensive questionnaires, utilizing control groups, and implementing statistical analyses (Wasserman, 2012).

James McKeen Cattell

James McKeen Cattell further advanced Galton's work and the Anthropometric Laboratory; however, he focused more on the physiological and psychological characteristics of an individual (Sokal, 2009). Cattell was a pioneer in reading instruction research and he diligently worked to establish psychology as a scientific field (Cattell, 1893). Despite his advancements in testing and measurement, Cattell never decisively stated his research focused on the concept of intelligence; rather, he posited that the results from his assessments more aligned with long-term student achievement and success (Wasserman, 2012).

Alfred Binet

Alfred Binet's work on intellectual assessment signaled the beginning of the contemporary testing movement. Binet was a French psychologist who many identify as the father of intellectual assessment (Wasserman, 2012). While Binet contributed greatly to furthering the advancement of intelligence tests, he was not well-known in the academic and research communities and preferred to work independently. As a result of his more introverted nature, Binet frequently assessed the cognitive functioning and personality of his two daughters (Wolf, 1961, 1964). This approach to studying intelligence paved the way for the future empirical study of intellectual assessment.

Binet originally posited a hierarchical approach to intelligence; however, his test did not reflect this theoretical hypothesis (Wasserman, 2012). In 1904, the Minister of Public Instruction appointed Binet with the task of determining a method to differentiate typical children from intellectually deficient children (Wolf, 1961). Binet worked alongside Theodore Simon and in 1905, they developed the first assessment tool for intelligence, the Binet-Simon Intelligence Scale. This test underwent multiple revisions over the years and included scales of language, auditory and visual processing, learning and memory, judgement, and problem solving (Boake, 2002; Wasserman, 2012).

Lewis Terman

Lewis Terman, an American psychologist who was employed at Stanford University, is credited with further advancing the Binet-Simon Intelligence Scale in the United States. Once the scale was translated into English by Henry Goddard and Elizabeth Kite Goddard, Terman used the Binet-Simon Intelligence Scale to assess a large sample of children at various cognitive levels. Through research and analyses, he adjusted and revised many of the scales present on the Binet-Simon assessment battery (Wasserman, 2012).

In 1916, after extensive research with a large standardization sample, Terman created the Stanford-Binet Tests of Intelligence (Terman, 1916), one of the most widely used intelligence tests of the first half of the 20th century. The original Stanford-Binet consisted of 90 items: over half of which were extracted from the original Binet-Simon Scale and the remainder were created directly from Terman's research. Furthermore, unlike its predecessors, the Stanford-Binet Tests of Intelligence provided a single, unitary score that represented overall intelligence (Boake, 2002; Terman & Merrill, 1937). The Stanford Binet Scales of Intelligence is now in its fifth iteration (SB-5; Roid, 2003).

Army Mental Tests: Alpha and Beta

While the discussion regarding intelligence test development has focused on various contributions by specific researchers, World War I provided the backdrop for large scale assessment. This approach to testing laid the foundation for future testing in the school system, a process that continues today. The specific goal of the Army mental tests was to differentiate between acceptable and unacceptable recruits. Additionally, the results allowed the Army to determine the future potential of acceptable recruits, such as officer or soldier (Wasserman, 2012).

Arthur Otis, a former student of Lewis Terman, led a group of psychologists in developing and norming the Group Examinations Alpha and Beta. Group Examination Alpha was geared toward recruits who spoke English and could read and write. Group Examination Beta was developed for non-English speaking individuals or those who were illiterate (Boake, 2002; Wasserman, 2012). While extremely controversial at the time, the Army mental tests greatly contributed to the process of intelligence testing and were crucial in developing norming and research procedures for subsequent assessment batteries (Boake, 2002; Kamphaus et al., 2012; Wasserman, 2012). The Army Alpha and Beta evolved over time into the current measure being used by the American military, the Armed Services Vocational Aptitude Battery (ASVAB).

David Wechsler

David Wechsler was another instrumental figure in the advancement of intelligence testing. He was an American psychologist who originally worked as an examiner for the Army mental tests. This position allowed Wechsler the ability to study and understand intelligence assessment, as well as make vital connections with other researchers in the field, while avoiding the draft for World War I (Wasserman, 2012). Throughout his early years as a psychologist, he worked in a variety of psychiatric hospitals and other clinical settings conducting diverse types of assessments. Wechsler noted there was a shortage of assessment instruments available for use with adults. Furthermore, the assessment tools that were available for adult populations, such as the Stanford-Binet, reflected poor norming and research procedures for that specific population (Boake, 2002).

As a result of his test development work, Wechsler was profoundly influential in advancing the study of intelligence. The Wechsler tests were predominately used in the latter part of the 20th century, eventually overtaking the Stanford-Binet (Boake, 2002). His first assessment battery, the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939), was an adult assessment that provided three broad measures of intellectual functioning: Full Scale, Verbal, and Performance Intelligence Quotients (IQ). Additionally, the test provided a new and more psychometrically sound IQ score known as the deviation IQ, which is the difference in scores from the standardized mean. Wechsler went on to create numerous assessment batteries for diagnosing and evaluating intelligence: the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949, 1991, 2003, 2014), Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955, 2008), and Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 1967, 1989, 2002, 2012).

Summary: Development of Intellectual Assessments

Intelligence assessment began with the first empirical study of intelligence in Francis Galton's Anthropometric Laboratory in England. Galton examined intellectual capacity from a series of tests measuring basic bodily dimensions and sensory discrimination. He also pioneered the use of testing batteries, questionnaires, control groups, and statistical analyses. James McKeen Cattell focused on the physiological and psychological makeup of an individual. He further advanced the field of psychology as a scientific discipline by assessing long-term academic success in students. Alfred Binet was significant to the more modern approach to intelligence assessment. He was assigned the task of differentiating typical and atypical children within the school system in France. Binet and his colleague, Theodore Simon, are credited with developing the first true intelligence test. Lewis Terman, an employee at Stanford University, is credited with the development of the American version of the Binet-Simon scale, known as the Stanford-Binet. The Stanford-Binet became one of the most widely used intelligence tests in the early 20^{th} century. Large scale assessment, still present today in the education and industrial systems, originated from the Army mental tests: Alpha and Beta. These assessments were designed to determine the acceptability of potential Army recruits. The Alpha test was designed for English recruits who were literate, while the Beta test was created for non-English speaking individuals or those who were illiterate. David Wechsler furthered the field of adult intelligence testing with his Wechsler intelligence tests. He greatly influenced the assessment of intelligence through his work in various psychiatric settings. The Wechsler scales continue to be used in modern day assessment.

Evolution of Intelligence Theory

Prior to discussing the evolution of intelligence theory, it is important to note that the concept of intelligence has long been a philosophical question. Early Greek, Roman, and Chinese philosophers debated the nature of humans and of humans' unique intellectual capabilities during the golden age of philosophy. Further, the nature of intellectual prowess among humankind continues to be discussed within the field of philosophy (Benjamin, 2009; Murdoch, 2007). However, this literature review will examine the evolution of intelligence theory as it strictly pertains to cognitive assessment.

Currently, there is not a single, agreed upon understanding of what defines a person's intelligence (Boake, 2002). Some posit any definition of intelligence is biased from the onset because researchers attempt to define intelligence by inadvertently examining their *own* personal strengths (Wasserman. 2012). In the middle of the 19th century, Herbert Spencer assimilated his understanding of intelligence with Charles Darwin's theories of evolution. Spencer suggested intelligence may be the process of internally adapting to the external environment for the sole purpose of survival (Spencer, 1855). While this understanding of intelligence was underdeveloped on many levels, it served as the inauguration of intelligence theory research. A review of relevant psychological research pertaining to the understanding of intelligence from a psychometric perspective follows.

Charles E. Spearman

Initially, intelligence was studied from a very qualitative framework; however, the past century has witnessed a transition to a more measurable, quantifiable approach known as psychometrics. Charles Spearman was a British psychologist who hypothesized a two-factor, psychometric theory of intelligence. While many researchers were attempting to conceptualize intelligence from a theoretical orientation, Spearman approached the concept of intelligence from a statistical framework and attempted to explain intelligence through empirical research (Boake, 2002; Wasserman, 2012). Specifically, Spearman posited there was a general factor, *g*, which accounted for the shared variance within a test and specific factors, *s*, which accounted for the unique variance within a test (Guilford, 1967; Spearman, 1904, 1923).

In 1904, Spearman published *General Intelligence, Objectively Determined and Measured* which served to lay the foundation for quantitatively measuring cognitive abilities. In this paper, Spearman discussed a series of complex experiments conducted on children in a village school and a preparatory school in Berkshire. Spearman examined several different cognitive areas in male and female students including sensory discrimination and memory. He attempted to identify positive intercorrelations between scores in four areas of recognized intelligence: present efficiency, native capacity, impressions from others, and common sense. Present efficiency included intelligence in areas of schooling such as foreign language and mathematics. Native capacity considered the difference in school ranking as compared to the child's age. Impressions examined the teacher's perceptions of the student's performance in the classroom. Finally, common sense related to the perceptions of a student's abilities outside of a school setting (Spearman, 1904).

This experiment and the subsequent articles resulting from this research indicated the genesis of quantitatively measuring intelligence. Spearman identified the importance of accounting for measurement error and identified there was a common shared variance among the cognitive tests, which he identified as *g* (Boake, 2002; Spearman, 1904; Wasserman, 2012). While Spearman suggested a general factor of intelligence existed through the psychometric *g*, he never hypothesized that *g* accounted for all the individual intricacies involved in intelligence (Spearman, 1923). Spearman laid the foundation for the empirical study of intelligence; however, his two-factor theory was challenged, altered, and evolved through the continued work of other researchers.

Louis L. Thurstone

While Spearman advanced the statistical study of intelligence from a very unidimensional approach with the concept of a single factor other researchers examined intelligence from a much broader framework. Louis Thurstone was an American psychologist and statistician who greatly contributed to the area of psychometrically studying intelligence. Thurstone challenged the suggested concept that intelligence could be singularly defined using Spearman's *g*; rather, he posited that intelligence was better described as a collection of cognitive abilities (Wasserman, 2012).

After the publication of Spearman's work on intelligence, Thurstone pioneered the use of multivariate statistics and factor analysis to better understand the psychometric structure of intelligence (Wasserman, 2012). Factor analysis allows for the evaluation of an ample collection of scores from a test and attempts to parsimoniously identify underlying constructs, or factors, present on the instrument (Fabrigar & Wegener, 2012). Thurstone further advanced factor analysis and targeted the extraction of factors that were independent, or orthogonal, from one another (Thurstone, 1947). Utilizing this extraction method, Thurstone conducted several studies and used the results to better understand the multidimensionality of intelligence. Through this line of research, he challenged the belief that intelligence was a single factor; rather, Thurstone hypothesized that intelligence was composed of several primary mental abilities (Thurstone, 1948).

Thurstone identified eight primary mental abilities (PMA) as significant contributors to one's overall intelligence: verbal comprehension (V), word fluency (W), number facility (N), memory (M), visualizing or space thinking (S), perceptual speed (P), induction (I), and speed of judgement (J; Thurstone, 1938; Wasserman, 2012). This categorization of mental abilities was groundbreaking at the time and served as the underpinning for subsequent research into broad and narrow abilities.

While he hypothesized intelligence was composed of multiple factors, Thurstone eventually came to acknowledge the possibility of Spearman's general intelligence factor (Thurstone, 1947). However, despite recognizing the existence of *g*, Thurstone argued that utilizing a single score to describe one's intelligence was limiting, and he proposed the importance of examining the primary mental abilities. Essentially, he pioneered the evaluation of individual strengths and weaknesses using scores obtained from the cognitive abilities profile (Thurstone, 1948).

Philip E. Vernon

Further expounding upon Spearman and Thurstone's work, Philip Vernon proposed the first hierarchical model of intelligence (Guilford, 1967). A hierarchical model of intelligence posits that while there is an overall general intelligence factor, *g*, there are also additional broad factors that are influenced by *g*. While this is a simplistic description of the hierarchical model of intelligence, a more detailed discussion will be presented in a subsequent section.

Philip Vernon was a Canadian psychologist who proposed that intelligence was composed of a single higher-order factor, *g*, and two specific lower-order group factors, verbal/educational (v:ed) ability and spatial/mechanical (k:m) ability (Wasserman, 2012). Through his research, Vernon posited that v:ed and k:m played an important role in more narrowly identified factors. Specifically, he stated that v:ed primarily influenced verbal, numerical, reasoning, attention, and fluency abilities, and k:m influenced spatial, mechanical, coordination, reaction, and other practical abilities (Vernon, 1950; Wasserman, 2012). Vernon was aware his dichotomous intelligence hypothesis required further elaboration and development (Vernon, 1969); however, his insight contributed to more current theoretical understandings of intelligence.

Summary: Evolution of Intelligence Theory

The debate of what defines intelligence has existed for centuries. While it has been traditionally viewed from a philosophical construct, the current study examines intelligence theory from a psychometric approach. Charles Spearman represented the genesis of the empirical study of intelligence and he was the first to propose a two-factor approach to intelligence. He posited that intelligence was composed of a general factor, *g*, and specific factors, *s*. Spearman stated *g* accounted for the shared variance on an instrument, while *s* accounted for the unique variance on an instrument. Louis Thurstone hypothesized that intelligence was not one single factor, *g*, but rather a collection of cognitive abilities which he identified as primary mental abilities. Thurstone pioneered the use of multivariate statistical analyses in his attempt to quantify intelligence. Furthermore, his work marked the commencement of a systematic evaluation of cognitive strengths and weaknesses garnered from a cognitive abilities profile. Philip Vernon is credited with being the first psychometrician to propose a hierarchical model of intelligence. He postulated there was an overall general intelligence factor (*g*) as well as broad factors that were influenced by *g*.

Current Theoretical Perspectives on Cognitive Abilities

While many researchers made significant contributions regarding the psychometric approach to understanding intelligence, Spearman, Thurstone, and Vernon were pivotal in laying the foundation for current perspectives regarding cognitive abilities. More recent and accepted theories of intelligence hypothesize it is composed of many related factors, yet the factors remain independent of each other. This is commonly known as the hierarchical model of intelligence (Carroll, 1993; Fabrigar & Wegener, 2012; Horn & Cattell, 1966).

A hierarchical model includes an overall general intelligence factor, *g*, as well as multiple cognitive abilities or factors. The foundation for this model was derived from the integration of Spearman's general factor of intelligence and Thurstone's identification of primary mental abilities (Spearman, 1923; Thurstone, 1938). To begin, the third, or highest order factor, is indicative of overall intelligence and is represented by Spearman's *g*. This is then followed by the second order factors which are known as broad abilities. These factors are independent of each other; however, they share a common variance due to the influence of *g*. Finally, the first order factors are located at the bottom and these are identified as narrow abilities. Again, these factors may be independent of each other, but they continue to be influenced by the second and third order factors (Wasserman, 2012).

Several models of intelligence have been posited over the past century; however, for this study, the evolution of the hierarchical model of intelligence as related to modern Cattell-Horn-Carroll (CHC) theory will be further explored. As previously stated, the WJ IV COG and WJ IV OL are based on the current interpretation of CHC theory of intelligence and it is important to review this theory from its inception.

Fluid-Crystallized Theory of Intelligence

Raymond B. Cattell further advanced the work of Thurstone and was pivotal in laying the foundation for current CHC theory. Cattell was a British psychologist who moved to the United States to conduct research. He initially proposed a dichotomous concept of intelligence known as fluid-crystalized (G*f*-G*c*) theory at the American Psychological Association (APA) conference in 1941. During his presentation, he did not identify a single factor of intelligence, rather Cattell hypothesized there were two highly correlated factors of intelligence (Brown, 2016; Flanagan & Dixon, 2013). It is important to note that at this same conference, the modern expression of general intelligence and broad factors was agreed upon and continues today. APA adopted that the general factor of intelligence would be identified by a lowercase *g*, while the broad factors of intelligence would be identified by an uppercase G followed by the appropriate letter representing the broad factor (Wasserman, 2012).

As previously stated, Cattell's fluid-crystallized theory initially conceptualized intelligence as a two-factor structure: fluid ability (G*f*) and crystalized ability (G*c*). Cattell acknowledged that various internal and external factors impacted the two components of intelligence, such as inheritability, neurological development and insult, and one's cultural exposure (Cattell, 1943, 1944, 1963, 1998; Cattell & Horn, 1978). While the definition and understanding of both factors has shifted over time, Cattell originally described G*f* as a general ability to differentiate between the relationships of objects that does not incorporate prior knowledge. He posited that fluid ability improved through adolescence and early adulthood, but then slowly declined with age. Additionally, Cattell suggested that G*f* represented the general factor of intelligence. Crystallized ability was hypothesized to be an extension of fluid ability, especially in adulthood; however, Cattell suggested G*c* evolved more from one's previous knowledge and required minimal personal insight. He posited that G*c* continued to develop through adulthood, but decline began around age 70 (Cattell, 1943, 1944). After the preliminary research on G*f*-G*c* theory in the early 1940's, Cattell did not revisit the theory until roughly twenty years later (Brown, 2016).

In 1963, Cattell published a seminal experiment which served to further enhance and develop his fluid-crystalized theory of intelligence (Cattell, 1963). Cattell conducted a factor analysis on the collected data from his experiment to determine if one or two factors were present on a variety of administered intelligence assessments: Thurstone Primaries, Institute of Personality and Ability Testing (IPAT) Culture Fair Intelligence Test, and IPAT High School Personality Questionnaire (HSPQ). His sample consisted of 277 male and female seventh and eighth grade students. Cattell's findings supported his original premise of two factors of intelligence: fluid and crystallized abilities. Cattell stated the conclusions endorsed his original supposition that fluid ability is "basic general intelligence" (p. 20) that continues to be impacted over time by age and cultural exposure, thus explaining the high intercorrelation with crystallized ability (Cattell, 1963).

In the later 1960s, John Horn, a student of Raymond Cattell, further expanded the fluid-crystalized theory of intelligence while working alongside his mentor. Horn's dissertation was seminal in reproducing previous results obtained from experiments involving Cattell's G*f*-G*c* theory of intelligence. His findings indicated there were more than two second-order factors that could be extracted from the data. Specifically, in addition to fluid and crystallized intelligence, Horn identified visualization (G*v*), speediness (G*s*), facility (F), carefulness (C), premsia (PRM), and positive self-image (PSI), as well as two additional factors that he did not readily identify (Flanagan & Dixon, 2013; Horn, 1965). Horn's dissertation served to lay the groundwork for the updated and expanded version of the G*f*-G*c* theory of intelligence; thus, eventually leading to contemporary CHC theory.

Throughout the next several decades, Horn and Cattell worked together to refine the G*f*-G*c* model and identify additional second-order factors. Over the course of their research, the theory of fluid-crystalized intelligence was extended to include nine broad, second-order abilities: fluid intelligence (G*f*), crystallized intelligence (G*c*), short-term acquisition and retrieval (G*sm*), visual processing (G*v*), long-term storage and retrieval (G*lr*), processing speed (G*s*), auditory processing (G*a*), decision speed (G*t*), and quantitative reasoning (G*q*; Cattell & Horn, 1978; Horn, 1968; Horn & Blankson, 2012). A broad reading-writing ability (G*rw*) was later identified; however, it originated from the work of Richard Woodcock (Flanagan & Dixon, 2013). While the G*f*-G*c* theory of intelligence was not the sole contributor to modern day CHC theory, it was pivotal in laying the foundation for the current understanding of intelligence.

Three Stratum Theory of Cognitive Abilities

While the G*f*-G*c* model of intelligence was considered seminal research regarding cognitive ability, another American statistician provided strong insight into the current conceptualization of intelligence. Unlike Horn and Cattell, John B. Carroll postulated that intelligence was best represented by a three-tiered, hierarchical model that consisted of a single general factor in addition to other specific cognitive abilities (Bickley et al., 1995). Carroll is credited with developing the three-stratum theory of intelligence, the other core component of CHC theory (Carroll, 1993).

In 1993, Carroll published a highly influential book that forever altered the understanding of intelligence. *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* (Carroll, 1993) was the first influential study providing a comprehensive psychometric analysis of data collected from previous published studies on cognitive abilities from a variety of researchers. Carroll selected the studies for his analysis based on several factors including the number of variables measured on cognitive tasks, the sample composition of the study, and the presence of correlation matrices. Once Carroll determined the compatibility of each study with his stringent requirements, he conducted a factor analysis on over 460 studies. Carroll identified different strata or levels for intelligence that consisted of a general factor of intelligence, *g*, as well as broad and narrow abilities (Carroll, 1993).

Carroll's hierarchical model of intelligence contained three different strata. Stratum III, the highest level, was conceptualized as general intelligence and identified by Spearman's *g*. This level was hypothesized to influence all other abilities subsumed underneath it. Stratum II, the middle level, represented broad cognitive abilities (Carroll, 1993). Broad cognitive abilities are inherent in an individual and strongly influence his or her overall behaviors (Flanagan & Dixon, 2013). Carroll labeled these broad abilities as fluid (G*f*) and crystallized (G*c*) intelligence, as well as general memory and learning (G*y*), broad visual perception (G*v*), broad auditory perception (G*a*), broad retrieval ability (G*r*), broad cognitive speediness (G*s*), and speed of processing (G*t*). While each of these broad factors is considered unique in themselves, Carroll specified they remained intercorrelated due to the presence and considerable influence of *g* (Bickley et al., 1995; Carroll, 1993). Additionally, in 1990, Richard Woodcock proposed the inclusion of broad reading and writing ability (G*rw*) and quantitative reasoning ability (G*q*) as factors that composed Stratum II. However, John Carroll did not accept Woodcock's addition, maintaining that G*rw* and G*q* were narrow abilities present in Stratum I (Carroll, 2003). Finally, Stratum I, the bottom level, was composed of over 70 narrow abilities. Narrow abilities are thought to be more specialized in nature and evolve through exposure and learning (Flanagan & Dixon, 2013). These narrow abilities were categorized into the Stratum II broad factors based on their shared factor loadings. Additionally, while there were intercorrelations among the narrow abilities due to the shared variance of the broad factors, the narrow abilities also demonstrated unique variance (Carroll, 1993).

Carroll's three stratum theory of intelligence provided a reference for other researchers in the field as a means of identifying and classifying various cognitive abilities. While Carroll did not believe his hierarchical model was static in nature, he suggested that his model was significant in laying the foundation for future theories regarding human intelligence (Carroll, 1993). Carroll proposed assessing intelligence using a variety of methods to garner the best understanding of a person's abilities. His research and taxonomy of intelligence was pivotal in establishing modern day theories of intelligence, specifically CHC theory.

A Merging of Theories: CHC Theory of Intelligence

While the fluid-crystalized theory and three-stratum theory both sought to psychometrically explain intelligence, there were some essential differences between the two concepts. Predominately, Cattell and Horn's G*f*-G*c* model did not posit the concept of a general intelligence factor (*g*) and this was an area of strong contention between the theories (Cattell, 1998; Horn & Blankson, 2012). Additionally, the debate of intelligence as a single factor versus an amalgamation of numerous factors continues to persist in modern day literature (Schneider & McGrew, 2012).

In 1997, Kevin McGrew suggested a single taxonomy, or classification system, that could be used to develop intelligence tests more effectively (Flanagan & Dixon, 2013). He sought to combine Cattell and Horn's fluid-crystallized theory with Carroll's three-stratum theory into a single hierarchical model. This merging of theories, combined with contributions from various researchers, later resulted in the Cattell-Horn-Carroll or CHC model of intelligence (McGrew, 1997, 2009; Newton & McGrew, 2010). The original CHC model included a general intelligence factor in addition to nine broad abilities with their associated narrow abilities: fluid reasoning (G*f*), crystalized ability (G*c*), visual processing (G*v*), auditory processing (G*a*), processing speed (G*s*), short-term memory (G*sm*), long-term memory storage and retrieval (G*lr*), reading and writing ability

(G*rw*), and quantitative knowledge (G*q*: Flanagan & McGrew, 1997). While CHC theory was not hypothesized to be static in nature, factor analytic studies provided empirical support for the existence of CHC theory in previously published tests (Schneider & McGrew, 2012). For example, the Woodcock-Johnson Psycho-Educational Battery – Revised (WJ-R; Woodcock & Johnson, 1989) was based on the expanded version of Cattell and Horn's G*f*-G*c* model of intelligence (McGrew et al., 1991); however, subsequent studies found early indications of CHC theory present on the instrument. John Carroll (2003) examined the data collected for the WJ-R and found evidence to support a hierarchical model encompassing an overall general ability and multiple broad factors; thus, reflective of modern CHC theory.

