AN EXPLORATORY FACTOR ANALYSIS OF THE WOODCOCK-JOHNSON IV TESTS OF COGNITIVE ABILITIES AND TESTS OF ORAL LANGUAGE FOR THE 9 TO 13-YEAR-OLD AGE RANGE

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY IN

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

DEPARTMENT OF PSYCHOLOGY AND PHILOSOPHY

COLLEGE OF ARTS AND SCIENCES

BY

ASHLEY S. ROWDEN, M.S.ED.

DENTON, TEXAS

DECEMBER 2020

DEDICATION

This dissertation is dedicated to Mom, and all of the “True Rowden’s” who came before us and taught us the value of education and perseverance; to Dad, for knowing exactly when to provide laughter and comfort food; to Dean, for encouraging me to explore new things and enduring my singing along the way; to Christiana, for being the sister I always wanted.

Finally, this dissertation is dedicated to Chris, for loving me and always reminding me to love myself.

ACKNOWLEDGEMENTS

So many people have encouraged and supported me throughout my time in this program, and particularly throughout the dissertation process.

First, I would like to thank my committee members. Dr. Denise Maricle, I would not have gotten to this point without your continued patience and support. Thank you for believing I could succeed and carrying me to the finish line. To Dr. Dan Miller and Dr. Wendi Johnson, thank you for reminding me of my capabilities and holding me to a higher standard.

I have received tremendous support from my coworkers and supervisors at the Dallas County Juvenile Department. I would especially like to thank Shelley Sullivan for always being willing to listen, advocate on my behalf, and offer stellar advice. I have experienced so much personal and professional growth over the past four years because of her leadership and guidance. I would also like to thank Dr. Leilani Hinton for believing in my potential and pushing me forward.

I would also like to acknowledge my Sarah Reed family: Dr. Eric Schwartz, Liza Pask, Brittany Hayden, and Maryanne DeSarro. I am so grateful to have had this experience and make lifelong friends along the way.

Finally, I would like to thank my family and friends for their persistent encouragement, as well as everyone who works hard to remind Black girls of their worth.

ABSTRACT

ASHLEY S. ROWDEN

AN EXPLORATORY FACTOR ANALYSIS OF THE WOODCOCK-JOHNSON IV TESTS OF COGNITIVE ABILITIES AND TESTS OF ORAL LANGUAGE FOR THE 9 TO 13-YEAR-OLD AGE RANGE

DECEMBER 2020

Scientific understanding of the nature of intelligence has steadily evolved over the years; however, the past century has seen an explosion of research aimed at understanding the construction of intelligence, the relationships between neuroscience and cognitive skills, and the best ways to measure intellectual abilities. While there is no clear consensus regarding the most accurate or all-encompassing theory of intelligence, Cattell-Horn-Carroll (CHC) theory has become one of the premiere guides to understanding the many facets of intelligence. The Woodcock-Johnson (WJ) series of tests have steadily incorporated CHC theory, aiming to provide practitioners with tangible measures of various cognitive skills. Two batteries from the most recent iteration of the WJ, the Woodcock-Johnson IV (WJ IV), are the WJ IV Tests of Cognitive Abilities (WJ IV COG) and the WJ IV Tests of Oral Language (WJ IV OL); WJ IV publishers purport that these two batteries provide a measure of general intelligence (*g*) as well as seven broad intelligence factors. However, research methods reportedly used to ensure adherence to CHC theory were both unorthodox and unclear. The purpose of this study is to use a commonly employed method of observing test structure — exploratory

factor analysis — to understand the factor structure of the WJ IV COG and WJ IV OL for the 9 to 13-year-old age group. A correlation matrix provided in the WJ IV *Technical Manual* was used for data analyses. Four subtests were removed from analyses due to weak or cross-loadings, thus the final solution was comprised of 23 subtests. Results indicated that the WJ IV COG and WJ IV OL are primarily measures of a single strong factor which coincides with comprehension-knowledge (G*c*). Four additional weaker but salient factors were also present and hypothesized to represent short-term working memory (G*wm*), perceptual reasoning, processing speed (G*s*), and auditory processing (G*a*). Perceptual reasoning was the only factor which did not clearly align with the factor structure reported in the *Technical Manual*, as it appeared to represent a blend of fluid reasoning (G*f*), long-term storage and retrieval (G*lr*), and visual processing (G*v*). These results closely mirrored the findings of other researchers examining the structural validity of the WJ IV.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The use of cognitive assessment measures to understand individual strengths and differences has become more commonplace over the past century. This is especially evident among school-aged children, as research has repeatedly informed of the benefits and importance of educating children within their zone of proximal development (Vygotsky, 1978). Currently, there are several theories of intelligence and a multitude of measures designed to ascertain cognitive abilities. One of the most commonly used tools over the past 40 years has been the Woodcock-Johnson (WJ; McGrew, LaForte, & Schrank, 2014). As the scientific understanding of intelligence has evolved, so have the WJ batteries. The most recent iteration of the WJ is the Woodcock-Johnson IV (WJ IV; Schrank, McGrew, & Mather, 2014). Given its ubiquitous use to determine student strengths and needs, it is crucial that the instrument is valid and provides an accurate assessment of the abilities it purports to measure. Failure to do so can result in incorrect diagnoses and academic recommendations, which can further hinder the very students evaluators are seeking to help.

This chapter of the study includes an abbreviated review of literature related to the development of cognitive assessment, the WJ, and theories of intelligence — particularly centering on the rise of the Cattell-Horn-Carroll theory (CHC; McGrew, 1997). In addition, this chapter briefly details the methodology and the rationale behind using exploratory factor analysis (EFA). It will conclude with a discussion of the purpose of the study and the research questions it aims to answer.

History of Intelligence Testing

Given that the current study seeks to explore the underlying structure of a measure of intelligence and oral language, understanding the history of intellectual assessment is imperative. French physician Edouard Sèguin developed one of the initial cognitive assessment techniques in the 1800s by using form boards with children believed to be intellectually delayed (Sèguin, 1856). The fact that variations of this measure are still in use (despite ever-evolving theories and measures of cognitive abilities) exemplifies the notion that common ideas about the definition and assessment of cognition have persisted for over a century.

In 1904, the work of Alfred Binet spurred modern-day intelligence testing (Urbina, 2004). Binet and Theodore Simon developed the Binet-Simon scale which was comprised of 30 tests aimed at measuring reasoning skills and judgment. This also marked the development of the intelligence quotient (IQ), which became especially common after the publication of the Stanford-Binet Intelligence Scale (Terman, 1916). Additionally, work by Lewis Terman contributed to the practice of comparing performance with other individuals of the same age. Development of the Stanford-Binet further contributed to the use of intelligence testing to ascertain suitability for the United States military. The Army Alpha test sought to assess verbal skills, reasoning abilities, judgment, and general knowledge. This was succeeded by the development of the Army

Beta test — a series of non-verbal tests designed for non-English speakers and those with limited literacy skills.

Ubiquitous use of the Army Alpha and Beta tests generated additional interest in intelligence testing for industrial and academic purposes (Boake, 2002). The Wechsler-Bellevue scale was one measure modeled on the Army tests, as its creator, David Wechsler, initially worked as a scorer of Alpha tests. He attributed his experiences doing this with the decision to merge verbal and non-verbal tasks into one scale. Wechsler also proposed a new strategy to determine IQ using standard scores and obtaining mean IQ scores and standard deviations at each age level. The resulting measure was the Wechsler-Bellevue Intelligence Scale (WBIS; Wechsler, 1939). Professionals in the field especially appreciated the measure's formatting and the availability of diagnostic profiles (Boake, 2002). Subsequent measures included the Wechsler Mental Ability Scale (Wechsler, 1946) followed by the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949) and the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955). These measures formed the foundation of intelligence testing and paved the way for newer scales, including the Kaufman Assessment Battery for Children (KABC; Kaufman & Kaufman, 1983), the Das-Naglieri Cognitive Assessment System (CAS; Naglieri & Das, 1997), the Reynolds Intellectual Assessment Scale (RIAS; Reynolds & Kamphaus, 2003), and the Woodcock-Johnson Psycho-Educational Battery (WJBEP; Woodcock & Johnson, 1977). This also includes revisions of the WJBEP, including the WJ IV (Schrank, McGrew, & Mather, 2014), which is the subject of the current study.

History of Factor Analysis and CHC Theory

The development and expansion of intelligence measures occurred alongside emergent theories of intelligence. In fact, Charles Spearman began studying intelligence prior to the appearance of cognitive assessment measures. His theory of general intelligence (1904) suggested that there is a factor underlying all abilities — *g*. Spearman also posited the existence of skills specific to each subtest — *s*. Spearman initially opposed the idea of group factors until work by Truman Kelley, a 20th century psychometrician, used partial correlational methods to provide evidence of group factors. Nowadays, these factors are typically referred to as broad ability factors, and they symbolize the shared variance among clusters of subtests. Louis Thurstone was another pioneer of factor analytic research. In 1934, Thurstone gave 56 tests to a group of college students and used factor analysis to obtain 13 ability factors (e.g., verbal comprehension, perceptual speed, associative memory). Thurstone did not find evidence of *g* until he developed higher-order factor analyses years later. Today, models of intelligence often include *g* along with group factors such as verbal comprehension and processing speed. Additional models of intelligence were presented by British researchers Vernon (1961) and Gustafsson (1984), both of whom supported a hierarchical model with *g* as the primary factor and well-defined broad abilities such as fluid and crystallized intelligence as lower-level abilities. This reflected the research of one of Spearman's students: Raymond Cattell.

In 1941, Cattell hypothesized that *g* might be comprised of fluid and crystallized abilities — G*f* and G*c*, respectively. He and John Horn described G*f* as a representation of how well individuals adjust to new situations. G*c* was defined as the ability to learn and retain factual knowledge. Their two-factor model eventually advanced to become nine factors: G*f*, G*c*, short-term acquisition and retrieval (G*sm*), visual intelligence (G*v*), auditory intelligence (G*a*), long-term storage and retrieval (G*lr*), cognitive processing speed (G*s*), correct decision speed (G*ds*), quantitative knowledge (G*q*), and reading/writing skills (G*rw*; Horn, 1991). Cattell proposed that these factors should be organized into three hierarchically arranged strata with *g* at the highest level, the nine broad ability factors in the middle, and narrow abilities in the third stratum (Wasserman & Tulsky, 2005).

Cattell and Horn's G*f*-G*c* theory gained notoriety among researchers, though practitioners struggled to understand the connections between the theory and actual measures of intelligence (McGrew, 2005). At a 1986 meeting to revise the WJ, John Horn presented the findings of an EFA of the WJ. This led to the incorporation of G*f*-G*c* theory into the Woodcock-Johnson – Revised (WJ-R; Woodcock & Johnson, 1989).

John Carroll was another researcher interested in structural models of intelligence. His 1993 publication, *Human Cognitive Abilities: A Survey of Factor-Analytic Studies*, reexamined 461 factor analyses of cognitive abilities and summarized the findings. He also current a three-tiered model of intelligence with *g* as the broadest stratum, eight broad abilities in the second stratum (G*c*, G*f*, general memory and learning [G*y*], broad auditory perception [G*a*], broad visual perception [G*v*], broad retrieval ability [G*r*], reaction time/decision speed [G*t*], and broad cognitive speediness [G*s*]). Carroll described his work as a blueprint intended to inform future research into cognitive abilities.

Given the similarities of the two theories, it is unsurprising that Cattell and Horn's G*f*-G*c* theory eventually combined with Carroll's three-stratum theory became what is now referred to as CHC theory (McGrew, 2005). The Woodcock-Johnson III is thought to be the first measure purportedly based on this underlying theory of intelligence. The most recent iteration of CHC theory includes several revisions, for example, new broad abilities, changes to nomenclature, the differentiation of some narrow abilities, and updated task requirements for certain subtests (Schneider & McGrew, 2018). Additional factors have been current as well, as the breadth of research into cognitive abilities continues to expand and understanding of intelligence becomes clearer.

The Woodcock-Johnson Tests

Eventually, the evolution of G*f*-G*c* and CHC theories became intertwined with the progression of the WJ batteries. However, before the WJ had any theoretical foundation, it was the vision of Richard Woodcock, who's work began as a graduate student at the University of Oregon (Schrank, 2010). Woodcock conducted experiments for his doctoral dissertation which ultimately led to the development of the Visual-Auditory Learning test to predict reading ability. As a postdoctoral fellow, Woodcock turned his sights toward bridging research in cognition and neuroscience with measures of cognitive ability. Along with his assistant, Mary Johnson, Woodcock developed the WJPEB (Woodcock & Johnson, 1977). The battery was comprised of three parts: the Tests of Cognitive Abilities, Tests of Achievement, and Tests of Interest Level. The Tests of Cognitive

Abilities included 12 subtests designed to assess an extensive range of verbal and nonverbal abilities; the Tests of Achievement included 10 subtests aimed at determining reading, writing, and mathematics skills as well as general knowledge; the Tests of Interest Level provided measures of Scholastic (i.e., academic) and Non-Scholastic (i.e., physical and social) interests (Schrank, 2010).

The structure of the battery was defined using factor and cluster analyses, which yielded four distinct groupings: knowledge-comprehension, memory-learning, reasoning-thinking, and discrimination-perception (Schrank, Decker, & Garruto, 2016). The addition of a Broad Cognitive Ability score helped to further distinguish the WJBEP from other test batteries such as the Wechsler Scales, as it was derived by calculating the weights of each subtest versus simply weighing all subtests equally; this allowed for a more nuanced reflection of one's cognitive abilities.

As previously mentioned, Horn's presentation of G*f*-G*c* theory at the 1986 conference (Sternberg, 1986) to revise the WJ led to Woodcock incorporating G*f*-G*c* as the theoretical basis of the WJ-R. The WJ III fully embraced CHC theory, including the three-tiered model which promoted *g*; the concept of general intelligence was represented by the new General Intellectual Ability Scale (GIA; Woodcock, McGrew, & Mather, 2001, 2007). As a result of the fusion of theory and practice, the WJ III COG became one of the most widely used tools to assess cognitive abilities.

The WJ IV (Schrank, McGrew, & Mather, 2014a) continues the practice of integrating theory and research with assessment. The WJ IV is a series of measures comprised of three independent and co-normed batteries, including tests of Cognitive

Abilities (COG), Academic Achievement (ACH), and Oral Language (OL). When used together, the WJ IV COG and WJ IV OL supposedly measure seven broad CHC abilities: comprehension-knowledge (G*c*), fluid reasoning (G*f*), short-term working memory (G*wm*), long-term storage and retrieval (G*lr*), auditory processing (G*a*), visual processing (G*v*), and processing speed (G*s*).

Current Research

Although a number of research studies have focused on the Woodcock-Johnson series of instruments, factor analytic studies of the most recent revisions of the series are especially important to discuss given their similarities to the current study. The structure of the WJ III (Woodcock et al., 2001) was determined using confirmatory factor analysis (CFA), a technique which forces the dataset to confine to a set number of factors, in this case the nine broad ability factors associated with CHC.

Dombrowski and Watkins (2013) conducted an EFA and higher-order factor analysis of the full WJ III among the 9–13 and 14 to19-year-old age groups. They found that six factors emerged among the younger group and five factors among the older group; this included a blended G*q*, G*f*, and G*v* factor among 9 to13-year-olds, the group which is the target of the current study. Dombrowski and Watkins noted that *g* accounted for most of the variance in their study.

Dombrowski, McGill, and Canivez (2017) conducted a similar study using the WJ IV and found similar results: three factors emerged for the 9 to13-year-old age group and four factors for the 14 to19-year-old age group. Again, researchers noted that the WJ IV

COG primarily provided a strong measure of *g*, though it was not consistent with CHC theory as described in the *Technical Manual*.

In 2018, Spurgin conducted an EFA of the WJ IV COG and WJ IV OL among the 14-19 year-old age group. Results of this study are particularly relevant to the current study since similar techniques were used, though the age group examined differed (9-13 year-olds instead of the older age group). Spurgin's findings were consistent with research by Dombrowski and Watkins (2013) and Dombrowski et al. (2017) in that the number of factors found (five) was inconsistent with the number of expected factors based on the WJ IV *Technical Manual*. Spurgin specifically noted that G*f* and G*lr* were not able to be clearly identified and that some subtests loaded onto several factors. Though the aforementioned studies utilized different methodologies than those reported in the *Technical Manual,* Spurgin agreed with Dombrowski et al.'s assertion that the WJ IV may be over-factored.

Rationale for the Study

Given the relatively recent publication of the newest version of the WJ series of instruments, a limited amount of research is available. As previously noted, the WJ series of tests have become increasingly more popular among practitioners, many of whom are using these instruments to determine the abilities and needs of school-aged students in order to provide them with the most appropriate educational settings and interventions. Therefore, it is imperative that the instruments used are theoretically aligned with the constructs practitioners aim to assess. Likewise, it is also critical for practitioners to be aware of the limitations of an instrument and to recognize that this particular tool is purported to be based on an ever-evolving theory of intelligence. Additionally, information provided in the *Technical Manual* about the factor structure of the WJ IV is based on cluster analysis, principal components analysis (PCA), and multidimensional scaling; although these methods give valuable insight and present unique ways of visualizing the data, EFA remains the most traditional means of gaining insight into the factor structure of the WJ IV. Findings from this study will also be useful in ascertaining whether the observed factor structure is consistent with the CHC-based structure presented by Schrank, McGrew, and Mather (2014).

This study seeks to ascertain the factor structure of the WJ IV COG and WJ IV OL batteries among the 9 to13-year-old group using EFA. The correlation matrix provided in the WJ IV *Technical Manual* (McGrew et al., 2014) was used to conduct the analysis. Exploratory factor analysis (EFA) was originally established as a method of determining whether intelligence is a singular or multifaceted concept (Spearman, 1904). It has been commonly used in the social sciences as a way to explain underlying structures in datasets, frequently data involving psychological batteries (Osborne, 2014). In EFA, the pairwise relationships among all variables are analyzed using various extraction and rotation techniques in order to extract commonalities or latent factors. For the current study, principal axis factoring was used to extract factors due to the size of the available dataset and information in the WJ IV *Technical Manual* which indicated the data is normally distributed (McGrew et al., 2014), as well as to extract weaker factors. The Kaiser (1960) criterion of eigenvalues greater than 1.0, an examination of the scree plot, and parallel analysis were utilized to determine the number of factors to retain.

Since the dataset was expected to be correlated, an oblique rotation method (Promax) was used to provide information about how factors are related to each other (structure matrix) as well as how each item is related to its factor (pattern matrix).

The purpose of the current study is to examine the underlying factor structure of the WJ IV COG and OL among a school-age population, which is frequently assessed for services based on their abilities — 9 to13 years old — using EFA. Knowledge gained from this study were useful in understanding the constructs the WJ IV COG and OL are actually measuring, as well as adding to the body of research on the measure. The following research questions and hypotheses are proposed:

1. When utilizing exploratory factor analysis, what is the factor structure of the WJ IV COG and WJ IV OL in the 9 to 13-year-old age group?
2. Does this obtained factor structure align with CHC theory, the various factor structures reported in the *Technical Manual*, or factor structures identified by other researchers (e.g., Dombrowski et al., 2017; Spurgin, 2018).

CHAPTER II

REVIEW OF THE LITERATURE

In order to explore the factor structure of the WJ IV COG and WJ IV OL batteries, an understanding of both measures and the theoretical framework on which they were established is necessary. This chapter will outline the parallel development of theories of intelligence, culminating in an overview of modern-day CHC theory, which provides the foundation for the WJ batteries (McGrew et al., 2014). Because the formulation and use of factor analytic techniques is intertwined with cognitive assessment, an overview of factor analysis — especially EFA due to its relevance to the current study — were addressed in the chapter. Additionally, this chapter will provide a broad synopsis of the development of cognitive assessment and the history of the WJ tests, including the changes in CHC theory and other rationales that led to each revision. The chapter will continue with a detailed look at the WJ IV, including the goals of the test, constructs measured, and psychometric properties. Current research related to the factor structure of the WJ IV will also be discussed.