Early conceptualization of CHC theory provided the foundation for several future assessment instruments. The Woodcock-Johnson III (WJ-III; Woodcock et al., 2001, 2007) is considered to be the first assessment tool based singularly on CHC theory. According to the technical manual for the WJ-III (McGrew & Woodcock, 2001), a confirmatory factor analysis of the standardization sample demonstrated evidence for a higher order general factor of intelligence in addition to nine second-order broad factors. Furthermore, the fourth and fifth edition of the Wechsler Intelligence Scale for Children (Wechsler, 2003, 2011) and the fourth edition of the Wechsler Adult Intelligence Scale (Wechsler, 2008) reference CHC theory in their manuals even though the instruments are not identified as explicitly based on the theory. The Differential Ability Scale – Second Edition (DAS-II; Elliot, 2007) is based on a multitude of intelligence theories, but there were several broad factors identified in the battery, such as fluid reasoning, crystallized ability, visual processing, short-term memory, and speed of processing. (Keith & Reynolds, 2010). The Kaufman Assessment Battery for Children – Second Edition (KABC-II; Kaufman & Kaufman, 2004) allows for interpretation of results from the Lurian model of intelligence or CHC theory. The fifth edition of the Stanford-Binet (Roid, 2003) is arranged based on a general intelligence factor in addition to five broad CHC factors: fluid reasoning, crystallized ability, visual processing, short-term memory, and quantitative reasoning.

Despite CHC being considered a prominent, psychometrically strong taxonomy for classifying various components of intelligence, the actual structure of the theory continues to evolve and increase in complexity. In 2012, Joel Schneider and Kevin McGrew presented a chapter that addressed a more updated conceptualization of CHC theory in the third edition of *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. CHC continued to be identified as a hierarchical model with *g* at the apex; however, there was acknowledgement that the existence of *g* continues to be an area of contention within psychology. Furthermore, Schneider and McGrew (2012) significantly shifted from the original nine broad CHC factors and proceeded to identify sixteen broad cognitive factors, which they classified into three theoretical categories: domain-free general capacities, acquired knowledge, and sensory- and motor-linked abilities.

The domain-free general capacities were recognized to not be associated with any specific sensory functions, but impacted by brain functioning. A total of six broad CHC factors as well as their accompanying narrow abilities were identified as belonging to this group: fluid reasoning (G*f*), short-term memory (G*sm*), long-term storage and retrieval (G*lr*), processing speed (G*s*), reaction and decision speed (G*t*), and psychomotor speed (G*ps*). Acquired knowledge was described as representative of *g* and "involve[d] the acquisition of useful knowledge" stored in long-term memory (Schneider & McGrew, 2012, p. 122). Four broad factors and their narrow abilities were identified within acquired knowledge: comprehension-knowledge (G*c*), domain-specific knowledge (G*kn*), reading and writing (G*rw*), and quantitative knowledge (G*q*). The final domain presented was sensory- and motor-linked abilities and the cornerstone of this domain was that each of the abilities were identified in specific regions of the brain. This group included six broad CHC factors and their respective narrow abilities: visual processing (G*v*), auditory processing (G*a*), olfactory abilities (G*o*), tactile abilities (G*h*), kinesthetic abilities (G*k*), and psychomotor abilities (G*p*). It is important to note that this latest conceptualization of CHC theory combined with recent neuropsychological research has been cited as guiding the development and factor structure of latest Woodcock Johnson assessment battery, the WJ IV (McGrew et al., 2014).

Interestingly, Joel Schneider's website (assessingpsyche.wordpress.com) provides his visual interpretation of CHC theory, which diverges from the work he completed with McGrew in 2012. It appears that Schneider divided CHC into five theoretical domains: motor, perception, controlled attention, knowledge, and speed. Under the motor domain, a single broad factor is listed, psychomotor abilities. The perception domain includes kinesthetic, tactile, and olfactory abilities in addition to visual and auditory processing. Controlled attention possesses fluid ability and short-term memory. The knowledge domain consists of domain-specific knowledge, comprehension-knowledge, reading and writing, and quantitative knowledge. The last domain, speed, includes psychomotor speed, reaction and decision speed, processing speed, and long-term storage and retrieval.

More recently, Woodcock, Miller, Maricle, and McGill (2017) proposed a modification to current CHC theory that provides a simpler approach for clinicians. Their theory, the Functional CHC Model (F-CHC), divides the CHC factors into three theoretical domains: acquired knowledge, thinking abilities, and cognitive efficiency. Further, Woodcock et al. modified the narrow abilities identified under each of the broad factors. The acquired knowledge field includes comprehension-knowledge and psychomotor abilities as well as broad reading, writing, and math abilities. The thinking abilities domain incorporates visual-spatial and auditory processing, learning-memory, and reasoning. Finally, the cognitive efficiency group consists of conscious memory and cognitive processing speed (Woodcock et al., 2017).

CHC theory continues to be one of the most researched theories examining intelligence. As evidenced from the three presented interpretations of CHC theory, this is a complicated model and is conceptualized in a variety of ways. Despite its multifaceted nature, CHC theory endures as a foundation for many assessment tools and interpretative processes employed in the field of school psychology. It is important for clinicians to understand the intricacies involved in attempting to assess latent constructs using these cognitive measures.

Summary: Current Theoretical Perspectives on Cognitive Abilities

The initial research on intelligence provided the foundation for modern approaches to intelligence theory. Raymond Cattell initially proposed a dichotomous concept of intelligence known as the fluid-crystallized (G*f*-G*c*) theory of intelligence. He identified G*f* as the general factor of intelligence and hypothesized it involved the ability to determine relationships between objects, while he posited that G*c* evolved from previously obtained knowledge. Cattell further stated intelligence is impacted by a range of factors imposed upon an individual. John Horn's research expanded Cattell's original G*f*-G*c* model to include other second-order factors. In addition to fluid and crystallized intelligence, Horn also identified visualization, speediness, facility, carefulness, premsia, and positive self-image as other components of intelligence. Through additional refinement, Cattell and Horn proposed a total of nine broad abilities of intelligence: fluid intelligence, crystallized intelligence, short-term acquisition and retrieval, visual processing, long-term storage and retrieval, processing speed, auditory processing, decision speed, and quantitative reasoning. John Carroll conducted a seminal factor analysis on over 460 published studies resulting in his conceptualization of intelligence as a three-stratum model. His model consisted of a single general factor (Stratum III), broad abilities (Stratum II), and narrow abilities (Stratum I). Carroll's broad abilities were comprised of fluid and crystallized intelligence as well as general memory and learning, visual perception, auditory perception, retrieval ability, cognitive speediness, and speed of processing. The narrow abilities consisted of more than 70 abilities that he identified. While Cattell and Horn's G*f*-G*c* theory and Carroll's three-stratum theory were highly influential in the study of intelligence, the merging of the two theories into the Cattell-Horn-Carroll (CHC) model of intelligence has dominated current intelligence theory. Through the evolution of CHC theory, a common taxonomy for identifying the various constructs of intelligence emerged. Numerous factor analytic studies have further substantiated the identification of intelligence based on the CHC model. The WJ III was the first testing battery to specifically operationalize CHC theory and this paved the way for subsequent measures to follow. CHC theory continues to be researched and reconceptualized. Because of this, CHC theory has become exceedingly more perplexing and complicated for current practitioners.

History of Test Interpretation

As previously discussed, beginning in the early 20th century, the ability to assess intelligence, as well as the evolution of intelligence theory, were essential in shaping psychological assessment. However, interpreting the meaning of test scores and applying the results to intelligence theory continues to dominate in the 21st century (Woodcock, 2002; Kamphaus et al., 2012). Test interpretation continues to play a significant role in diagnosis and treatment of individuals, especially within the education setting. To date, there is not one universally accepted approach to score interpretation.

The genesis of test interpretation began with the development of the first formal measure of intelligence by Binet and Simon (Boake, 2002; Kamphaus, 2001). Early test interpretation began as a more qualitative approach, but has become more sophisticated through the years. Originally, test interpretation attempted to define intelligence based on the specific construct each assessment battery was purported to measure. However, as intelligence testing has become more advanced and grounded in theory, interpretation has become increasingly more complex (Kamphaus et al., 2012). There are numerous methods utilized for score interpretation and the following discussion examines the four major waves of test interpretation presented by Kamphaus and colleagues (2012).

The First Wave

The initial wave of test interpretation began with the utilization of the Binet-Simon Intelligence Scale in the early 1900's. This assessment tool targeted specific areas of cognitive ability: thinking and reasoning, problem solving, and one's approach to new experiences (Pintner, 1923). The Binet-Simon Intelligence Scale was identified as criterion referenced and provided the first pragmatic approach to intelligence testing. A criterion-referenced test examines a person's raw score and compares it to some predetermined cutoff score. This score is then interpreted as one's cognitive ability for an individual at their age (Wasserman, 2012). Additionally, the Binet-Simon Intelligence Scale paved the way for using the results from intelligence testing as a means of classification within the education system (Kamphaus et al., 2012).

James McKeen Cattell also contributed to the first wave of intelligence test interpretation. His focus of assessment targeted the perception and motor ability of an individual (Pintner, 1923). More importantly, Cattell's work laid the foundation for standardization procedures regarding testing. He posited that for a measure to be scientifically interpreted and comparable to others, it was important to administer the testing instrument under the same environment and conditions (Anastasi, 1988).

Finally, the first wave of test interpretation witnessed the beginning of the classification system of intelligence. While classification continues to be used in modern-day interpretation, this system has contributed to much of the stigmatization associated with intelligence scores (Kamphaus et al., 2012). For example, Levine and Marks (1928) used the following terms to classify intelligence based on an individual's IQ range: idiots (0-24), imbeciles (25-49), morons (50-74), borderline (75-84), dull (85-94), average (95-104), bright (105-114), very bright (115-124), superior (125-149), very superior (150-174), and precocious (175 and up). Similarly, David Wechsler (1944) proposed the following classification system and IQ ranges: defective (65 and below), borderline (66-79), dull normal (80-90), average (91-110), bright normal (111-119), superior (120-127), and very superior (128 and higher). While classification based on ranges of intelligence scores continue to be utilized today, the terms have become more objective in nature and tend to focus on terms centered around the mean (Kamphaus et al., 2012).

The first wave of test interpretation witnessed a strong focus on the actual global intelligence score derived from the testing battery. Interpretation tended to focus on a person's ability level based on the administrator's analyses of test performance. Goodenough (1949) and Wechsler (1958) proposed that there was more to intelligence assessment beyond the single score. They both speculated the scores obtained from a single measure did not represent absoluteness; rather, the global score contributed a small portion to the large, overall picture of an individual.

The Second Wave

The next wave of test interpretation moved away from a single score interpretative model to a more multilevel model of understanding. This period witnessed the introduction of subtest score comparison, individual test item interpretation, and specific diagnoses based on test performance (Kamphaus et al., 2012).

Rapaport, Gil, and Schafer (1945-1946) proposed a system of evaluating the general intelligence score as well as the profile of subtest scores. They stated that comparing the high and low subtest scores may provide insightful information regarding appropriate diagnostic decisions. The system proposed by Rapaport and colleagues (1945-1946) had five major components. The first element of their model included an evaluation of each individual item response. Then, a comparison of each individual item within a specific subtest. The third element relied on within-subject comparison of an individual's subtest scores. The fourth element consisted of a comparison between broad groups of scores (such as the Verbal and Performance scales) on an instrument. Finally, the last element involved comparing an individual's performance on an intelligence test with other forms of testing.

This novel approach to test interpretation offered clinicians a step-by-step approach to a more detailed analysis of test findings. Rather than focusing solely on the single intelligence score, it allowed for comparison between multiple levels of intelligence. Additionally, this approach was the first methodology proposed to examine intra-individual cognitive strengths and weaknesses (Kamphaus et al., 2012). While this second wave marked an essential shift in the approach to test interpretation with profile analysis, it lacked normative guidelines and evidence of validity for the proposed methods.

The Third Wave

This wave of test interpretation was credited with the emergence of factor analysis and psychometric profile analysis, approaches that continue to be used. With the advent of powerful statistical software packages in the 1960's, researchers were able to conduct more detailed factor analyses of testing instruments. This allowed for further interpretation resulting from the quantitative identification of cognitive factors (Kamphaus et al., 2012).

Factor analysis is a mathematical approach used to assess the relationship of variables or factors present on an instrument (Fruchter, 1954). Two common factor analytic approaches consist of exploratory and confirmatory procedures. Exploratory factor analysis (EFA) is utilized when the overall goal is to determine the relationship between the measured variables on an instrument without any preconceived hypotheses. Confirmatory factor analysis (CFA) is used when there is already a hypothesis regarding the specific factor structure of an instrument (Fabrigar & Wegener, 2012). While EFA and CFA target different questions, they are viewed as complementary procedures to one another (Carroll, 1993). A more detailed discussion regarding factor analysis will be provided in the subsequent chapter.

Factor analysis became known with Cohen's 1959 study of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949). Through his analysis, Cohen found widespread support for three factors on the WISC: Full-Scale Intelligence Quotient, Verbal Ability, and Performance Ability (Cohen, 1959). Cohen also introduced the concept of subtest specificity and strongly cautioned against solely interpreting individual subtests. Subtest specificity refers to the fact that each subtest present on a testing battery shares a certain amount of variance with other subtests. Therefore, one is unable to deduce whether the scores of one specific subtest are indicative of that subtest alone, the specific variance, or a result of the shared variance between all of the subtests (Kamphaus et al., 2012).

Psychometric profile analysis, or ipsative analysis, involves comparing an individual's performance on a subtest to their average performance across all subtests. Scores found to be below the mean are identified as cognitive weaknesses, while scores found to be higher than the mean are considered cognitive strengths (Kamphaus et al., 2012). While profile analysis from a clinical perspective first became known during the second wave of test interpretation, many felt that psychometric profile analysis was stronger due to its quantitative foundation. However, Alan Kaufman (1979) purported that subtest score comparison was not a valid interpretation of test scores and strongly cautioned against this form of evaluation. Kaufman urged clinicians to examine all the data gathered for an individual to garner a workable hypothesis (Kaufman, 1990).

The Fourth Wave

Presently, test interpretation continues to be at the forefront of intelligence research. Modern testing instruments are being developed on more contemporary theories of intelligence and neuropsychological research (Miller & Maricle, 2012). Additionally, test interpretation tends to be guided through data collected from a working, testable hypothesis concerning cognitive abilities, rather than a single factor (Kamphaus et al., 2012). In the previous interpretation wave, Kaufman stated great concern for test validity and addressed the difficulty associated with test interpretation, especially when there lacks sufficient empirical research (Kaufman, 1990). As a result, the fourth wave of test interpretation has witnessed a more selective approach to assessment, with an emphasis being placed on the broad cognitive abilities (Horn & Blankson, 2012; Kamphaus, 2001).

Beginning in the 1990's, the development of intelligence tests began to be founded more on a theoretical understanding of intelligence and the constructs that represent numerous factors of intelligence (Kamphaus et al., 2012). Specifically, the emergence of Cattell and Horn's (Horn & Cattell, 1966) fluid-crystallized theory of intelligence and Carroll's (Carroll, 1993) three-stratum theory of intelligence paved the way for the current conceptualization of intelligence. During the most recent wave of test interpretation and through much research and debate, these two theories were merged into a hierarchical model of intelligence known as the Cattell-Horn-Carroll (CHC) model of intelligence (McGrew, 1997, 2009; Newton & McGrew, 2010).

While test interpretation appears to have become stronger because of rigorous analyses and research, the fourth wave of interpretation posits the importance of more critically examining the collected data. Specifically, Kamphaus (2001) recommends when conducting an evaluation for diagnostic purposes, collecting data from multiple sources as well as actively utilizing a minimum of two additional pieces of data is crucial. Moreover, this wave of interpretation has suggested the benefit of examining alternate hypotheses. Clinicians tend to overestimate their assessment and interpretative abilities while maintaining a certain unawareness of their own biases; therefore, it is important to maintain objectivity during interpretation and collect data beyond the single administered assessment (McDermott et al., 2014; Watkins, 2009).

Summary: History of Test Interpretation

Test interpretation continues to dominate current research in intelligence testing. Initially, test interpretation was basic in nature with the advent of the Simon-Binet scale. However, over the years, testing instruments have increased in complexity and test interpretation has become more sophisticated. The first wave of test interpretation witnessed the emergence of criterion referenced testing, standardized norming procedures, and a classification system based on scores. There was strong emphasis given to the global score obtained and interpretation relied heavily on the examiner's analysis of the examinee's testing performance. The second wave relied more on subtest score comparison and test item interpretation. Additionally, this era observed the beginning of diagnoses being identified based on testing performance. The third wave experienced the emergence of statistical analyses regarding test interpretation. Specifically, the use of factor analysis and psychometric profile analysis were employed to further identify cognitive abilities. More recently, the fourth wave considers the actual development and construction of the testing instruments. Assessment batteries are founded in theory and this guides the overall interpretative process. This wave has been subjected to more rigorous analyses and an emphasis has been placed on critical evaluation of the collected data. The cornerstone of this era is grounded in moving beyond the scores provided on a testing battery to include a broad and objective interpretation of the individual.

The Woodcock-Johnson

This study examines the factor structure of the WJ IV COG and OL; therefore, it is important to trace the evolution of the Woodcock assessment batteries from the beginning. This section includes an introduction to each iteration of the WJ as well as a discussion on the theoretical basis for each. Additionally, an analysis of relevant research on each testing instrument is addressed.

Woodcock-Johnson Psycho-Educational Battery: 1977

The Woodcock-Johnson Psycho-Educational Battery (WJPEB; Woodcock & Johnson, 1977) was the first generation of the Woodcock-Johnson testing instruments and consisted of three targeted areas of measurement. Specifically, the WJPEB purported to assess cognitive and academic ability, as well as interest, for children as young as three to senior age. Additionally, this instrument was the first to provide a standardized achievement and cognitive measure on the same standardization sample (Woodcock & Johnson, 1977).

The WJPEB differentiated itself from the traditional Wechsler and Binet scales by providing several unique cognitive subtests. These subtests included a component of direct examiner training to the examinee during the administration process. The purpose was to assess the examinees ability to learn, such as reading and mathematics, and consisted of Analysis-Synthesis, Concept Formation, and Visual-Auditory Learning. Additionally, the WJPEB incorporated a more detailed evaluation of academic writing through the Written Language subtest. There were no manipulative measures present on the WJPEB, another deviation from the Wechsler and Binet assessment batteries (Estabrook, 1983).

The development of the WJPEB was not originally based on any theoretical orientation; rather, it was created through numerous experiments conducted by Dr. Richard Woodcock that examined learning in individuals (Schrank et al., 2016). Each of the measured abilities on the battery ranged from lower mental processes to higher mental processes (Woodcock & Johnson, 1977). After completion of the norming process for the WJPEB, factor and cluster analyses were conducted to determine the broad functions measured by the test. The four broad functions identified were Knowledge-Comprehension, Reasoning-Thinking, Memory-Learning, and Discrimination-Perception. Rather than a specific intelligence quotient (IQ) score, the WJPEB provided a Broad

Cognitive Ability (BCA), which served as a measure of one's intellectual ability. Additionally, test administrators were provided four additional scores for broad functioning: Knowledge-Comprehension, Reasoning-Thinking, Memory-Learning, and Discrimination-Perception (McGue, Shinn, & Ysseldyke, 1982; Schrank et al., 2016; Woodcock & Johnson, 1977).

Woodcock-Johnson Psycho-Educational Battery – Revised: 1989

The Woodcock-Johnson Psycho-Educational Battery - Revised (WJ-R; Woodcock & Johnson, 1989) was the next generation in the WJ testing series and consisted of the Woodcock-Johnson Tests of Cognitive Ability (WJ-R COG) and the Woodcock-Johnson Tests of Achievement (WJ-R ACH). The WJ-R assessed cognitive, scholastic, and academic abilities in preschool children through geriatric adults. Similar to the norming procedures for the WJPEB, the norming sample for the WJ-R was administered the cognitive and academic measures in order to compare a single person's scores across abilities (Hicks & Bolen, 1996; Woodcock, 1990). Furthermore, like its predecessor, the WJ-R identified Broad Cognitive Ability rather than a specific IQ score; however, the scores obtained from the WJ-R also allowed the administrator the ability to examine intra-cognitive discrepancies (McGrew & Murphy, 1995; Woodcock, 1990).

The WJ-R COG was the first assessment battery in the sequence to be based on a specific theoretical foundation. Dr. Woodcock (1990) identified the WJ-R COG as an "operational representation" (p. 233) of Cattell and Horn's fluid-crystallized (G*f*-G*c*) theory of intelligence (Horn & Cattell, 1966). The revision of the WJPEB was spearheaded by Kevin McGrew, who conducted most of the statistical analyses and was paramount in helping align the WJ-R with G*f*-G*c* theory (Schrank et al., 2016). McGrew conducted an extensive review of the published research on exploratory and confirmatory factor analyses of the WJPEB. From there, the foundation of the WJ-R was modeled to fit the theory of fluid-crystallized intelligence (McGrew et al., 1991).

Overall, the WJ-R assessed eight of the cognitive abilities from the G*f*-G*c* model of intelligence: fluid reasoning (G*f*), comprehension-knowledge (G*c*), visual processing (G*v*), auditory processing (G*a*), processing speed (G*s*), short-term memory (G*sm*), long-term retrieval (G*lr*), and quantitative ability (G*q*; McGrew et al., 1991; Woodcock & Johnson, 1989). Each of the broad factors measured by the WJ-R were derived from the cluster scores of two subtests that measured the narrow abilities of that broad factor. Seven of the broad factors were identified through the administration of the WJ-R COG, while the eighth factor, G*q*, was assessed when using the WJ-R ACH (McGrew & Murphy, 1995; Woodcock, 1990).

In 1990, Dr. Woodcock published an influential article detailing various factor analytic studies of the WJPEB and WJ-R. His goal was to provide empirically supported evidence for the proposed factor structure of the revised WJ. Dr. Woodcock analyzed nine data sets from the norming and validity studies from 1977-1989. Additionally, he examined concurrent studies of the WJ-R with other cognitive tests in the field. In total, 15 sets of exploratory and confirmatory factor analyses were investigated. Dr. Woodcock concluded that because of these comprehensive analyses, there was adequate support indicating that the WJ-R appropriately measured the eight factors of the G*f*-G*c* theory of intelligence. Specifically, he stated that each of the eight factors were composed of a minimum of two "clean" measures of that factor (Woodcock, 1990, p. 253).

Finally, it is important to note that Dr. Woodcock acknowledged all the factor analytic studies utilized for his comprehensive analysis, apart from two, were derived from confirmatory factor analysis (CFA). He stated that exploratory analyses were not needed because the factor structure of the instrument was already known to the investigator. Therefore, confirmatory procedures were utilized due to the *a priori* hypothesis of the factor structure of the WJ-R (Woodcock, 1990). As evidenced by the large amount of research available on the third iteration of the WJ tests, the flaw of solely relying on CFA procedures when determining the factor structure of an instrument is further discussed in the subsequent section.

Woodcock-Johnson III: 2001

The Woodcock-Johnson III (WJ III; Woodcock et al., 2001, 2007) was the third assessment battery in the WJ series. Similar to previous versions, the WJ III included Tests of Cognitive Abilities (COG) and Tests of Achievement (ACH). There were additional testing items added in the Diagnostic Supplement (DS) to provide further assistance to the examiner (Woodcock, McGrew, Mather, & Schrank, 2003, 2007). The WJ III COG, ACH, and DS were co-normed together and the full battery was recommended for preschool to senior age. Utilization of the complete WJ III battery allowed the examiner to determine cognitive and academic abilities in addition to oral language ability (Schrank et al., 2010).

Unlike previous WJ versions, the WJ III was purportedly grounded in CHC theory and this served to guide the interpretation of the WJ III (McGrew & Woodcock, 2001; Schrank et al., 2010; Schrank & Wendling, 2012). Examiner interpretation involved a hierarchical approach and included three distinct levels: general intellectual ability, broad cognitive abilities, and narrow cognitive abilities (McGrew & Woodcock, 2001). Previously, the WJ-R did not subscribe to a general intelligence factor, rather it provided the Broad Cognitive Ability score. However, with the transition of test theory shifting from fluid-crystallized theory to CHC theory, the WJ III witnessed the emergence of the General Intellectual Ability (GIA) score. The GIA score was posited to be representative of Spearman's *g* and was at the apex of the CHC hierarchy (McGrew & Woodcock, 2001). The overall score was described as an extraction score consisting of multiple factors. To be more precise, the GIA was not hypothesized to represent a single, distinct ability rather it embodied an amalgamation of different cognitive functions. Furthermore, the GIA was reported to be responsible for the shared variance observed across all given tests (Schrank et al., 2010). The WJ III permitted four different options for calculating the GIA score, each dependent upon the selection of subtests given by the testing administrator.

At the second level of the CHC model are the broad factors or abilities. The broad cognitive abilities hypothesized to be measured by the WJ III included comprehension-knowledge (G*c*), long-term retrieval (G*lr*), visual-spatial (G*v*), auditory processing (G*a*), fluid reasoning (G*f*), processing speed (G*s*), and short-term memory (G*sm*). Additionally, these broad abilities were further classified into three cognitive category clusters: verbal ability, thinking ability, and cognitive efficiency (McGrew & Woodcock, 2001; Schrank et al., 2010). The lowermost level of the model was composed of the narrow cognitive abilities. These were reported to be directly measured by the subtests on the WJ III battery. The narrow abilities were ostensibly grouped together based on shared similarity or factor loading; thus, giving rise to the second-order broad factors. The narrow ability clusters included the following: phonemic awareness, associative memory, working memory, quantitative reasoning, visualization, sound discrimination, auditory memory span, and perceptual speed (Schrank et al., 2010). The technical manual posits that a minimum of two narrow abilities were measured for each broad ability (McGrew & Woodcock, 2001).

Since its publication, the WJ III has been extensively evaluated and subjected to intense scrutiny (Keith & Reynolds, 2010). Upon its initial release, the technical manual specified there was wide-spread corroboration through a myriad of studies indicating the WJ III strongly aligned with CHC theory (Taub & McGrew, 2014). McGrew and Woodcock (2001) indicated there were numerous CFA studies comparing the CHC model to other theoretical models. They stated these studies provided verification for the presence of *g* plus seven broad factors derived from the cognitive assessment and two additional broad factors from the academic assessment on the WJ III. Several studies have also established that the WJ III has been found to be invariant across multiple age ranges; therefore, positing that the CHC model remains stable across a wide age range (Floyd et al., 2009; Taub & McGrew, 2004). Finally, Keith and Reynolds (2010) provided a comprehensive review of a multitude of studies examining the WJ III in cross battery confirmatory factor analysis (CFA). They reported the factor structure of the WJ III is further documented when analyzed with the Wechsler Intelligence Scale for Children – Third Edition, Cognitive Assessment System, Differential Abilities Scales, and the Kaufman Assessment Battery for Children – Second Edition.

Despite the appearance of significant support regarding the factor structure of the WJ III, there have been numerous studies hypothesizing an opposing view of the assessment battery. As previously stated, a substantial amount of evidence provided through numerous studies indicated support for the reported factor structure of the WJ III; however, they were based primarily on CFA procedures. The predominate concern regarding only using CFA is the strong potential for over-factoring (Keith & Reynolds, 2010; Strickland et al., 2015). Frazier and Youngstrom (2007) posit that the sole utilization of CFA procedures allows for a “heavy reliance on liberal statistical criteria for determining the number of factors measured by a test” (p. 170). Furthermore, Carroll (1993, 2003) promoted the use of EFA *and* CFA procedures to be used concomitantly.

Exploratory factor analysis has traditionally been used to determine the structural validity of testing instruments by identifying the latent constructs present (Fabrigar & Wegener, 2012; Fabrigar, Wegener, MacCallum, & Strahan, 1999). Essentially, EFA permits the data to speak for itself during the analyses without outside (researcher) interference. On the other hand, when using CFA procedures, the researcher drives the analyses by determining the number of factors *a priori*. Therefore, because of this predetermination of factors, the potential for over-factoring significantly increases (Frazier & Youngstrom, 2007; Strickland et al., 2015). While this procedure of factor extraction via CFA procedures has been evident across a large variety of cognitive testing batteries (Keith & Reynolds, 2010), a review of the research only involving the WJ III will be analyzed.

A study by Dombrowski (2013) conducted an EFA on the correlation matrices for the 9- to 13- and 14- to 19- year old age groups from the norming sample of the WJ III COG. His findings suggested the 9 to 13 age group presented with a total of four factors that could be retained, while the 14 to 19 age range data suggested three factors for retention. These findings are in stark contrast to the reported seven factor structure for the COG specified by the technical manual (McGrew & Woodcock, 2001). Another EFA study conducted on the full battery of the WJ III analyzed the correlation matrices for school-aged children on all the subtests. This study found support for six factors in the younger school-age range and five factors in the older population (Dombrowski & Watkins, 2013). While this study indicated more factors compared to the subsequent study, there remains a deviation from the suggested nine broad factors for the full battery reported in the technical manual. Strickland and colleagues (2015) examined the structure of the WJ III COG in a clinical sample of school-aged children and found evidence of two to three factors, a significant decrease in the proposed seven factors found in CFA studies. Finally, Dombrowski (2015) conducted an exploratory bifactor analysis of the WJ III ACH and found support for two to three factors for school-aged children.

While a large majority of CFA studies on the WJ III have traditionally supported the proposed factor structure (Floyd et al., 2009; Keith & Reynolds, 2010; McGrew & Woodcock, 2001; Taub & McGrew, 2004), there has been a markedly increased number of EFA studies proposing a significantly reduced number of factors (Dombrowski, 2013, 2015; Dombrowski & Watkins, 2013; Strickland et al., 2015). Considering that each WJ appears to build on its predecessor, this is cause for strong skepticism moving forward with the recently released testing battery in the series, the WJ IV.

Woodcock Johnson IV: 2014

The fourth and most recent version in the Woodcock-Johnson testing series is the Woodcock-Johnson IV (WJ IV; Schrank, McGrew, & Mather, 2014a). This is the first revision of the WJ to not include its original developer, Dr. Woodcock. The complete battery includes the WJ IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014b), Tests of Academic Abilities (WJ IV ACH; Schrank, Mather, & McGrew, 2014a), and Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b). All three tests in the battery were co-normed with each other to allow for appropriate comparisons between them and are suitable for administration with preschool children to senior adults.

When the complete WJ IV battery is used together and in its entirety, test publishers identified a general measure of intelligence (*g*) in addition to nine CHC factors: fluid reasoning (G*f*), comprehension-knowledge (G*c*), short-term working memory (G*wm*), processing speed (G*s*), auditory processing (G*a*), long-term storage and retrieval (G*lr*), visual-spatial processing (G*v*), reading-writing (G*rw*), and quantitative reasoning (G*q*) (McGrew et al., 2014). Seven of the factors are derived from the WJ IV COG and WJ IV OL (G*f*, G*c*, G*wm*, G*s*, G*a*, G*lr*, and G*v*), while two additional factors are available when utilizing the WJ IV ACH (G*rw* and G*q*).

According to McGrew and colleagues (2014), their modern conceptualization of CHC theory contributed to amending some of the broad cognitive factors. Fluid reasoning (G*f*) is the ability to solve novel problems without reliance on previously acquired knowledge. Comprehension-knowledge (G*c*) involves the application of previously learned knowledge to solve practical problems and demonstrates the complexity and extent of one's acquired knowledge through personal experience. Short-term working memory (G*wm*) is an extension from the previously used term short-term memory (G*sm*) from the Woodcock-Johnson III (WJ III; Woodcock et al., 2001, 2007) and refers to one's ability to hold and manipulate information that is in temporary storage. Processing speed (G*s*) is the ability to perform simple tasks in a relatively brief period with speed and efficiency. Auditory processing (G*a*) includes detecting and processing information through sounds. Long-term storage and retrieval (G*lr*) involves taking in and learning information, then retrieving this information from memory storage after a period of elapsed time. Visual-spatial processing (G*v*) is the ability to mentally perceive information to solve problems which are visual-spatial in nature. Reading-writing (G*rw*) is the complexity of word knowledge a person possesses, including: spelling, comprehension, and English language usage. Quantitative reasoning (G*q*) is the understanding of relationships involving quantitative concepts and the ability to manipulate these concepts (Schneider & McGrew, 2012; McGrew et al., 2014).

Additionally, changes to the WJ IV were implemented regarding cluster scores, subtests, and interpretative procedures based on the most current conceptualization of CHC theory and neuroscience research (McGrew et al., 2014). The changes to the WJ IV purportedly allow for a more concise alignment with CHC theory. For example, short-term working memory (G*wm*) is considered an extension to short-term memory from the WJ III (McGill et al., 2014) and measures the temporary storage of information as well as active manipulation of information. Speed of lexical access and memory for sound patterns are two new narrow abilities introduced on the WJ IV. Also, increasing measures of cognitive complexity were targeted for the WJ IV, specifically with auditory processing and reading fluency (McGrew et al., 2014). McGrew et al. acknowledge that all composite clusters and CHC factor clusters available on the WJ IV are altered in some capacity when compared to the WJ III. Therefore, it is crucial for test administrators to understand that performance interpretation does not directly correlate from the third edition to the fourth edition of the WJ.

According to the publishers, interpretation of the WJ IV is based on four distinct levels. Level one involves qualitative observations from the testing administrator, level two examines the developmental scores, level three addresses proficiency, and level four allows for score comparisons (McGrew et al., 2014; Reynolds & Niileksela, 2015). The standard battery, tests one through ten, provides three general intelligence composites: General Intellectual Ability (GIA), Brief Intellectual Ability (BIA), and a Fluid-Crystallized (G*f*-G*c*) composite. The GIA is derived from the first seven tests on the WJ COG and include one test measuring each of the broad factors (G*c*, G*f*, G*wm*, G*s*, G*a*, G*lr*, and G*v*). The GIA-extended from the WJ III has been eliminated with the release of the WJ IV (McGrew et al., 2014). The BIA is calculated from the first three tests on the WJ IV COG and consists of one test from each of the broad factors (G*c*, G*f*, and G*wm*). The G*f*-G*c* composite is a new addition to the WJ IV and is derived from two subtests purporting to measure G*c* and two measuring G*f*. The Technical Manual suggests that in some cases, at the discretion of the examiner, the G*f*-G*c* composite may be utilized rather than the GIA (McGrew et al., 2014).

Additionally, the seven CHC broad ability clusters measured on the WJ III have been retained for the WJ IV COG, with two specific differences noted. Each of the ability clusters are reported to be measured based on a minimum of two administered subtests. The clusters include Comprehension Knowledge, Fluid Reasoning, Short-Term Working Memory, Long-Term Retrieval, Auditory Processing, Cognitive Processing Speed, and Visual Processing. The Short-Term Working Memory and Cognitive Processing Speed clusters are the two noted changes to the WJ IV. Furthermore, the WJ IV includes three additional narrow abilities cluster scores (Perceptual Speed, Quantitative Reasoning, and Number Facility) and one clinical cluster score (Cognitive Efficiency; Schrank, McGrew, & Mather, 2014a). The WJ IV also witnessed the return of a collection of scholastic scores from the WJPEB and WJ-R. The WJ IV provides two measures each for the areas of reading, writing, and mathematics (McGrew et al., 2014; Reynolds & Niileksela, 2015). When administering the WJ IV OL (Schrank, Mather, & McGrew, 2014b), a new component to the WJ family, additional interpretive information is provided. The first two subtests on the WJ IV OL provide the English Oral Language cluster, while administration of the entire eight English subtests provide five additional cluster scores: Broad Oral Language, Listening Comprehension, Oral Expression, Phonetic Coding, and Speed of Lexical Access (McGrew et al., 2014).

McGrew and colleagues provide support for the factor structure of the WJ IV through several statistical analyses. With respect to exploratory procedures, McGrew et al. (2014) conducted a principal component analysis (PCA) with varimax rotation as well as principal axis factoring (PAF) with promax rotation. While many in the field assume PCA is the same and equal to EFA, there are some noted differences that contribute to overall dissimilar findings. The two statistical approaches are based on fundamentally diverse mathematical models. Most notably, PCA does not allow for the distinction between common and shared variance among variables and is typically implicated in serving primarily as a data reduction technique (Fabrigar & Wegener, 2012). Also, while the technical manual indicates that PAF was used in the analysis of the WJ IV factor structure, there were no results reported in the manual due to Heywood cases (i.e., communalities greater than 1). This may indicate a lack of theoretical convergence, the potential of over-factoring, and/or poor model fit indices in the analyses of the WJ IV (Fabrigar & Wegener, 2012; McGrew et al., 2014).

Additionally, McGrew et al. reported that the results from the principal component analyses indicated there were two strong factor structures present on the WJ IV: a five-factor and a three-factor model. The publishers stated the five-factor model included comprehension-knowledge, processing speed, reading and writing ability, quantitative-fluid ability, and short-term working memory. The three-factor model was also described as equally acceptable and included factors identified as auditory (auditory processing, comprehension-knowledge, and portions of short-term working memory), visual (visual processing, fluid reasoning, and long-term storage and retrieval), and quantitative (quantitative reasoning and portions of short-term working memory).

McGrew et al. further provided support for the proposed factor structure of the WJ IV via confirmatory analytic procedures. Through this process, the authors postulated three possible models to describe the data. The first model proposed that "all 51 subtests loaded on a single latent *g* factor" (McGrew et al., 2014, p. 163). The second model was identified as a top-down model and included only two levels: a general intelligence factor and nine broad CHC factors. The third model was described as a bottom-up approach and was indicative of the traditional hierarchical model of intelligence. This model included a general intelligence factor, broad factors, and narrow abilities.

With respect to the three testing batteries available in the WJ IV, analysis of the technical manual does not reveal the rationale behind the factor structure specific to the WJ IV COG, WJ IV OL, or WJ IV ACH independently of each other (Schrank, McGrew, & Mather, 2014b). Unlike the WJ III, there were no factor analytic studies, exploratory or confirmatory, reported regarding the internal structure specific to either the COG, ACH, or OL battery. Rather, McGrew et al. (2014) identified the factor structure of the WJ IV COG through extrapolation procedures from the full WJ IV battery. Extrapolation procedures suggest that the current factor structure identified for the WJ IV will automatically extend downward to the three specific testing batteries for the WJ IV.

As previously noted, the WJ III underwent intense scrutiny concerning its overall factor structure, especially for the cognitive battery (Dombrowski, 2013, 2015; Dombrowski & Watkins, 2013; Strickland et al., 2015). The authors of the WJ IV appear to have overlooked the possibility that the WJ III was profoundly over-factored. When utilizing EFA or CFA procedures to determine the factor structure for the WJ IV, McGrew and colleagues (2014) did not employ the gold standard of factor analysis identified by many in the field (Fabrigar et al., 1999; Fabrigar & Wegener, 2012; Frazier & Youngstrom, 2007; Henson & Roberts, 2006). Moreover, the interpretative discussion concerning the WJ IV and modern CHC theory combined with the presented analyses in the Technical Manual are profoundly arduous and confusing to the common practitioner.

Due to the recent release of the WJ IV, there have been limited studies examining the factor structure of the WJ IV instruments. One study by Dombrowski, McGill, and Canivez (2016) conducted an EFA and hierarchical factor analysis of the WJ IV COG on the two school age correlation matrices. Their EFA results found support for two to four factors for school-aged children, a strong deviation from the publisher's supposition of seven. Another study by Dombrowski, McGill, and Canivez (2017) examined the factor structure of the full battery of the WJ IV on school aged children through utilization of hierarchical EFA. Their results indicated that the WJ IV appears to measure more factors than its predecessor, the WJ III; however, the EFA only identified seven of the nine CHC factors for the full battery.

Resulting from limited statistical analyses combined with the obfuscated results presented in the Technical Manual, the current study seeks to further advance the general understanding of the factor structure on the WJ IV COG and WJ IV OL for a school-aged population.

Summary: The Woodcock-Johnson

The introduction of the Woodcock-Johnson into the world of assessment began almost forty years ago. In general, all versions of the WJ assessment family target cognitive, scholastic, and academic abilities and assess a broad age range from preschool to senior.

The WJPEB was the first generation of the WJ series and was not based on any single theoretical orientation. Rather, Dr. Woodcock developed the testing battery after conducting a large variety of experiments on learning. This test separated itself from the other instruments in the field by introducing unique subtests. It was the first assessment tool that directly evaluated the process of learning through examiner training of the examinee. Furthermore, the WJPEB extended the evaluation process for writing to include more academically appropriate measures. This version of the WJ did not provide a specific intelligence score, but provided a Broad Cognitive Ability score that assessed overall cognitive ability.

The WJ-R was introduced roughly a decade later. The BCA score continued to be provided rather than an IQ score, but examiners were also able to analyze intra-cognitive discrepancies. The most significant difference in this version of the WJ series was that the WJ-R was exclusively based on Cattell and Horn's fluid-crystallized theory of intelligence. In general, the WJ-R assessed eight broad cognitive abilities from the G*f*-G*c* model: fluid reasoning, comprehension-knowledge, visual processing, auditory processing, processing speed, short-term memory, long-term retrieval, and quantitative ability.

The WJ III was introduced in 2001 and significantly deviated from its predecessors: this version was solely grounded in CHC theory. Test interpretation was guided by the hierarchical model of CHC theory and included general intellectual ability as well as broad and narrow cognitive abilities. The broad factors the WJ III assessed were comprehension-knowledge, long-term retrieval, visual-spatial, auditory processing, fluid reasoning, processing speed, and short-term memory.

The most recent iteration of the WJ testing family is the WJ IV released in 2014. This version is reported to be grounded in the most current conceptualization of CHC theory and recent neuropsychological research. While this battery includes the traditional cognitive and academic tests witnessed in previous versions, it also includes an assessment of oral language. The general intelligence score includes evaluation of seven broad factors: comprehension knowledge, fluid reasoning, short-term working memory, cognitive processing speed, auditory processing, long-term retrieval and visual processing. Additionally, the examiner has another option for interpreting cognitive ability beyond the general score by using the G*f*-G*c* composite score.

Like all assessment batteries in the field, the Woodcock-Johnson tests have been subjected to extensive evaluation. There have been numerous statistical analyses conducted on the various batteries and there appears to be concern regarding the latent factor structure of the instruments. A majority of the research done from the test publishers involves confirmatory factor analyses which tends to be more liberal in nature and may contribute to confirmation biases of the test developers.

Rationale for the Proposed Study

This section provides a discussion on the purpose and importance of the current study.

Purpose of the Current Study

A large majority of professionals view the role of school psychologists as gatekeepers to special education. The Individuals with Disabilities Education

Improvement Act of 2004 states that any child who is being considered for special education must undergo a comprehensive evaluation. In many states, a portion of this evaluation involves an assessment of the individual's intellectual functioning. Interestingly, Dr. Woodcock stated, "the primary purpose for cognitive testing should be to find out more about the problem, not to obtain an IQ" (2002, p.6).

To delineate cognitive or academic problems a child is experiencing, one must be able to understand the vast components of intellectual functioning. Furthermore, to assist students in an educational setting, it is important to develop an understanding regarding potential delays in cognitive functioning. Specifically, adolescence (12 to 20 years of age) is a time of significant cognitive change within an individual; therefore, one must possess the ability to distinguish between typical and atypical cognitive performance. Adolescents face increased pressures at academic and social levels (Watson & Gable, 2013). This period of development witnesses' students cultivating a stronger sense of abstract concepts and a more developed sense of executive functioning. Additionally, their processing speed, short-term memory capacity, and reasoning abilities significantly increase. There is also marked improvement in their verbal, mathematical, and visual-spatial ability (Feldman, 2003). With momentous changes occurring for adolescent individuals, school psychologist must be able to differentiate between a weak skill and true cognitive deficits in a skill (Watson & Gable, 2013). The difficulty in assessing the factors that compose intelligence lies in the fact that each of these constructs are theoretical in nature. To address this, professionals rely heavily on cognitive tests to provide insight to these latent constructs. Therefore, it is exceedingly important that cognitive tests are appropriately measuring the cognitive factors they purport to measure.

One theory that has garnered significant support in the intelligence community is CHC theory. This theory provides a common taxonomy and general understanding for the various mechanisms of intelligence (Newton & McGrew, 2010). Fortunately, intelligence testing instruments are currently developed based on extensive research and strong theoretical foundations (Keith & Reynolds, 2010). The Woodcock-Johnson IV is purported to be solely grounded in modern CHC theory and current neuropsychological research (McGrew et al., 2014). In the education setting, the WJ IV (Schrank, McGrew, & Mather, 2014a) is a commonly used instrument employed to garner an understanding of a child's cognitive functioning (Schrank & Wendling, 2012).

When used together, the Woodcock-Johnson IV Tests of Cognitive Abilities (Schrank, McGrew, & Mather, 2014b) and Tests of Oral Language (Schrank, Mather, & McGrew, 2014b) indicate they measure a total of seven broad CHC factors: comprehension-knowledge, fluid reasoning, short-term working memory, long-term retrieval, visual processing, auditory processing, and processing speed. The purpose of the current study is to evaluate the factor structure of the WJ IV COG and OL for the 14- to 19- year old age group through exploratory factor analysis to confirm the test publishers' factor structure of the two instruments.

Importance of the Current Study

Decisions for special education should never be taken lightly by professionals in education. Appropriate test interpretation leads to strong interventions and this remains paramount for student success. The research conducted on the WJ assessment family, as well as other intelligence tests currently used in the field, have indicated a significant concern regarding the reported factor structure of these instruments. Specifically, the uneasiness lies in the fact that many testing instruments are hypothesized to be over-factoring; in turn, potentially leading to one's overreliance on a test's abilities (Dombrowski et al., 2016; Frazier & Youngstrom, 2007).

The publishers of the WJ IV COG and OL provide support for the factor structure through a variety of statistical procedures. However, there was not an initial exploratory factor analysis conducted prior to their statistical analyses. The utilization of an EFA allows a non-biased approach to determining the factor structure of an instrument *a priori* (Henson & Roberts, 2006). Specifically, "EFA is used when a researcher wishes to identify a set of latent constructs underlying a battery of measured variables" (Fabrigar et al., 1999, p. 275). Rather than perform an EFA, McGrew and colleagues (2014) employed a principal component analysis (PCA) of the norming data to provide support for the factor structure of the test battery. This is potentially problematic because PCA analyzes the data without appropriate consideration to the latent factors. To be more precise, PCA does not differentiate between the shared variance and unique variance between manifest variables; therefore, there is an increased chance of inflating the variance of the components which potentially leads to over-factoring (Costello & Osborne, 2005).

The results of the various analyses laid out by McGrew and colleagues (2014) in the technical manual for the WJ IV are disjointed and obfuscated. There is not a concise discussion on the factor structure of the testing battery nor the limitations professionals may encounter during interpretation. Considering the WJ IV is frequently used in an education setting, it is imperative that testing administrators are aware of potential limitations in the instruments they elect to utilize.

CHAPTER III

METHOD

The purpose of this chapter is to outline the methodology of the current study for examining the factor structure of the Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014b) and the Woodcock-Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b). This chapter includes a discussion of research participants and procedures as well as information about the WJ IV COG and OL. Additionally, the data analyses used for this study are identified and discussed in detail.

Research Participants and Procedures

The data for this study was extracted from the norming sample data reported in the Woodcock-Johnson IV (WJ IV) Technical Manual (McGrew et al., 2014). The current study utilized the 14- to 19- year-old subtest correlation matrix derived from the norming group for the Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014b) and the Woodcock-Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b).

According to the information presented in the manual, the total norming sample consisted of 7, 416 individuals ranging in age from 2 to greater than 80 years of age. The sample represented participants from diverse populations and included people from 46

States and the District of Columbia. The authors used a stratified sampling design to randomly select examinees from the following groups: geographic region, sex, birthplace, race, ethnicity, location of community, parental education, type of school or college, level of education, status of employment, and occupation.