History of Intelligence Testing

To fully appreciate the development of the WJ IV, it is important to acknowledge its predecessors and discuss the history and origins of mental testing in general. Some of the earliest known attempts at assessing cognition date back to French physician Edouard Sèguin's use of form boards in the late 1800s to train children suspected of having cognitive deficits (Sèguin, 1856). The boards consisted of geometrically-shaped cutouts into which corresponding blocks were to be inserted. This measure now has several variations still in use by psychologists today, as it is easy to use, appealing to young children, and provides a generic understanding of a child's processing speed, visual-spatial skills, knowledge of shapes, and cognitive ability. The longevity of this measure demonstrates that while measures of intelligence have generally evolved over the last century, a common thread has persisted about what constitutes cognitive ability and how that ability is measured.

The work of Alfred Binet became one of the primary catalysts for modern-day intelligence testing. In 1904, Binet was part of a commission tasked with devising a way to assess children with developmental delays or intellectual disabilities in order for them to receive specialized educational services (Urbina, 2004). The following year, he and Theodore Simon published the Binet-Simon Scale, a collection of 30 tests primarily aimed at assessing judgment and reasoning. Specific tasks evaluated comprehension, vocabulary, memory, understanding directions, and other skills which might be indicators of school performance. Because the measure covered a broad range of skills, it was deemed highly useful and became a popular method of assessing a child's general intellectual ability in multiple countries. Revisions of the scale occurred in 1908 and 1911; chiefly, the scoring system was updated to provide a mental age score. This score was then divided by the child's chronological age and multiplied by 100 to obtain the IQ. This method of using a ratio to determine IQ was made more commonplace with the 1916 publication of the Stanford-Binet Intelligence Scale (Terman, 1916) and persisted for several decades. In addition to adapting the measure for use in the United States, Lewis Terman, a Stanford University professor studying giftedness, addressed the need for internal consistency and standardized instructions. While the method of determining overall intelligence has changed in recent years, the concept of comparing abilities based on the performance of others in the same age classification has remained a fundamental practice of most modern-day batteries.

The history of factor analysis is closely tied to the concept and study of intelligence. Francis Galton, inspired by Darwin's *The Origin of the Species* (Darwin, Duthie, & Hopkins, 1859), began studying the hereditary nature of intellect in an attempt to promote eugenics – selective breeding to improve the intelligence of the human race. Though his controversial goal was unsuccessful, Galton's work nevertheless contributed to the formation of theories of intelligence and the research methods used to devise them, including correlation and regression analyses (Urbina, 2004). His work also inspired subsequent researchers, including Charles Spearman, Raymond Cattell, and James McKeen Cattell. Spearman expounded upon Galton's use of correlation and laid the groundwork for modern day factor analysis; he sought to study and organize group data in a way that would reveal relationship patterns (McCredie, 2018). By examining patterns among correlations, Spearman aimed to understand the nature of intelligence, including whether or not there are different types of intellect. Ultimately, he theorized that there is a generalized intelligence (*g*) underlying all cognitive skills. While Spearman's beliefs about intelligence differed from those of Binet, their work simultaneously propagated the field of psychological testing.

The development of the Stanford-Binet propelled the use of intelligence testing for numerous purposes. One of the primary uses was as part of a core battery aimed at determining suitability for the United States military — the Army Alpha test (Boake, 2002). A panel of psychologists developed the procedures and included measures of verbal and reasoning abilities, as well as general knowledge and judgment. The success and ubiquitous use of the easily scored Army Alpha test led to the development of the Army Beta test, which included a series of nonverbal subtests such as Cube Counting, Pictorial Completion, and Mazes to assess the cognitive skills of non-English speakers and individuals with limited reading ability. Information regarding military intelligence testing was recorded in a lengthy document which yielded controversial ideas about ethnic intellectual differences, and concerns regarding the appropriateness of intelligence measures with different ethnic groups have persisted in the field (Berry, Clark, & McClure, 2011; Kwate, 2001). In addition to taking their cues from the Stanford-Binet, tasks included in the Group Beta Examination were at least partially adapted from performance measures used to assess intellect in individuals with limited English skills or formal education. Nonverbal tasks involving puzzles and form boards were eventually used to screen immigrants arriving at Ellis Island for physical and mental disorders (Boake, 2002). Work by Scott (1913), Otis (1918), Pressey and Pressey (1918), and Thorndike (1919), and others were used as models for the Army Alpha and Beta tests. The assessment of immigrants at Ellis Island contributed to the development of the Pintner-Paterson Performance Scale (Pintner & Paterson, 1917) to assess children with hearing impairments, and portions of the scale have emerged in modern-day measures of cognitive ability (Boake, 2002).

The 1920s and 1930s saw a massive expansion of intelligence testing following the widespread use of the Army tests. In 1921, the Psychological Corporation was founded; it incorporated tests used by the Army as well as those developed in conjunction with Columbia University to provide psychological testing services for academic and industrial purposes (Boake, 2002). Performance measures such as the Block Design test (Kohs, 1923), Leiter International Performance Scale (Leiter, 1936), and the Cornell-Coxe Performance Ability Scale (Cornell & Coxe, 1934) were increasingly used to supplement more common measures such as the Stanford-Binet (Boake, 2002).

The Alpha and Beta tests were also used as models for academic intelligence testing, which became exceedingly more commonplace (Urbina, 2004). The Wechsler-Bellevue scale is one such measure, with items related to math problem solving, information, and judgment, which originated directly from the Alpha test (Boake, 2002). David Wechsler began his career scoring Army Alpha tests in 1917. The following year, he began training as a psychological examiner at the School for Military Psychology. Wechsler was assigned to conduct brief interviews and administer commonly used intelligence and performance measures to soldiers who failed the Alpha and Beta tests. He attributed this experience to his idea to merge these verbal and nonverbal measures into a single scale, as he noted that many individuals who appeared to lead productive civilian lives were determined to have low intellectual functioning when administered the Stanford-Binet for military purposes. Wechsler went on to study with Spearman and

Pearson as well as with psychologists studying emotionality. He eventually became the chief psychologist at the Bellevue Psychiatric Hospital, an experience which strengthened his belief that the Stanford-Binet was not a sufficient or statistically sound method of determining intellectual functioning. Wechsler proposed to discard the ratio method of determining IQ and instead calculated intelligence by deriving a standard score from the sum of subtest scores. He also proposed to determine the mean IQ scores and standard deviations for every age level. The result was the WBIS (Wechsler, 1939), a test which not only incorporated verbal and performance measures, but which also allowed for the inclusion of qualitative data on an individual's temperament and motivation to complete tasks. This was viewed as a precursor to measures of executive functions.

Despite Wechsler's goals of creating a measure more relevant and accurate than the Stanford-Binet, the WBIS fell short in several areas. Wechsler acknowledged, for example, that not all subtests included in the new measure were statistically sound or selected based on empirical evidence (Boake, 2002). These tasks (e.g., Object Assembly) were instead included because of the potential for clinical interpretation. As a result, later factor analyses indicated that the structure of the WBIS differed from Wechsler's current classifications. Additionally, standardization procedures relied on a narrow sample, as all the data was collected from participants in New York, and the entire adult sample was comprised of white Americans who understood and wrote in English. Furthermore, although Wechsler borrowed from other scales already in existence, it appeared that he did not properly credit previous test creators or those who helped assemble the WBIS. Additionally, Paterson expressed concern that Wechsler's test was developed and used without consideration for individual differences, thus its use would lead to clinicians assessing cognitive deficits in a narrower fashion in accordance with the predetermined tests of the WBIS (Boake, 2002).

While the WBIS was far from perfect, the cohesiveness of the scale and the availability of diagnostic profiles was alluring to professionals in the rapidly expanding fields of psychology and adult psychiatry (Boake, 2002). The measure grew in popularity among practitioners and researchers. The Wechsler Mental Ability Scale (Wechsler, 1946) was established soon after as an alternative screening measure for those entering the armed services during World War II, thus continuing the legacy of intelligence testing for military purposes (Boake, 2002). The Wechsler Mental Ability Scale — or the Wechsler-Bellevue Form II as it was also known — was revised in 1949 to be suitable for children and renamed the WISC (Wechsler, 1949). While several of the items from Form II remained the same, the measure was normed on children aged 5 to 15 years old instead of adults (Boake, 2002). Successive revisions of the WBIS became the WAIS (Wechsler, 1955).

The WBIS and initial revisions of the scale were not overtly based on any theoretical model; however, a four-factor structure was reported in the *WISC-IV Technical and Interpretive Manual* (Wechsler, 2003): verbal comprehension, perceptual reasoning, working memory, and processing speed. Subsequent analyses yielded five factors; the verbal comprehension, working memory, and processing speed factors remained consistent, while perceptual reasoning was divided into fluid reasoning and visual processing (Reynolds, Keith, Flanagan, & Alfonso, 2013). Canivez, Watkins, and

Dombrowski (2016) evaluated the factor structure of the most recent Wechsler measure, the Wechsler Intelligence Scales for Children – Fifth Edition (WISC-V; Wechsler, 2014a) using an EFA. While the *WISC-V Technical and Interpretive Manual* (Wechsler, 2014b) presented evidence of a five-factor model similar to previous versions of the measure, Dombrowski et al. (2017, 2018) found no evidence of five factors, and instead asserted that the WISC-V primarily provided a strong measure of *g* (i.e., global or general intelligence).

Along with the aforementioned evolutions in the Wechsler Scales and the Stanford-Binet, additional measures of intelligence were borne from the preliminary work of Wechsler and others. These included the KABC (Kaufman & Kaufman, 1983), the CAS (Naglieri & Das, 1997), and the RIAS (Reynolds & Kamphaus, 2003), and latter editions of these measures are currently in use. The WJPEB (Woodcock & Johnson, 1977) was another measure which can be traced back to the work of early cognitive researchers, along with its subsequent revisions, the WJ-R, WJ III NU, and the WJ IV, which is the subject of the current study.

History of Factor Analysis and CHC Theory

As measures of intelligence were developed and expanded, theories of intelligence were simultaneously being explored. Spearman (1904) offered a theory of general intelligence prior to the establishment of any cognitive assessment measures. As a graduate student, Spearman proposed that there was a general intelligence, or *g* factor, which underlied all abilities (Wasserman & Tulsky, 2005). This was initially named as a two-factor theory, as it described *g* as an indicator of shared variance across measures while also allotting for specific skills, *s*, related to individual subtests (Spearman, 1904). Spearman's concept of *g* has permeated psychological testing despite early controversy, as each leader in the field at the time seemed to hold a different viewpoint on the definition of intelligence (Wasserman & Tulsky, 2005). Current models tend to incorporate the idea of a *g* factor but also allude to group factors, a notion which Spearman initially opposed but reluctantly acknowledged later in his career. This was in part due to the work of Truman Kelley, a 20th century psychometrician, who provided evidence of group factors using partial correlational methods. These factors refer to variance shared between clusters of subtests and are routinely referred to as broad ability factors. Additionally, modern models acknowledge that some degree of variance is subtest specific, while other variance is associated with error.

The foundation of these models may also be partially attributed to Louis Thurstone's work, as he used factor analyses to determine relationships among groups of variables. Using scores obtained from 56 tests given to 240 college students, Thurstone (1934), used a factor analysis to obtain 13 factors, including seven which he determined were primary abilities: reasoning, word fluency, verbal comprehension, associative memory, numerical facility, perceptual speed, and spatial visualization. No evidence was found for a general intelligence at the time; however, Thurstone later developed higher-order factor analyses which yielded the possibility of *g*.

The idea of *g* was further promoted by British researchers, including P. E. Vernon and G. Gustafsson, who proposed their own models of intelligence. Vernon (1961) suggested *g* as the primary factor with verbal-educational ability and mechanical-spatial ability as subordinate factors. These subordinate factors were further divided into more specific skills such as verbal and spatial abilities. Gustafsson (1984) also suggested a hierarchy which placed *g* at the top. He posited that crystallized intelligence and general visualization should be categorized as lower-level abilities (Wasserman & Tulsky, 2005).

Meanwhile in the United States, Raymond Cattell, who obtained his degree under the supervision of Spearman, was also interested in factor analysis and intelligence (Wasserman & Tulsky, 2005). He posited that *g* could be divided into two general factors: fluid ability (G*f*) and crystallized ability (G*c*; Cattell, 1941). Cattell and student-turned-colleague John Horn defined G*f* as a skill critical for reasoning, as it allows individuals to adapt to novel situations and does not require factual knowledge. Alternatively, crystallized intelligence refers expressly to factual information and the ability to acquire this kind of knowledge. These skills tend to be measured by observing lexical knowledge, language comprehension, quantitative abilities, and recollection of facts (Wasserman & Tulsky, 2005). Cattell and Horn later added visualization, cognitive speed, and retrieval capacity to the list of factors. Eventually, the model grew to nine factors, including G*f*, G*c*, short-term acquisition and retrieval (G*sm*), visual intelligence (G*v*), auditory intelligence (G*a*), long-term storage and retrieval (G*lr*), cognitive processing speed (G*s*), correct decision speed (GDS), quantitative knowledge (G*q*), and reading/writing skills (G*rw*; Horn, 1991). Regarding the organization of these factors, Cattell proposed a three-stratum model arranged hierarchically. Given that *g* was presumed to underlie all other abilities and included the largest amount of shared variance, it was placed in the highest stratum. The broad ability factors constituted the second stratum, while narrow ability factors comprised the third stratum (Wasserman & Tulsky, 2005).

While G*f*-G*c* theory gained prominence among cognitive researchers, individuals conducting intelligence tests still relied on measures which lacked significant theoretical backing (McGrew, 2005). This changed at a meeting to revise the WJ in 1986, as educational psychologist John Horn presented the results of an EFA of the WJ. Horn's work enabled practitioners to more clearly understand the connection between G*f*-G*c* theory and measures of intelligence (McGrew, 2005). After conducting and reviewing additional factor analytic studies, the Woodcock-Johnson – Revised (WJ-R; Woodcock & Johnson, 1989) was developed, and it was the first intelligence test to merge theory and practice. While some subtests from the previous edition remained or were generally updated, additional subtests were developed to fully assess each of the seven broad abilities proposed by G*f*-G*c*, and the theory began to gain prominence among practitioners.

In 1993, Carroll's publication of *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* further propelled intelligence theory by summarizing and reexamining 461 factor analyses related to cognitive abilities. His three-tiered model was conceptually similar and based in part on Horn and Cattell's G*f*-G*c* theory. Carroll agreed with the notion of a general intelligence factor (*g*) and positioned this as the third and broadest stratum. The second stratum consisted of eight broad abilities: crystallized intelligence (G*c*), fluid intelligence (G*f*), general memory and learning (G*y*), broad auditory perception (G*a*), broad visual perception (G*v*), broad retrieval ability (G*r*), reaction time/decision speed (G*t*), and broad cognitive speediness (G*s*). The first stratum incorporated several narrow abilities such as induction, lexical knowledge, and visualization; these were categorized as “level,” “speed,” “speed and level,” and “rate” factors. According to Carroll (2005), “level” factors provide information about an individual’s level of mastery. “Speed” factors indicate the quickness with which a task is performed, while “rate” factors provide a measure of learning rate in relation to memory and learning tasks. Carroll noted that his three-stratum theory was primarily intended to inform further research and as a blueprint of cognitive abilities for practitioners.

The Woodcock-Johnson III *Technical Manual* is believed to be the first published convergence of Cattell and Horn’s theory with Carroll’s (1993) model (McGrew, 2005). One reason for this merger was that G*f*-G*c* theory was still confusing to some practitioners, who either found the terms too cryptic or erroneously believed the theory only included the two factors associated with its name. Therefore, individuals associated with Riverside Publishing, Richard Woodcock, and Gale Roid, author of the Stanford-Binet Intelligence Scales, Fifth Edition (2003), convened and agreed to rename the model the Cattell-Horn-Carroll theory of cognitive abilities, which was eventually shortened to its common label, CHC theory (McGrew, 2005).

The most recent rendition of CHC theory is included in Flanagan and McDonough’s (2018) *Contemporary Intellectual Assessment*, *Fourth Edition.* A chapter by Schneider and McGrew (2018) notes that several revisions have been made to the current version of CHC theory such as changes to the nomenclature of certain broad and narrow abilities. For example, G*wm* now refers to working memory capacity in lieu of short-term working memory. Memory span has been reconfigured to reflect the different tasks it attempted to measure; it is now segmented into auditory short-term storage (*Wa*) and visual-spatial short-term storage (*Wv*). Another important change includes the segmentation of long-term storage and retrieval (G*lr*) into learning efficiency (G*l*) and retrieval fluency (G*r*). As Schneider and McGrew explain, the tasks within the former G*lr* broad ability appeared to assess different, unrelated skills; some tests called for individuals to learn information efficiently, while others required individuals to recall information from long-term memory.

Additional changes to CHC theory have been current by Schneider and McGrew (2018), such as the inclusion of emotional intelligence (G*ei*), changes to narrow abilities within G*f* (e.g., reasoning speed and Piagetian reasoning), and the removal of certain elements within the G*s* domain (for example, removing the rate-of-test-taking). As the use of CHC theory becomes more ubiquitous around the world, more research is being devoted to fully understanding these factors and defining them more clearly. Undoubtedly, CHC theory will continue to evolve; the current CHC integrated model includes 17 broad abilities and over 80 narrow abilities. The definitions of broad abilities and their associated narrow abilities are detailed in Table 1.

Table 1

Current CHC Broad and Narrow Abilities

Broad ability	Definition	Associated narrow abilities
Fluid Reasoning (G*f*)	The ability to reason, form concepts, and solve problems using unfamiliar information or new procedures	Induction, general sequential reasoning, quantitative reasoning

Broad ability	Definition	Associated narrow abilities
Comprehension-knowledge (G*c*)	Acquired knowledge, the ability to communicate knowledge, and the ability to reason using previously learned experiences or procedures	General knowledge, language development, lexical knowledge, listening ability, communication ability, grammatical sensitivity
Working Memory Capacity (G*wm*)	The ability to capture and preserve information in immediate awareness before manipulating or using it to complete a task	Auditory short-term storage, visual-spatial short-term storage, attentional control, working memory capacity
Learning Efficiency (G*l*)	The ability to learn and encode information over time	Associative memory, meaningful memory, free recall memory
Retrieval Fluency (G*r*)	The ability to efficiently recall information from long-term memory	Ideational fluency, associational fluency, expressional fluency, sensitivity to problems/alternative solution fluency, originality/creativity, naming facility, word fluency, speed of lexical access, figural fluency, figural flexibility
Processing Speed (G*s*)	The ability to quickly and accurately perform simple automatic mental tasks	Perceptual speed, perceptual speed – search, perceptual speed – compare, number facility, reading speed, writing speed
Reaction and Decision Speed (G*t*)	The ability to make basic judgments and decisions quickly	Simple reaction time, choice reaction time, semantic processing speed, mental comparison speed, inspection time
Psychomotor Speed (G*ps*)	Refers to the agility of body movements	Speed of limb movement, writing speed, speed of articulation, movement time
Domain-Specific Knowledge (G*kn*)	The degree to which one has mastered specialized	Foreign language proficiency, knowledge of signing, skill in lip reading, general science

Broad ability	Definition	Associated narrow abilities
	knowledge related to work, interests, or hobbies	information, knowledge of culture, mechanical knowledge
Reading and Writing (G*rw*)	Refers to the comprehension and usage of written language	Verbal language comprehension, reading decoding, reading comprehension, reading speed, English usage, writing speed
Quantitative Knowledge (G*q*)	The ability to comprehend mathematics	Mathematical knowledge, mathematical achievement
Visual Processing (G*v*)	The ability to perceive, analyze, synthesize, and think with visual configurations	Visualization, speeded rotation, closure speed, flexibility of closure, visual memory, spatial scanning, serial perceptual integration, length estimation, perceptual illusions, perceptual alternations, imagery, perceptual speed
Auditory Processing (G*a*)	The ability to analyze, synthesize, and discriminate sounds, including sounds presented under distorted conditions	Phonetic coding, speech sound discrimination, resistance to auditory stimulus distortion, memory for sound patterns, maintaining and judging rhythm, musical discrimination and judgment, absolute pitch, sound localization
Olfactory Abilities (G*o*)	The ability to sense and process relevant information related to smell	Olfactory memory
Tactile Abilities (G*h*)	The ability to sense and process relevant information related to touch	No narrow abilities have been well-supported to date
Kinesthetic Abilities (G*k*)	Skill in identifying and processing proprioceptive information	No narrow abilities have been well-supported to date
Psychomotor Abilities (G*p*)	The ability to make physical movements in an accurate or coordinated manner	Aiming, manual dexterity, finger dexterity, static strength, gross body

Broad ability	Definition	Associated narrow abilities
		equilibrium, multi-limb coordination, arm-hand steadiness, control precision

History of the WJ

The history and development of the WJ series of measures is closely aligned with the development of CHC theory; however, it is important to note the environment and rationales which contributed to the test's origination. The initial development of the WJ began in the 1950s as a series of experiments aimed at assessing differences in learning ability (Schrank, 2010). Visual-auditory learning was the first test developed (Woodcock, 1958); it arose from Richard Woodcock's dissertation as a doctoral student at the University of Oregon and was designed as an experiment involving visual-auditory association, encoding, and retrieval to predict reading ability. Similarly, analysis-synthesis was later developed to assess mathematics reasoning and deduction. As a postdoctoral fellow at the Tufts New England Medical Center in 1974, Woodcock established a goal of developing a full battery of tests to measure cognitive abilities based on available cognition and neuroscience research (Schrank, 2010). The first iteration of the WJ included three components: Tests of Cognitive Ability, Tests of Achievement, and Tests of Interest Level. In all, 12 tests assessing verbal and nonverbal abilities were included in the Tests of Cognitive Ability, and they were intended to measure a range of cognitive processes, for example, lower-level tasks such as basic arithmetic to higher-level tasks such as advanced calculations (Schrank et al., 2016). Though the WJ incorporated relevant research, the test was not steered by any theoretical model.