The total school-aged norming group consisted of 3, 891 total participants ranging from kindergarten to twelfth grade. Demographics are provided for the K through 12 normative group. Of that group, 49.4 percent were male and 50.6 percent female. This sub-sample was 63.2 percent White/Non-Hispanic, 13.8 percent Black/Non-Hispanic, 15.2 percent White/Hispanic, 4.2 percent Asian/Pacific Islander/Non-Hispanic, and 13.6 percent from other racial and/or ethnic backgrounds (McGrew et al., 2014). The 14- to 19- year old group included a total of 1, 685 participants and specific demographic breakdown for this sub-sample was not provided in the technical manual. Further detailed information regarding the norming sample and employed procedures are located in the Technical Manual.

Regarding clinical populations, the norming process included participants from an assortment of educational and clinical settings. These participants were not randomly selected, rather their participation relied solely on inclusion criteria for each study group. As a result, the technical manual states that the results for the clinical populations provided in the technical manual are not to be considered statistically precise (McGrew et al., 2014). Selective subtests were administered from a total of nine preselected populations: gifted (n = 53), intellectual disabilities (n = 50), learning disability in reading (*n* = 79), learning disability in mathematics (*n* = 73), learning disability in writing (*n* = 75), language delay (*n* = 75), attention deficit/hyperactivity disorder (*n* = 50), head injury (*n* = 12), and autism spectrum disorder (*n* = 50: McGrew et al., 2014).

It is important to note that all sample participants were not administered the full WJ IV testing battery during the norming process. As a result of this, the test publishers employed a planned incomplete data collection method known as multiple matrix sampling design (McGrew et al., 2014). According to the Technical Manual, the test developers administered a set of core tests to the norming participants and then matrix sampled the remaining subtests. Therefore, all results presented in the Technical Manual, including the correlation matrices, specifically relied on imputed data from the norming sample.

Measurement Instrumentation

The Woodcock-Johnson IV (WJ IV) is a comprehensive assessment battery purported to measure cognitive functioning, language, and academic achievement (Schrank, McGrew, & Mather, 2014a). The WJ IV is theoretically based on the Cattell-Horn-Carroll (CHC) theory of broad and narrow abilities (McGrew et al. 2014). In addition to the WJ IV COG and WJ IV OL, the WJ IV series includes the Woodcock-Johnson IV Tests of Achievement (WJ IV ACH; Schrank, Mather, & McGrew, 2014a). These three instruments can be used in combination together or independently.

Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG) and Woodcock-Johnson IV Tests of Oral Language (WJ IV OL)

The WJ IV COG consists of 18 subtests separated into two distinct batteries: Standard Battery (Tests 1-10) and Extended Battery (Tests 11-18). The standard and extended versions of the test, as well as the subtests, can be used independently or separately based on the needs of the examiner. The authors identify a variety of interpretive data available to the administrator based on the selected subtests administered. Three potential composites gleaned from testing administration include a General Intellectual Ability (GIA), the Brief Intellectual Ability, and a G*f*-G*c* composite. Seven broad CHC factors are available depending on the administered subtests and include G*c*, G*f*, G*wm*, G*s*, G*a*, G*lr*, and G*v*. The narrow abilities incorporated in the WJ IV COG include quantitative reasoning, auditory memory span, number facility, perceptual speed, vocabulary, and cognitive efficiency (McGrew et al., 2014; Schrank et al., 2016). The current research study examines data from all eighteen subtests on the WJ IV COG.

The WJ IV OL consists of 12 subtests which include nine subtests presented in English and three subtests in Spanish. This optional testing battery can be administered individually or in combination with the WJ IV COG or WJ IV ACH. The English oral language clusters available upon testing include Oral Language, Broad Oral Language, Oral Expression, Listening Comprehension, Phonetic Coding, and Speed of Lexical Access. Additionally, when the WJ IV OL is combined with the WJ IV COG, two additional narrow abilities are available: vocabulary and auditory memory span (McGrew et al., 2014). The current study examines data from the nine English subtests on the WJ IV OL.

Reliability and Validity of the WJ IV COG and WJ IV OL

Reliability statistics for the WJ IV for the various age ranges are found in the Technical Manual (McGrew et al., 2014). Reliability involves the replication of performance across different administrations with similar testing environments (Field, 2009).

Internal-consistency reliability is a theoretical estimate that measures the homogeneity of test items. Specifically, internal-consistency refers to how well the testing items within a subtest are measuring the same latent construct. Demonstrating internal-consistency is extremely important when scores from subtests are combined to produce a composite score (Henson, 2001). Internal-consistency reliabilities for the untimed subtests producing dichotomous scores on the WJ IV were calculated by separating the test questions into odd and even numbers through the split-half procedure and using the Spearman-Brown correction formula (McGrew et al., 2014). Per Cronbach (1947), the split-half method is commonly used and involves assessing the correlation between half of the items within a specific subtest with the remaining items in that same subtest. The Spearman-Brown correction formula is utilized afterwards to address the adjusted length of the published test compared to the original piloted test (Cronbach, 1947). Internal-consistency reliability for timed subtests and subtests that required more subjective scoring was calculated using an item analysis model known as the Rasch model (Anderson, 1973). The reliability of cluster and composite scores was calculated using Mosier's formula and the reliabilities obtained from the weighted scores from the split-half procedure or the Rasch model (McGrew et al., 2014). Mosier's formula allows for the weighted composite or cluster score to be determined through the estimation of weighted scores (Mosier, 1943). McGrew and colleagues (2014) reported that apart from one subtest, all median test reliabilities were greater than .80.

Test-retest reliability is an evaluation of an individual's performance on the same assessment measure over two different time points. Strong test-retest reliability indicates that an individual's performance does not vary significantly between the two different testing times (Field, 2009). Test-retest reliability was calculated for the timed tests on the WJ IV for three different age groups: 7-11 years, 14-17 years, and 26-79 years. The participants were given each of the timed tests and administered the same test one day later (McGrew et al., 2014). Apart from two timed subtests, all test-retest reliabilities for the timed measures demonstrated reliability coefficients ranging from .80 to .90; thus, suggesting strong test-retest reliability (McGrew et al., 2014). Table 1 provides information from the test manual on the test-retest reliability, cluster summary, and reliability statistics for the 14- to 19- year old age group for the speeded tests.

Table 1

Summary Statistics and Test-Retest Reliability Coefficients for the WJ IV Speeded Test

		Mean		Standard Deviation				
Test	***n***	**Test**	**Retest**	**Test**	**Retest**	r_{12}	**Mean Difference**	**Difference in *SD* units**
Letter-Pattern Matching	49	554.40	564.41	10.13	11.65	0.88	10.01	0.99
Number-Pattern Matching	49	491.91	496.32	11.20	13.22	0.84	4.41	0.39
Pair Cancellation	49	550.94	563.87	18.33	16.32	0.89	12.93	0.71
Rapid Picture Naming	49	535.09	544.08	21.58	20.69	0.79	8.99	0.42
Sentence Reading Fluency	49	560.39	583.19	23.78	33.23	0.93	22.80	0.96
Math Facts Fluency	49	540.69	545.88	18.38	19.98	0.97	5.19	0.28
Sentence Writing Fluency	49	547.15	557.92	11.58	15.73	0.76	10.77	0.93
Word Reading Fluency	49	550.38	573.82	21.93	22.30	0.91	23.44	1.07
Median						0.88	10.39	0.82

Validity, in a broad sense, specifies whether a test is measuring what it purports to measure (Field, 2009). Furthermore, validity is an extremely important concept in test development and must be continuously evaluated throughout the life of a testing instrument (Cronbach, 1988). The WJ IV Technical Manual provides a comprehensive and lengthy discussion on a variety of validity measures evaluated in the WJ IV battery; however, for the current study, special consideration will be given to content, construct, and concurrent validity.

Content validity is demonstrated when the specific content of a test directly aligns to the latent construct the test is attempting to measure (Field, 2009). As stated in chapter two, there are significant concerns regarding the factor structure of the WJ IV battery. McGrew et al. (2014) initially established the content validity for the WJ IV by acknowledging this most recent version of the test was grounded in theory based on analysis of the three previous editions of the Woodcock-Johnson assessment batteries: the Woodcock-Johnson Psycho-Educational Battery (WJ; Woodcock & Johnson, 1977), the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R; Woodcock & Johnson, 1989), and the Woodcock-Johnson III (WJ III; Woodcock et al., 2001). Specifically, the WJ assessment tools were based on the Cattell-Horn Extended G*f*-G*c* and the Cattell-Horn-Carroll (CHC) theories of general intelligence (Ding & Alfonso, 2016; McGrew et al., 2014). Additionally, the authors reported there were low correlations between the cognitive cluster scores and the achievement cluster scores; thus, identifying this as further support that the WJ IV is measuring the identified theoretical CHC constructs. In conjunction with evaluating the intercorrelations between tests and cluster scores, the test publishers also completed further multivariate statistical procedures. These additional analyses were broken into three separate stages: split-sample random sample generation, exploratory structural model generation and evaluation, and confirmatory structural model cross validation (McGrew et al., 2014). In the split-sample random sample generation stage, the six different age groups were equally divided between the model development and model cross-validation samples. The model development stage utilized cluster analysis, exploratory principal components analysis, and multidimensional scaling analysis to evaluate the factor structure of the WJ IV. During this stage of modeling, the authors stated the solutions ranged from eight to ten components or factors; therefore, these solutions were identified as the most interpretable and were retained for further analysis (McGrew et al., 2014). Finally, the last stage consisted of confirmatory structural modeling to evaluate the goodness of fit indices. Two final models were identified as appropriately representing the construct validity of the WJ IV: a top-down model consisting of nine broad CHC factors and a bottom-up model consisting of thirteen broad and narrow CHC factors (McGrew et al., 2014; Ding & Alfonso, 2016).

Concurrent validity addresses how closely an assessment tool correlates with other established testing instruments (Field, 2009). The relationship between the WJ IV battery and other conventional instruments was examined to evaluate the concurrent validity. Specifically, the WJ IV COG was compared against other intelligence measures and the WJ IV OL was compared against other language measures. These analyses were primarily conducted at the cluster or composite levels for all comparisons (McGrew et al., 2014). The Technical Manual provides a detailed breakdown of the demographics for the different comparison groups.

With regard to the WJ IV COG, the 14- to 19- year old age range was compared to four different assessments of cognitive intelligence in order to determine concurrent validity: the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003), the Wechsler Adult Intelligence Scale – Fourth Edition (WAIS-IV; Wechsler, 2008), the Kaufman Assessment Battery for Children – Second Edition (KABC-II; Kaufman & Kaufman, 2004), and the Stanford-Binet Intelligence Scales, Fifth Edition (SB5: Roid, 2003). Measures of overall general intelligence for the WJ IV COG were found to be highly correlated with the Full Scale Intelligence Quotient (FSIQ) from the WISC-IV (.83 to .86) and the WAIS-IV (.74 to .84). The KABC-II Fluid-

Crystallized Index (FCI) and the WJ IV COG clusters for intelligence were found to be highly correlated (.71 to .77). Finally, the FSIQ for the SB5 was also found to be highly correlated with the intelligence clusters for the WJ IV COG (.79 to .82). A more detailed discussion regarding correlations between the various cluster and composite scores for the cognitive measures can be found in the Technical Manual (McGrew et al., 2014).

The WJ IV OL was compared to four additional measures of language abilities for the 14- to 19- year old age range to evaluate concurrent validity: the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4: Semel, Wiig, & Secord, 2003), the Peabody Picture Vocabulary Test – Fourth Edition (PPVT-4: Dunn & Dunn, 2007), the Oral and Written Language Scales: Listening Comprehension/Oral Expression (OWLS: Carrow-Woolfolk, 1995), and the Comprehensive Assessment of Spoken Language (CASL: Carrow-Woolfolk, 1999). In general, there were high correlations observed between the primary WJ IV OL clusters and the PPVT-4 and CELF-4 composites. Additionally, the CASL Core Composite and the OWLS Oral Composite were found to be highly correlated with the WJ IV clusters that measured oral language abilities (McGrew et al., 2014). Additional information regarding the specific breakdown of the correlations for the assessment of language is available in the WJ IV OL Technical Manual.

Data Analysis

The purpose of the current study is to objectively assess the underlying factor structure of the WJ IV COG and WJ IV OL for the 14- to 19- year old age group. As previously discussed, the information provided in the Technical Manual (McGrew et al., 2014) regarding the statistical analyses and factor structure of the COG and OL are difficult to disentangle for the lay reader.

This study seeks to corroborate the hypothesized factor structure presented by the test publishers using exploratory factor analysis. The data needed for the EFA is provided in the correlation matrix listed in the Technical Manual for this specific age range. The findings garnered from this study will be useful in examining whether the currently proposed factor structure for the WJ IV COG and WJ IV OL battery is supported through an objective empirical evaluation of the test.

Overview

Factor analysis is an analytical method utilized to parsimoniously determine the interrelationship between a set of variables (Gorsuch, 1983). More specifically, factor analysis is a quantitative scientific approach used to estimate the latent constructs present on an instrument through statistical analyses of the correlation patterns observed between measured variables (Fabrigar & Wegener, 2012). Latent variables are factors that are unable to be directly measured, such as intelligence, while manifest or observed variables are measured directly, such as test scores (Keith, 2015).

The current understanding of factor analysis is based on the common factor model originally proposed by L. L. Thurstone. This model posits that correlated measured variables are a result of an underlying common factor (Thurstone, 1947). The common factor model utilizes matrices containing the correlation coefficients between two measured variables. These matrices are a visual representation of how closely related two observed variables appear and serve as the foundation for further analyses (Keith, 2015).

There are three primary types of factor analysis currently utilized: exploratory factor analysis, confirmatory factor analysis, and structural equation modeling (Meyers, Gamst, & Guarina, 2013). Exploratory factor analysis (EFA) is conducted when there is not a clear hypothesis as to the number of latent or common factors present. Confirmatory factor analysis (CFA) is utilized when there exists a predetermined supposition as to the number of latent factors expected. Structural equation modeling (SEM) estimates the relationship between variables and factors and utilizes a linear equation model to assess these relationships (Meyers et al., 2013).

This study examined the factor structure of the WJ IV COG and WJ IV OL without previous consideration of the publisher's hypothesized factor structure; therefore, an exploratory factor analysis was conducted. The following discussion systematically outlines the research methodology employed for this study.

Exploratory Factor Analysis

At the genesis of test development, test publishers typically hypothesize the presence of latent constructs on the instruments through qualitative processes. For example, test publishers heavily rely on theoretical rationale, such as CHC theory, for test conceptualization and interpretation (Fabrigar & Wegener, 2012). Additionally, when cultivating subsequent editions of a testing battery, publishers traditionally utilize data from the previously published tests in their developmental process.

As formerly discussed, exploratory factor analysis (EFA) is a quantitative approach used to empirically determine the number of common factors present in a testing instrument. Specifically, EFA is a common methodology employed to analyze preliminary data to determine how items naturally vary together. Using EFA allows for analysis of correlations between measured variables and provides an explanation regarding the variation and covariation between observed variables (Meyers et al., 2013). While on the surface, EFA appears to be a straightforward approach to evaluating latent constructs, great care must be utilized in conducting and interpreting the results from this method of evaluation (Frazier & Youngstrom, 2007; Henson & Roberts, 2006; Preacher & MacCallum, 2003).

Assumptions

Preceding the data analyses for this study, it was important to determine if conducting an EFA was a viable option (Fabrigar & Wegener, 2012). An initial evaluation of general assumptions was performed *a priori* to determine the suitability of the data for an EFA.

Originally, the proposed data analyses for this study included an evaluation of sample size. When conducting an EFA, it is important to ensure that the sample size is large enough to adequately represent the population (Fabrigar & Wegener, 2012). According to the technical manual, the norming sample for the WJ IV COG and WJ IV OL for the 14- to 19- year old age range consisted of 1,685 participants, which the authors identified as representative of the population (Fabrigar et al., 1999; McGrew et al.

2014). This study utilized the correlation matrix provided in the technical manual; therefore, the assumption of sample size was not further evaluated.

Bartlett's test of sphericity (Bartlett, 1950) was implemented to determine the variance that exists between variables within the correlation matrix. For the purposes of EFA, the variables must exhibit sufficient correlation with each other and not be random in nature. In order to establish whether EFA is an appropriate analysis, the Bartlett's test of sphericity must be significant at the $p < 0.05$ level. If $p > 0.05$ on Bartlett's test, an EFA is deemed inappropriate for examining the factor structure of an instrument (Bartlett, 1950).

Lastly, when examining the suitability of EFA, it is necessary to determine the degree of common variance through the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy. The KMO test examines whether there are enough items present for each factor to be identified (Kaiser, 1974). The closer the value of KMO is to 1.00 indicates that the unique factors lack correlation (Fabrigar & Wegener, 2012). The reported index for the KMO ranges from 0.00 to 1.00, with 0.60 or higher generally considered an appropriate KMO value to conduct an EFA (Meyers et al., 2013).

Model Fitting Procedures

After test assumptions are evaluated and the data is deemed appropriate for conducting an EFA, a model fitting (i.e. factor extraction) procedure must be determined (Fabrigar & Wegener, 2012). Appropriate model fitting is important because an EFA does not simply assume the factor loadings between the latent and measured variables.

Factor loadings are the correlation coefficients that result from the variables and the factors (Fabrigar & Wegener, 2012). These loadings are projected in an EFA "when the model is fit to the data" (Fabrigar & Wegner, 2012, p. 8).

Factor extraction is the process of mathematically evaluating the correlations identified between the measured variables to determine the underlying factor structure of an instrument (Fabrigar et al., 1999; Keith, Caemmerer, & Reynolds, 2016). There are an extensive variety of different methods currently used for factor extraction and selecting the appropriate method for fitting the common factor model to the data can be profoundly perplexing (Henson & Roberts, 2006; Thompson & Daniel, 1996). The literature is deficient on identifying strengths and weaknesses of the various factor extraction methods available and oftentimes a selection occurs based on the default setting in different statistical packages (Costello & Osborne, 2005). Additionally, it is important to select a factor extraction method that neither overestimates nor underestimates the number of factors (Keith et al., 2016; Frazier & Youngstrom, 2007).

The current study originally proposed to employ multiple model fitting procedures to hypothesize the underlying factor structure of the WJ IV COG and WJ IV OL. The reasoning behind this decision was based on a branch of literature that identified how multiple estimation methods allow for an investigation regarding substantial differences between the models (de Winter & Dodou, 2012). The proposal for this study indicated that non-iterated principal axis factor analysis, iterated principal axis factor analysis, and maximum likelihood would all be conducted (Fabrigar et al., 1999; Fabrigar

& Wegener, 2012). These three model fitting procedures each have strengths and limitations associated with their usage; however, upon further evaluation of the nature and purpose of this study, only iterated principal axis factor analysis was performed.

The general goal of principal axis factor analysis is to parsimoniously identify common factors that account for the common variance among a set of variables. Using principal axis factor analysis also provides less chance of producing improper solutions when compared to other model fitting techniques because it does not assume multivariate normality. Additionally, implementing principal factoring analysis removes the potential random error from the factors; thus, allowing the relationship among the factors to more closely approximate the population. It is important to note that despite the conservative nature of conducting principal factor analysis, this methodology does not result in detailed goodness of fit indices and fails to provide confidence intervals or significance testing (Fabrigar et al., 1999; Fabrigar & Wegener, 2012; Gorsuch, 1983). Principal axis factor analysis consists of two specific model fitting procedures: non-iterated and iterated.

Iterated principal axis factor analysis, sometimes simply referred to as principal factor analysis, is a method for determining how well the common factor model fits the data (Fabrigar & Wegener, 2012). This model fitting procedure calculates the squared-multiple correlations for each variable to estimate the communalities. The squared-multiple correlations for "each measured variable refers to the proportion of variance in that variable that is accounted for by the remaining measured variables in the battery"

(Fabrigar Wegener, 2012, p. 43). Iterative principal axis factor analysis involves the *a priori* specification regarding the number of factors, estimates communalities, and then concludes with a repetitive procedure to determine a new factor loading matrix (Fabrigar & Wegener, 2012).

As stated previously, determining goodness of fit for principal factor analysis is somewhat more limited when compared to other factor extraction methods. The ordinary least squares function (OLS), also known as the discrepancy function, allows for the determination of the fit of the model to the data. Larger values indicate a poorer fit of the model to the data. The OLS is calculated by examining the discrepancy between the model and the reduced correlation matrix.

Factor Retention

Through the process of fitting the common factor model to the data, an important area of consideration involves deciding the numbers of factors to retain from the extraction process (Preacher & MacCallum, 2003). The goal of factor retention is to arrive at a parsimonious and reasonable decision regarding the underlying factor structure of a testing instrument (Fabrigar et al., 1999). For this step in an EFA, great care must be utilized to not over- or under-factor a test battery; thus, seeking to avoid interpretation errors. Best practice in the process of retaining factors is to employ multiple factor retention techniques to ascertain converging data regarding the underlying factor structure (Courtney, 2013). For this analysis, two extraction methods were implemented to determine the factor structure of the WJ IV COG and WJ IV OL (Frazier & Youngstrom, 2007): Kaiser criterion (Kaiser, 1960) and visual scree (Cattell, 1966).

Kaiser criterion, also known as the eigenvalue-greater-than-one rule (Kaiser, 1960) is a method of factor retention commonly employed by researchers. This approach uses the correlation matrix to calculate the eigenvalues for each variable. Specifically, when examining the total variance explained in an analysis, any eigenvalues below the value of one are essentially dropped. Therefore, the number of factors identified for the EFA are based on eigenvalues greater than one (Kaiser, 1960). While this method should not be used exclusively in determining the number of factors to retain, it is an important component in initially determining the number for factors to retain.

Visual scree is a factor retention approach originally postulated by Cattell (1966). This approach provides a visual representation of the analyzed data and is frequently used in current research (Fabrigar & Wegener, 2012). Specifically, Cattell's visual scree test involves plotting the eigenvalues obtained during extraction and connecting the lowest values obtained with a straight line. The eigenvalues are acquired from either the reduced or unreduced correlation matrix; however, for this study, the reduced values for the matrix was utilized. The eigenvalues located above the specified line are then retained as identified factors (Cattell, 1966; Frazier & Youngstrom, 2007). The visual scree test has been cited as a reliable approach to factor retention when the common factors are fairly strong (Courtney, 2013).

Factor retention has been cited as one of the most important, yet subjective processes in EFA (Henson & Roberts, 2006). As previously stated employing a single method for factor retention is unadvisable and does not represent best practice when conducting factor analysis (Courtney, 2013; Fabrigar et al., 1999; Frazier & Youngstrom, 2007). While the presented factor retention processes independently have significant discussion and support in the literature, the strongest approach to objective factor analysis of the WJ IV COG and WJ IV OL includes the utilization of Kaiser criterion and Cattell's visual scree test (1966).

Factor Rotation

Based on the nature of factor extraction and retention, an infinite number of solutions are possible when conducting an EFA for two or more factors; therefore, a selection of the single, best-fitting solution must be identified (Fabrigar & Wegener, 2012). Per Thurstone (1947), the simplest structure is achieved by rotating the factors in multidimensional space, otherwise known as factor rotation.

There are two overarching rotations used in this specific process: orthogonal and oblique rotations. Orthogonal rotation is primarily used when the data is considered uncorrelated, while oblique rotation is typically utilized when the data is correlated (Meyers et al., 2013). Furthermore, there are a variety of methods available when considering the use of an orthogonal or oblique rotation. With regard to oblique factor rotation, promax rotation produces a simpler structure by allowing for lower cross-loadings while maintaining the original larger factor loadings (Hendrickson & White, 1964).

Oblique rotation is frequently identified in research relating to psychological testing batteries due to the expected nature of correlation among identified factors (Fabrigar et al., 1999). Therefore, the proposed study employed an oblique, promax factor rotation in the final data analyses.

Chapter Summary

This chapter outlined the research design and methodology utilized to examine the factor structure of the WJ IV COG and WJ IV OL through an exploratory factor analysis. The data for the analysis was extracted from the subtest correlation matrix for the 14- to 19- year old nationally represented norming sample located in the Technical Manual (McGrew et al., 2014). This type of analyses allows for an objective and impartial approach to determining the factor structure underlying the WJ battery for this age range. A detailed breakdown of each step was provided in this chapter: assumptions, model fitting procedures, factor retention, and factor rotation.