After it was normed around 1976, factor and cluster analyses helped to better define the structure of the WJ (Schrank et al., 2016). Four specific categories emerged and were labeled as knowledge-comprehension, memory-learning, reasoning-thinking, and discrimination-perception. Additionally, a Broad Cognitive Ability (BCA) was included due to the perceived need for a total cognitive score. The BCA was calculated by determining different weights of each of the 12 subtests, as test developers believed this would provide a more accurate reflection of one's cognitive abilities than weighing each subtest equally. The structure of the WJ-R (Woodcock & Johnson, 1989) was further defined following Horn's aforementioned presentation at the 1986 conference in which G*f*-G*c* was introduced as a potential theoretical outline. McGrew, Werder, and Woodcock (1991) then worked to conduct statistical analyses and merge exploratory and confirmatory analyses of the WJ that were available at the time. This work provided a foundation for organizing and developing the next revision of the WJ (i.e., the WJ -R) in alignment with G*f*-G*c* theory.

As previously noted, it is difficult to separate the development of the WJ from the progression of CHC theory. The introduction of Carroll's three-tiered model of cognitive abilities combined with Horn and Cattell's G*f*-G*c* theory provided the theoretical underpinnings of the next iteration of the WJ — the Woodcock-Johnson III Tests of Cognitive Abilities (Woodcock et al., 2001, 2007). In addition to being based on the newly formed CHC theory, a primary difference between the WJ III COG and previous editions of the test was the inclusion of the GIA score, a representation of *g*. The integration of theory and practice spurred the WJ III COG to become one of the most commonly used tools to measure cognitive abilities.

The WJ IV (Schrank et al., 2014a) incorporated foundational CHC theory with more recent reconfigurations of aspects of the theory (referred to as CHC 2.0; Schneider & McGrew, 2012); however, an objective of this rendition of the WJ IV was to also infuse current research on neuroscience and human cognition, as a considerable amount of literature has been produced on these topics since the introduction of the WJ III COG. For example, the separation of working memory and short-term memory was initiated by neuroscience researchers who noted processing differences required by each skill (i.e., working memory involves more complex processing of information in immediate awareness, whereas short-term memory emphasizes storage more than processing; Gathercole & Alloway, 2008). The WJ IV authors incorporated these definitional changes into the CHC model; in lieu of the short-term memory cluster, McGrew et al. (2014) determined that G*wm* better encapsulated the constructs they are attempting to describe and measure. They have since proposed renaming the G*wm* factor working memory Capacity to better depict what these tasks are measuring (Schneider & McGrew, 2018).

Psychometric Properties of the WJ IV: Standardization, Reliability, and Validity

McGrew et al. (2014) maintain that norming and technical analyses of the WJ IV were steered by *The Standards for Educational and Psychological Testing* produced by the American Educational Research Association, the American Psychological

Association, and the National Council on Measurement in Education. These organizations have collaborated on a set of guidelines since 1966 to promote valid, reliable, and fair testing practices, as well as to ensure accurate assessment and scoring practices and the appropriate application of assessment results. Normative data for the WJ IV was obtained via a single sample of 7,416 participants from the United States, including 664 children ages 2 to 5 years, 3,891 school-age children, 775 college-level adults, and 2,086 additional adults (McGrew et al., 2014). At least 200 participants were included for each age with the exception of the 2-year-old group. The sample also included individuals from 46 states and the District of Columbia. Stratified sampling was used to control for variables such as sex, race, ethnicity, region, birth country, type of community, parent education, type of college/school, educational level, and employment for adults; the goal of this was to match the demographic characteristics of the 2010 U.S. Census projections. Despite these efforts, categories related to adults such as college type and country of birth among 20 to 29-year-olds differed from the national population. Therefore, McGrew et al. utilized examinee weighting to reduce sampling bias. This involved determining partial weights based on whether a participant was a member of an over- or under- represented group in the sample and then calculating the overall weight of each examinee.

When examining the psychometric properties of an instrument it is important to review and understand the reliability and validity evidence provided in the *Technical Manual*. Thus, a brief review of the validity and reliability of the WJ IV COG and WJ IV OL batteries is provided and further discussion can be found in Chapter 3. Reliability refers to the consistency of a measure. The reliability of the WJ IV (McGrew et al., 2014) was assessed differently based on the type of scoring involved. Internal-consistency reliability refers to the consistency with which test items measure the same construct (Urbina, 2004). To determine internal consistency, split-half reliability was calculated for untimed subtests with dichotomously scored items (McGrew et al., 2014). Split-half reliability is determined by segmenting the assessment measure into two components and comparing individual scores for both halves (Urbina, 2004). Reliability for subtests with multiple-point scoring was determined using the Rasch model which is the basis for the WJ IV's *W* scale (McGrew et al., 2014). This model gives a standard error of measurement (*SEM*) related to the ability estimate for each individual in the norm sample, which in turn yields a *W* score. Test-retest reliability, which is determined by giving the measure to a group of participants on two occasions and conducting correlations between the sets of scores, was used to assess the reliability of speeded subtests. The *Technical Manual* specifically provides reliability coefficients with a 68% confidence interval for both speeded and non-speeded tests across age groups, as well as a median reliability coefficient for each subtest. Additionally, cluster score reliabilities were calculated using Mosier's (1943) equation. Information in the WJ IV *Technical Manual* (McGrew et al., 2014) indicates there was a goal of a reliability coefficient of 0.90 or greater for cluster scores and 0.80 or greater for individual subtest reliability. About 91% of cluster scores, 98% of non-speeded tests, and 92% of speeded tests met this goal. Generally, the manual reports medium to high reliability coefficients across subtests; however, lower reliability statistics were reported for Picture Recognition (0.74) and on specific subtests at certain ages (e.g., Sound Awareness among adolescents) (McGrew et al., 2014). The manual explains that variation in reliability among different age ranges may be ascribed to limited variability in that particular sample. Median reliability coefficients for each subtest to be included in the current study are reported along with subtest descriptions in Chapter 3.

Validity refers to the extent to which a tool measures what it is intended to measure (Urbina, 2004). While the current study aims to explore the underlying factor structure of the WJ IV COG and WJ IV OL batteries, results may ultimately provide a source for comparison with the CHC structure on which the WJ IV is based. While there are multiple types of validity, an examination of content and construct validity are particularly relevant to this study. Content validity refers to the extent to which measurement items are relevant and accurately reflect the constructs they are intended to measure. The WJ IV *Technical Manual* reports the use of multidimensional scaling (MDS) to determine content validity (McGrew et al., 2014). In MDS, points are plotted along a one- or two-dimensional matrix in order to ascertain the underlying structure of a group of data (Cox & Cox, 2000). It provides a spatial representation of the relationships among variables, as those with more commonalities are plotted closer together and those with fewer commonalities are plotted further apart. While this method is more subjective than EFA, the qualitative information gained may prove valuable in determining the validity of the content as well as the processes which impact performance on cognitive tasks (McGrew et al., 2014). The MDS procedure was also supplemented with cluster and correlational analyses. In reporting results of MDS for the group relevant to the current study — ages 9 to 13 — the WJ IV *Technical Manual* posits that there is preliminary evidence for shared content characteristics among four broad factors: auditory-linguistic, figural-visual, quantitative-numeric, and reading-writing. While a number of subtests maintain the same classifications across various age groups, Numbers Reversed and Object-Number Sequencing on the WJ IV COG shifted classifications in different age groups. This was attributed to either developmental differences or an incomplete understanding of test content which could potentially be resolved using three-dimensional MDS. The WJ IV OL subtests were exclusively classified as auditory-linguistic, and the three separate batteries of the WJ IV generally clustered together.

Construct validity refers to how well results of a measure (e.g., test scores) relate to relevant research and underlying theories (Urbina, 2004). The structural validity of the WJ IV was assessed using a three-stage analysis involving split-sample random sample generation, exploratory multivariate methods, and confirmatory structural model cross-validation (McGrew et al., 2014). The second stage incorporated the use of exploratory methods, including cluster analysis, exploratory PCA, and multidimensional scaling. As the current study involves an EFA, a review of the PCA detailed in the *Technical Manual* is particularly relevant, though it is important to note that the PCA was conducted for the full battery of tests, whereas the current study excludes analysis of the WJ IV ACH. After a varimax rotation, a scree plot and eigenvalues were used to determine the number of components to extract, and eight, nine, and 10-component solutions were retained and reported. Each solution included five broad CHC abilities: comprehension-knowledge, processing speed, reading-writing, a combination of fluid reasoning and quantitative reasoning, and short-term working memory. Results of the three-stage analysis reported in the manual support the structural validity of the WJ IV, indicating that the structure is generalizable to a top-down CHC model with nine broad factors as well as a bottom-up CHC model with 13 broad and narrow factors.

Finally, the WJ IV was compared to other measures of intelligence, oral language, and achievement abilities to provide a measure of concurrent validity (McGrew et al., 2014). High correlations (.83 to .86) were reported between the WJ IV general intelligence clusters and the Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV; Wechsler, 2003) Full Scale Intelligence Quotient (FSIQ). Moderately high correlations (.71 to .77) were reported between the WJ IV general intelligence clusters and the Kaufman Assessment Battery for Children – Second Edition (KABC-II; Kaufman & Kaufman, 2004) Fluid-Crystallized Index (FCI). When compared to the Stanford-Binet Intelligence Scales, Fifth Edition (SB-5; Roid, 2003) FSIQ correlations of .79 to .82 were reported. Oral language batteries compared with the WJ IV OL included 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). Correlations with the CELF-4 Core Language Composite ranged from .42 to .83, while correlations with the PPVT-4 ranged from .14 to .76 for individuals age 10–18. Notably, the Speed of Lexical Access cluster of the WJ IV OL was least correlated with the other batteries. Correlations with the

CASL Core Composite ranged from .72 to .85 for the 7–17 age group, and the primary WJ IV OL clusters correlated moderately highly (.62–.68) with the OWLS Oral Composite. Additional correlations with other measures of cognitive, oral language, and achievement measures can be found in the WJ IV *Technical Manual* (McGrew et al., 2014).

Factor Analytic Studies of the WJ III

Dombrowski and Watkins (2013) conducted an EFA and higher order factor analysis of the full WJ III (Woodcock et al., 2001, 2007) battery to assess the test's alignment with CHC theory. Woodcock and colleagues utilized CFA to determine the structure of the battery when forced into nine factors in accordance with CHC. Dombrowski and Watkins (2013) suggested that while CFA provides important information about the theoretical basis of the battery, EFA would reveal whether the full battery organically aligned with CHC. After analyzing 42 subtests among two school-age groups (9-13 and 14-19), researchers found that six factors emerged for the younger group while five factors emerged for the older group. Factors for both groups were labeled crystallized ability/comprehension-knowledge (G*c*), processing speed/fluency (G*s*), and long-term retrieval (G*lr*). A blended G*q*, G*f*, and G*v* factor was found among 9 to 13-year-old group, while a general G*q* factor was found among 14 to 19-year-old group. Analysis of the younger age group also yielded a reading-writing factor (G*rw*); an auditory processing (G*a*) factor was found among the older age group. Dombrowski and Watkins noted the absence of clear G*v* and G*sm* factors. Overall, the authors of the study described a hierarchical structure with *g* accounting for most of the variance. While some elements were consistent with CHC theory (i.e. G*rw* and G*q* among the 14 to 19 age group), findings presented clear differences from the underlying theory proposed by the WJ III authors.

Another study of the WJ III Tests of Cognitive Abilities (WJ III COG) by Strickland, Watkins, and Caterino (2015) sought to determine whether the structure of the test remained consistent among a population of elementary school students referred for special education services. Researchers conducted exploratory and confirmatory factor analyses using the WJ III COG extended battery scores of 529 students. For the EFA, parallel analysis and examination of the scree plot yielded two distinct factors: one comprised of tests purporting to measure *Gf*, G*c*, G*v*, G*lr*, G*a*, and G*sm*; the other comprised of two G*s* tests. Researchers proposed the resulting data was under factored because of a salient general factor; therefore, they found that G*s* and G*c* combined to form a factor in a three-factor solution. When subjected to CFA, results were more consistent with the seven-factor model presented by McGrew and Woodcock (2001). As with the previously discussed study by Dombrowski and Watkins (2013), *g* accounted for most of the variance and was the only reliable factor. As a result, Strickland et al. (2015) concluded that interpretation of the WJ III COG is best restricted to *g* and G*s*, which demonstrated the most independence and reliability.

Factor Analytic Studies of the WJ IV

Given the relative newness of the WJ IV's release (2014), there is a paucity of research involving use of the WJ IV in clinical and academic settings. There is also little examination of the theoretical underpinnings of the WJ IV, making the current study a necessary contribution to the literature. While McGrew et al. (2014) presented the structural analyses on the full WJ IV in the *Technical Manual*, Dombrowski et al. (2017) argued that these analyses were insufficient for a multitude of reasons, and some of the statistics chosen were inappropriate for the battery, resulting in additional limits to the interpretation of WJ IV data.

Among their concerns, Dombrowski et al. (2017) primarily noted that the underlying structure of the individual tests of the WJ IV was never examined or reported; this was instead determined by extrapolating data based on the analyses of the full battery. This presents a problem for practitioners seeking to assess the validity and utility of the measure, as it is common for individual subtests or portions of the WJ IV to be administered in lieu of the full battery. Prior research regarding the structural validity of the WJ III COG concluded that the test yielded too many factors, as Dombrowski (2014) previously found that the WJ III COG produced only three or four factors among the 14-19 and 9-13 age groups, respectively. Different factor analytic techniques (i.e., principal axis factoring, PCA) also produced differing results indicating that the WJ III COG was over-extracted. Furthermore, these analyses revealed considerable concerns with the data, including poor convergence with CHC theory, Heywood cases (in which there is so much shared variance among variables that the correlation coefficient (r) is equal to one), and impermissible factors (e.g., factors without any relevant loadings; Dombrowski et al., 2017).

Additional concerns noted with the statistical methods reported in the WJ IV *Technical Manual* (McGrew et al., 2014) include the use of a varimax rotation and the subsequent lack of higher-order factor analyses (Dombrowski et al., 2017). According to Dombrowski et al., the presumption of a general factor, *g*, makes the use of an orthogonal rotation method (e.g., varimax) inappropriate, as the presence of a general factor insinuates that the factors may be highly correlated. While some researchers argue that the rotation method is inconsequential as long as a simple structure is achieved (Brown, 2009; Kim & Mueller, 1978), others believe in strong adherence to the traditional rules of factor rotation. Furthermore, because intelligence theoretically follows a hierarchical model (again, the presumption of *g*, broad abilities, and narrow abilities as delineated by CHC theory), a hierarchical factor analysis (HFA) can help to provide a broader, more generalized understanding of the underlying structure (e.g., Evans, 1999).

In addition to the aforementioned concerns, Dombrowski et al. (2017) took issue with the omission of the Schmid-Leiman (SL; Schmid & Leiman, 1957) procedure. This procedure has been commonly used in other cognitive assessment measures to clarify the relationships between lower and higher-level factors. The SL procedure was developed specifically for use with EFA, though it can be generalized to other factor analytic techniques. Within this procedure, higher-order factors explain most of the variance; any remaining variance is attributed to first-order factors (Brown, 2014). This helps to clarify the uniqueness of the first-order factors, as they are then not correlated with higher-order factors. Additionally, Carroll (1993) utilized this method when developing the three stratum theory of cognitive abilities; therefore, it stands to reason that it might be beneficial to use again when considering the link between the WJ IV and CHC theory (Dombrowski et al., 2017).

To further understand the factor structure of the WJ IV Tests of Cognitive Abilities (Schrank, McGrew, & Mather, 2014c), Dombrowski et al. (2017) conducted their own EFA and HFA using data from two school age groups (9–13 and 14–19) included in the normative sample. The researchers initially determined suitability for factor analysis by assessing the intercorrelation matrices for each age group. They then used Principal Axis Factoring (PAF) with an oblique rotation (Promax); factor extraction was determined using multiple methods, including an examination of the scree plot and the standard error of scree, Horn's parallel analysis (HPA; Horn, 1965), and minimum average partials (MAP; Velicer, 1976). Following these procedures, an HFA of the rotated factors was conducted along with the SL procedure (Dombrowski et al., 2017).

Results of both the HPA and standard error of scree resulted in three factors for the younger age group and four for the older group. Velicer's (1976) MAP criterion suggested that only two factors be retained for the younger group and one for the older group. An observation of the scree plot suggested that four factors should be retained for each age group, suggesting that the WJ IV COG has a maximum of four first-order factors, not seven in accordance with CHC theory (Dombrowski et al., 2017). Among the younger age group, researchers associated these factors with short-term working memory (G*wm*), processing speed (G*s*), comprehension-knowledge (G*c*), and a factor they labeled as perceptual reasoning. Six subtests did not have salient loadings on any factor (Story Recall, Non-Word Repetition, Phonological Processing, Concept Formation, Number Series, and Numbers Reversed). Among the older age group, the factors appeared to represent processing speed (G*s*), short-term working memory (G*wm*), perceptual reasoning, and crystallized ability.

Since none of these methods yielded seven factors as suggested by the WJ IV authors, Dombrowski et al. (2017) forced the extraction of seven factors to further investigate whether subtests conformed to the current theory. Results mimic aforementioned studies of the WJ (Dombrowski, 2014; Dombrowski & Watkins, 2013): the WJ IV COG provides a strong measure of *g*; however, it is generally not consistent with CHC theory. Researchers therefore caution against interpretation based on lower order or CHC factors due to poor structural validity.

Another study by Cormier, McGrew, Bulut, and Funamoto (2017) sought to study the correlations between the broad cognitive abilities outlined in CHC theory and different measures of reading achievement, including reading comprehension, reading fluency, reading rate, and basic reading skills. Additionally, researchers aimed to examine correlations between CHC and reading achievement for individual age groups (6-19) as opposed to the age groupings presented in the WJ IV *Technical Manual*. Results indicated that the G*f* cluster is the clearest predictor of all four measurements of reading achievement among all ages, and this is particularly fueled by the number series subtest. G*c* was also found to moderately predict reading achievement on the WJ IV Tests of Achievement, which is different from previous versions of the WJ measures in which G*c* was more highly correlated with reading achievement among 9 to 19-year-old participants. Cormier et al. (2017) suggest that more variance is now accounted for by G*f* than by other clusters such as G*c* and G*wm*, as was previously the case. Results also indicated that G*s* strongly predicts reading rate and reading fluency; however, since these clusters were not present in previous versions of the WJ instruments, no comparisons could be made. Changes to the G*a* cluster in the WJ IV (i.e., evaluating auditory processing at a more cognitively complex level) may be related to the cluster being more consistently associated with reading skills. Researchers noted that G*v* continues to be poorly associated with reading skills, similar to findings with the WJ-R and WJ III.