CHAPTER IV

RESULTS

The primary focus of this study was to objectively determine the number of factors present on the Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank, McGrew, & Mather, 2014b) and the Woodcock-Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014b) for the 14- to 19- year old standardization sample. Exploratory factor analysis permitted the data to speak for itself; thus, allowing the factors to freely load without influence from the investigator (Fabrigar & Wegener, 2012). The statistical software used for the analyses was employed through the syntax editor provided in the Statistical Package for the Social Sciences (SPSS) 25. The results presented include data from four separate iterated principal axis factor analyses with oblique promax rotation.

Descriptive Discussion

The results obtained from this study are based on EFA using the correlation matrix provided in the Technical Manual (McGrew et al., 2014). This group included a total of 1, 685 participants and specific demographic breakdown for this sub-sample was not provided in the manual. The analyses examined the twenty-seven subtests on the WJ IV COG and WJ IV OL. See Table 2 for information on the subtests examined in the EFA. Table 3 provides the correlation matrix.

Table 2

Subtests, Abbreviations, and Descriptions for the WJ IV COG and WJ IV OL

Subtest (CHC Factor)	Abbreviations	Description
Oral Vocabulary (Gc)	OV	Identifying synonyms and antonyms
Phonological Processing (Ga)	PP	Examining phonological processes
Object-Number Sequencing (Gwm)	ONS	Recalling sequential information
Oral Comprehension (Gc)	OC	Providing words based on syntax cues
Picture Vocabulary (Gc)	PV	Identifying pictures of objects
Sound Awareness (Ga)	SA	Identifying parts of presented words
Concept Formation (Gf)	CF	Identifying rules composing a stimulus
Verbal Attention (Gwm)	VA	Recalling a variety of oral information
Understanding Directions (Gwm)	UD	Recalling oral instructions sequentially
Numbers Reversed (Gwm)	NR	Recalling numbers in a specific order
Number Series (Gf)	NS	Identifying missing numbers in series
Memory for Words (Gwm)	MFW	Recalling a variety of unrelated words
General Information (Gc)	GI	Providing general information
Analysis-Synthesis (Gf)	AS	Completing a series of complex tasks
Segmentation (Ga)	SEG	Identifying parts of words
Visualization (Gv)	VIS	Identifying parts of a shape or pattern
Sentence Repetition (Gwm)	SENR	Recalling presented material
Story Recall (Glr)	STOR	Recalling details from oral stories
Retrieval Fluency (Glr)	RF	Providing examples from a category
Letter-Pattern Matching (Gs)	LPM	Identifying letter patterns in rows
Nonword Repetition (Ga)	NWR	Recalling nonsense words provided
Sound Blending (Ga)	SB	Blending sounds to make words
Pair Cancelation (Gs)	PC	Identifying repeated patterns
Number-Pattern Matching (Gs)	NPM	Identify number patterns in rows
Visual-Auditory Learning (Glr)	VAL	Recalling visual-auditory associations
Rapid Picture Naming (Glr)	RPN	Timed naming of pictures
Picture Recognition (Gv)	PR	Recalling pictures despite distractors

Table 3

*Correlation matrix for 14- to 19- year old on WJ IV COG and WJ IV OL (*n *= 1,685, 27 subtests)*

Subtest	OV	NS	VA	LPM	PP	STOR	VIS	GI	CF	NR	NPM	NWR	VAL	PR	AS	ONS	PC	MFW	PV	OC	SEG	RPN	SENR	UF	SB	RF	SA
OV	1.00	0.51	0.50	0.37	0.58	0.42	0.36	0.74	0.49	0.42	0.33	0.41	0.38	0.17	0.40	0.45	0.36	0.36	0.72	0.67	0.41	0.27	0.48	0.41	0.40	0.43	0.47
NS	0.51	1.00	0.41	0.40	0.49	0.41	0.41	0.36	0.50	0.47	0.45	0.26	0.23	0.09	0.46	0.41	0.29	0.28	0.39	0.41	0.35	0.18	0.27	0.48	0.26	0.27	0.50
VA	0.50	0.41	1.00	0.30	0.47	0.33	0.25	0.40	0.30	0.49	0.36	0.45	0.21	0.13	0.37	0.57	0.30	0.53	0.39	0.47	0.33	0.31	0.51	0.46	0.29	0.31	0.42
LPM	0.37	0.40	0.30	1.00	0.39	0.29	0.30	0.29	0.30	0.45	0.57	0.22	0.30	0.26	0.37	0.50	0.58	0.30	0.31	0.17	0.28	0.33	0.28	0.34	0.26	0.33	0.36
PP	0.58	0.49	0.47	0.39	1.00	0.25	0.37	0.45	0.45	0.45	0.32	0.36	0.37	0.12	0.36	0.48	0.40	0.48	0.51	0.53	0.58	0.28	0.50	0.36	0.52	0.47	0.59
STOR	0.42	0.41	0.33	0.29	0.25	1.00	0.41	0.35	0.39	0.32	0.31	0.34	0.34	0.33	0.43	0.45	0.20	0.38	0.40	0.45	0.31	0.18	0.29	0.43	0.24	0.27	0.39
VIS	0.36	0.41	0.25	0.30	0.37	0.41	1.00	0.35	0.48	0.39	0.29	0.34	0.39	0.39	0.51	0.41	0.25	0.37	0.39	0.33	0.44	0.20	0.31	0.40	0.34	0.27	0.42
GI	0.74	0.36	0.40	0.29	0.45	0.35	0.35	1.00	0.36	0.37	0.21	0.29	0.28	0.19	0.35	0.36	0.26	0.28	0.72	0.56	0.31	0.24	0.32	0.26	0.34	0.30	0.38
CF	0.49	0.50	0.30	0.30	0.45	0.39	0.48	0.36	1.00	0.40	0.26	0.35	0.46	0.20	0.52	0.46	0.33	0.37	0.39	0.40	0.43	0.32	0.33	0.45	0.40	0.31	0.43
NR	0.42	0.47	0.49	0.45	0.45	0.32	0.39	0.37	0.40	1.00	0.36	0.30	0.31	0.23	0.37	0.49	0.33	0.47	0.36	0.40	0.37	0.26	0.31	0.32	0.34	0.31	0.42
NPM	0.33	0.45	0.36	0.57	0.32	0.31	0.29	0.21	0.26	0.36	1.00	0.22	0.11	0.16	0.34	0.40	0.56	0.25	0.22	0.33	0.32	0.30	0.21	0.25	0.17	0.37	0.29
NWR	0.41	0.26	0.45	0.22	0.36	0.34	0.34	0.29	0.35	0.30	0.22	1.00	0.23	0.25	0.22	0.45	0.21	0.44	0.32	0.32	0.37	0.30	0.50	0.46	0.36	0.22	0.43
VAL	0.38	0.23	0.21	0.30	0.37	0.34	0.39	0.28	0.46	0.31	0.11	0.23	1.00	0.27	0.34	0.34	0.21	0.27	0.34	0.31	0.42	0.20	0.20	0.31	0.41	0.21	0.36
PR	0.17	0.09	0.13	0.26	0.12	0.33	0.39	0.19	0.20	0.23	0.16	0.25	0.27	1.00	0.34	0.30	0.11	0.26	0.19	0.32	0.25	0.35	0.24	0.31	0.17	0.29	0.24
AS	0.40	0.46	0.37	0.37	0.36	0.43	0.51	0.35	0.52	0.37	0.34	0.22	0.34	0.34	1.00	0.45	0.32	0.38	0.31	0.34	0.39	0.21	0.22	0.39	0.37	0.45	0.31
ONS	0.45	0.41	0.57	0.50	0.48	0.45	0.41	0.36	0.46	0.49	0.40	0.45	0.34	0.30	0.45	1.00	0.38	0.52	0.39	0.40	0.37	0.40	0.42	0.45	0.38	0.45	0.37
PC	0.36	0.29	0.30	0.58	0.40	0.20	0.25	0.26	0.33	0.33	0.56	0.21	0.21	0.11	0.32	0.38	1.00	0.22	0.28	0.26	0.28	0.37	0.26	0.30	0.27	0.36	0.31
MFW	0.36	0.28	0.53	0.30	0.48	0.38	0.37	0.28	0.37	0.47	0.25	0.44	0.27	0.26	0.38	0.52	0.22	1.00	0.34	0.34	0.40	0.30	0.49	0.39	0.40	0.46	0.36
PV	0.72	0.39	0.39	0.31	0.51	0.40	0.39	0.72	0.39	0.36	0.22	0.32	0.34	0.19	0.31	0.39	0.28	0.34	1.00	0.67	0.29	0.36	0.46	0.39	0.34	0.42	0.42
OC	0.67	0.41	0.47	0.17	0.53	0.45	0.33	0.56	0.40	0.40	0.33	0.32	0.31	0.32	0.34	0.40	0.26	0.34	0.67	1.00	0.37	0.33	0.52	0.42	0.37	0.44	0.44
SEG	0.41	0.35	0.33	0.28	0.58	0.31	0.44	0.31	0.43	0.37	0.32	0.37	0.42	0.25	0.39	0.37	0.28	0.40	0.29	0.37	1.00	0.23	0.34	0.33	0.45	0.26	0.49
RPN	0.27	0.18	0.31	0.33	0.28	0.18	0.20	0.24	0.32	0.26	0.30	0.30	0.20	0.35	0.21	0.40	0.37	0.30	0.36	0.33	0.23	1.00	0.28	0.38	0.20	0.44	0.26
SENR	0.48	0.27	0.51	0.28	0.50	0.29	0.31	0.32	0.33	0.31	0.21	0.50	0.20	0.24	0.22	0.42	0.26	0.49	0.46	0.52	0.34	0.28	1.00	0.49	0.19	0.29	0.40
UD	0.41	0.48	0.46	0.34	0.36	0.43	0.40	0.26	0.45	0.32	0.25	0.46	0.31	0.31	0.39	0.45	0.30	0.39	0.39	0.42	0.33	0.38	0.49	1.00	0.34	0.28	0.44
SB	0.40	0.26	0.29	0.26	0.52	0.24	0.34	0.34	0.40	0.34	0.17	0.36	0.41	0.17	0.37	0.38	0.27	0.40	0.34	0.37	0.45	0.20	0.19	0.34	1.00	0.28	0.45
RF	0.43	0.27	0.31	0.33	0.47	0.27	0.27	0.30	0.31	0.31	0.37	0.22	0.21	0.29	0.45	0.45	0.36	0.46	0.42	0.44	0.26	0.44	0.29	0.28	0.28	1.00	0.32
SA	0.47	0.50	0.42	0.36	0.59	0.39	0.42	0.38	0.43	0.42	0.29	0.43	0.36	0.24	0.31	0.37	0.31	0.36	0.42	0.44	0.49	0.26	0.40	0.44	0.45	0.32	1.00

Exploratory Factor Analyses

As previously indicated, the discussion regarding the results of the EFA for this study include data from four separate factor analyses. The purpose of multiple EFAs was to address the multi- or cross-loadings as well as the weak loadings identified in the pattern matrix for each analysis while striving to account for the largest amount of variance. This section provides four separate discussions regarding each EFA. Immediately following the discussion of results for each factor analysis, tables are provided for additional support.

Exploratory Factor Analysis I

Prior to employing an EFA, a brief analysis of assumptions was performed. The results from Bartlett's Test of Sphericity (Bartlett, 1950) identified that the correlation matrix was not random in nature ($\chi^2 = 25{,}256.93$, $df = 351$, $p < .000$). The degree of common variance was determined through the Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) measure of sampling adequacy and found to be appropriate at .898. Therefore, based on the results from Bartlett's and KMO, the correlation matrix was deemed to be suitable for the exploratory factor analysis.

Communality estimates allow the researcher to examine the relationship amongst the studied variables. Communality is the variance in a manifest variable that is explained by all the common factors within the model. Reporting this information allows one to determine if there is an association between all variables and factors or outliers that require further consideration (Fabrigar & Wegener, 2012). An interpretation of the factors loadings for each manifest variable was important in assisting to determine the underlying factor structure of the WJ IV COG and WJ IV OL. See Table 4.

Table 4

Communalities for EFA I

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	0.75	0.79
Number Series (NS)	0.60	0.64
Verbal Attention (VA)	0.60	0.60
Letter-Pattern Matching (LPM)	0.63	0.59
Phonological Processing (PP)	0.69	0.75
Story Recall (STOR)	0.47	0.47
Visualization (VIS)	0.50	0.50
General Information (GI)	0.66	0.64
Concept Formation (CF)	0.54	0.50
Numbers Reversed (NR)	0.47	0.41
Number-Pattern Matching (NPM)	0.57	0.58
Nonword Repetition (NWR)	0.46	0.45
Visual-Auditory Learning (VAL)	0.40	0.40
Picture Recognition (PR)	0.43	0.45
Analysis-Synthesis (AS)	0.55	0.50
Object-Number Sequencing (ONS)	0.59	0.56
Pair Cancellation (PC)	0.50	0.53
Memory for Words (MFW)	0.56	0.52
Picture Vocabulary (PV)	0.71	0.75
Oral Comprehension (OC)	0.70	0.64
Segmentation (SEG)	0.50	0.50
Rapid Picture Naming (RPN)	0.43	0.41
Sentence Repetition (SENR)	0.59	0.56
Understanding Directions (UD)	0.53	0.49
Sound Blending (SB)	0.48	0.47
Retrieval Fluency (RF)	0.54	0.45
Sound Awareness (SA)	0.54	0.49

Extraction Method: Principal Axis Factoring.

The factor extraction method used was iterated principal factor analysis with an oblique, promax rotation. For this study, the number of factors identified for the EFA were based on eigenvalues greater than one (Kaiser, 1960) and visual scree (Cattell, 1966). A total of six factors were retained based on the Kaiser criterion. Factor one accounted for 38.73% of the variance, factor two accounted for 6.22% of the variance, factor three accounted for 5.48% of the variance, factor four accounted for 5.09% of the variance, factor five accounted for 4.61% of the variance, and factor six accounted for 4.08% of the variance. Cumulatively, the total amount of variance explained by the six factors was 64.21%. The visual scree also aids in visually identifying the number of factors present by plotting the number of factors to the eigenvalues. Upon examination of the visual scree plot, the initial EFA appeared to identify between six and eight factors respectively. See Table 5 and Figure 1.

Table 5

Total Variance Explained for EFA I

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation of Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	10.46	38.73	38.73	10.02	37.10	37.10	6.97
2	1.68	6.22	44.95	1.29	4.78	41.88	7.70
3	1.48	5.48	50.43	1.02	3.78	45.66	7.42
4	1.37	5.09	55.52	0.90	3.34	49.00	6.51
5	1.24	4.61	60.13	0.78	2.88	51.88	6.00
6	1.10	4.08	64.21	0.63	2.33	54.21	1.96

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

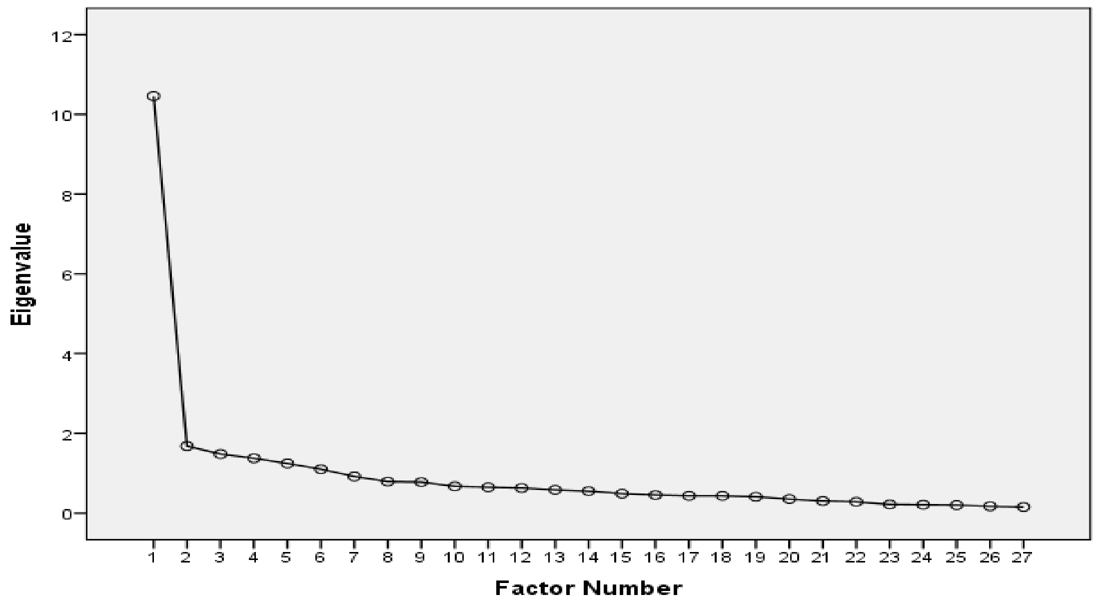

Figure 1. Visual Scree Plot for EFA I

Factor loadings allow for a numerical value to identify the relationship each observed variable has regarding the latent factor(s) present. This assisted in determining if the latent factors were independent of each other or if there exists overlap between them (Fabrigar & Wegener, 2012). Similar to the correlation coefficients, values for factor loadings range from -1 to 1 and the higher the number, the more salient the loading. A negative number indicates an inverse impact on the specific factor (Gorsuch, 1983). Factor loadings were considered to be salient at 0.30 and above (Fabrigar & Wegener, 2012). For this initial EFA, the pattern matrix converged after fifteen iterations and the subtests loaded onto six factors with four multi- or cross-loadings (see Table 6).

Table 6

Pattern Matrix for EFA I

Subtest	Factor 1	2	3	4	5	6
PV	0.89	-0.01	-0.03	-0.05	0.02	0.08
GI	0.84	-0.14	0.03	-0.03	0.09	-0.01
OV	0.77	0.03	0.06	0.04	0.10	-0.10
OC	0.68	0.20	-0.04	-0.10	0.07	0.10
SENR	0.14	0.80	-0.07	-0.12	-0.08	0.03
VA	0.05	0.76	-0.13	0.10	0.04	-0.12
NWR	-0.09	0.68	0.12	-0.17	0.07	0.06
MFW	-0.16	0.61	0.23	-0.02	-0.01	0.14
UD	-0.04	0.49	-0.04	-0.04	0.34	0.12
ONS	-0.07	0.40	0.04	0.27	0.15	0.16
SB	0.01	-0.01	0.74	-0.06	-0.03	0.00
SEG	-0.11	0.09	0.66	-0.02	0.09	-0.03
PP	0.17	0.20	0.65	0.15	-0.23	-0.21
VAL	0.06	-0.18	0.56	-0.11	0.21	0.15
SA	0.05	0.23	0.41	0.00	0.15	-0.11
NPM	-0.06	-0.02	-0.20	0.83	0.16	-0.02
PC	0.00	-0.11	0.07	0.79	-0.10	0.02
LPM	-0.09	-0.10	0.02	0.79	0.11	0.07
RF	0.22	0.03	0.08	**0.37**	-0.14	**0.30**
NR	0.00	**0.20**	0.16	**0.24**	0.19	-0.04
NS	0.11	0.06	-0.02	0.24	0.56	-0.30
STOR	0.15	0.14	-0.12	-0.06	0.56	0.17
VIS	-0.01	-0.05	0.30	-0.06	0.50	0.19
AS	0.01	-0.15	0.19	0.19	0.48	0.19
CF	0.06	-0.02	**0.37**	0.01	**0.38**	0.04
PR	-0.01	0.04	0.00	-0.04	0.25	0.58
RPN	0.10	0.19	-0.06	**0.34**	-0.16	**0.38**

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.[a]
a. Rotation converged in 15 iterations.

For the first factor, four subtests produced salient loadings: picture vocabulary (0.89), general information (0.84), oral vocabulary (0.77), and oral comprehension (0.68). Six subtests loaded onto the second factor: sentence repetition (0.80), verbal attention

(0.76), nonword repetition (0.68), memory for words (0.61), understanding directions (0.49), and object-number sequencing (0.40). The third factor consisted of five subtests: sound blending (0.74), segmentation (0.66), phonological processing (0.65), visual-auditory learning (0.56), and sound awareness (0.41). Factor four obtained three subtest loadings: number-pattern matching (0.83), pair cancelation (0.79), and letter-pattern matching (0.79). The fifth factor consisted of four subtests: number series (0.56), story recall (0.56), visualization (0.50), and analysis-synthesis (0.48). The final factor only demonstrated a single salient loading, picture recognition (0.58).

There was a total of four subtests that cross-loaded onto multiple factors, thus demonstrating a difference between the factors of less than .10. Retrieval fluency loaded onto factor four (0.37) and factor six (0.30) for a difference of 0.07. Numbers reversed loaded onto factors two (0.20) and four (0.24) for a total difference of 0.04. Concept formation loaded onto factor three (0.37) and factor five (0.38) for a difference of 0.01. Rapid picture naming loaded onto factors four (0.34) and six (0.38) for a difference of 0.04.

Exploratory Factor Analysis II

Based on the results from the initial EFA, a second EFA was performed after removing the subtest *Numbers Reversed* due to a multi-loading difference of < .100 across two factors. Additionally, the values obtained in the pattern matrix indicated a weak factor loading for *Numbers Reversed.* The second EFA results from Bartlett's Test of Sphericity (Bartlett, 1950) identified that the correlation matrix was not random in nature (χ^2 = 24,191.16, df = 325, p < .000). The degree of common variance was determined through the Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) measure of sampling

Table 7

Communalities for EFA II (Numbers Reversed Removed)

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	0.75	0.79
Number Series (NS)	0.59	0.63
Verbal Attention (VA)	0.59	0.57
Letter-Pattern Matching (LPM)	0.61	0.57
Phonological Processing (PP)	0.69	0.75
Story Recall (STOR)	0.46	0.47
Visualization (VIS)	0.49	0.50
General Information (GI)	0.66	0.65
Concept Formation (CF)	0.54	0.50
Number-Pattern Matching (NPM)	0.57	0.59
Nonword Repetition (NWR)	0.46	0.46
Visual-Auditory Learning (VAL)	0.40	0.40
Picture Recognition (PR)	0.43	0.45
Analysis-Synthesis (AS)	0.55	0.51
Object-Number Sequencing (ONS)	0.58	0.56
Pair Cancellation (PC)	0.50	0.54
Memory for Words (MFW)	0.54	0.50
Picture Vocabulary (PV)	0.71	0.75
Oral Comprehension (OC)	0.70	0.64
Segmentation (SEG)	0.50	0.50
Rapid Picture Naming (RPN)	0.43	0.41
Sentence Repetition (SENR)	0.59	0.57
Understanding Directions (UD)	0.53	0.50
Sound Blending (SB)	0.48	0.47
Retrieval Fluency (RF)	0.53	0.46
Sound Awareness (SA)	0.53	0.49

Extraction Method: Principal Axis Factoring.

adequacy and found to be an appropriate standard at .894. Therefore, based on the results from Bartlett's and KMO, the adjusted correlation matrix was determined to be appropriate for the second exploratory factor analysis. The second EFA identified the number of factors based on eigenvalues greater than one (Kaiser, 1960) and visual scree (Cattell, 1966). See Tables 7 and 8 and Figure 2. There was a total of six factors retained based on the Kaiser criterion. Factor on accounted for 38.72% of the variance, factor two accounted for 6.42% of the variance, factor three accounted for 5.69% of the variance, factor four accounted for 5.28% of the variance, factor five accounted for 4.74% of the variance, and factor six accounted for 4.22% of the variance. The total amount of variance explained by the six factors was 65.06%. Upon examination of the visual scree plot, the second EFA appeared to identify between six and seven factors.