Though the study by Cormier et al. (2017) is primarily focused on associations between the WJ IV COG and ACH, this research is relevant to the current study because it reflects the evolution of how the Woodcock Johnson series has incorporated CHC theory, and how these changes may impact the utility of the instrument. As a result, more research is needed regarding recent changes to the structure of the WJ IV to assist practitioners with using the measure in the most accurate and efficient way possible.

Recent research conducted by Spurgin (2018) is especially relevant to the current study, as it provided an examination of the WJ IV COG and WJ IV OL among the 14 to 19-year-old age group using EFA. Spurgin used iterated principal axis factoring with a Promax rotation to identify the factor structure of the aforementioned measures. Results indicated that instead of the seven broad CHC factors measured by the WJ IV COG and WJ IV OL as reported in the *Technical Manual*, five factors emerged: comprehension-knowledge (G*c*), short-term working memory (G*wm*), auditory processing (G*a*), a blended processing speed (G*s*) and fluid reasoning (G*f*) factor hypothesized to represent attention, and a blended long-term storage and retrieval (G*lr*), visual processing (G*v*), and G*f* factor posited to represent cognitive reasoning. This aligns with previous research conducted by Dombrowski and Watkins (2013) and Dombrowski et al. (2017) which failed to confirm the theoretical structure presented in the WJ IV *Technical Manual*.

Chapter Summary

This chapter provided an overview of literature pertinent to the current study, including the history of intellectual assessment and factor analysis, as well as the evolution of CHC theory as a model of intelligence. It also detailed research studies relevant to the current study, namely studies of the WJ III and WJ IV which also examined factor structures in order to identify whether the tests measured what WJ authors purported to measure. The aim of the current study is to add to the growing body of literature surrounding the WJ series of tests, which are some of the most commonly used assessment measures among school-aged children.

CHAPTER III

METHOD

The purpose of this chapter is to outline the methodology of the study to explore the factor structure of the Woodcock-Johnson IV Tests of Cognitive Abilities (WJ IV COG; Schrank et al., 2014c) and the Woodcock-Johnson IV Tests of Oral Language (WJ IV OL; Schrank, Mather, & McGrew, 2014) in the 9 to 13 year-old age group. This chapter includes information regarding research participants, study procedures, necessary materials, and the psychometric properties of the WJ IV COG and WJ IV OL batteries. Additionally, the research questions posed by the study along with the statistical analyses used are detailed.

Research Participants and Procedures

The correlation matrix for the 9 to 13-year-old age group provided in the WJ IV *Technical Manual* (McGrew et al., 2014) was extracted and used to conduct the EFA. This group was selected for the current study due to the frequency of psychoeducational assessments of children in this age range for diagnostic and classification purposes.

The WJ IV *Technical Manual* (McGrew et al., 2014) reports data for a nationally representative sample of 7,416 individuals ranging from age 2 to over 90. Participants were divided into six age groups: 3 to 5 years, 6 to 8 years, 9 to 13 years, 14 to 19 years, 20 to 39 years, and 40 years and older. Information reported in the *Technical Manual* indicates the norming sample was selected to represent the U.S. population according to the 2010 U.S. Census projections. Participants were randomly selected using a stratified sampling design controlling for geographic region, sex, country of birth, race, ethnicity, community size, parent education, type of school, type of college, educational attainment, employment status, and occupational level. The total school-aged norm group consisted of 3,891 children enrolled in Kindergarten through 12th grade. Among the Kindergarten through 12th grade norming sample, 49.4% of participants were male and 50.6% of participants were female; 63.2% of participants were White, Non-Hispanic, 13.8% were Black, Non-Hispanic, 15.2% were White, Hispanic, 4.2% were Asian or Pacific Islander, Non-Hispanic, and the remaining 13.6% of participants were of other racial and ethnic backgrounds. In this group, 85.4% of participants resided in metropolitan areas and 89.5% attended public school. Additional demographic information can be found in the *Technical Manual*. The study will use data for the 9 to 13-year-old group, which included 1,572 participants (McGrew et al., 2014). The *Technical Manual* does not provide further delineation of the demographic characteristics for this sub-sample. This group was chosen to reflect the age at which cognitive assessments are commonly conducted among school-aged students.

An important point regarding the WJ IV norming process for the reader to be aware of is the procedures utilized in the original norming process of the WJ IV batteries. Sample participants were not administered the entire WJ IV battery of tests during the norming process (e.g., no one participant took all the subtests of the Cognitive, Achievement, or Oral Language batteries). The test publishers utilized a multiple matrix sampling design, which is a planned incomplete data collection method (McGrew et al., 2014). In other words, the test developers administered sets of tests to the participants and then matrix sampled the remaining subtests. As a result, all the data presented in the *Technical Manual* including the correlation tables relied on for this study consisted of imputed data from the normative sample.

Measures

Woodcock-Johnson IV Tests of Cognitive Abilities and Woodcock-Johnson IV Tests of Oral Language

The WJ IV is comprised of three co-normed batteries of cognitive, oral language, and academic abilities (Schrank et al., 2014b; Schrank et al., 2014c; Schrank, Mather, & McGrew, 2014). The data for this study will include 18 subtests from the WJ IV COG and nine subtests from the WJ IV OL for a total of 27 subtests. Descriptions of each subtest and associated reliability coefficients are detailed in the following section.

Subtest descriptions and psychometric properties

Oral vocabulary. This subtest provides a broad measure of comprehension-knowledge while narrowly assessing lexical knowledge and language development. Individuals are tasked with listening to a word and then providing a synonym (part A of the subtest) or antonym (part B). Cognitive processes involved in this subtest include matching, accessing information, and semantic activation. The median reliability coefficient reported for this subtest is .89.

Number series. Number Series provides a broad measure of fluid reasoning. Narrowly, this subtest is a measure of quantitative reasoning and induction skills. Individuals are tasked with determining increasingly difficult numerical sequences. They must mentally manipulate points on a number line and then apply an underlying rule to complete the sequence. This subtest has a median reliability coefficient of .91.

Verbal attention. According to the WJ IV authors, this subtest broadly measures short-term working memory while narrowly assessing working memory capacity and attentional control. This task requires individuals to listen to a mixed string of animals and numbers, and then answer questions regarding sequence. A median reliability of .86 was reported for Verbal Attention.

Letter-pattern matching. This subtest broadly assesses processing speed while narrowly measuring perceptual speed. Individuals are tasked with quickly identifying and circling matching letters or letter patterns. This task also requires visual discrimination, orthographic processing, and divided attention. Letter-Pattern Matching has a median reliability of .91 for the 7–11 age group.

Phonological processing. The Phonological Processing subtest is further subdivided into three parts: Word Access, Word Fluency, and Substitution. This subtest broadly provides a measure of auditory processing while narrowly assessing phonetic coding, word fluency, and speed of lexical access. For Word Access, individuals must provide a word that follows a specific phonetic rule. Word Fluency requires naming as many words as one can that begin with a certain sound. Substitution requires replacing part of one word to create a new one. Phonological Processing has a median reliability of .84.

Story recall. This subtest provides a broad measure of long-term storage and retrieval while narrowly measuring meaningful memory and listening ability. The Story

Recall task requires individuals to listen to and then recall story details. A median reliability coefficient of .93 was reported for this subtest.

Visualization. This subtest is further divided into two components: Spatial Relations and Block Rotation. Spatial Relations requires individuals to identify picture pieces to form a shape, while Block Rotation requires mentally rotating three-dimensional patterns to match a design. The Visualization subtest broadly measures visual processing and narrowly measures visualization abilities. The median reliability reported for this subtest is .85.

General information. The first part of the General Information subtest requires individuals to ascertain where an object is found; the second part requires identification of how an object is typically used. The subtest provides a broad measure of comprehension-knowledge and a narrow measure of general information skills. General Information has a median reliability of .88.

Concept formation. Concept Formation broadly measures fluid reasoning and narrowly measures induction skills. The subtest requires using two-dimensional representations to determine rules and categories. A median reliability coefficient of .93 was reported for this subtest.

Numbers reversed. The Numbers Reversed subtest provides a broad measure of short-term working memory and a measure of working memory capacity and attentional control as narrow abilities. In this task, individuals must listen to a string of numbers and then recall them in reversed order. Numbers Reversed has a median reliability coefficient of .88.

Number-pattern matching. This subtest is a new addition from the previous versions of the WJ (McGrew et al., 2014). It broadly provides a measure of processing speed and measures perceptual speed as a narrow ability. The task requires individuals to quickly identify and circle matching numbers in a set. Number-Pattern Matching has a median reliability coefficient of .85 for the 7–11 age group.

Nonword repetition. The Nonword Repetition subtest broadly measures auditory processing and assesses phonetic coding, memory for sound patterns, and memory span as narrow abilities. This subtest requires individuals to listen to and then repeat a nonsense word. Nonword Repetition has a high median reliability coefficient of .91 across all age groups.

Visual-auditory learning. This subtest provides a measure of long-term storage and retrieval and associative memory as a narrow ability. Individuals must learn and then recall strings of symbolic representations of words. This subtest has a high median reliability of .97.

Picture recognition. The broad construct measured by Picture Recognition is visual processing, while visual memory is the narrow ability assessed. This subtest requires individuals to select previously presented pictures while avoiding similar but incongruent pictures. Picture Recognition has a median reliability of .74.

Analysis-synthesis. Analysis-synthesis broadly assesses fluid reasoning and narrowly assesses general sequential reasoning. For this subtest, individuals use symbolic formulas to identify missing components of puzzles. A high median reliability coefficient of .93 was reported for Analysis-Synthesis.

Object-number sequencing. Broadly, this subtest provides a measure of short-term working memory. Narrowly, it assesses working memory capacity. Object-Number Sequencing requires listening to a blended group of numbers and words and then sequentially recalling the separated groups. This subtest has a median reliability coefficient of .89.

Pair cancellation. Pair cancellation broadly measures processing speed while assessing perceptual speed, spatial scanning, and attentional control as narrow abilities. The subtest requires individuals to quickly find and mark a repeated pattern. Pair Cancellation has a median reliability of .89 for the 7–11 age range.

Memory for words. This subtest provides an assessment of short-term working memory and narrowly assesses memory span. It requires individuals to listen to a string of unrelated words and then repeat them. Memory for Words has a median reliability of .82.

Picture vocabulary. Picture Vocabulary broadly assesses comprehension-knowledge and narrowly assesses lexical knowledge and language development. For this subtest, individuals must simply identify pictured objects. A median reliability coefficient of .88 was reported for this subtest.

Oral comprehension. This subtest also provides a broad measurement of comprehension-knowledge while narrowly assessing listening ability. The Oral Comprehension subtest requires individuals to listen to a passage and then determine an appropriate missing word based on contextual clues. This subtest has a median reliability of .82.

Segmentation. The Segmentation subtest provides a measure of auditory processing broadly and phonetic coding narrowly. Individuals are required to listen to a word and then divide that word into phonemes or syllables. Segmentation has a high median reliability of .94.

Rapid picture naming. Broadly, Rapid Picture Naming provides a measure of long-term storage and retrieval; narrowly, it assesses naming facility and speed of lexical access. The subtest requires individuals to identify objects and quickly recall and name them. A median reliability coefficient of .90 was reported for Rapid Picture Naming.

Sentence repetition. This subtest provides a broad measure of both short-term working memory and comprehension-knowledge. Narrowly, it assesses memory span and listening ability. Individuals must listen to words, phrases, or sentences and then repeat them in sequential order. Sentence Repetition has a median reliability of .83.

Understanding directions. The Understanding Directions subtest broadly assesses short-term working memory as well as comprehension-knowledge. It narrowly measures working memory capacity and listening ability. Individuals are required to briefly study a picture and then follow a sequence of instructions to point to various depicted items. A median reliability coefficient of .87 was reported for Understanding Directions.

Sound Blending. This subtest broadly assesses auditory processing and narrowly assesses phonetic coding. Sound Blending requires individuals to merge phonemes in order to say a complete word. A median reliability coefficient of .89 was reported for the subtest.

Retrieval Fluency. Broadly, the Retrieval Fluency subtest measures long-term storage and retrieval. Narrowly, it assesses speed of lexical access and ideational fluency. This task requires individuals to use up to one minute providing as many examples as possible of items in a given category. This subtest has a median reliability of .80.

Sound Awareness. The final subtest of the WJ IV OL provides a broad measure of auditory processing and a narrow measure of phonetic coding. Individuals are required to give rhyming words as well as remove word parts to create new words. Sound Awareness has a median reliability coefficient of .82.

Research Rationale, Significance, and Question

Given the relatively new release of the WJ IV, there is a considerable shortage of available research regarding the measure. Information provided in the *Technical Manual* about the factor structure of the WJ IV is based on cluster analysis, PCA, and multidimensional scaling; while these methods provide valuable insight and unique ways of visualizing the data, an EFA can present additional insight about the factor structure in a way that is commonly used in this line of research. Findings from this study will also be useful in ascertaining whether the observed factor structure is consistent with the CHC-based structure proposed by Schrank, McGrew, and Mather (2014a). The purpose of the current study is to determine the underlying factor structure of the WJ IV COG and WJ IV OL subtests among 9 to 13-year-old children. The following research questions are posed:

1. When utilizing exploratory factor analysis, what is the factor structure of the WJ IV COG and WJ IV OL in the 9 to 13-year-old age group?

2. Does this obtained factor structure align with CHC theory, the various factor structures reported in the *Technical Manual*, or factor structures identified by other researchers (e.g. Dombrowski et al., 2017; Spurgin, 2018).

Data Analysis

Primary Analysis

EFA was initially developed as a method of understanding whether intelligence is a singular or multifaceted concept (Spearman, 1904). This method of analysis has continued to be used in social sciences as a way to clarify underlying structures in datasets, frequently data involving psychological batteries (Osborne, 2014). In EFA, the pairwise relationships among all variables are analyzed using various extraction and rotation techniques in order to extract commonalities or latent factors. CFA is another method of factor extraction that seeks to group correlated variables based on an assumed underlying theory. While these alternative methods of factor extraction are commonly used, EFA is particularly useful for the current study because it can offer a more thorough exploration of the factors underlying a widely used instrument. The aim of the current study is to ascertain the factor structure of the WJ IV COG and WJ IV OL batteries among the 9 to 13-year-old group using EFA. A correlation matrix provided in the WJ IV manual was used to conduct the analysis. Additionally, the analysis was conducted using the Statistical Package for the Social Sciences (SPSS) version 25.0.

Tests of assumptions. It is pertinent to confirm that certain assumptions about the data have been met prior to conducting the EFA in order to promote valid results. One such assumption involves ensuring that the sample size is large enough.

Recommendations vary regarding what constitutes the minimum sample for an EFA; some have suggested making this determination based on an absolute number [e.g., 100 (Gorsuch, 1983); at least five times the number of variables (Hatcher, 1994); 51 more cases than the number of variables (Lawley & Maxwell, 1971), etc.]. Others have proposed a ratio method in which the number of cases is proportional to the number of variables (e.g., Bryant & Yarnold, 1995; Suhr, 2006). The consensus is that the appropriate sample size should correspond to the strength of the factor loadings; weaker factor loadings will require higher sample sizes — at least 300 — while stronger factors (e.g., at least four items with loadings greater than .60) may use smaller samples (Beavers et al., 2013). The current study includes data from 1,572 individuals, a sufficient sample size by any of the aforementioned criteria.

The strength and linearity of relationships in the data must also be assessed; this typically involves examining a correlation matrix of the data and ensuring that correlations are strong enough (above .30) to justify segmenting them into factors (Beavers et al., 2013). While correlations should be strong enough to conduct a factor analysis, it is important that correlations are not so strong that it is difficult to assess differences among groups of data — this refers to the need for there to be an absence of excessive multicollinearity in order to proceed with a research study. Finally, consistency of the data (e.g., absence of outliers and relatively complete data sets) was evaluated prior to moving forward with a factor analysis.

The Keiser-Meyer-Olkin Test of Sampling Adequacy (KMO; Kaiser, 1974) was used to test for basic assumptions and indicate whether the correlations were adequate for factor analysis. The purpose of KMO is to assess shared variance in data; a KMO value of .70 or higher will indicate that an EFA may proceed. Bartlett's (1950) test of sphericity is another measure which checks assumptions by testing the null hypothesis that no variables are significantly correlated; however, given the large number of participants in this study, this test may over-identify statistical significance and would not be as useful in determining whether to proceed with an EFA (Henson & Roberts, 2006).

Factor extraction. Extraction is a method of examining the correlation among all variables and pulling out the underlying factors (Osborne, 2014). In an initial extraction, there is an assumption that all factors are independent of each other (Beavers et al., 2013). The first factor attempts to account for the largest percent of the variance in the data; the second factor accounts for the largest percent of the remaining variance, and so forth until all variance can be explained. While this process is generally referred to as factor analysis, it may indicate two different techniques: component analysis and common factor analysis. The overarching distinction between the two is that component analysis presumes that the scores of individual items cause the component, while common factor analysis presumes that a common factor is contributing to item scores. Other differences relate to the type of variance included in the analyses. Component analysis incorporates shared, specific, and error variance, while common factor analysis only incorporates shared variance in the factor solution (Beavers et al., 2013).

Understanding the different factor analytic techniques is important in understanding various extraction methods and determining which method is the best fit given the data set and research question. Common extraction methods used in EFA include PCA, principal axis factoring, generalized least squares factoring, maximum likelihood extraction, alpha factoring, and image factoring (Meyers, Gamst, & Guarino, 2006). Principal component analysis is the most commonly used type of component analysis. The purpose of PCA is to "reduce the number of items to a smaller number of representative components" (Beavers et al., 2013, p. 6). Because the goal of the current study is to ascertain an underlying factor structure, common factor analysis extraction methods are more fitting. Frequently used extraction techniques for common factor analysis include principal axis factoring (PAF) and maximum likelihood extraction. Principal axis factoring may be iterated or non-iterated, meaning the shared variance may be estimated in one (non-iterated) or multiple (iterated) steps. Principal axis factoring can be used if data is not normally distributed, while maximum likelihood requires normality. However, maximum likelihood also provides additional information, including information regarding the significance of each item and the fitness of the factor structure. In addition to the aforementioned purposes for maximum likelihood extraction, the overall goal is to maximize the likelihood that sample data is an approximation of the population correlation matrix, which has the added benefit of more generalizable results (Osborne, 2014). There are certain limitations associated with maximum likelihood extraction, as this method runs the risk of estimation problems and is less useful than PAF for extracting weaker factors. Given the size of the available dataset and information in the WJ IV *Technical Manual* which indicates the data is normally distributed (McGrew et al., 2014), as well as the need to examine potentially weak factors, principal axis factoring was used to extract factors.

Factor retention. As some factors may not contribute significantly to the EFA solution, decisions need to be made regarding which factors to retain (Henson & Roberts, 2006). Methods of determining factor retention include Bartlett's chi-square test, examining the scree plot which provides a graph of eigenvalues, parallel analysis, and the minimum average partial method. Bartlett's chi-square test has been described as inconsistent and overly influenced by the size of the sample. According to Henson and Roberts (2006), the scree test is a more accurate metric; however, reliance on this method commonly results in over extracting factors. Parallel analysis and minimum average partial correlations tend to be most accurate. Parallel analysis uses randomly generated uncorrelated data and compares eigenvalues from the EFA (Osborne, 2014). As a result, only factors with eigenvalues which are significantly greater than the mean of eigenvalues from the randomly generated sample are retained. Minimum average partial criteria is useful for PCA; it entails separating out shared variance with each successive component until only unique variance remains. Researchers make the case for using multiple methods to determine factor retention in order to achieve the most accurate result (Henson & Roberts, 2006; Osborne, 2014). Therefore, the Kaiser (1960) criterion of eigenvalues greater than 1.0 and an examination of the scree plot were utilized to determine the number of factors to retain.