Table 8

Total Variance Explained for EFA II (Numbers Reversed Removed)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation of Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	10.07	38.72	38.72	9.63	37.05	37.05	6.79
2	1.67	6.42	45.13	1.28	4.91	41.96	7.30
3	1.48	5.69	50.83	1.02	3.92	45.88	7.11
4	1.37	5.28	56.10	0.90	3.46	49.34	6.09
5	1.23	4.74	60.84	0.77	2.95	52.30	5.54
6	1.10	4.22	65.06	0.62	2.39	54.69	2.27

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

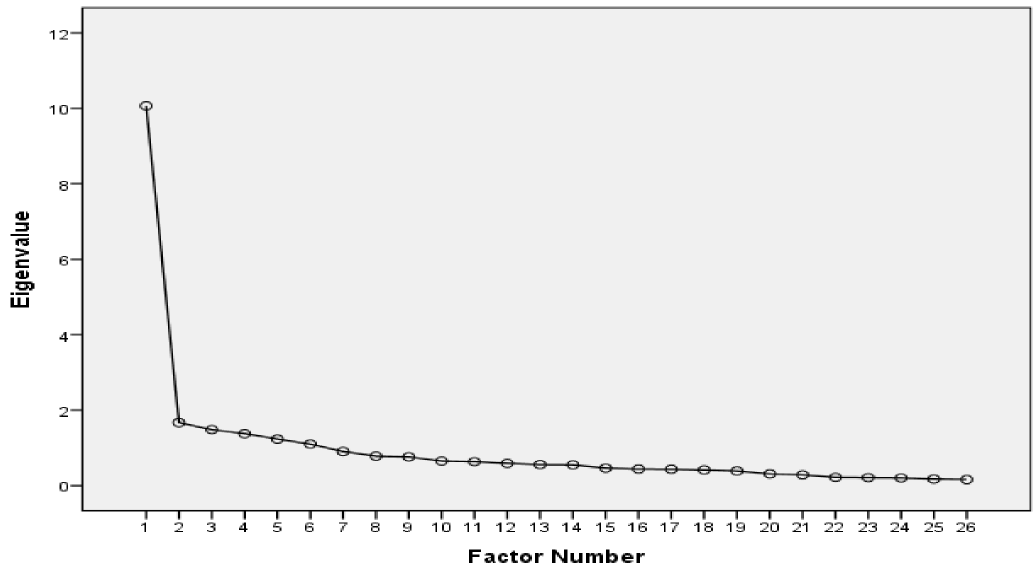

Figure 2. Visual Scree Plot for EFA II (Numbers Reversed Removed)

The pattern matrix converged after thirteen iterations on the second EFA and the subtests loaded onto a total of six factors with three multi-loadings (see Table 9). The first factor identified four salient subtest loadings: picture vocabulary (0.89), general information (0.86), oral vocabulary (0.76), and oral comprehension (0.67). The second factor was comprised of six subtests: sentence repetition (0.80), verbal attention (0.73), nonword repetition (0.68), memory for words (0.58), understanding directions (0.50), and object-number sequencing (0.38). The third factor consisted of five subtests: sound blending (0.74), segmentation (0.66), phonological processing (0.66), visual-auditory learning (0.56), and sound awareness (0.41). Three salient subtest loadings were in the fourth factor: number-pattern matching (0.82), pair cancelation (0.78), and letter-pattern matching (0.76). The fifth factor identified four subtests: number series (0.56), story recall (0.54), visualization (0.48), and analysis-synthesis (0.47). The sixth factor only demonstrated a single salient loading, picture recognition (0.59).

Table 9

Pattern Matrix for EFA II (Numbers Reversed Removed)

	Factor					
Subtest	1	2	3	4	5	6
PV	0.89	-0.01	-0.04	-0.05	0.01	0.08
GI	0.86	-0.14	0.03	-0.04	0.08	-0.01
OV	0.76	0.04	0.06	0.04	0.10	-0.11
OC	0.67	0.19	-0.04	-0.09	0.07	0.10
SENR	0.13	0.80	-0.07	-0.11	-0.06	0.02
VA	0.06	0.73	-0.12	0.10	0.04	-0.11
NWR	-0.10	0.68	0.12	-0.16	0.08	0.05
MFW	-0.14	0.58	0.23	-0.02	-0.02	0.16
UD	-0.06	0.50	-0.04	-0.02	0.35	0.10
ONS	-0.06	0.38	0.05	0.27	0.15	0.17
SB	0.01	-0.01	0.74	-0.06	-0.03	0.00
SEG	-0.11	0.10	0.66	-0.01	0.10	-0.03
PP	0.16	0.20	0.66	0.15	-0.21	-0.22
VAL	0.06	-0.18	0.56	-0.11	0.20	0.16
SA	0.05	0.23	0.41	0.01	0.16	-0.12
NPM	-0.07	-0.01	-0.19	0.82	0.17	-0.04
PC	-0.01	-0.10	0.08	0.78	-0.08	0.01
LPM	-0.09	-0.09	0.02	0.76	0.11	0.07
RF	0.23	0.02	0.08	**0.37**	-0.15	**0.31**
NS	0.11	0.07	-0.01	0.25	0.56	-0.30
STOR	0.15	0.14	-0.11	-0.05	0.54	0.17
VIS	-0.01	-0.04	0.30	-0.05	0.48	0.19
AS	0.01	-0.15	0.20	0.19	0.47	0.19
CF	0.05	-0.02	**0.37**	0.02	**0.38**	0.04
PR	-0.01	0.04	0.00	-0.05	0.23	0.59
RPN	0.09	0.19	-0.06	**0.32**	-0.15	**0.38**

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.[a]
a. Rotation converged in 13 iterations.

For this EFA, there was a total of three subtests that cross-loaded onto multiple factors and obtained a difference between the factors of less than .10. Retrieval fluency loaded onto factor four (0.37) and factor six (0.31) for a difference of 0.06. Concept formation loaded onto factor three (0.37) and factor five (0.38) for a difference of 0.01. Rapid picture naming loaded onto factors four (0.32) and six (0.38) for a difference of 0.06.

Exploratory Factor Analysis III

Based on the results from the second EFA, a third EFA was performed with the removal of two subtests, *Numbers Reversed* and *Retrieval Fluency.* Again, the removal of the two subtests were due to a multi-loading difference of < .100 across two factors as well as weak factor loadings. The third EFA results from Bartlett's Test of Sphericity (Bartlett, 1950) identified that the correlation matrix was not random in nature (χ^2 = 22,914.84, *df* = 300, *p* < .000). The degree of common variance was determined through the Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) measure of sampling adequacy and found to be an appropriate standard at .899. Therefore, based on the results from Bartlett's and KMO, the adjusted correlation matrix was determined to be appropriate for the third exploratory factor analysis.

This EFA also identified the number of factors based on eigenvalues greater than one (Kaiser, 1960) and visual scree (Cattell, 1966). See Tables 10 and 11 and Figure 3.

Table 10

Communalities for EFA III (Numbers Reversed and Retrieval Fluency Removed)

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	0.74	0.78
Number Series (NS)	0.58	0.65
Verbal Attention (VA)	0.57	0.58
Letter-Pattern Matching (LPM)	0.61	0.62
Phonological Processing (PP)	0.67	0.74
Story Recall (STOR)	0.46	0.47
Visualization (VIS)	0.49	0.50
General Information (GI)	0.66	0.66
Concept Formation (CF)	0.53	0.50
Number-Pattern Matching (NPM)	0.56	0.58
Nonword Repetition (NWR)	0.46	0.45
Visual-Auditory Learning (VAL)	0.40	0.41
Picture Recognition (PR)	0.42	0.46
Analysis-Synthesis (AS)	0.50	0.51
Object-Number Sequencing (ONS)	0.57	0.55
Pair Cancellation (PC)	0.50	0.56
Memory for Words (MFW)	0.50	0.49
Picture Vocabulary (PV)	0.70	0.75
Oral Comprehension (OC)	0.69	0.63
Segmentation (SEG)	0.49	0.50
Rapid Picture Naming (RPN)	0.39	0.38
Sentence Repetition (SENR)	0.58	0.57
Understanding Directions (UD)	0.53	0.49
Sound Blending (SB)	0.47	0.47
Sound Awareness (SA)	0.53	0.49

Extraction Method: Principal Axis Factoring.

There was a total of six factors retained based on the Kaiser criterion. Factor one accounted for 38.98% of the variance, factor two accounted for 6.60% of the variance, factor three accounted for 5.85% of the variance, factor four accounted for 5.47% of the variance, factor five accounted for 4.83% of the variance, and factor six accounted for

Table 11

Total Variance Explained for EFA III (Numbers Reversed and Retrieval Fluency Removed)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation of Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.75	38.98	38.98	9.31	37.25	37.25	7.18
2	1.65	6.60	45.59	1.27	5.09	42.34	6.40
3	1.46	5.85	51.44	1.01	4.04	46.38	6.82
4	1.37	5.47	56.91	0.89	3.57	49.95	5.21
5	1.21	4.83	61.74	0.75	2.99	52.94	6.50
6	1.05	4.20	65.94	0.57	2.29	55.23	2.11

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

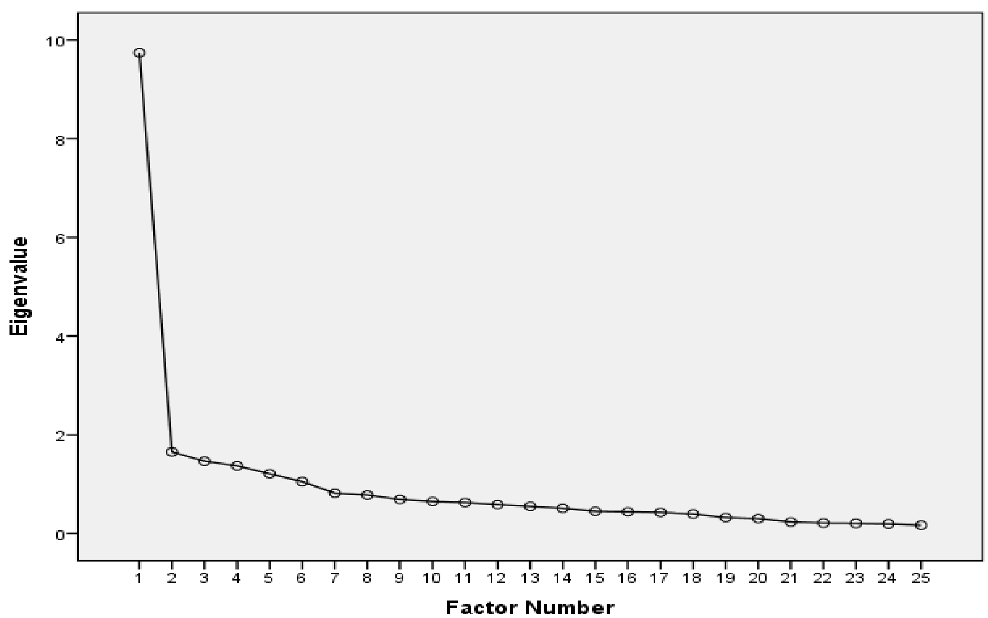

Figure 3. Visual Scree Plot for EFA III (Numbers Reversed and Retrieval Fluency Removed)

4.20% of the variance. The total amount of variance explained by the six factors was 65.94%. Upon examination of the visual scree plot, the third EFA appeared to identify six factors.

The pattern matrix converged after seven iterations and the subtests loaded onto a total of six factors with two multi-loadings (see Table 12). Six salient subtest loadings comprised the first factor: sentence repetition (0.80), verbal attention (0.73), nonword repetition (0.67), memory for words (0.63), understanding directions (0.48), and object-number sequencing (0.42). The second factor indicated four strong subtest loadings: picture vocabulary (0.88), general information (0.86), oral vocabulary (0.74), and oral comprehension (0.64). The third factor was comprised of five subtests: sound blending (0.73), segmentation (0.65), phonological processing (0.63), visual-auditory learning (0.58), and sound awareness (0.41). Three salient subtest loadings were in the fourth factor: pair cancellation (0.78), letter-pattern matching (0.77), and number-pattern matching (0.72). The fifth factor identified four subtests: number series (0.77), analysis-synthesis (0.55), story recall (0.53), and visualization (0.41). The sixth factor only demonstrated a single salient loading, picture recognition (0.62).

For the third EFA, a total of two subtests cross-loaded onto multiple factors and obtained a difference between the factors of less than .10. Rapid picture naming loaded onto factors four (0.36) and six (0.33) for a difference of 0.03. Concept formation loaded onto factor three (0.36) and factor five (0.38) for a difference of 0.02.

Table 12

Pattern Matrix for EFA III (Numbers Reversed and Retrieval Fluency Removed)

	Factor					
Subtest	1	2	3	4	5	6
SENR	0.80	0.14	-0.06	-0.07	-0.13	0.03
VA	0.73	0.05	-0.13	0.06	0.12	-0.13
NWR	0.67	-0.08	0.13	-0.11	-0.03	0.09
MFW	0.62	-0.15	0.21	-0.05	0.01	0.09
UD	0.48	-0.05	-0.03	-0.02	0.29	0.15
ONS	0.42	-0.06	0.05	0.24	0.14	0.14
PV	0.00	0.88	-0.02	0.00	-0.04	0.09
GI	-0.15	0.85	0.04	0.01	0.03	0.02
OV	0.03	0.74	0.07	0.04	0.13	-0.10
OC	0.22	0.64	-0.04	-0.09	0.07	0.07
SB	0.02	0.02	0.73	-0.03	-0.08	0.00
SEG	0.10	-0.10	0.65	0.01	0.06	-0.01
PP	0.24	0.14	0.63	0.12	-0.11	-0.27
VAL	-0.18	0.09	0.58	-0.04	0.07	0.22
SA	0.22	0.05	0.41	0.00	0.16	-0.08
PC	-0.08	0.04	0.10	0.78	-0.11	0.01
LPM	-0.09	-0.03	0.05	0.77	0.05	0.11
NPM	0.00	-0.05	-0.18	0.72	0.26	-0.05
RPN	0.26	0.12	-0.03	**0.36**	-0.26	**0.33**
NS	0.02	0.06	-0.04	0.11	0.77	-0.29
AS	-0.12	-0.03	0.18	0.10	0.55	0.16
STOR	0.12	0.13	-0.11	-0.08	0.53	0.21
VIS	-0.05	0.00	0.30	-0.05	0.41	0.24
CF	-0.02	0.04	**0.36**	-0.01	**0.38**	0.06
PR	0.08	0.03	0.02	0.03	0.02	0.62

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.[a]
a. Rotation converged in 7 iterations.

Exploratory Factor Analysis IV

Based on the results from the third EFA after removing *Numbers Reversed* and *Retrieval Fluency*, a fourth and final EFA was conducted. This EFA was performed after removing the subtest *Rapid Picture Naming* due to a multi-loading difference of < .100 across two factors as well as weak factor loading. Bartlett's Test of Sphericity

Table 13

Communalities for EFA IV (Numbers Reversed, Retrieval Fluency, and Rapid Picture Naming Removed)

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	0.74	0.79
Number Series (NS)	0.58	0.43
Verbal Attention (VA)	0.57	0.56
Letter-Pattern Matching (LPM)	0.61	0.59
Phonological Processing (PP)	0.67	0.74
Story Recall (STOR)	0.45	0.46
Visualization (VIS)	0.48	0.50
General Information (GI)	0.66	0.66
Concept Formation (CF)	0.52	0.49
Number-Pattern Matching (NPM)	0.56	0.59
Nonword Repetition (NWR)	0.46	0.45
Visual-Auditory Learning (VAL)	0.40	0.39
Picture Recognition (PR)	0.38	0.32
Analysis-Synthesis (AS)	0.50	0.50
Object-Number Sequencing (ONS)	0.57	0.55
Pair Cancellation (PC)	0.48	0.50
Memory for Words (MFW)	0.50	0.49
Picture Vocabulary (PV)	0.69	0.73
Oral Comprehension (OC)	0.69	0.63
Segmentation (SEG)	0.49	0.50
Sentence Repetition (SENR)	0.58	0.58
Understanding Directions (UD)	0.51	0.47
Sound Blending (SB)	0.47	0.46
Sound Awareness (SA)	0.53	0.48

Extraction Method: Principal Axis Factoring.

(Bartlett, 1950) identified that the correlation matrix was not random in nature (χ^2 = 22,081.86, df = 276, p < .000). The degree of common variance was determined through the Kaiser-Meyer-Olkin (KMO; Kaiser, 1974) measure of sampling adequacy and found to be an appropriate standard at .901. Therefore, based on the results from Bartlett's and KMO, the adjusted correlation matrix was determined to be appropriate for the fourth and final exploratory factor analysis.

Table 14

Total Variance Explained for EFA IV (Numbers Reversed, Retrieval Fluency, and Rapid Picture Naming Removed)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation of Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.53	39.71	39.71	9.08	37.85	37.85	6.62
2	1.63	6.80	46.51	1.24	5.17	43.02	6.66
3	1.46	6.08	52.59	0.98	4.10	47.12	7.06
4	1.33	5.53	58.12	0.85	3.56	50.68	5.81
5	1.18	4.93	63.05	0.70	2.92	53.60	5.52

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

This EFA also identified the number of factors based on eigenvalues greater than one (Kaiser, 1960) and visual scree (Cattell, 1966). There was a total of five factors retained based on the Kaiser criterion. Factor one accounted for 39.71% of the variance, factor two accounted for 6.80% of the variance, factor three accounted for 6.08% of the variance, factor four accounted for 5.53% of the variance, and factor five accounted for 4.93% of the variance. The total amount of variance explained by the five factors was 63.05%. Upon examination of the visual scree plot, the final EFA appeared to identify five factors. See Tables 13 and 14 and Figure 4.

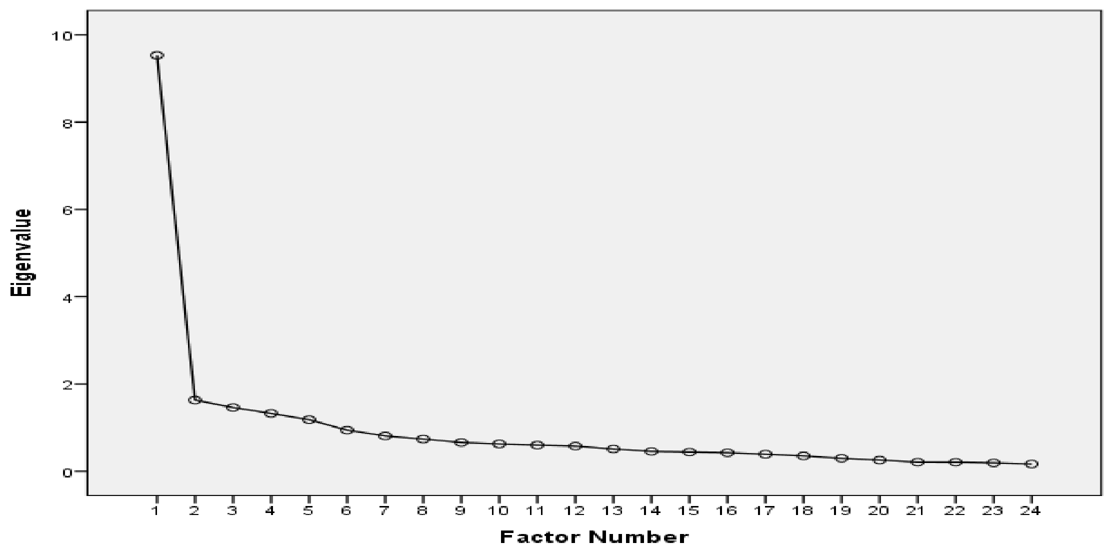

Figure 4. Visual Scree Plot for EFA IV (Numbers Reversed, Retrieval Fluency, and Rapid Picture Naming Removed)

The pattern matrix converged after six iterations and the subtests loaded onto a total of five factors with no multi- or cross-loadings (see Table 15). Four salient subtest loadings comprised the first factor: picture vocabulary (0.88), general information (0.88), oral vocabulary (0.80), and oral comprehension (0.67). The second factor indicated six strong subtest loadings: sentence repetition (0.79), verbal attention (0.68), nonword repetition (0.64), memory for words (0.60), understanding directions (0.43), and object-number sequencing (0.40). The third factor consisted of five subtests: sound blending (0.73), phonological processing (0.72), segmentation (0.69), visual-auditory learning

Table 15

Pattern Matrix for Final EFA (Numbers Reversed, Retrieval Fluency, and Rapid Picture Naming Removed)

	Factor				
Subtests	1	2	3	4	5
PV	0.88	0.01	-0.04	-0.06	0.04
GI	0.88	-0.13	0.01	-0.01	0.02
OV	0.80	0.03	0.08	0.08	-0.06
OC	0.67	0.21	-0.04	-0.10	0.09
SENR	0.14	0.79	-0.08	-0.08	-0.04
VA	0.10	0.68	-0.08	0.16	-0.08
NWR	-0.09	0.64	0.13	-0.13	0.11
MFW	-0.16	0.60	0.20	-0.05	0.13
UD	-0.01	0.43	0.01	0.03	0.33
ONS	-0.06	0.40	0.03	0.27	0.23
SB	-0.02	0.01	0.74	-0.09	0.00
PP	0.15	0.20	0.72	0.12	-0.36
SEG	-0.11	0.08	0.69	-0.01	0.06
VAL	0.05	-0.16	0.54	-0.12	0.31
SA	0.07	0.18	0.48	0.04	0.01
CF	0.09	-0.06	0.42	0.04	0.32
NPM	-0.03	0.01	-0.18	0.87	0.01
LPM	-0.07	-0.04	-0.01	0.80	0.07
PC	0.00	-0.05	0.08	0.75	-0.13
NS	0.20	-0.02	0.13	0.34	0.13
PR	-0.05	0.12	-0.09	-0.09	0.61
STOR	0.19	0.13	-0.14	0.03	0.55
VIS	0.01	-0.05	0.28	-0.01	0.54
AS	0.02	-0.12	0.18	0.20	0.50

Extraction Method: Principal Axis Factoring.
Rotation Method: Promax with Kaiser Normalization.[a]
a. Rotation converged in 6 iterations.

(0.54), sound awareness (0.48) and concept formation (0.42). Four salient subtest loadings were indicated in the fourth factor: number-pattern matching (0.87), letter-pattern matching (0.80), pair cancellation (0.75), and number series (0.35). The fifth and final factor identified four subtests: picture recognition (0.61), story recall (0.55), visualization (0.54), and analysis-synthesis (0.50).

Chapter Summary

This chapter discussed the results from separate iterated principal factor analyses conducted with promax rotation. Four factor analyses were conducted due to multi- or cross-loadings of subtests onto multiple factors. Multi-loading was determined based on a difference in the subtests between the factors equaling less than 0.10. Subtests were systematically removed from subsequent analyses based on their cross-loading status as well as weak factor loading.

The first EFA was ran on all twenty-seven subtests. The results indicated the proposed model accounted for 64.21% of the variance and identified six separate factors. There were four identified subtests that loaded onto more than one factor. The second EFA was conducted with the removal of *Numbers Reversed* being identified as a multi-loader and its weak factor loading. This model accounted for 65.06% of the variance and also identified six factors. There were three subtests in the second EFA that loaded onto multiple factors. The third EFA was ran after removing *Numbers Reversed* and *Retrieval Fluency* due to loading on multiple factors and weak factor loadings. This model accounted for 65.94% of the variance and identified six factors with two multi-loaders. The final EFA was performed upon the removal of *Numbers Reversed*, *Retrieval Fluency*, and *Rapid Picture Naming* resulting from cross-loading and weak factor loading. The final model accounted for 63.05% of the variance. This model identified a total of five factors and contained zero cross-loadings.

CHAPTER V

DISCUSSION

The most recent iteration in the Woodcock-Johnson family, the WJ IV, was developed based on contemporary understanding of Cattell-Horn-Carroll (CHC) theory and current neuropsychological research (McGrew et al., 2014). The WJ IV series provides three different assessments for administration: the WJ IV COG, WJ IV OL, and WJ IV ACH. McGrew and colleagues (2014) identified seven broad CHC factors present when the WJ IV COG and WJ IV OL are administered together. These seven factors were generalized to the WJ IV COG and WJ IV OL from the results of the analysis of the entire WJ battery. To be more specific, the factor structure of the WJ IV COG and WJ IV OL has yet to be evaluated. The purpose of the present study was to examine the factor structure of the WJ IV COG and WJ IV OL for the 14- to 19- year old age range.

Through the utilization of exploratory factor analysis (EFA) using the correlation matrix provided in the technical manual (McGrew et al., 2014), the results from the current study found data to support the presence of five broad CHC factors for the WJ IV COG and WJ IV OL. This chapter includes a discussion regarding the current findings, implications, limitations, and future directions.

Explanation of Findings

As discussed in previous chapters, the test publishers identified there was variation found regarding the factor structure of the WJ IV during the initial analyses of the battery. Furthermore, the norming procedures for the WJ IV only examined the factor structure of the entire test (i.e., cognitive, academic, and oral language) rather than each test independently. It is important to note that the information provided in the technical manual for the initial factor structure of the WJ IV only presents information concerning the 9- to 13- year old age range. The procedures utilized by the test publishers during the study of the factor structure included a variety of analyses: exploratory cluster, exploratory factor, and confirmatory factor (McGrew et al., 2014).