Factor rotation. Rotation is used to clarify the results of the factor analysis for interpretation (Osborne, 2014). This process occurs after the number of factors which were retained has been determined; other factors are removed and another factor analysis is conducted in order to force items into the remaining factors (Beavers et al., 2013). The goal of factor rotation is to achieve a simple structure — each factor only contains at least three items which load strongly (.70 or greater) and only on that factor (no higher than .40 on other factors). The recommended strength of loadings varies, as some researchers believe a loading of .50 is strong enough to be part of the factor solution; however, the general aim is to have the factor account for as much variance in an item as possible.

The primary rotation methods are orthogonal and oblique; orthogonal rotation methods force factors to be uncorrelated while oblique rotations allow for correlated factors and control for shared variance (Beavers et al., 2013). Additionally, oblique rotations provide information regarding how factors relate to each other (structure matrix) as well as how each item relates to its factor (pattern matrix). Given the nature of intelligence tests, some correlation is expected; therefore, an oblique rotation method proved more suitable for the purposes of this study. Methods of oblique rotation include Direct Oblimin and Promax. Using Direct Oblimin rotation allows for the implementation of a ceiling on the amount of correlation between factors, which helps to prevent factors from overlapping and increases clarity of the factor structure. Promax rotation performs a similar function; however, it simplifies this procedure for larger data sets. Given the size of the data, a Promax rotation would be more practical.

Chapter Summary

This chapter outlined information regarding research participants as well as the reliability and validity as reported in the WJ IV *Technical Manual*. There were also descriptions of each subtest and related psychometric data. The chapter concluded with a discussion of different analyses required in an EFA and the tests of assumptions (KMO and Bartlett's test of sphericity), factor extraction (PAF), factor retention (visual scree and Kaiser criterion), and factor rotation (Promax) techniques used in this study.

CHAPTER IV

RESULTS

The aim of this study is to use EFA to examine how many factors exist on the WJ IV COG (Schrank et al., 2014c) and the WJ IV OL (Schrank, Mather, & McGrew, 2014) among the 9 to 13-year-old norming group. As mentioned in Chapter II of this study, Schrank, Mather, and McGrew (2014) relied on unconventional and somewhat subjective methods of observing factor structure (i.e., multidimensional scaling, cluster analysis, and PCA). EFA was the preferred method of data reduction in this study because it provides a more objective view of data and is used more often in this line of research; therefore, factors are easier to identify. The Statistical Package for the Social Sciences (SPSS) version 25.0 was used to conduct statistical analyses. Results include information from multiple iterated principal axis factor analyses with oblique Promax rotation.

Descriptive Discussion

An EFA was conducted using the correlation matrix found in the WJ IV *Technical Manual* (McGrew et al., 2014). Twenty-seven subtests from the WJ IV COG and WJ IV OL were included in this study; subtest abbreviations and descriptions are provided in Table 2. A total of 1,572 participants were included in this sub-sample. It should be noted that while demographic information is provided for the entire standardization sample in the *Technical Manual*, specific information pertaining to the 9 to 13-year-old group was unavailable. The correlation matrix which was used for data analyses is presented in Table 3.

Table 2

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

Subtests (CHC broad ability)	Abbreviations	Descriptions
Oral Vocabulary (G*c*)	OV	Listening to words and identifying synonyms/antonyms
Phonological Processing (G*a*)	PP	Identifying words that have specified phonemes, words that begin with a specified sound, and exchanging parts of words to create new words
Object-Number Sequencing (G*wm*)	ONS	Recalling sequences of numbers mixed with words
Oral Comprehension (G*c*)	OC	Identifying words which correctly fit with a passage which is read orally
Picture Vocabulary (G*c*)	PV	Looking at pictures of objects and identifying them.
Sound Awareness (G*a*)	SA	Identifying rhyming words as well as subtracting parts of words to create new words
Concept Formation (G*f*)	CF	Using inductive reasoning to identify and categorize rules in order to solve a puzzle
Verbal Attention (G*wm*)	VA	Identifying information after listening to a mixed sequence of numbers and animals
Understanding Directions (G*wm*)	UD	Observing a picture and then listening to and following a sequence of directions using the picture
Numbers Reversed (G*wm*)	NUMR	Attending to a series of numbers and then repeating the sequence in reverse order
Number Series (G*f*)	NS	Using reasoning skills to complete a numerical sequence
Memory for Words (G*wm*)	MW	Repeating a series of unconnected words
General Information (G*c*)	GI	Determining the use of objects and where objects may be located
Analysis-Synthesis (G*f*)	AS	Using reasoning skills to solve visual, symbolic puzzles
Segmentation (G*a*)	SEG	Hearing words and breaking them into segments (i.e. phonemes, syllables)
Visualization (G*v*)	VIS	Mentally manipulating two- and three-dimensional objects to solve puzzles
Sentence Repetition (G*wm*)	SENR	Repeating words, phrases, and sentences in the correct order
Story Recall (G*wm*)	STOR	Attending to a story and recalling information from it
Retrieval Fluency (G*lr*)	RF	Quickly identifying examples within various categories
Letter-Pattern Matching (G*s*)	LPM	Quickly identifying and circling letters and letter patterns
Nonword Repetition (G*a*)	NONR	Repeating a made-up word
Sound Blending (G*a*)	SB	Merging phonemes to create a word
Pair Cancellation (G*s*)	PC	Quickly finding and circling repeated patterns
Number-Pattern Matching (G*s*)	NPM	Quickly finding and circling matching numbers
Visual-Auditory Learning (G*lr*)	VAL	Learning symbolic representations of words and recalling them at a later point
Rapid Picture Naming (G*lr*)	RPN	Quickly identifying and saying the names of objects
Picture Repetition (G*v*)	PR	Studying a set of pictures and then identifying correct pictures in a set mixed with distracting/similar objects

Table 3

Correlation Matrix for the WJ IV COG and WJ IV OL Among the 9 to 13-Year-Old Age Group (n=1,572, 27 subtests)

Subtest	OV	NS	VA	LPM	PP	STO	VIS	GI	CF	NR	NPM	NWR	VAL	PR	AS	ONS	PC	MFW	PV	OC	SEG	RPN	SENR	UD	SB	RF	SA
OV	1.00	0.44	0.45	0.32	0.48	0.38	0.32	0.71	0.43	0.35	0.26	0.40	0.27	0.24	0.39	0.38	0.27	0.33	0.70	0.62	0.31	0.32	0.46	0.41	0.28	0.41	0.38
NS	0.44	1.00	0.41	0.40	0.45	0.36	0.33	0.27	0.42	0.41	0.48	0.24	0.20	0.14	0.43	0.33	0.27	0.23	0.32	0.35	0.30	0.19	0.25	0.45	0.13	0.28	0.48
VA	0.45	0.41	1.00	0.25	0.46	0.31	0.23	0.33	0.30	0.43	0.32	0.44	0.22	0.15	0.34	0.55	0.23	0.50	0.34	0.42	0.31	0.34	0.47	0.44	0.19	0.28	0.42
LPM	0.32	0.40	0.25	1.00	0.32	0.22	0.22	0.22	0.23	0.41	0.60	0.14	0.23	0.28	0.33	0.43	0.59	0.18	0.23	0.13	0.16	0.30	0.22	0.28	0.15	0.28	0.30
PP	0.48	0.45	0.46	0.32	1.00	0.24	0.31	0.34	0.38	0.36	0.29	0.34	0.33	0.18	0.33	0.43	0.30	0.51	0.39	0.39	0.58	0.26	0.44	0.32	0.43	0.45	0.56
STOR	0.38	0.36	0.31	0.22	0.24	1.00	0.33	0.29	0.32	0.24	0.28	0.30	0.32	0.31	0.38	0.39	0.20	0.34	0.37	0.45	0.26	0.19	0.25	0.41	0.18	0.28	0.35
VIS	0.32	0.33	0.23	0.22	0.31	0.33	1.00	0.30	0.42	0.32	0.28	0.33	0.40	0.43	0.46	0.36	0.23	0.34	0.31	0.32	0.36	0.19	0.28	0.38	0.30	0.18	0.37
GI	0.71	0.27	0.33	0.22	0.34	0.29	0.30	1.00	0.33	0.31	0.14	0.29	0.21	0.24	0.31	0.30	0.16	0.24	0.68	0.53	0.20	0.26	0.30	0.28	0.25	0.27	0.32
CF	0.43	0.42	0.30	0.23	0.38	0.32	0.42	0.33	1.00	0.34	0.25	0.36	0.42	0.21	0.52	0.42	0.27	0.36	0.33	0.34	0.37	0.31	0.27	0.42	0.31	0.24	0.35
NUMR	0.35	0.41	0.43	0.41	0.36	0.24	0.32	0.31	0.34	1.00	0.31	0.24	0.28	0.26	0.34	0.45	0.25	0.39	0.27	0.30	0.24	0.23	0.26	0.29	0.24	0.26	0.35
NPM	0.26	0.48	0.32	0.60	0.29	0.28	0.28	0.14	0.25	0.31	1.00	0.18	0.15	0.24	0.34	0.35	0.57	0.20	0.16	0.30	0.26	0.32	0.17	0.24	0.08	0.34	0.29
NONR	0.40	0.24	0.44	0.14	0.34	0.30	0.33	0.29	0.36	0.24	0.18	1.00	0.21	0.26	0.19	0.45	0.19	0.42	0.29	0.31	0.32	0.29	0.51	0.44	0.28	0.18	0.40
VAL	0.27	0.20	0.22	0.23	0.33	0.32	0.40	0.21	0.42	0.28	0.15	0.21	1.00	0.30	0.37	0.30	0.11	0.30	0.24	0.26	0.36	0.16	0.16	0.31	0.39	0.14	0.32
PR	0.24	0.14	0.15	0.28	0.18	0.31	0.43	0.24	0.21	0.26	0.24	0.26	0.30	1.00	0.34	0.32	0.18	0.20	0.20	0.39	0.21	0.35	0.25	0.32	0.21	0.30	0.28
AS	0.39	0.43	0.34	0.33	0.33	0.38	0.46	0.31	0.52	0.34	0.34	0.19	0.37	0.34	1.00	0.38	0.28	0.38	0.25	0.30	0.33	0.20	0.18	0.34	0.32	0.39	0.26
ONS	0.38	0.33	0.55	0.43	0.43	0.39	0.36	0.30	0.42	0.45	0.35	0.45	0.30	0.32	0.38	1.00	0.35	0.47	0.34	0.36	0.31	0.38	0.37	0.40	0.29	0.39	0.31
PC	0.27	0.27	0.23	0.59	0.30	0.20	0.23	0.16	0.27	0.25	0.57	0.19	0.11	0.18	0.28	0.35	1.00	0.13	0.21	0.20	0.18	0.38	0.22	0.24	0.10	0.34	0.24
MFW	0.33	0.23	0.50	0.18	0.51	0.34	0.34	0.24	0.36	0.39	0.20	0.42	0.30	0.20	0.38	0.47	0.13	1.00	0.29	0.28	0.39	0.28	0.48	0.35	0.37	0.41	0.35
PV	0.70	0.32	0.34	0.23	0.39	0.37	0.31	0.68	0.33	0.27	0.16	0.29	0.24	0.20	0.25	0.34	0.21	0.29	1.00	0.64	0.18	0.40	0.42	0.40	0.22	0.39	0.31
OC	0.62	0.35	0.42	0.13	0.39	0.45	0.32	0.53	0.34	0.30	0.30	0.31	0.26	0.39	0.30	0.36	0.20	0.28	0.64	1.00	0.28	0.39	0.48	0.44	0.27	0.40	0.39
SEG	0.31	0.30	0.31	0.16	0.58	0.26	0.36	0.20	0.37	0.24	0.26	0.32	0.36	0.21	0.33	0.31	0.18	0.39	0.18	0.28	1.00	0.15	0.27	0.28	0.42	0.21	0.44
RPN	0.32	0.19	0.34	0.30	0.26	0.19	0.19	0.26	0.31	0.23	0.32	0.29	0.16	0.35	0.20	0.38	0.38	0.28	0.40	0.39	0.15	1.00	0.29	0.37	0.12	0.46	0.23
SENR	0.46	0.25	0.47	0.22	0.44	0.25	0.28	0.30	0.27	0.26	0.17	0.51	0.16	0.25	0.18	0.37	0.22	0.48	0.42	0.48	0.27	0.29	1.00	0.48	0.12	0.26	0.40
UD	0.41	0.45	0.44	0.28	0.32	0.41	0.38	0.28	0.42	0.29	0.24	0.44	0.31	0.32	0.34	0.40	0.24	0.35	0.40	0.44	0.28	0.37	0.48	1.00	0.25	0.26	0.41
SB	0.28	0.13	0.19	0.15	0.43	0.18	0.30	0.25	0.31	0.24	0.08	0.28	0.39	0.21	0.32	0.29	0.10	0.37	0.22	0.27	0.42	0.12	0.12	0.25	1.00	0.17	0.40
RF	0.41	0.28	0.28	0.28	0.45	0.28	0.18	0.27	0.24	0.26	0.34	0.18	0.14	0.30	0.39	0.39	0.34	0.41	0.39	0.40	0.21	0.46	0.26	0.26	0.17	1.00	0.24
SA	0.38	0.48	0.42	0.30	0.56	0.35	0.37	0.32	0.35	0.35	0.29	0.40	0.32	0.28	0.26	0.31	0.24	0.35	0.31	0.39	0.44	0.23	0.40	0.41	0.40	0.24	1.00

Exploratory Factor Analysis

Upon examining the initial EFA, certain subtests appeared to account for a fair portion of the variance on multiple factors. Conversely, some subtests loaded weakly (<.30) across factors. Therefore, multiple factor analyses were conducted in order to ultimately observe a clear factor structure. Results of each EFA are presented in the following section, along with tables and charts.

Exploratory Factor Analysis I

Assumptions regarding the data (i.e., absence of multicollinearity, adequate sample size, and dataset consistency) were assessed using the Keiser-Meyer-Olkin Test of Sampling Adequacy (KMO; Kaiser, 1974) and Bartlett's (1950) test of sphericity. The KMO test was used to examine the shared variance within the data; the degree of common variance was .880, which indicated fitness to proceed. Bartlett's test of sphericity, which checks assumptions by testing the null hypothesis that no variables are significantly correlated, was significant (x^2 = 21,357.34, *df* = 351, *p* < .001). Results of these two measures indicate appropriateness to proceed with the factor analysis.

Table 4 provides a list of the communalities within the dataset. Communalities, or common variance, is the degree to which one variable is impacted by the other variables and common factors (Vogt, 1999). This provided insight into each variable's value within the data and showed that the variables and factors are related — an important component of examining the factor structure of the WJ IV COG and WJ IV OL. The information also demonstrated that the dataset is devoid of outliers, thus an EFA may proceed.

Table 4

EFA I: Communalities

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	.707	.511
Number Series (NS)	.578	.345
Verbal Attention (VA)	.568	.400
Letter-Pattern Matching (LPM)	.632	.241
Phonological Processing (PP)	.645	.457
Story Recall (STOR)	.411	.293
Visualization (VIS)	.454	.312
General Information (GI)	.611	.315
Concept Formation (CF)	.497	.369
Numbers Reversed (NUMR)	.394	.304
Number-Pattern Matching (NPM)	.587	.241
Nonword Repetition (NONR)	.473	.301
Visual-Auditory Learning (VAL)	.367	.219
Picture Recognition (PR)	.440	.204
Analysis-Synthesis (AS)	.535	.341
Object-Number Sequencing (ONS)	.552	.444
Pair Cancellation (PC)	.505	.197
Memory for Words (MW)	.566	.355
Picture Vocabulary (PV)	.674	.379
Oral Comprehension (OC)	.663	.432
Segmentation (SEG)	.455	.271
Rapid Picture Naming (RPN)	.447	.246
Sentence Repetition (SENR)	.567	.328
Understanding Directions (UD)	.495	.396
Sound Blending (SB)	.430	.192
Retrieval Fluency (RF)	.506	.281
Sound Awareness (SA)	.525	.388

Extraction Method: Principal Axis Factoring.

Iterated principal axis factoring with a Promax rotation was used to extract factors. Observation of the scree plot — a chart of factors (x-axis) plotted against eigenvalues (y-axis) — and eigenvalues greater than one were additional methods used to confirm the number of factors (Kaiser, 1960). The scree plot appeared to present one strong factor. Examination of the Kaiser criterion resulted in the retention of six factors with the first factor accounting for the largest proportion of the variance (34.89%); this was consistent with the visual scree. The second factor comprised 7.11% of the variance, the third accounted for 6.47%, the fourth accounted for 5.13%, the fifth accounted for 4.50%, and the sixth factor accounted for 4.10% of the variance. Together, the six factors explained 62.20% of the total variance. Thus, the first EFA produced six factors; however, given the strength of the first factor, this EFA could be interpreted as a one-factor solution. See Table 5 and Figure 1.

Table 5

EFA I: Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.419	34.885	34.885	8.957	33.175	33.175	6.087
2	1.919	7.109	41.994	1.503	5.566	38.741	7.021
3	1.748	6.474	48.468	1.296	4.799	43.540	5.492
4	1.384	5.126	53.593	.900	3.334	46.875	6.013
5	1.216	4.503	58.097	.770	2.851	49.726	5.070
6	1.107	4.102	62.199	.642	2.378	52.104	2.124

Extraction Method: Principal Axis Factoring.

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

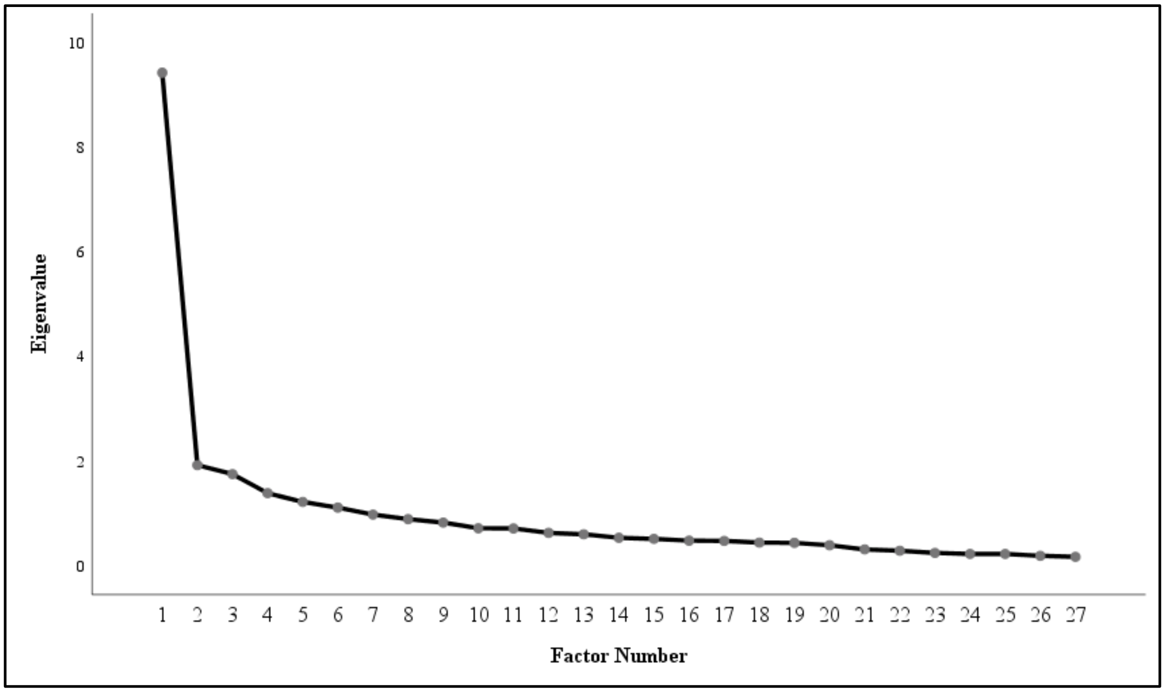

Figure 1. EFA I: Visual scree plot.