First, the results from the cluster analysis identified nine CHC factors: G*wm*, G*a*, G*v*, G*f*, G*lr*, G*q*, G*c*, G*rw*, and G*s*. Then the exploratory principal components analysis found an eight, nine, and ten factor structure. There were only five broad CHC factors consistently across the three different models: G*c*, G*s*, G*rw*, G*q*+G*f*, and G*wm*. The exploratory multidimensional scaling analysis discovered seven factors: G*a*, G*c*, G*rw*, G*q*/G*f*-RQ, G*wm*, G*f*, and G*v*. The test developers then utilized the results from all analyses to identify three models for confirmatory factor analysis. The first model identified that all CHC factors loaded onto a single intelligence factor. The second model posited nine CHC factors in addition to general intelligence: G*c*, G*rw*, G*f*, G*s*, G*q*, G*v*, G*lr*, G*wm*, and G*a*. The final model included a general intelligence factor as well as thirteen broad factors: G*c*, G*rw*, G*f*-Vbl, G*f*, G*f*-RQ, G*s*, G*v*, G*q*, G*lr*, MA, LA, G*wm*, and G*a*. Through model fitting procedures, the test publishers then conducted model fit comparisons across the different age groups in the standardization sample. The second model was identified to be “the preferred model per the parsimony principal” (McGrew et al., 2014, p. 170) to represent the factor structure of the entire WJ IV test battery. Figure 5 provides the visual representation McGrew and colleagues (2014) defined regarding the CHC factor structure for the WJ IV COG and WJ IV OL as well as the subtests that are subsumed under each factor.

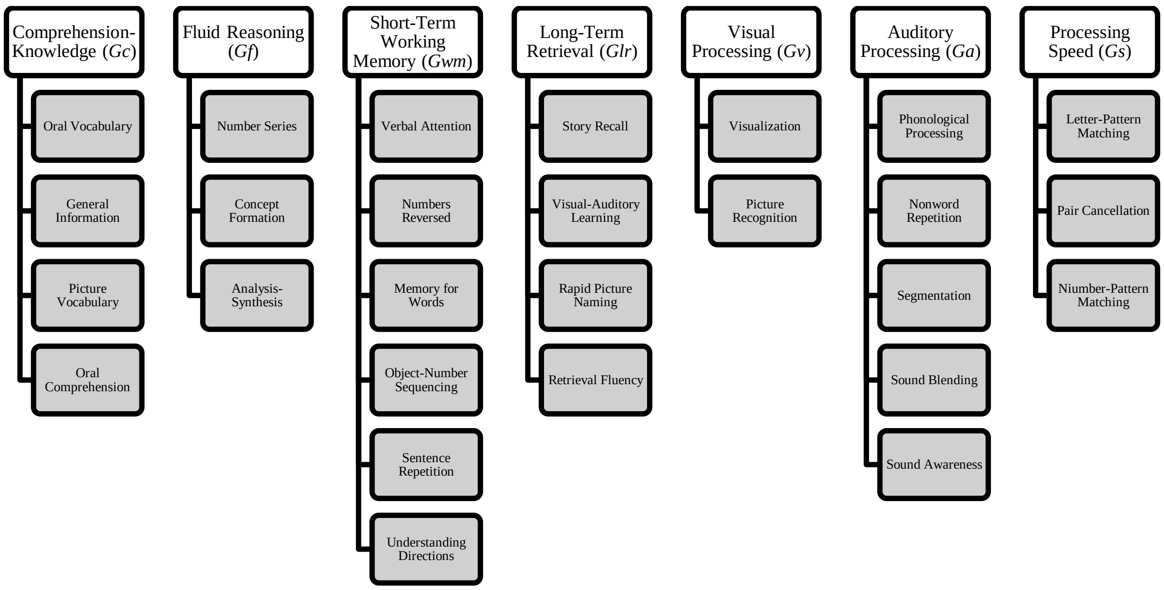

Figure 5. Test publishers' defined factor structure for the WJ IV COG and WJ IV OL and the subtests hypothesized to measure each CHC factor. Adapted from *Technical Manual: Woodcock-Johnson IV,* by K. S. McGrew, E. M. LaForte, and F. A. Schrank, 2014, Rolling Meadows, IL: Riverside.

Comprehension-knowledge (G*c*) involves the application of previously learned knowledge to solve practical problems and demonstrates the complexity and extent of one's acquired knowledge through personal experience. The subtests identified as representing G*c* include oral vocabulary, general information, picture vocabulary, and oral comprehension. Fluid reasoning (G*f*) is the ability to solve novel problems without reliance on previously acquired knowledge and includes three subtests: number series, concept formation, and analysis-synthesis. Short-term working memory (G*wm*) refers to one's ability to hold and manipulate information in temporary storage. There are six subtests posited to represent G*wm*: verbal attention, numbers reversed, memory for words, object-number sequencing, sentence repetition, and understanding directions. Long-term storage and retrieval (G*lr*) involves taking in and learning information, then retrieving this information from memory after a period of elapsed time. Story recall, visual-auditory learning, rapid picture naming, and retrieval fluency are incorporated under G*lr*. Visual-spatial processing (G*v*) is the ability to mentally perceive information to solve problems that are visual-spatial in nature and is assessed through two subtests, visualization and picture recognition. Auditory processing (G*a*) includes detecting and processing information through sounds. Five subtests support G*a*, phonological processing, nonword repetition, segmentation, sound blending, and sound awareness. Processing speed (G*s*) is the ability to perform simple tasks in a relatively brief period with speed and efficiency. Letter-pattern matching, pair cancellation, and number-pattern matching are the measures of G*s* for the WJ IV (Schneider & McGrew, 2012; McGrew et al., 2014).

Resulting from the newness of the WJ IV, there has been minimal independent analyses conducted regarding the factor structure of the battery. One study by Dombrowski et al. (2017) examined the factor structure of the full battery of the WJ IV on two school aged populations (9- to 13- years old and 14- to 19- years old).

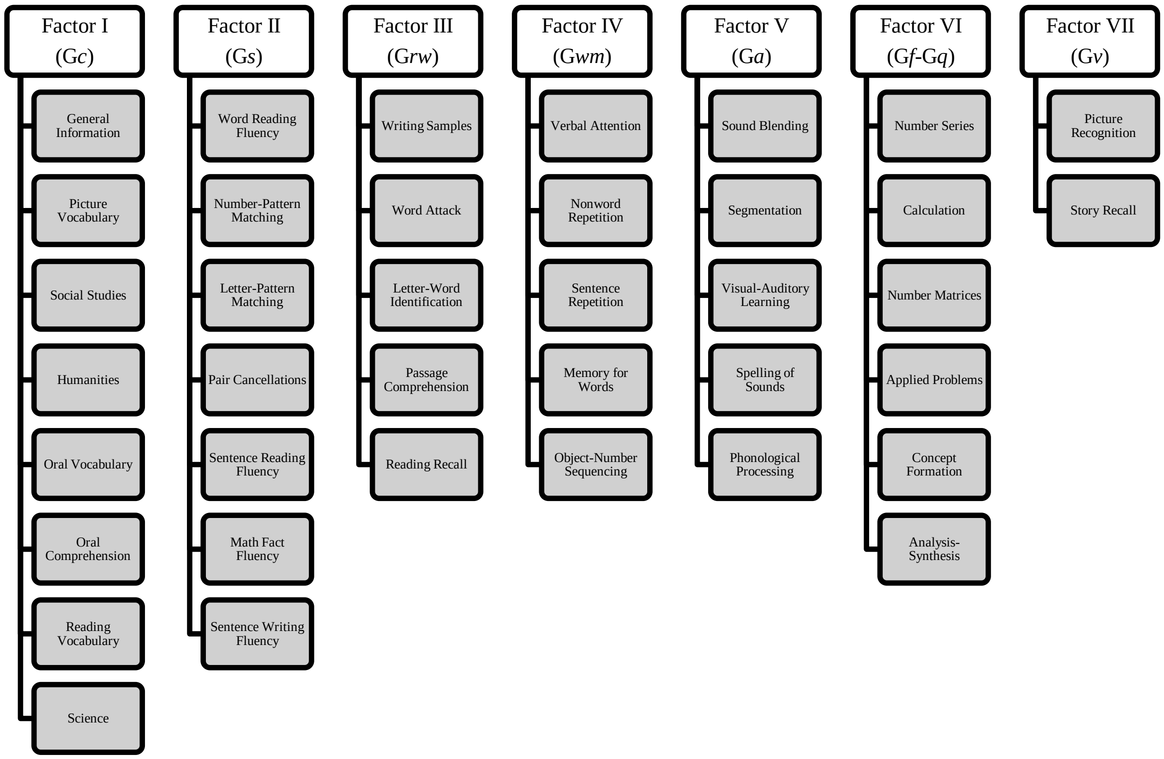

Figure 6. The seven CHC factors identified by Dombrowski and colleagues for the 14-to 19-year old age group for the WJ IV full battery. Adapted from "Hierarchical Exploratory Factor Analysis of the Woodcock-Johnson IV Full Test Battery: Implications for CHC Application in School Psychology," by S. C. Dombrowski, R. J. McGill, and G. I. Canivez, 2017, *School Psychology Quarterly,* advanced online publication.

This study utilized hierarchical EFA based on the correlation matrices provided in the technical manual. Dombrowski and colleagues (2017) found several instances of cross- or multi-loading of subtests between numerous factors as well as some factors not clearly being identified. Specific to the 14-to 19-year old standardization sample, the results of the analyses identified seven of the nine CHC factors when the WJ IV COG, ACH, and OL are administered together. As seen in Figure 6, the factors identified include G*c*, G*rw*, G*s*, G*wm*, G*a*, G*v*, and a combined G*q*-G*f* (Dombrowski et al., 2017). The end results were unable to clearly identify two of the broad CHC factors originally proposed by McGrew et al. (2014) for the WJ IV, G*f* and G*lr*.

There has only been a single independent study examining an individual test within the WJ IV series. Dombrowski et al. (2016) conducted an EFA and hierarchical factor analysis of the WJ IV COG using the correlation matrices in the technical manual for the 9- to 13- year old and 14- to 19- year old standardization samples.

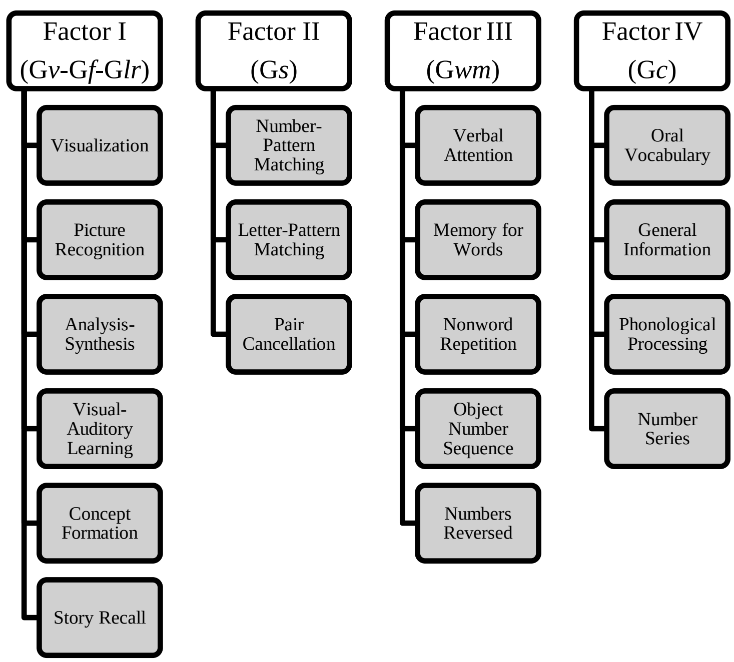

Figure 7. The four CHC factors identified by Dombrowski and colleagues for the 14- to 19- year old age group for the WJ IV COG. Adapted from “Exploratory and Hierarchical Factor Analysis of the WJ-IV Cognitive at School Age,” by S. C. Dombrowski, R. J. McGill, and G. L. Canivez, 2016, *Psychological Assessment*, 28, p. 403.

The EFA results indicated support for four factors in the school-aged population, a deviation from the publisher's proposed seven. As seen in Figure 7, Dombrowski and colleagues (2016) identified three clear factors, G*c*, G*wm*, and G*s*, as well as a single mixed factor containing G*f*-G*v*-G*lr*. They were unable to establish a separate G*f*, G*lr*, G*a*, or G*v* factor for the WJ IV COG.

The present study performed iterated principal axis factoring with promax rotation on the correlation matrix provided in the technical manual for the 14- to 19- year old standardization sample (McGrew et al., 2014). The EFA was run four separate times due to low loadings in addition to multi- or cross-loadings of subtests across more than one factor. The final EFA accounted for 63.05% of the total variance after removing the following subtests: numbers reversed, retrieval fluency, and rapid picture naming. While this model structure accounted for a large majority of the total explained variance (63.05%), the remaining percentage of total variance that was not explained by the factor structure is based on several sources. First, the five factor loadings that contributed to a majority of the explained variance demonstrated overall salient loadings; however, there were other potential factors that did not provide salient enough loadings to be included in the overall model. Additionally, one must consider that the total variance explained does not include potential model or sampling error found within the proposed factor structure (Costello & Osborne, 2005).

When comparing the factor structure of the WJ IV COG and WJ IV OL presented by the test authors (McGrew et al., 2014) and the results from this independent study, there are distinguishable differences with the most evident being the number of factors found. While the WJ IV publishers identified seven CHC factors (G*c*, G*f*, G*wm*, G*lr*, G*v*, G*a*, and G*s*), the current study only identified five CHC factors. Figure 8 identifies the subtests that load on each of the five factors for this study.

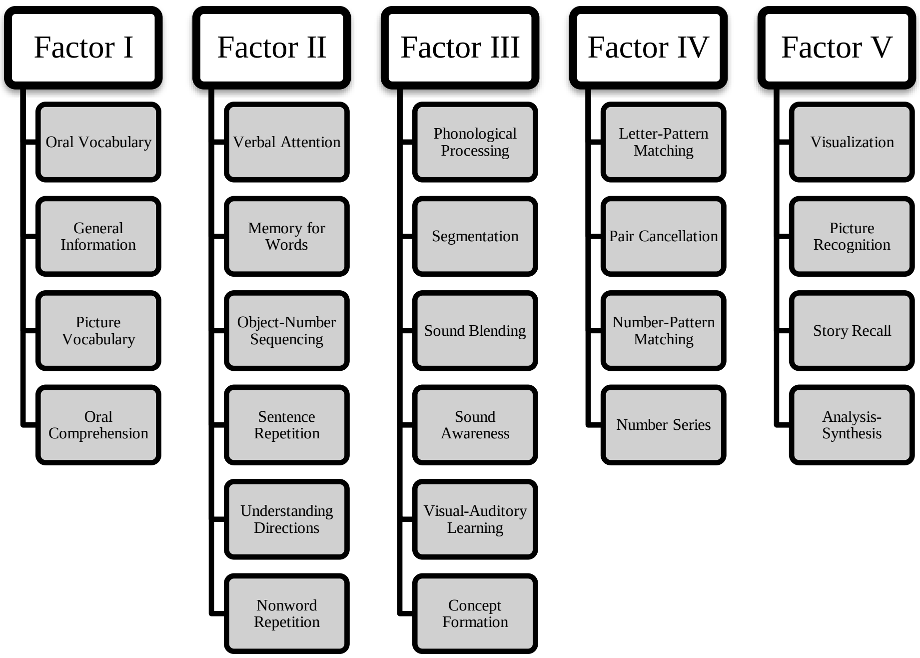

Figure 8. The subtests listed under each factor found by current study.

After determining the number of factors present on the WJ IV COG and WJ IV OL through exploratory analyses, the next step was to hypothesize the CHC ability represented by the five proposed factors. Accounting for the removal of the three subtests previously identified, six subtests appear to have loaded on several factors when compared to the test publishers' factor structure, while eighteen of the subtests maintained their positions under their respective factors. Table 16 presents the proposed

Table 16

Hypothesized CHC factors for the current study along with the subtests comprising each factor

Factor	Traditional Broad CHC Factors	Hypothesized Broad CHC Factor	Subtests *Narrow Ability*					
I	G*c*	G*c*	Oral Vocabulary	General Information	Picture Vocabulary	Oral Comprehension		
			Lexical knowledge *Language development*	*General verbal knowledge*	*Lexical knowledge* *Language development*	*Listening ability*		
II	G*wm*	G*wm*	Verbal Attention	Memory for Words	Object-Number Sequencing	Sentence Repetition	Understanding Directions	Nonword Repetition
			Working memory capacity *Attentional control*	*Memory span*	*Working memory capacity*	*Memory span* *Listening ability*	*Working memory capacity* *Listening ability*	*Phonetic coding* *Memory for sound patterns* *Memory span*
III	G*a*/G*lr*/G*f*	G*a*	Phonological Processing	Segmentation	Sound Blending	Sound Awareness	Visual-Auditory Learning	Concept Formation
			Phonetic coding *Word fluency* *Speed of lexical access*	*Phonetic coding*	*Phonetic coding*	*Phonetic coding*	*Associative memory*	*Induction*
IV	G*s*/G*f*	Attention	Letter-Pattern Matching	Pair Cancellation	Number-Pattern Matching	Number Series		
			Perceptual speed	*Perceptual speed* *Spatial scanning* *Attentional control*	*Perceptual speed*	*Quantitative reasoning* *Induction*		
V	G*v*/G*lr*/G*f*	Cognitive Reasoning	Visualization	Picture Recognition	Story Recall	Analysis-Synthesis		
			Visualization	*Visual memory*	*Meaningful memory* *Listening ability*	*General sequential reasoning*		

broad CHC factors for this study as well as the subtests incorporated within each of the factors.

The subtests identified under comprehension knowledge (G*c*; oral vocabulary, general information, picture vocabulary, and oral comprehension) emerged as the only factor that exactly aligned with the test publishers. The second proposed factor, short-term working memory (G*wm*), retained many of the same subtests identified by McGrew and colleagues (2014): verbal attention, memory for words, object-number sequencing, sentence repetition, and understanding directions. Numbers reversed was removed because of cross loading onto more than one factor and nonword repetition was relocated to G*wm*. Nonword repetition may potentially have loaded under G*wm* rather that G*a* due to the demand of maintaining unique auditorily presented stimuli within short-term memory to repeat the word. The next proposed factor is hypothesized to represent auditory processing (G*a*). While nonword repetition relocated to G*wm*, this factor preserved the same subtests as the technical manual: phonological processing, segmentation, sound blending, and sound awareness. Two additional subtests were incorporated under G*a*, visual-auditory learning and concept formation. The shifting of these subtests may have been based on the fact that both of these subtests require a certain amount of auditory processing and reasoning.

The fourth proposed factor was reconceptualized for the present study. The subtests that loaded on this factor resulting from the EFA represented processing speed (G*s*) and fluid reasoning (G*f*). According to the technical manual, the subtests representing G*s* included letter-pattern matching, pair cancellation, and number-pattern matching, while number series represented G*f*. When examining each of the individual subtests loading onto this factor, attention emerges as a shared connecting concept. Each of these subtests require the examinee to sustain or divide his or her attention to complete the presented tasks. Letter-pattern matching, pair cancellation, and number-pattern matching entail the examinee to locate items or patterns within a specific timeframe. To complete these tasks efficiently, one must demonstrate appropriate attentional control between the presented and targeted stimuli. Number series directs the examinee to identify a numerical sequence in the stimuli; however, it also requires a certain amount of attentional control. The examinee is required to shift attention between the visually presented information and the mental demands of accessing a number line in addition to the underlying principal guiding the number series.

The final factor was the most unclear of the factors identified by the EFA; essentially, producing a combination of G*v*-G*lr*-G*f*. This factor included visualization and picture recognition, both of which the publishers identified as measuring G*v*. However, story recall (G*lr*) and analysis-synthesis (G*f*) were also included under this fifth factor. Visualization utilizes mental rotations to solve problem, picture recognition requires the examinee to access visually stored representations, story recall relies on the cognitive visualization of an orally presented story, and analysis-synthesis utilizes mental symbolic formulations in order to solve complex tasks. Each of these subtests demonstrate a need for the examinee to complete tasks utilizing either a form of mental problem-solving or symbolic representation. In general, these four subtests require an increased demand for higher cognitive reasoning abilities.

The current results failed to clearly delineate the two broad factors of fluid reasoning (G*f*) and long-term storage and retrieval (G*lr*). The subtests the publishers identified as measures of G*f* and G*lr* either spread throughout numerous factors postulated by the current EFA or were eliminated due to cross loadings. McGrew and colleagues (2014) acknowledged this latest iteration of the WJ included more cognitive complexity within its measures; therefore, it may be hypothesized that when there are measures of greater cognitive complexity, more than one latent factor may be measured (Dombrowski et al., 2016, 2017); thus, explaining why this study found multiple cross loadings of various subtests.

The results of the current study and the analyses by Dombrowski et al. (2017) and Dombrowski et al., (2016) share some similarities regarding the factor structure of the WJ IV. All three of the independent analyses failed to confirm the factor structure as proposed by McGrew et al. (2014). Specifically, each study was unable to clearly identify G*f* and G*lr* as separate and distinct CHC factors on the testing instrument. Moreover, there were consistent findings of multi- or cross-loadings of various subtests across several factors. When the results from the three studies are examined, it may be suggested that the WJ IV is potentially overfactored.

It is important to note that Dombrowski et al. (2016, 2017) and the current study approached the examination of the factor structure of the WJ IV with different methodologies. The current study examined the factor structure of the WJ IV COG and WJ IV OL using multiple exploratory factor analyses while systematically removing subtests that cross loaded onto multiple factors. The two studies completed by Dombrowski and colleagues (2016, 2017) used principal axis factoring as well as Schmid-Leiman procedures to analyze the factor structure of the WJ IV COG and the full battery of the WJ IV. Essentially, the procedures outlined in these two investigations extracted the variance explained by the second-order factors by allowing first-order factors to be orthogonal to the second-order factors. Using the Schmid-Leiman procedure allows one to differentiate between the variance of higher and lower order factors (Dombrowski et al., 2016, 2017). The present study only employed EFA on the WJ IV COG and the WJ IV OL for the 14- to 19- year old standardization sample; therefore, the results from all three studies would not be expected to produce the same findings.

Implications

While the results from the present study do not align perfectly with the publishers proposed factor structure, there is some convergence for the factor structure of the WJ IV COG and WJ IV OL. The current findings, while obtained via different statistical analyses, share many similarities and overlap with the findings from the two studies conducted by Dombrowski and colleagues (2016, 2017). When examining the limited research on the WJ IV, it appears this latest iteration of the assessment tool does provide quantifiable measures of broad CHC factors; however, there is some deviation from the results obtained in the norming process provided in the Technical Manual (McGrew et al., 2014). Specifically, the three analyses that have been conducted on the WJ IV appear to be unable to categorize fluid reasoning (G*f*) and long-term storage and retrieval (G*lr*) as separate and distinct CHC factors. Rather, the subtests that load on these factors scatter to other hypothesized CHC factors. It stands to reason that the results from this study combined with two independent studies (Dombrowski et al., 2016, 2017) may indicate that the WJ IV COG and WJ IV OL are potentially overfactored.

When combined with the current literature on the WJ IV, the most salient implication of this study is to understand the WJ IV has inherent limitations. The cognitive complexity employed for the WJ IV may have unintentionally contributed to the muddying of the broad CHC factors. For practicing school psychologists, the WJ IV is an assessment tool that can easily be employed within a school-based setting; however, when one evaluates the implications associated with the limited analyses of the WJ IV, practitioners are cautioned to understand the complexities associated with test score interpretation. The WJ IV allows a glimpse into the cognitive abilities of individuals, yet the testing results may not necessarily transfer to true cognitive strengths and deficits. As demonstrated by the current study, there is not always a clear relationship between subtests and their proposed CHC factors. Schrank and colleagues (2016) identified three modes of test interpretation for the WJ IV: test-level performance, cluster-level performance, and general intellectual ability. Due to the findings of this study, as well as two previously published analyses (Dombrowski et al., 2016, 2017), it is advisable to not interpret the WJ IV solely on subtest performance. There are multiple subtests that cross load onto more than one factor and it is difficult to pinpoint the broad CHC factor represented by the subtests. Interpretation should occur at a more global level by focusing on the cluster scores and general intellectual ability (GIA) score to demonstrate a best-practice approach to test interpretation (Watkins et al., 2005; Watkins, 2009). Furthermore, it is important to understand that any form of assessment has limitations. This is imperative to acknowledge to ensure best practice in the field. Test development hinges on a theoretical understanding of the concepts it strives to measure. Developing quantifiable measures for latent constructs is not easily achieved. Theory is not perfect, neither is assessment.

Limitations and Considerations

As with any study, there are concerns to be aware of when conducting complex data analyses. While an EFA does not possess a preconceived hypothesis regarding the latent variables present on an assessment tool, there are several limitations one must be aware of to ensure competent practice.

Traditionally, EFA is considered a complicated, multivariate mathematical approach to understanding the underlying factor structure of an instrument containing a large set of variables (Fabrigar & Wegener, 2012). For example, there are numerous sequential steps integrated within an EFA and significant responsibility is placed on the judgements of the researcher conducting the EFA. There are a variety of options available regarding the type of EFA used and each option may lead to an array of different results (Preacher & MacCallum, 2003). Furthermore, interpretation of the results obtained from the EFA heavily relies upon the conceptualization and acumen of the researcher. Therefore, strong caution must always be utilized when conducting and interpreting results from an EFA (Henson & Roberts, 2006).

Other limitations intrinsic to the results from this study center around "the soundness of the design of the study from which the data [were] collected" (Fabrigar et al., 1999, p. 273). The present study utilized the correlation matrix provided in the Technical Manual for the WJ IV COG and WJ IV OL (McGrew et al., 2014). As a result of this, the data was not collected or managed by the present researcher; therefore, there is no way to control for potential errors or inaccuracies that may be present in the Technical Manual. Moreover, the data that was utilized in developing the correlation matrices relied on imputed data (McGrew et al., 2014). The current researcher was unable to demonstrate any control over the method of data imputation. Testing participants were not administered the full battery of the WJ IV. Rather, the results obtained in the testing manual were based upon imputed data using a partial matrix sampling plan. While this is an acceptable approach to imputing data, the concern remains regarding the fact that results presented in the technical manual are derived from estimated scores or ability as opposed to actual testing subjects.

Additionally, there is widespread support indicating a common factor must be represented by multiple manifest variables. However, because the EFA is conducted from the information provided by the test publishers, there is no way to objectively determine if this occurred for the WJ IV battery. McGrew et al. (2014), specify each of the factors underlying the test are supported by a minimum of two measured variables, yet best practice indicates that a minimum of three to five variables should be used to evaluate the underlying factor structure (Fabrigar et al., 1999; Henson & Roberts, 2006; Preacher & MacCallum, 2003; Watkins, 2009).

Furthermore, the results of the EFA are primarily based on the conceptualization of the factors by the researcher and the potential for confirmation bias is present. This interpretive portion of the analyses heavily relies on the modern theories of intelligence available in the literature. For the current study, CHC theory was utilized during the identification of the factor structure that results from the EFA. However, there are numerous theories hypothesized in the field and each conceptualizes the components of intelligence differently (Frazier & Youngstrom, 2007).

With regard to the data analyses, there are identified concerns when working with a visual scree test that must be acknowledged. The visual scree is somewhat subjective based on the researcher's interpretation (Fabrigar & Wegener, 2012). While there is usually clear distinction in strong common factors, sometimes the interpretation of delineating large eigenvalues from small eigenvalues can be challenging. The data may not provide a clear breaking point in the eigenvalues or may require multiple lines to be drawn through the smaller eigenvalues (Frazier & Youngstrom, 2007). While the visual scree is a common evaluation procedure for factor extraction, it should be used with additional analyses to ensure best practice.

Future Directions

While this study was unable to fully replicate the findings from the test publishers, it found support for five CHC factors within the testing battery using an EFA. Moving forward, it will be important to follow-up this study with a confirmatory factor analysis (CFA). When examining the factor structure of a testing instrument, a CFA is employed when there is an *a priori* hypothesis regarding the specific factor structure of an instrument (Fabrigar & Wegener, 2012). According to Carroll (1993) an EFA and CFA are viewed as complementary procedures to one another. Employing a CFA in future studies would allow one to determine if the proposed factor structure of this study is viable.

Furthermore, this study provided an independent analysis regarding the factor structure of the WJ IV COG and WJ IV OL. Unfortunately, this study relied on the data obtained from the norming procedures for the WJ IV. As a result, there was limited standardized testing completed on clinical populations (e.g., specific learning disability, traumatic brain injury, attention deficit hyperactivity disorder, etc.) as well as racially diverse populations. Considering that most clinicians primarily work with clinical populations (i.e., schools and private practice), it will be important to collect data on the factor structure of the WJ IV when utilized on these populations. Additionally, the racial and ethnic make-up of the United States is continuously changing; therefore, it is imperative to ensure that testing instruments are culturally sensitive to a variety of populations.

Finally, this study focused on a small population of the norming sample (14- to 19- year-old individuals). Moving forward, it will be important for future studies to be conducted to examine the factor structure of the WJ IV in all the age populations for the entire standardization sample. Additionally, further analyses should target the different testing instruments within the entire testing battery (i.e., cognitive, academic, and oral language) separately and together.

Chapter Summary

In summation, the findings from this study support a five-factor model for the WJ IV COG and WJ IV OL. When incorporating the CHC model of intelligence as well as the test publishers research, it is hypothesized that the WJ IV COG and WJ IV OL for the 14- to 19- year old age range identifies the following CHC factors: comprehension-knowledge (G*c*), short-term working memory (G*wm*), auditory processing (G*a*), processing speed (G*s*), and visual processing (G*v*). Based on a deficiency of strong factor loading, fluid reasoning (G*f*) and long-term storage and retrieval (G*lr*) were removed from the hypothesized factor structure. These findings, while not in complete agreement, demonstrate comparable results when compared to two other independent evaluations of the WJ IV (Dombrowski et al., 2016, 2017). The importance of these findings lies in the fact that clinicians must understand the limitations of testing instruments. Cognitive and academic testing is foundationally grounded in theory; thus, presenting the difficulty of measuring latent constructs with quantifiable variables. As with all studies, there are inherent limitations. Specifically, an EFA requires the researcher to perform extensive decision-making steps which can be easily influenced. Also, the data provided by the WJ publishers relied heavily on imputed data, potentially confounding any results obtained. Moving forward it is recommended to perform a CFA on the results from this study to determine if the factor structure is further supported through additional analyses. Also, it is recommended to expand the research of the WJ IV to include a greater number of clinical populations as well as culturally diverse individuals. There is no single testing instrument that is without fault; however, the WJ IV appears to measure several CHC factors and is a strong instrument that could be utilized in the global assessment of an individual.

REFERENCES

Anastasi, A. (1988). *Psychological testing* (6th ed.). New York: Macmillan.

Anderson, E. B. (1973). A goodness of fit test for the Rasch model. *Psychometrika, 38,* 123-140. doi: 10.1007/BF02291180

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology, 3,* 77-85. doi: 10.1111/j.2044-8317.1950.tb00285.x

Benjamin, L. T., Jr. (2009). *A history of psychology: Original sources and contemporary research* (3rd ed.). Malden, MA: Blackwell Publishing.

Bickley, P. G., Keith, T. Z., & Wolfe, L. M. (1995). The three-stratum theory of cognitive abilities: Tests of the structure of intelligence across the life span. *Intelligence, 20,* 309-328.

Boake, C. (2002). From the Binet-Simon to the Wechsler-Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology, 24,* 383-405. doi: 10.1079/jcen.24.3.383.981

Brown, R. E. (2016). Hebb and Cattell: The genesis of the theory of fluid and crystallized intelligence. *Frontiers in Human Neuroscience, 10*(606), 1-11. doi: 10.3389/fnhum.2016.00606

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies.* New York: Cambridge University Press.

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

Carrow-Woolfolk, E. (1995). *Oral and Written Language Scales: Listening Comprehension/Oral Expression*. Torrance, CA: Western Psychological Services.

Carrow-Woolfolk, E. (1999). *Comprehensive Assessment of Spoken Language*. Torrance, CA: Western Psychological Services.

Cattell, J. M. (1893). Tests of the senses and faculties. *Educational Review, 5,* 257-268.

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

Cattell, R. B. (1944). Psychological measurement: Normative, ipsative, and interactive. *Psychological Review, 51,* 292-303. doi: 10.1037/h0057299

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

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1,* 245-276. doi: 10.1207/s15327906mbr0102_10

Cattell, R. B. (1998). What is intelligence? Some answers from the triadic theory. In J. J. McArdle & R. W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 37-47). Mahwah, NJ: Psychology Press.

Cattell, R. B., & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational*

Measurement, 15, 139-164. Retrieved from http://ezproxy.twu.edu:2069/stable/1433661

Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology, 23,* 285-299. doi: 10.1037/h0043898

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research & Evaluation, 10*(7), 1-9. Retrieved from http://pareonline.net/getvn.asp?v=10&n=7

Courtney, M. G. R. (2013). Determining the number of factors to retain in EFA: Using the APSS R-Menu v2.0 to make more judicious estimations. *Practical Assessment, Research & Evaluation, 18*(8), 1-14. Retrieved from http://pareonline.net/getvn.asp?v=18&n=8

Cronbach, L. J. (1947). Test "reliability": Its meaning and determination. *Psychometrika, 12*(1), 1-16. doi: 10.1007/BF02289289

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

de Winter, J. C. F., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39,* 695-710. doi: 10.1080/02664763.2011.610445

Ding, Y., & Alfonso, V. C. (2016). Overview of the Woodcock-Johnson IV: Organization, content, and psychometric properties. In D. P. Flanagan & V. C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 1-30). San Diego, CA: Academic Press.

Dombrowski, S. C. (2013). Investigating the structure of the WJ-III cognitive at school age. *School Psychology Quarterly, 28,* 154-169. doi: 10.1037/spq0000010

Dombrowski, S. C. (2015). Exploratory bifactor analysis of the WJ-III achievement at school age via the Schmid-Leiman orthogonalization procedure. *Canadian Journal of School Psychology, 30,* 34-50. doi: 10.1177/0829573514560529

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2016). Exploratory and hierarchical factor analysis of the WJ-IV cognitive at school age. *Psychological Assessment, 28,* 394-407. doi: 10.1037/pas0000350

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Hierarchical exploratory factor analyses of the Woodcock-Johnson IV full test battery: Implications for CHC application in school psychology. *School Psychology Quarterly.* Advanced online publication. http://dx.doi.org/10.1037/spq0000221

Dombrowski, S. C., & Watkins, M. W. (2013). Exploratory and higher order factor analysis of the WJ-III full test battery: A school-aged analysis. *Psychological Assessment, 25,* 442-455. doi: 10.1037/a0031335

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test* (4th ed.). San Antonio, TX: Pearson.

Estabrook, G. (1983). Test review: Woodcock-Johnson Psycho-Educational Battery. *Journal of Psychoeducational Assessment, 1,* 315-319.

Elliot, C. D. (2007). *Differential Ability Scale* (2nd ed.). San Antonia, TX: Harcourt Assessment.

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis: Understanding statistics.* New York, NY: Oxford University Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological methods, 4,* 272-299. doi: 10.1037/1082-989X.4.3.272

Feldman, R. S. (2003). *Development across the life span* (3rd ed.). New Jersey, NJ: Prentice Hall.

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). Los Angeles, CA: Sage.

Flanagan, D. P., & Dixon, S. G. (2013). The Cattell-Horn-Carroll theory of cognitive abilities. In C. R. Reynolds, K. J. Vannest, & E. Fletcher-Janzen (Eds.), *Encyclopedia of Special Education* (Vol. 381, pp. 368-382). Hoboken, NJ: John Wiley & Sons.

Flanagan, D. P., & McGrew, K. S. (1997). A cross-battery approach to assessing and interpreting cognitive abilities: Narrowing the gap between practice and cognitive science. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 314-325. New York: NY. Guilford Press.

Floyd, R. G., McGrew, K. S., Barry, A., Rafael, F., & Rogers, J. (2009). General and specific effects on Cattell-Horn-Carroll broad ability composites: Analysis of the Woodcock-Johnson III Normative Update Cattell-Horn-Carroll factor clusters across development. *School Psychology Review, 38,* 249-265.

Frazier, T. W., & Youngstrom, E. A. (2007). Historical increase in the number of factors measured by commercial tests of cognitive ability: Are we overfactoring? *Intelligence, 35,* 169-182. doi: 10.1016/j.intell.2006.07.002

Fruchter, B. (1954). *Introduction to Factor Analysis.* Princeton, NY: D. Van Nostrand Company, Inc.

Goodenough, F. L. (1949). *Mental testing: Its history, principles, and applications.* New York: Rinehart.

Gorsuch, R. L. (1983). *Factor Analysis* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Guilford, J. P. (1967). *The nature of human intelligence.* New York; McGraw-Hill.

Hendrickson, A. E., & White, P. O. (1964). Promax: A quick method for rotation to oblique simple structure. *British Journal of Statistical Psychology, 17,* 65-70. doi: 10.1111/j.2044-8317.1964.tb00244.x

Henson, R. K. (2001). Understanding internal consistency reliability estimates: A conceptual primer on coefficient alpha. *Measurement and Evaluation in Counseling and Development, 34,* 177-189.

Henson, R. K., & Roberts, J. K. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*, 393-416. doi: 10.1177/0013164405282485

Hicks, P., & Bolen, L. M. (1996). Review of the Woodcock-Johnson Psycho-Educational Battery – Revised. *Journal of School Psychology, 34,* 93-102. doi: 10.1016/0022-4405(95)0027-5

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign. (AAT 6507113).

Horn, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review, 75,* 242-259. doi: 10.1037/h0025662

Horn, J. L., & Blankson, A. N. (2012). Foundations for better understanding of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 73-98). New York, NY: The Guilford Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology, 57*, 253-270. doi: 10.1037/h0023816

Individuals with Disabilities Education Improvement Act of 2004, 20 U.S.C. § 1400 (2004).

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141–151. doi.10.1177/001316446002000116

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*, 31–36. doi:10.1007/BF02291575

Kamphaus, R. W. (2001). *Clinical assessment of children's intelligence.* Needham Heights, MA: Allyn & Bacon.

Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2012). A history of intelligence testing interpretation. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 56-70). New York, NY: The Guilford Press.

Kaufman, A. S. (1979). *Intelligent testing with the WISC-R.* New York: Wiley-Interscience.

Kaufman, A. S. (1990). *Assessing adolescent and adult intelligence.* Needham Heights, MA: Allyn & Bacon.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children* (2nd ed.). San Antonio, TX: Pearson.

Keith, T. Z. (2015). *Multiple regression and beyond* (2nd ed.). New York, NY: Routledge.

Keith, T. Z., Caemmerer, J. M., & Reynolds, M. R. (2016). Comparison of methods for factor extraction for cognitive test-like data: Which overfactor, which underfactor. *Intelligence, 54,* 37-54. doi: 10.1016/j.intell.2015.11.003

Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47,* 635-650. doi: 10.1002/pits.20496

Levine, A. J., & Marks, L. (1928). *Testing intelligence and achievement*. New York: Macmillan.

McDermott, P. A., Watkins, M. A., & Rhoad, A. M. (2014). Whose IQ is it? – Assessor bias variance in high-stakes psychological assessment. *Psychological Assessment, 26*, 207-214. doi: 10.1037/a0034832

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.). *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-180). New York, NY: Guilford Press.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37,* 1-10. doi: 10.1016/j.intell.2008.08.004

McGrew, K. S., LaForte, E. M., & Schrank, F. A. (2014). *Technical Manual: Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.

McGrew, K. S., & Murphy, S. (1995). Uniqueness and general factor characteristics of the Woodcock-Johnson Tests of Cognitive Ability – Revised. *Journal of School Psychology, 33,* 235-245. doi: 10.1016/0022-4405(95)00011-A

McGrew, K. S., & Wendling, B. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47,* 651-675. doi: 10.1002/pits.20497

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *WJ-R Technical Manual*. Chicago: Riverside.

McGrew, K. S., & Woodcock, R. W. (2001). *Technical manual: Woodcock-Johnson III*. Itasca, IL: Riverside.

McGue, M., Shinn, M., & Ysseldyke, J. (1982). Use of cluster scores on the Woodcock-Johnson Psycho-Educational Battery with learning disabled students. *Learning Disability Quarterly, 5,* 274-287. doi: 10.2307/1510294

Meyers, L. S., Gamst, G., & Guarina, A. J. (2013). *Applied multivariate research: Design and interpretation* (2nd ed.). Los Angeles, CA: Sage.

Miller, D. C., & Maricle, D. E. (2012). The emergence of neuropsychological constructs into tests of intelligence and cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 800-819). New York, NY: The Guilford Press.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika, 8,* 161-168. doi: 10.1007/BF02288700

Murdoch, S. (2007). IQ: *A smart history of a failed idea.* Hoboken, NJ: John Wiley & Sons.

Newton, J. H., & McGrew K. S. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll-based assessment. *Psychology in the Schools, 47,* 621-634. doi: 10.1002/pits.20495

Pintner, R. (1923). *Intelligence testing.* New York: Holt, Rinehart, & Winston.

Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics, 2,* 13-43. doi: 10.1207/S15328031US0201_02

Rapaport, D., Gil, M., & Schafer, R. (1945-1946). *Diagnostic psychological testing* (2 vols.). Chicago, IL: Year Book Medical.

Reynolds, M. R., & Niileksela, C. R. (2015). Test review: Schrank, F. A., McGrew, K. S., & Mather, N. (2014). Woodcock-Johnson IV Tests of Cognitive Abilities. *Journal of Psychoeducational Assessment, 33,* 381-390. doi: 10.1177/0734282915571408

Roid, G. H. (2003). *Stanford Binet Intelligence Scales* (5th ed.). Austin, TX: PRO-ED.

Schneider, W. J. (2017, April 6). *Assessing psyche, engaging Gauss, seeking Sophia: Psychological assessment and psychometrics.* Retrieved from http://assessingpsyche.wordpress.com

Schneider, W. J., & McGrew, K. A. (2012). The Cattell-Horn-Carroll model of intelligence. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual* *assessment: Theories, tests, and issues* (3rd ed., pp. 99-144). New York, NY: The Guilford Press.

Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV Cognitive Abilities Assessment.* A. S. Kaufman & N. L. Kaufman (Eds.). Hoboken, NJ: Wiley.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014a). *Woodcock-Johnson IV Tests of Achievement*. Rolling Meadows, IL: Riverside.

Schrank, F. A., Mather, N., & McGrew, K. S. (2014b). *Woodcock-Johnson IV Tests of Oral Language*. Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014a). *Woodcock-Johnson IV*. Rolling Meadows, IL: Riverside.

Schrank, F. A., McGrew, K. S., & Mather, N. (2014b). *Woodcock-Johnson IV Tests of Cognitive Abilities*. Rolling Meadows, IL: Riverside.

Schrank, F. A., Miller, D. C., Wendling, B. J., & Woodcock, R. W. (2010). *Essentials of WJ III Cognitive Abilities Assessment* (4th ed.). Hoboken, NJ: John Wiley & Sons.

Schrank, F. A., & Wendling, B. J. (2012). The Woodcock-Johnson III Normative Update: Tests of cognitive abilities and tests of achievement. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 297-335). New York, NY: The Guilford Press.

Semel, W., Wiig, W. H., & Secord, W. A. (2003). *Clinical Evaluation of Language Fundamentals* (4th ed.). San Antonio, TX: Pearson.

Sokal, M. M. (2009). James McKeen Cattell and the failure of Anthropometric Mental Testing, 1890-1901. In L. T. Benjamin (Ed.), *A history of psychology,* (3rd ed., pp. 106-113). Malden, MA: Blackwell Publishing.

Spearman, C. (1904). “General Intelligence,” objectively determined and measured. *The American Journal of Psychology, 15,* 201-292. Retrieved from http://www.jstor.org/stable/1412107

Spearman, C. (1923). *The nature of ‘intelligence’ and the principles of cognition.* London: Macmillan.

Spencer, H. (1855). *The principles of psychology.* London: Longman, Brown, Green & Longmans.

Strickland, T., Watkins, M. W., & Caterino, L. C. (2015). Structure of the Woodcock-Johnson III cognitive tests in a referral sample of elementary school students. *Psychological Assessment, 27,* 689-697. doi: 10.1037/pas0000052

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities. *School Psychology Quarterly, 19,* 72-87. doi: 10.1521/scpq.19.1.72.29409

Taub, G. E., & McGrew, K. S. (2014). The Woodcock-Johnson Tests of Cognitive Abilities III’s cognitive performance model: Empirical support for intermediate factors within CHC theory. *Journal of Psychoeducational Assessment, 32*, 187-201. doi: 10.1177/0734282913504808

Terman, L. M. (1916). *The measurement of intelligence: An explanation and a complete guide for the use of the Stanford revision and extensions of the Binet-Simon Intelligence Scale.* Boston, MA: Houghton Mifflin.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence: A guide to the administration of the new Revised Stanford-Binet Tests of Intelligence*. Boston, MA: Houghton Mifflin.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208. doi: 10.1177/0013164496056002001

Thurstone, L. L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Thurstone, L. L. (1947). *Multiple-factor analysis: A development and expansion of the vectors of the mind.* Chicago, IL: The University of Chicago Press.

Thurstone, L. L. (1948). Psychological implications of factor analysis. *American Psychologist, 3*, 402-408. doi: 10.1037/h0058069

Vernon, P. E. (1950). *The structure of human abilities*. London: Methuen.

Vernon, P. E. (1969). *Intelligence and cultural environment*. London: Methuen.

Walsh, J., Kemerer, F., & Maniotis, L. (2010). *The educator's guide to Texas school law* (7th ed.). Austin, TX: University of Texas Press.

Wasserman, J. D. (2012). A history of intelligence assessment: The unfinished tapestry. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 3-55). New York, NY: The Guilford Press.

Watkins, M. W. (2009). Errors in diagnostic decision making and clinical judgment. In T. B. Gutkin & C. R. Reynolds (Eds.), *Handbook of School Psychology,* (4th ed., pp. 210-229). Hoboken, NY: Wiley.

Watkins, M. W., Glutting, J. J., & Youngstrom, E. A. (2005). Issues in subtest profile analysis. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues,* (2nd ed., pp. 251-268). New York: Guilford.

Watson, S. M. R., & Gable, R. (2013). Cognitive development of adolescents at risk or with learning and/or emotional problems: Implications for teachers. *Intervention in School and Clinic, 49,* 108-112. doi: 10.1177/1053451213493171

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

Wechsler, D. (1944). *The measurement of adult intelligence* (3rd ed.). Baltimore, MD: Williams & Wilkins.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children.* New York: Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale - Restandardized.* New York: Psychological Corporation.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Baltimore, MD: Williams & Wilkins.

Wechsler, D. (1967). *Wechsler Preschool and Primary Scale of Intelligence.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (1989). *Wechsler Preschool and Primary Scale of Intelligence — Revised.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence* (3rd ed.). San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). San Antonio, TX: Pearson.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). San Antonio, TX: NCS Pearson.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). San Antonio, TX: NCS Pearson.

Wolf, T. (1961). An individual who made a difference. *American Psychologist, 16,* 245-248. doi: 10.1037/h0044555

Wolf, T. (1964). Alfred Binet: A time of crisis. *American Psychologist, 19*, 762-771. doi: 10.1037/h0046878

Woodcock, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8,* 231-258.

Woodcock, R. W. (2002). New looks in the assessment of cognitive ability. *Peabody Journal of Education, 77,* 66-22.

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho-Educational Battery.* Rolling Meadows, IL: Riverside.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery – Revised.* Rolling Meadows, IL: Riverside.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001, 2007). *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside.

Woodcock, R. W., McGrew, K. S., Mather, N., & Schrank, F. A. (2003, 2007). *The Diagnostic Supplement to the Woodcock-Johnson III Tests of Cognitive Abilities.* Rolling Meadows, IL: Riverside.

Woodcock, R. W., Miller, D. C., Maricle, D. E., & McGill, R. J. (2017). *Evidenced-based selective assessment for academic disorders*. Middletown, MD: School Neuropsych Press.