Pattern matrices show the weight of each variable relative to each factor, and observing the pattern matrix provided insight into whether factors were clear and fairly independent of each other or if they contained considerable degrees of overlap. Factor loadings depict the amount of influence each variable has on a factor, with 1 and -1 indicating the strongest impact toward or away from the factor, respectively. The general consensus among researchers is that factor loadings should not be below .30 to be considered salient, though some contend that a floor of .40 provides a more practical solution (Gorsuch, 1983; Stevens, 2002). As a compromise, variables with factor loadings greater than 0.35 are considered to be relevant to the factor for the purposes of the present study. For EFA I, the pattern matrix converged after seven iterations. The variables loaded onto six factors with two variables without salient loadings on any particular factor. See Table 6.

Table 6

EFA I: Pattern Matrix

	Factor					
Subtest	1	2	3	4	5	6
PV	.92	.01	-.05	-.03	-.04	-.02
GI	.84	-.15	-.09	.07	.04	.04
OV	.80	.01	.00	-.03	.08	.14
OC	.61	.20	-.06	.13	-.07	-.01
SENR	.10	.82	-.10	-.17	-.01	.00
NONR	-.08	.74	-.17	.09	.03	.00
VA	-.01	.68	.06	-.15	.09	.15
UD	.03	.56	-.04	.27	-.18	.15
MW	-.14	.53	-.05	.10	.39	-.15
ONS	-.09	.42	.27	.17	.09	-.07
LPM	-.06	-.15	.80	.05	.00	.14
NPM	-.12	-.06	.80	.03	-.06	.22
PC	-.04	-.07	.79	-.04	-.01	.01
RF	.25	.01	.42	-.02	.18	-.23
RPN	.16	.28	.39	.04	-.11	-.27
NUMR	.01	.13	**.23**	.14	.12	.13
VIS	-.01	.03	-.03	.64	.04	.05
VAL	-.02	-.12	-.09	.61	.26	-.03
PR	.03	.07	.14	.57	-.12	-.22
AS	.05	-.18	.20	.55	.13	.09
CF	.08	.05	.01	.43	.13	.12
STOR	.15	.18	.01	.37	-.11	.11
PP	.12	.09	.14	-.22	.80	.05
SEG	-.10	.06	-.05	.22	.56	.05
SB	.03	-.13	-.15	.38	.56	-.13
SA	.02	**.26**	-.01	.10	**.30**	**.22**
NS	.10	.08	.25	.02	-.02	.65

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

The first factor included four subtests with factor loadings greater than .35: Picture Vocabulary (.92), General Information (.84), Oral Vocabulary (.80), and Oral Comprehension (.61). The second factor included Sentence Repetition (.82), Nonword Repetition (.74), Verbal Attention (.68), Understanding Directions (.56), Memory for Words (.53), and Object-Number Sequencing (.42). Factor three included five subtests with salient factor loadings: Letter-Pattern Matching (.80), Number-Pattern Matching (.80), Pair Cancellation (.79), Retrieval Fluency (.42) and Rapid Picture Naming (.39). The fourth factor was comprised of Visualization (.64), Visual-Auditory Learning (.61), Picture Recognition (.57), Analysis-Synthesis (.55), Concept Formation (.43) and Story Recall (.37). Factor five included three subtests: Phonological Processing (.80), Segmentation (.56), and Sound Blending (.56). Only Number Series produced a salient loading for factor six (.65).

The Sound Awareness subtest cross-loaded onto multiple factors: factor two (.26), factor five (.30), and factor six (.22). The Numbers Reversed subtest loaded weakly across factors two through six, with factor three holding its highest loading (.23).

Exploratory Factor Analysis II

A second EFA was conducted after removing the Numbers Reversed subtest due to weak loadings across factors two, three, four, five, and six. For this EFA, Bartlett's (1950) test of sphericity indicated once again that the variables are related enough to proceed with the EFA (x^2 = 20,580.35, *df* = 325, $p < .000$). The KMO test was used to examine the shared variance within the data; the degree of common variance was .876, which indicated fitness to proceed. Communalities are presented in Table 7.

Table 7

EFA II: Communalities (Numbers Reversed Removed)

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	.707	.753
Number Series (NS)	.566	.665
Verbal Attention (VA)	.558	.517
Letter-Pattern Matching (LPM)	.617	.560
Phonological Processing (PP)	.644	.825
Story Recall (SR)	.407	.352
Visualization (VIS)	.452	.463
General Information (GI)	.608	.611
Concept Formation (CF)	.495	.427
Number-Pattern Matching (NPM)	.586	.628
Nonword Repetition (NONR)	.471	.468
Visual-Auditory Learning (VAL)	.366	.407
Picture Recognition (PR)	.437	.384
Analysis-Synthesis (AS)	.534	.477
Object-Number Sequencing (ONS)	.546	.500
Pair Cancellation (PC)	.505	.527
Memory for Words (MW)	.553	.524
Picture Vocabulary (PV)	.673	.750
Oral Comprehension (OC)	.662	.587
Segmentation (SEG)	.455	.473
Rapid Picture Naming (RPN)	.446	.448
Sentence Repetition (SENR)	.566	.559
Understanding Directions (UD)	.493	.516
Sound Blending (SB)	.430	.423
Retrieval Fluency (RF)	.506	.433
Sound Awareness (SA)	.525	.466

Extraction Method: Principal Axis Factoring.

Iterated principal axis factoring with a Promax rotation was again used to extract factors. Observation of the scree plot and eigenvalues greater than one were additional methods used to confirm the number of factors (Kaiser, 1960). The scree plot again appeared to present one strong factor. Examination of the Kaiser criterion resulted in the retention of six factors with the first factor accounting for the largest proportion of the variance (35.05%); this was consistent with the visual scree. The second factor comprised 7.32% of the variance, the third accounted for 6.70%, the fourth accounted for 5.32%, the fifth accounted for 4.66%, and the sixth factor accounted for 4.25% of the variance. Together, the six factors explained 63.3% of the total variance. Thus, the second EFA produced six factors, as presented in Table 8 and Figure 2.

Table 8

EFA II: Total Variance Explained (Numbers Reversed Removed)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.114	35.052	35.052	8.659	33.305	33.305	5.981
2	1.903	7.320	42.372	1.485	5.713	39.018	6.789
3	1.742	6.699	49.071	1.295	4.982	44.000	5.207
4	1.383	5.318	54.389	.900	3.460	47.461	5.815
5	1.211	4.658	59.047	.768	2.955	50.416	4.804
6	1.106	4.254	63.302	.636	2.446	52.862	1.547

Extraction Method: Principal Axis Factoring.

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

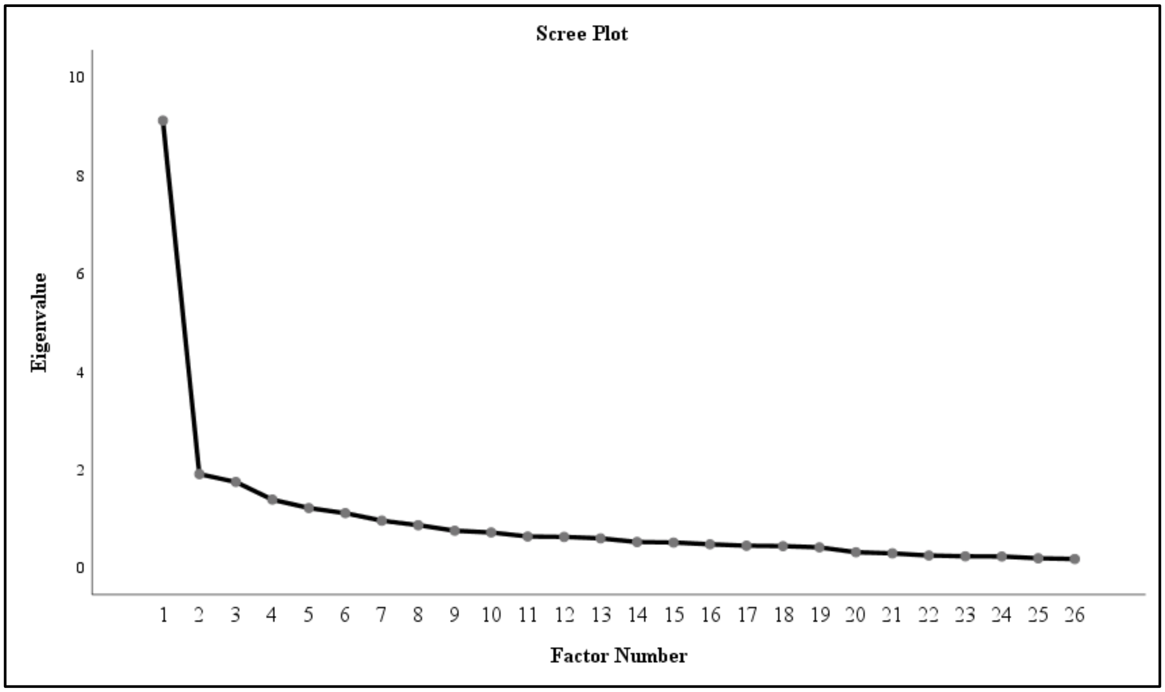

Figure 2. EFA II: Visual scree plot (Numbers Reversed removed).

The pattern matrix, which shows the weight of each variable relative to each factor, converged after seven iterations. The variables loaded onto six factors with only the Sound Awareness subtest showing cross-loadings onto factors 2 (.269), 5 (.303), and 6 (.230). See Table 9.

Table 9

EFA II: Pattern Matrix (Numbers Reversed Removed)

	Factor					
Subtest	1	2	3	4	5	6
PV	.920	.009	-.054	-.029	-.042	-.021
GI	.853	-.146	-.099	.072	.035	.034
OV	.801	.013	.003	-.025	.079	.136
OC	.607	.200	-.053	.126	-.073	-.004
SENR	.091	.828	-.097	-.177	.000	.009
NONR	-.082	.741	-.159	.087	.029	.005
VA	.004	.665	.065	-.146	.101	.132

	Factor					
Subtest	1	2	3	4	5	6
UD	.027	.566	-.030	.265	-.172	.159
MW	-.134	.514	-.062	.105	.381	-.159
ONS	-.079	.408	.259	.179	.088	-.080
NPM	-.129	-.053	.815	.034	-.051	.225
PC	-.050	-.061	.795	-.036	-.008	.023
LPM	-.054	-.131	.782	.051	.002	.134
RF	.248	.004	.410	-.017	.178	-.228
RPN	.157	.271	.381	.035	-.106	-.265
VIS	-.007	.038	-.023	.639	.035	.053
VAL	-.014	-.117	-.095	.618	.241	-.026
PR	.024	.068	.131	.556	-.131	-.211
AS	.052	-.183	.205	.556	.125	.076
CF	.079	.053	.023	.439	.129	.118
STOR	.144	.182	.019	.369	-.106	.112
PP	.114	.094	.139	-.206	.808	.058
SEG	-.113	.061	-.044	.228	.560	.069
SB	.031	-.129	-.157	.386	.535	-.117
SA	.020	**.269**	.002	.104	**.303**	**.230**
NS	.103	.094	.284	.030	.009	.609

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

The first factor included four subtests with factor loadings greater than .35: Picture Vocabulary (.92), General Information (.85), Oral Vocabulary (.80), and Oral Comprehension (.61). The second factor included Sentence Repetition (.83), Nonword Repetition (.74), Verbal Attention (.67), Understanding Directions (.57), Memory for Words (.51), and Object-Number Sequencing (.41). Factor 3 included five subtests with salient factor loadings: Number-Pattern Matching (.82), Pair Cancellation (.80), Letter-Pattern Matching (.78), Retrieval Fluency (.41) and Rapid Picture Naming (.38). The fourth factor was comprised of Visualization (.64), Visual-Auditory Learning (.62), Picture Recognition (.56), Analysis-Synthesis (.56), Concept Formation (.44) and Story Recall (.37). Factor 5 included three subtests: Phonological Processing (.81), Segmentation (.56), and Sound Blending (.54). Only Number Series produced a salient loading for Factor 6 (.61).

Exploratory Factor Analysis III

Because all variables in the second EFA converged into six factors with the exception of Sound Awareness, which cross-loaded onto multiple factors, a third EFA was conducted without the Numbers Reversed or Sound Awareness subtests. Bartlett's (1950) test of sphericity showed that the variables are not random in nature and that the third EFA could proceed (x^2 = 19,423.56, df = 300, p < .000). The KMO test was used to examine the shared variance within the data; the degree of common variance was found to be appropriate at .873, which indicated suitability to conduct the EFA. Table 10 provides a list of the communalities within the dataset. This showed that the variables and factors are related and continue to be devoid of outliers, thus an EFA may proceed.

Table 10

EFA III: Communalities (Numbers Reversed and Sound Awareness Removed)

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	.703	.766
Number Series (NS)	.538	.619
Verbal Attention (VA)	.548	.537
Letter-Pattern Matching (LPM)	.614	.562
Phonological Processing (PP)	.628	.757

Subtest (Abbreviation)	Initial	Extraction
Story Recall (STOR)	.394	.350
Visualization (VIS)	.447	.464
General Information (GI)	.604	.611
Concept Formation (CF)	.494	.440
Number-Pattern Matching (NPM)	.585	.628
Nonword Repetition (NONR)	.465	.458
Visual-Auditory Learning (VAL)	.365	.404
Picture Recognition (PR)	.433	.454
Analysis-Synthesis (AS)	.521	.489
Object-Number Sequencing (ONS)	.531	.502
Pair Cancellation (PC)	.505	.531
Memory for Words (MW)	.551	.542
Picture Vocabulary (PV)	.672	.748
Oral Comprehension (OC)	.661	.593
Segmentation (SEG)	.452	.461
Rapid Picture Naming (RPN)	.446	.454
Sentence Repetition (SENR)	.560	.546
Understanding Directions (UD)	.493	.513
Sound Blending (SB)	.400	.415
Retrieval Fluency (RF)	.506	.419

Extraction Method: Principal Axis Factoring.

For EFA III, the scree plot again appeared to present one strong factor. Examination of the Kaiser criterion resulted in the retention of six factors with the first factor accounting for the largest proportion of the variance (34.91%); see Table 11. This was consistent with the visual scree. The second factor comprised 7.59% of the variance, the third accounted for 6.86%, the fourth accounted for 5.49%, the fifth accounted for 4.76%, and the sixth factor accounted for 4.38% of the variance. Together, the six factors explained 64% of the total variance. See Table 11 and Figure 3. The pattern matrix converged after 13 iterations, and the variables loaded onto six factors. See Table 12.

Table 11

EFA III: Total Variance Explained (Numbers Reversed and Sound Awareness Removed)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	8.728	34.914	34.914	8.276	33.104	33.104	5.873
2	1.898	7.593	42.507	1.485	5.941	39.045	6.511
3	1.715	6.862	49.369	1.257	5.026	44.071	5.274
4	1.372	5.486	54.855	.897	3.588	47.659	5.516
5	1.190	4.760	59.615	.742	2.967	50.627	4.097
6	1.096	4.384	64.000	.606	2.423	53.050	.837

Extraction Method: Principal Axis Factoring.

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

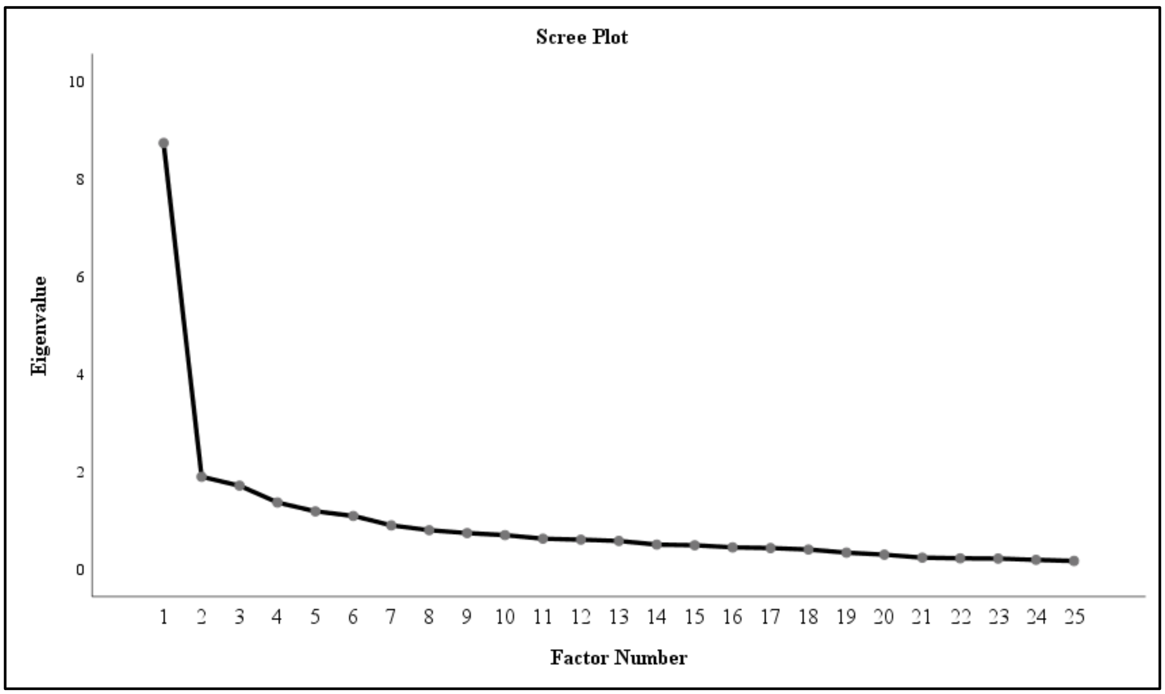

Figure 3. EFA III: Visual scree plot (Numbers Reversed and Sound Awareness removed).

Table 12

EFA III: Pattern Matrix (Numbers Reversed and Sound Awareness Removed)

Subtest	Factor 1	2	3	4	5	6
PV	.921	.006	-.060	-.037	-.034	-.039
GI	.859	-.144	-.097	.069	.036	.022
OV	.812	.026	.004	-.019	.074	.137
OC	.613	.187	-.055	.137	-.071	-.072
SENR	.099	.807	-.102	-.150	-.007	-.023
NR	-.074	.726	-.164	.102	.036	-.022
VA	.003	.698	.060	-.137	.088	.147
UD	.030	.565	-.028	.295	-.165	.084
MW	-.145	.540	-.089	.075	.406	-.097
ONS	-.090	.423	.243	.151	.118	-.075
NPM	-.126	-.044	.843	.045	-.059	.132
PC	-.047	-.069	.815	-.055	-.004	-.047
LPM	-.049	-.133	.810	.050	-.003	.049
RF	.238	.007	.392	-.059	.195	-.222
RPN	.151	.246	.368	-.008	-.074	-.323
VIS	-.007	.038	-.021	.649	.055	-.012
VAL	-.014	-.109	-.096	.609	.260	-.043
PR	.016	.025	.127	.572	-.109	-.353
AS	.038	-.149	.198	.543	.157	.085
CF	.076	.079	.019	.441	.146	.122
STOR	.144	.188	.021	.380	-.095	.057
PP	.128	.126	.150	-.156	.712	.084
SEG	-.100	.078	-.032	.247	.534	.072
SB	.043	-.129	-.153	.371	.531	-.107
NS	.116	.132	.320	.111	-.038	.510

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

The first factor included four subtests with factor loadings greater than .35: Picture Vocabulary (.92), General Information (.86), Oral Vocabulary (.81), and Oral

Comprehension (.61). The second factor included Sentence Repetition (.81), Nonword Repetition (.73), Verbal Attention (.70), Understanding Directions (.57), Memory for Words (.54), and Object-Number Sequencing (.42). Factor three included five subtests with salient factor loadings: Number-Pattern Matching (.84), Pair Cancellation (.82), Letter-Pattern Matching (.81), Retrieval Fluency (.39) and Rapid Picture Naming (.37). The fourth factor was comprised of Visualization (.65), Visual-Auditory Learning (.61), Picture Recognition (.57), Analysis-Synthesis (.54), Concept Formation (.44) and Story Recall (.38). Factor 5 included three subtests: Phonological Processing (.71), Segmentation (.53), and Sound Blending (.53). Again, only Number Series produced a salient loading for Factor 6 (.51).

Exploratory Factor Analysis IV

Though the third EFA appeared to clearly establish six factors, the sixth factor only contained the Number Series subtest. According to Raubenheimer (2004), a factor should be comprised of at least three items in order to maximize validity. Therefore, a fourth and final EFA was conducted with the removal of Numbers Reversed, Sound Awareness, and Number Series. Bartlett's (1950) test of sphericity showed that the variables are not random in nature and that the fourth EFA could proceed (x^2 = 18220.54, df = 276, p < .000). The KMO test was used to examine the shared variance within the data; the degree of common variance was found to be appropriate at .872, which indicated suitability to conduct the EFA. Table 13 provides a list of the communalities within the dataset. This showed that the variables and factors are related and continue to be devoid of outliers, thus an EFA may proceed.

Table 13

EFA IV: Communalities (Numbers Reversed, Sound Awareness Removed, and Number Series Removed)

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	.700	.742
Verbal Attention (VA)	.541	.499
Letter-Pattern Matching (LPM)	.606	.575
Phonological Processing (PP)	.598	.762
Story Recall (STOR)	.386	.340
Visualization (VIS)	.442	.463
General Information (GI)	.602	.620
Concept Formation (CF)	.480	.421
Number-Pattern Matching (NPM)	.556	.573
Nonword Repetition (NONR)	.465	.455
Visual-Auditory Learning (VAL)	.356	.410
Picture Recognition (PR)	.426	.336
Analysis-Synthesis (AS)	.512	.477
Object-Number Sequencing (ONS)	.529	.505
Pair Cancellation (PC)	.489	.571
Memory for Words (MW)	.543	.524
Picture Vocabulary (PV)	.672	.749
Oral Comprehension (OC)	.661	.592
Segmentation (SEG)	.452	.467
Rapid Picture Naming (RPN)	.442	.356
Sentence Repetition (SENR)	.556	.552
Understanding Directions (UD)	.458	.480
Sound Blending (SB)	.389	.382
Retrieval Fluency (RF)	.506	.344

Extraction Method: Principal Axis Factoring.

For EFA IV, the scree plot depicted one strong factor. Examination of the Kaiser criterion resulted in the retention of five factors with the first factor accounting for the largest proportion of the variance (35.01%); this was consistent with the visual scree. The second factor comprised 7.77% of the variance, the third accounted for 7.15%, the fourth accounted for 5.71%, and the fifth accounted for 4.87% of the variance. Together, the five factors explained 60.5% of the total variance. See Table 14 and Figure 4.

Table 14

EFA IV: Total Variance Explained (Numbers Reversed, Sound Awareness, and Number Series Removed)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	8.402	35.010	35.010	7.930	33.040	33.040	5.525
2	1.865	7.770	42.780	1.445	6.022	39.062	6.146
3	1.715	7.148	49.928	1.244	5.185	44.247	4.578
4	1.370	5.709	55.637	.880	3.666	47.913	5.251
5	1.168	4.866	60.504	.696	2.900	50.813	4.197

Extraction Method: Principal Axis Factoring.

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

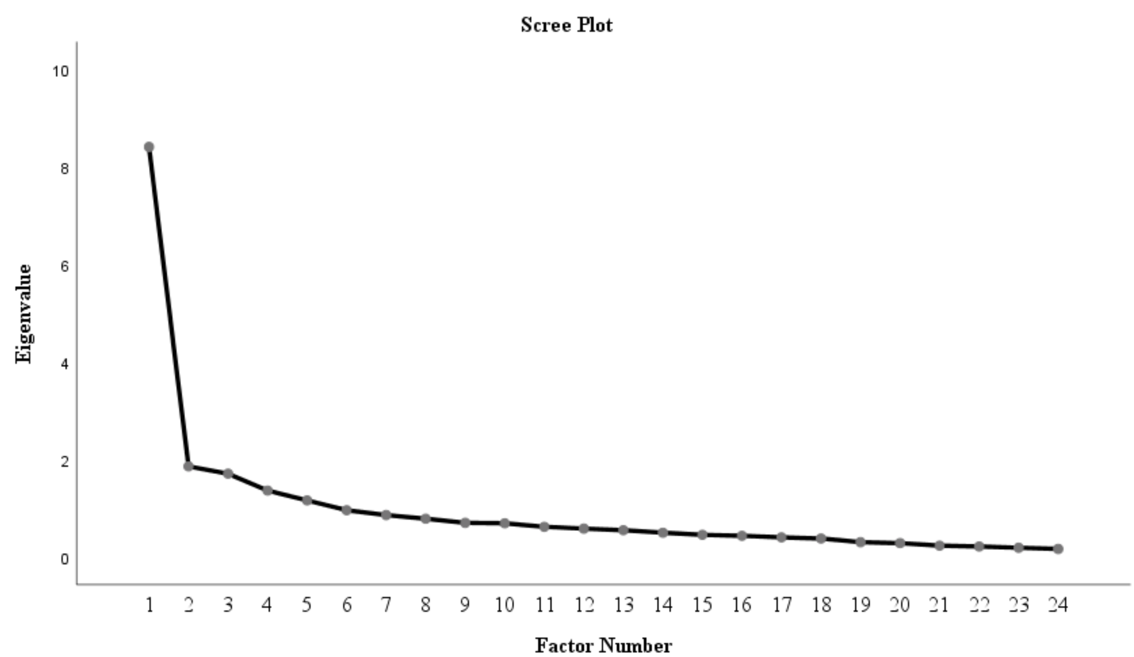

Figure 4. EFA IV: Visual scree plot (Numbers Reversed, Sound Awareness, and Number Series removed).

Table 15

EFA IV: Pattern Matrix (Numbers Reversed, Sound Awareness, and Number Series Removed)

	Factor				
Subtest	1	2	3	4	5
PV	.901	.022	-.040	-.027	-.051
GI	.850	-.153	-.065	.062	.051
OV	.789	.010	.034	-.013	.122
OC	.595	.208	-.049	.147	-.093
SENR	.092	.796	-.080	-.138	-.011
NR	-.076	.698	-.140	.110	.039
VA	.012	.638	.071	-.105	.135
MW	-.141	.554	-.084	.061	.369
UD	.031	.512	-.016	.319	-.120
ONS	-.084	.435	.240	.162	.095
PC	-.026	-.043	.812	-.042	-.022
LPM	-.025	-.122	.791	.063	.016
NPM	-.096	-.035	.781	.077	-.015
RF	.228	.103	.339	-.049	.104
RPN	.150	**.309**	**.323**	.019	-.165
VIS	-.007	.028	-.014	.638	.070
VAL	-.012	-.113	-.083	.582	.265
AS	.043	-.152	.190	.535	.191
PR	.021	.096	.109	.517	-.176
CF	.077	.049	.029	.436	.195
STOR	.138	.167	.026	.397	-.070
PP	.138	.151	.140	-.215	.747
SEG	-.091	.072	-.027	.197	.578
SB	.044	-.092	-.137	.309	.497

The first factor included four subtests with factor loadings greater than .30: Picture Vocabulary (.90), General Information (.85), Oral Vocabulary (.79), and Oral Comprehension (.56). The second factor included Sentence Repetition (.80), Nonword

Repetition (.70), Verbal Attention (.64), Memory for Words (.55), Understanding Directions (.51) and Object-Number Sequencing (.44). Factor 3 included five subtests with salient factor loadings: Pair Cancellation (.81), Letter-Pattern Matching (.79), Number-Pattern Matching (.78), Retrieval Fluency (.34) and Rapid Picture Naming (.32). The fourth factor was comprised of Visualization (.64), Visual-Auditory Learning (.58), Analysis-Synthesis (.54), Picture Recognition (.52), Concept Formation (.44) and Story Recall (.40). Factor 5 included three subtests: Phonological Processing (.75), Segmentation (.58), and Sound Blending (.50). The pattern matrix converged in nine iterations, and the variables loaded onto five factors. See Table 15.

The fourth EFA appeared to present the clearest solution thus far. However, the Rapid Picture Naming subtest, which loaded onto factor four (.323) also loaded onto the third factor (.309). Therefore, a fifth EFA was conducted with the removal of Numbers Reversed, Sound Awareness, Number Series, and Rapid Picture Naming. Bartlett's (1950) test of sphericity showed that the variables are not random in nature and that the fourth EFA could proceed (x^2 = 17312.58, *df* = 253, *p* < .001).

Exploratory Factor Analysis V

The KMO test was used to examine the shared variance within the data; the degree of common variance was found to be appropriate at .871, which indicated suitability to conduct the EFA. Table 16 provides a list of the communalities within the dataset. This showed that the variables and factors are related and continue to be devoid of outliers, thus an EFA may proceed.

For EFA V, the scree plot depicted one strong factor. Examination of the Kaiser criterion resulted in the retention of five factors with the first factor accounting for the largest proportion of the variance (35.37%); this was consistent with the visual scree. See Table 16 and Figure 5. The second factor comprised 8.01% of the variance, the third accounted for 7.18%, the fourth accounted for 5.94%, and the fifth accounted for 5.03% of the variance. Together, the five factors explained 61.53% of the total variance. See Table 17 and Figure 5. The pattern matrix converged in nine iterations, and the variables loaded onto five factors. See Table 18.

Table 16

EFA V: Communalities (Numbers Reversed, Sound Awareness, Number Series, and Rapid Picture Naming Removed)

Subtest (Abbreviation)	Initial	Extraction
Oral Vocabulary (OV)	.697	.751
Verbal Attention (VA)	.533	.501
Letter-Pattern Matching (LPM)	.605	.598
Phonological Processing (PP)	.594	.776
Story Recall (STOR)	.372	.352
Visualization (VIS)	.437	.462
General Information (GI)	.601	.621
Concept Formation (CF)	.466	.420
Number-Pattern Matching (NPM)	.555	.591
Nonword Repetition (NONR)	.463	.458
Visual-Auditory Learning (VAL)	.356	.411

Subtest (Abbreviation)	Initial	Extraction
Picture Recognition (PR)	.396	.314
Analysis-Synthesis (AS)	.502	.478
Object-Number Sequencing (ONS)	.529	.501
Pair Cancellation (PC)	.471	.552
Memory for Words (MW)	.541	.522
Picture Vocabulary (PV)	.658	.738
Oral Comprehension (OC)	.659	.585
Segmentation (SEG)	.452	.460
Sentence Repetition (SENR)	.554	.578
Understanding Directions (UD)	.451	.476
Sound Blending (SB)	.388	.391
Retrieval Fluency (RF)	.474	.324

Extraction Method: Principal Axis Factoring.

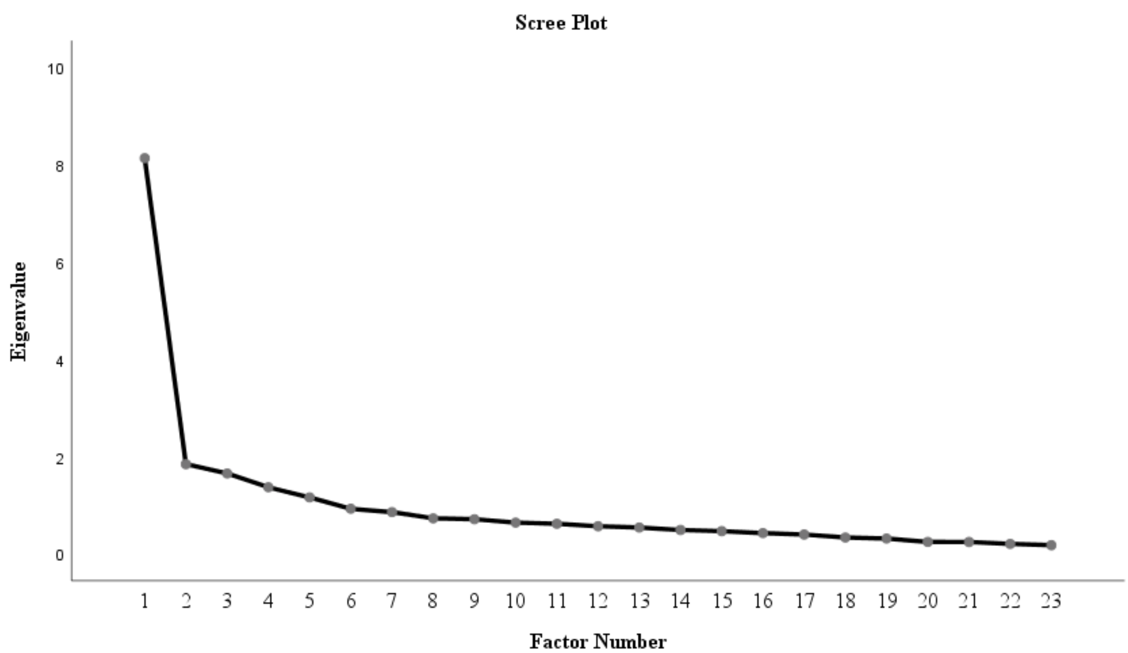

Figure 5. EFA V: Visual scree plot (Numbers Reversed, Sound Awareness, Number Series, and Rapid Picture Naming removed).

Table 17

EFA V: Total Variance Explained (Numbers Reversed, Sound Awareness, Number Series, and Rapid Picture Naming Removed)

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings [a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	8.136	35.374	35.374	7.671	33.350	33.350	5.332
2	1.841	8.006	43.380	1.439	6.255	39.606	5.821
3	1.650	7.176	50.556	1.190	5.175	44.780	5.137
4	1.367	5.942	56.498	.875	3.804	48.585	4.219
5	1.157	5.032	61.530	.686	2.982	51.566	4.548

Extraction Method: Principal Axis Factoring.

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

Table 18

EFA V: Pattern Matrix (Numbers Reversed, Sound Awareness, Number Series, and Rapid Picture Naming Removed)

	Factor				
Subtest	1	2	3	4	5
PV	.894	.018	-.021	-.041	-.047
GI	.842	-.135	.060	-.059	.033
OV	.788	.032	-.017	.046	.091
OC	.600	.195	.150	-.043	-.088
SENR	.104	.814	-.142	-.052	-.051
NONR	-.067	.688	.116	-.122	.020
VA	.023	.633	-.099	.085	.111
MW	-.138	.520	.066	-.086	.382
UD	.044	.501	.325	-.003	-.133
ONS	-.069	.422	.170	.241	.088
VIS	-.007	.038	.634	-.004	.055
VAL	-.022	-.124	.580	-.092	.281
AS	.042	-.141	.533	.188	.180
PR	.035	.069	.513	.098	-.144

	Factor				
Subtest	1	2	3	4	5
CF	.074	.037	.441	.017	.204
SR	.145	.185	.400	.049	-.108
LPM	-.013	-.087	.060	.799	-.022
NPM	-.082	-.006	.078	.785	-.050
PC	-.010	-.024	-.033	.785	-.035
RF	.238	.058	-.035	.301	.153
PP	.128	.126	-.228	.121	.784
SEG	-.103	.059	.192	-.038	.590
SB	.030	-.120	.307	-.156	.535

The first factor included four subtests with factor loadings greater than .30: Picture Vocabulary (.89), General Information (.84), Oral Vocabulary (.79), and Oral Comprehension (.60). The second factor included Sentence Repetition (.81), Nonword Repetition (.69), Verbal Attention (.63), Memory for Words (.52), Understanding Directions (.50) and Object-Number Sequencing (.42).

Factor 3 included six subtests with salient factor loadings: Visualization (.63), Visual-Auditory Learning (.58), Analysis-Synthesis (.53), Picture Recognition (.51), Concept Formation (.44) and Story Recall (.40). The fourth factor was comprised of Letter-Pattern Matching (.80), Number-Pattern Matching (.79), Pair Cancellation (.79), and Retrieval Fluency (.30). Factor 5 included three subtests: Phonological Processing (.78), Segmentation (.59), and Sound Blending (.40).

Chapter Summary

This chapter provided the results of five separate iterated principal factor analyses. All were conducted using a Promax rotation. Five analyses were included in order to arrive at a clear five-factor solution since some subtests had weak loadings or cross-loadings across different factors for the first four EFAs. Additionally, this aided in ensuring factors contained more than one item (Number Series continuously constituted the sixth factor) and were thus interpretable.

The first EFA included 27 subtests, and results showed that a six-factor solution accounted for 62.2% of the variance. Two subtests — Numbers Reversed and Sound Awareness — did not load cleanly onto a single factor. Instead, Numbers Reversed demonstrated weak loadings (< .30) across all factors, while Sound Awareness cross-loaded onto Factors 2, 5, and 6 (all ≤ .30). As a result of these weak loadings, the second EFA was conducted without the Numbers Reversed subtest. With this removal, the model produced six factors which explained 63.3% of the variance. For the second EFA, one subtest — Sound Awareness — cross-loaded onto three factors. Therefore, a third EFA was conducted with the removal of both the Numbers Reversed and Sound Awareness subtests. This yielded another six-factor solution which accounted for 64% of the variance, and there were no cross-loadings. The fourth EFA was conducted with the removal of Numbers Reversed, Sound Blending, and Number Series. Results produced a five-factor solution which accounted for 60.5% of the variance. Because the Rapid Picture Naming Subtest cross-loaded onto Factors 2 (.31) and 3 (.32), the fifth and final EFA was conducted without this subtest in addition to the subtests which were previously removed. The fifth EFA yielded a five-factor solution which accounted for 61.53% of the variance. Throughout each factor analysis, the first factor consistently constituted the largest portion of the variance.

CHAPTER V

DISCUSSION

In 1986, intelligence researcher John Horn presented the results of an EFA based on the Woodcock-Johnson Psychoeducational Battery (Woodcock & Johnson, 1977) and proposed a marriage between the practice of intellectual assessment and G*f*-G*c* theory (Cattell, 1941). G*f*-G*c* theory eventually merged with the work of John Carroll to become CHC theory, a three-tiered model of intelligence comprised of broad and narrow abilities as well as the all-encompassing *g*, or general ability factor. Future renditions of the WJ tests incorporated this three-tiered model and purported to provide measures of CHC abilities.

As research has progressed, CHC theory has expanded to incorporate additional broad and narrow abilities, and developers of the most recent revision of the WJ, the WJ IV (Schrank et al., 2014a), claim to include these theoretical updates. Specifically, WJ IV COG (Schrank et al., 2014c), Tests of Achievement (WJ IV ACH; Schrank et al., 2014b), and WJ IV OL (Schrank, Mather, & McGrew, 2014) were reported to provide measures of nine broad ability factors: comprehension-knowledge (G*c*), fluid reasoning G*f*), short-term working memory (G*wm*), cognitive processing speed (G*s*), auditory processing (G*a*), long-term retrieval (G*lr*), visual processing (G*v*), quantitative reasoning (G*q*), and reading/writing ability (G*rw*). It should be noted, however, that this information was based on statistical analyses of all three WJ IV batteries together rather than assessing each battery individually. The current study aimed to assess the factor structure of the WJ IV COG and WJ IV OL among a school-aged group (9-13- years old) based on the correlation matrix provided in the WJ IV *Technical Manual* (McGrew et al., 2014).

The present study used EFA to examine the underlying factor structure of the WJ IV COG and WJ IV OL. Results pointed to the presence of five broad CHC factors instead of the current nine. This chapter provides explanation and implications of these results. Additionally, the chapter will include limitations of the current study and directions for future research.

Explanation of Findings

The WJ IV test publishers used a variety of analyses to identify the underlying factor structure of the entire WJ IV battery before arriving at a three-tiered, nine-factor solution purporting to measure CHC abilities. The methods used by McGrew et al. (2014) presented several opportunities for further research. For one, all three batteries were included in the norming process rather than norming each battery separately (WJ IV COG, WJ IV ACH, and WJ IV OL); given that practitioners typically do not administer the full test, information regarding individual batteries would likely be useful for test interpretation. Secondly, McGrew at el. utilized cluster analyses, multidimensional scaling, and confirmatory factor analyses to identify CHC factors. While these methods provide valuable information, they can also be more subjective in nature than EFAs, leaving room for bias among test publishers. Finally, McGrew et al. only presented information from the 9 to 13-year-old norming group as the basis of determining factor structure. This inhibits the generalizability of their factor solutions.

McGrew et al. (2014) used a three-stage model fitting procedure to identify the factor structure of the WJ IV: the first stage involved randomly generating a split sample among each age group. Half of the sample was used for model determination while the other half was reserved for cross-validation of the selected model. Stage two involved examining factor structure using cluster, exploratory principal components, and multidimensional scaling analyses, followed by converging the results of each procedure with theory and neuropsychological research. Results of the cluster analysis alluded to nine broad CHC factor clusters: G*wm*, G*a*, G*v*, G*f*, G*lr*, G*q*, G*c*, G*rw*, and G*s*, as well as an overarching *g* factor.

The exploratory PCA provided evidence for eight, nine, and 10-factor solutions. It should be noted that the test publishers reported using an orthogonal rotation versus the expected oblique rotation which is typically used with correlated data. McGrew et al. (2014) justified this by arguing that the varimax rotation was simpler and would not result in the loss of subtests. The eight-factor PCA included G*c*, G*s*, a blended G*v*/G*f*/G*a* factor, G*rw*, G*lr* (with the removal of the Retrieval Fluency subtest), a blended G*q*/G*f* factor, G*wm*, and G*v*. The nine-factor solution included G*c*, G*s*, a blended G*v*/G*f* factor, G*rw*, G*lr*, a blended G*q*/G*f* factor, G*wm*, G*v*, and G*a*. In each solution, clear G*c*, G*s*, G*rw*, G*lr*, and G*wm* factors emerged; G*f* frequently appeared mixed in with other broad abilities. Additionally, several subtests loaded across multiple factors, and in the first two solutions, Numbers Reversed lacked significant loadings onto any factor. Results of the

MDS analysis provided evidence for seven broad ability factors: G*a*, G*c*, G*rw*, a blended G*q*/G*f* factor, G*wm*, G*f*, and G*v*. Test publishers noted that the Memory for Names subtest appeared to be an outlier and was not correlated with any other subtests.

After converging results of exploratory analyses, McGrew et al. (2014) identified three potential models of the underlying structure of the WJ IV. Limited information was provided in the test manual regarding the first model other than all 51 subtests loaded onto one factor — *g*. The second model provided a factor structure most aligned with current CHC theory in that it included a top-down model representing *g* and nine broad ability areas: G*c*, G*rw*, G*f*, G*s*, G*q*, G*v*, G*lr*, G*wm*, and G*a*. Although the test publishers believed the third model was more complex due to providing information on broad and narrow abilities, it was ultimately rejected since results contained inappropriate statistical outcomes such as Haywood cases and negative factor loadings. Additionally, this model did not include data from the preschool population and thus was deemed less generalizable.

In stage three, each model was cross-validated across age groups using various CFA model fit indices, including minimum discrepancy, degrees of freedom, adjusted goodness-of-fit, comparative fit index, the Tucker-Lewis non-normed fit index, parsimony adjustment, and the root mean squared error of approximation (RMSEA). The closer each of these indices is to 1.00, the better the fit of the model; an exception is the RMSEA in which a number closer to 0.00 is ideal. Results showed that while models two and three were useful for the school-age population and older, model two was also a good fit for the preschool age group. Additionally, model two was simpler and thus preferred according to the parsimony principal (Kline, 2011); therefore, model two was confirmed as the best representation of the structural validity of the WJ IV. Table 19 provides a visual representation of the chosen factor structure for the WJ IV COG and WJ IV OL tests.

Table 19

CHC Factors and Related Subtests According to the WJ IV Technical Manual

Comprehension-knowledge (G*c*)	Fluid Reasoning (G*f*)	Short-Term Working Memory (G*wm*)	Cognitive Processing Speed (G*s*)	Auditory Processing (G*a*)	Long-Term Retrieval (G*lr*)	Visual Processing (G*v*)
Oral Vocabulary	Number Series	Verbal Attention	Letter-Pattern Matching	Phonological Processing	Story Recall	Visualization
General Information	Concept Formation	Numbers Reversed	Pair Cancellation	Segmentation	Visual-Auditory Learning	Picture Recognition
Picture Vocabulary	Analysis-Synthesis	Object-Number Sequencing	Number Pattern Matching	Sound Blending	Rapid Picture Naming	
Oral Comprehension		Sentence Repetition		Nonword Repetition	Retrieval Fluency	
		Understanding Directions		Sound Awareness		
		Memory for Words				

Comprehension-knowledge (G*c*) consists of one's acquired knowledge, the ability to communicate knowledge, and the ability to reason using previously learned experiences or procedures. There are four subtests presumed to provide a measure of G*c*: Oral Vocabulary, General Information, Picture Vocabulary, and Oral Comprehension. Fluid reasoning (G*f*) is the ability to reason, form concepts, and solve problems using unfamiliar information or new procedures. Number Series, Concept Formation, and Analysis-Synthesis are believed to measure G*f*. Short-term working memory (G*wm*) includes the ability to remember and use auditory information within a short time period. This factor includes the subtests Verbal Attention, Numbers Reversed, Object-Number Sequencing, Sentence Repetition, Understanding Directions, and Memory for Words. Cognitive processing speed (G*s*) is the ability to quickly and accurately perform automatic mental tasks. This factor includes the subtests Letter-Pattern Matching, Pair Cancellation, and Number-Pattern Matching. Auditory Processing (G*a*) involves the ability to analyze, synthesize, and discriminate sounds, including sounds presented under distorted conditions. Phonological Processing, Segmentation, Sound Blending, Nonword Repetition, and Sound Awareness are believed to represent G*a*. Long-term retrieval (G*lr*) refers to the ability to store information and retrieve it fluently at a later point. Subtests included under G*lr* include Story Recall, Visual-Auditory Learning, Rapid Picture Naming, and Retrieval Fluency. Finally, visual processing (G*v*) includes the ability to perceive, analyze, synthesize, and think with visual patterns, and is assessed by the Visualization and Picture Recognition subtests (McGrew et al., 2014; Schrank & Wendling, 2018). It should also be noted that for three subtests — Number-Pattern Matching, Memory for Words, and Sound Awareness — no factor loadings were reported. Additionally, Rapid Picture Naming cross-loaded onto both G*lr* and G*s*, and Nonword Repetition loaded higher (0.59) on the G*wm* factor than the G*a* factor (0.18). The test publishers justify placing the subtest under G*a* by noting that G*a* may be more complex than previously believed and may contain working memory elements.

Though the WJ IV has been available since 2014, relatively little research has been conducted regarding the structural validity of the test. Per McGrew et al. (2014), the

WJ IV COG yields seven factors across all ages. A study by Dombrowski et al. (2017) challenged this by examining the factor structure of the WJ IV COG using EFA and hierarchical factor analysis (HFA) for two school age groups (9–13 and 14–19). Data for this study were derived from the correlation matrices provided in the WJ IV *Technical Manual*. For the first order EFA, researchers extracted four factors for the 9 to 13-year-old age group (G*wm*, G*s*, G*c*, and a factor believed to measure perceptual reasoning) and four for the 14 to 19-year-old age group (G*s*, G*wm*, perceptual reasoning, and crystallized ability). Additionally, the following subtests did not appear to load clearly onto any factor: Story Recall, Nonword Repetition, Phonological Processing, Concept Formation, Number Series, and Numbers Reversed. After performing the higher-order EFA, Dombrowski et al. (2017) reported that 61.0% of the common variance appeared to be accounted for by a single factor which likely represents *g*. The factor structure as determined by Dombrowski et al. (2017) using EFA for the 9 to 13-year-old age group is presented in Table 20.

Table 20

Proposed Factor Structure of WJ IV COG for the 9 to 13- Year-Old Age Group by Dombrowski et al. (2017)

Comprehension-knowledge (G*c*)	Short-Term Working Memory (G*wm*)	Cognitive Processing Speed (G*s*)	Perceptual Reasoning
Oral Vocabulary	Verbal Attention	Letter-Pattern Matching	Visualization
General Information	Numbers Reversed	Pair Cancellation	Picture Recognition
Picture Vocabulary	Object-Number Sequencing	Number Pattern Matching	

Comprehension-knowledge (G*c*)	Short-Term Working Memory (G*wm*)	Cognitive Processing Speed (G*s*)	Perceptual Reasoning
Oral Comprehension	Sentence Repetition		
	Understanding Directions		
	Memory for Words		

An additional study by Dombrowski, McGill, and Canivez (2018) examined the factor structure of the entire WJ IV battery using EFA and HFA. While McGrew et al. (2014) promoted a nine-factor solution, Dombrowski et al.(2018) found evidence for seven factors among both school aged groups (9–13 and 14–19): G*c*, G*rw*, G*s*, G*wm*, G*a*, G*q*/G*f*, and G*v*. Additionally, the first factor — *g* — accounted for 38.94% of the variance, while the second factor only accounted 6.78% of variance. Several subtests cross-loaded onto multiple factors or loaded onto a factor inconsistent with what the subtest was purported to measure, though most subtests were generally aligned according to the current model in the *Technical Manual* (McGrew et al., 2014).

It should be noted that neither Numbers Reversed or Story Recall loaded saliently onto any factors. Dombrowski et al. (2018) concluded that the WJ IV may be over-factored, and that the present factors may be more complex than originally reported in the *Technical Manual*. Table 21 provides a visualization of the current factor structure for the 9 to 13-year-old age group according to Dombrowski et al. (2018). A study by Spurgin (2018) also examined the structural validity of the WJ IV COG and WJ IV OL using EFA, though this study focused on the 14 to 19-year-old age group. Spurgin’s final model accounted for 63.05% of the total variance; however, this was after the removal of the Numbers Reversed, Retrieval Fluency, and Rapid Picture Naming subtests due to low factor loadings.

Table 21

Proposed Factor Structure of WJ IV for the 9 to 13-Year-Old Age Group by Dombrowski et al. (2018)

Factor I (G*rw*)	Factor II (G*c*)	Factor III (G*s*)	Factor IV (G*wm*)	Factor V (G*f*/G*q*)	Factor VI (G*a*)	Factor VII (G*v*)
Writing Samples	Picture Vocabulary	Pair Cancellation	Nonword Repetition	Number Series	Sound Blending	Picture Recognition
Reading Recall	General Information	Letter-Pattern Matching	Sentence Repetition	Applied Problems	Segmentation	Visualization
Letter-Word Identification	Social Studies	Number-Pattern Matching	Verbal Attention	Calculation	Phonological Processing	
Passage Comprehension	Humanities	Word Reading Fluency	Memory for Words	Analysis-Synthesis	Spelling of Sounds	
Word Attack	Oral Vocabulary	Math Fact Fluency	Object-Number Sequencing	Number Matrices	Visual-Auditory Learning	
Oral Reading	Oral Comprehension	Sentence Reading Fluency	Understanding Directions	Concept Formation		
Spelling	Science	Rapid Picture Naming		Story Recall		
Editing	Reading Vocabulary					
Sentence Writing Fluency						

The EFA found evidence for five CHC factors believed to provide measures of G*c*, G*wm*, G*a*, attention, and cognitive reasoning. The attention factor was comprised of the WJ IV G*s* and G*f* factors. The cognitive reasoning factor included the WJ IV G*v*, G*lr*, and G*f* factors. Spurgin's hypothesized factor structure with accompanying subtests is presented in Table 22. The present study conducted an EFA using iterated principal axis factoring with Promax rotation, and was based on the 9 to 13-year-old standardization sample reported in a correlation matrix in the WJ IV *Technical Manual* (McGrew et al., 2014). Five separate EFAs were conducted in order to obtain a clear factor structure.It should be noted that examination of the visual scree and eigenvalues for each EFA appeared to support a single strong factor which accounted for most of the variance (ranging from 33.35% to 35.05%) and with eigenvalues over 8.00 while the rest of the factors had eigenvalues generally between 1.00 and 2.00. This is consistent with the works of Dombrowski et al. (2017; 2018) and Spurgin (2018); these studies alluded to the single-factor as a likely representation of *g*.

Table 22

Proposed Factor Structure for the WJ IV COG and WJ IV OL for the 14 to 19-Year-Old Age Group by Spurgin (2018)

Factor I (G*c*)	Factor II (G*wm*)	Factor III (G*a*)	Factor IV (Attention)	Factor V (Cognitive Reasoning)
Oral Vocabulary	Verbal Attention	Phonological Processing	Letter-Pattern Matching	Visualization
General Information	Memory for Words	Segmentation	Pair Cancellation	Picture Recognition
Picture Vocabulary	Object-Number Sequencing	Sound Blending	Number-Pattern Matching	Story Recall
Oral Comprehension	Sentence Repetition	Sound Awareness	Number Series	Analysis-Synthesis
	Understanding Directions	Visual-Auditory Learning		
	Nonword Repetition	Concept Formation		

The first EFA included all subtests; however, the Sound Awareness and Numbers Reversed subtests cross-loaded onto multiple factors. Because Numbers Reversed did not have factor loadings greater than 0.35 on any factors, this variable was excluded from the second EFA. The third EFA excluded Numbers Reversed and Sound Awareness, since Sound Awareness continued to load across multiple factors. Results yielded a fairly converged six-factor model; however, the sixth factor appeared to only include one item (Number Series) and was thus deemed uninterpretable. The fourth EFA excluded Numbers Reversed, Sound Awareness, and Number Series. In this five-factor model, Rapid Picture Naming cross-loaded onto factors two and three. Therefore, a fifth and final EFA was conducted with the omission of Numbers Reversed, Sound Awareness, Number Series, and Rapid Picture Naming. The result was a five-factor solution which accounted for 61.53% of common variance, though this figure excludes any sampling or modeling errors. This solution contradicts the test publishers' current seven-factor structure for the WJ IV COG and WJ IV OL, and certain subtests appear to be associated with different factors than originally reported. The hypothesized factor structure as well as possible CHC classifications as determined by the present study is shown in Table 23.

Table 23

Hypothesized Factor Structure as Found by the Present Study, Along with Subtest Loadings

Factor I (G*c*)	Factor II (G*wm*)	Factor III (Perceptual Reasoning)	Factor IV (G*s*)	Factor V (G*a*)
Picture Vocabulary	Sentence Repetition	Visualization	Letter-Pattern Matching	Phonological Processing
General Information	Nonword Repetition	Visual-Auditory Learning	Number-Pattern Matching	Segmentation
Oral Vocabulary	Verbal Attention	Analysis-Synthesis	Pair Cancellation	Sound Blending
Oral Comprehension	Memory for Words	Picture Recognition	Retrieval Fluency	
	Understanding Directions	Concept Formation		
	Object-Number Sequencing	Story Recall		

Table 23 presents the five-factor solution obtained from the exploratory analyses, as well as the subtests which loaded onto each factor and the hypothesized broad CHC abilities each factor represents. The first factor likely represents comprehension-knowledge (G*c*) and includes the Picture Vocabulary, General Information, Oral Vocabulary, and Oral Comprehension subtests. This is consistent with the G*c* factor presented in the *Technical Manual* (McGrew et al., 2014), though subtest loadings are somewhat different (i.e., Picture Vocabulary had the strongest relationship to the factor for the current study whereas McGrew et al. presented Oral Vocabulary as having the strongest relationship).

The second factor is believed to represent short-term working memory (G*wm*) and includes the following subtests: Sentence Repetition, Nonword Repetition, Verbal Attention, Memory for Words, Understanding Directions, and Object-Number Sequencing. While most of these variables were also consistent with the *Technical Manual*, Nonword Repetition was determined to provide a measure of auditory processing (G*a*) rather than G*wm*. Additionally, Numbers Reversed, which was removed after the first analysis due to weak loadings across multiple factors, was included under G*wm* in the *Technical Manual*.

Factor three contained the Visualization, Visual-Auditory Learning, Analysis-Synthesis, Picture Recognition, and Concept Formation subtests. This factor combined visual processing (G*v*) and fluid reasoning (G*f*), and it is hypothesized to represent perceptual reasoning. Dombrowski et al. (2017, 2018) alluded to a perceptual reasoning factor in their work, and Wechsler (2003) also defined a Perceptual Reasoning Index comprised of fluid reasoning and visual processing.

The fourth factor was hypothesized to represent cognitive processing speed (G*s*) comprised of the Letter-Pattern Matching, Number-Pattern Matching, Pair Cancellation, and Retrieval Fluency subtests. This mostly corresponds to *Gs* as presented in the *Technical Manual*, with the exception of the Retrieval Fluency subtest, which was included with long-term retrieval (G*lr*) according to McGrew et al. (2014). Given that the task of Retrieval Fluency involves quickly recalling and naming items within a given category, it is unsurprising that the subtest would include components of G*s* and G*lr*.

The final factor included the following subtests: Phonological Processing, Segmentation, and Sound Blending. This parallels the auditory processing (G*a*) factor as presented by McGrew et al. (2014) with the exception Nonword Repetition and Sound Awareness which were both omitted from the final factor structure of the current study. Therefore, this factor was hypothesized to also represent G*a*.

In all, four subtests were excluded from the final analysis: Numbers Reversed, Sound Awareness, Number Series, and Rapid Picture Naming. Though Numbers Reversed did not have strong loadings on any factor, it was purported to narrowly provide a measure of attentional control and working memory capacity. Sound Awareness narrowly measures phonetic coding; for the current study, this subtest loaded onto factors associated with G*wm* and G*a*. The Number Series subtest was the sole variable on a sixth, unnamed factor, though it also loaded (<.30) onto the factor associated with G*s*. A study by Cormier et al. (2017) acknowledged the Number Series subtest appeared have a unique contribution to the factor structure of the WJ IV. Cormier et al. posit the subtest is more cognitively complex than previously thought, as it requires understanding patterns, problem solving, and working memory. They further note that Number Series is a strong predictor of reading achievement. The Rapid Picture Naming subtest loaded saliently onto factors associated with G*wm* and G*s*; this is consistent with narrow abilities measured by the subtest (ability to retrieve names and speed of lexical access).

Of the remaining subtests, 15 aligned under the same broad factors presented in the *Technical Manual*, two subtests switched factors (Nonword Repetition loaded under G*wm* instead of G*a*; Retrieval Fluency loaded under G*s* instead of G*lr*), and six subtests appeared to merge into a blended G*v*/G*f* factor hypothesized to represent perceptual reasoning. The present study identified five broad ability factors as opposed to the seven CHC factors identified by WJ IV publishers. Table 24 illustrates the comparisons between WJ IV factor structure and the findings of this study.

Table 24

Traditional versus hypothesized CHC factor structure and subtest loadings of the WJ IV COG and WJ IV OL

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

Notably, the WJ IV authors understood the utility of designing tasks to require higher cognitive processing abilities, and several subtests require the combination of several cognitive skills to determine solutions. An appreciation of child and adolescent development is also pertinent here, as there are significant differences in cognitive and language skills between 9-year-old children and 13-year-old individuals. Merging results of participants from ages 9 to 13 years old likely obscured specific age-based normative data, making interpretation of scores more challenging for practitioners.

Results of this study mirrored the outcomes reported by Dombrowski et al. (2017, 2018) and Spurgin (2018) in that each study yielded fewer than the number of factors reported in the *Technical Manual*. Additionally, each study identified mixed broad ability factors. None of these studies clearly delineated a G*f* factor, instead finding that G*f* was a component of other broad abilities (e.g., G*v*). Each study also identified a strong single factor which likely represents *g*, indicating that the WJ IV mainly provides a measure of general intelligence and is possibly over-factored.

Though the current study shared several similarities with the other EFA-based analyses of the WJ IV, there are a few differences worth noting. First, the statistical procedures employed by Dombrowski et al. (2017, 2018) differed from the present study; each of the studies by Dombrowski et al. (2017) included higher-order factor analyses using the Schmid-Leiman procedure which enabled them to parse higher-order factors from first-order factors. Second, these studies focused on either the full WJ IV battery or solely on the WJ IV COG, whereas the current study omitted the WJ IV ACH. Third, several subtests were removed from the current study in order to clarify the factor structure; Spurgin (2018) also removed subtests, though the omitted subtests were not comparable. Dombrowski et al. (2017, 2018) also noted that some subtests did not fit neatly into a singular factor, and McGrew et al. (2014) also acknowledged that running certain statistical analyses might result in sacrificing subtests for a clear structural validity. Lastly, this study focused on the 9 to 13-year-old standardization sample; research by Spurgin addressed the 14 to 19-year-old group. While results were broadly similar, differences between the four studies were to be expected.

Since Spurgin's (2018) research was conceptually and statistically similar to the current study, it is worth further examining possible rationales for differing results. In both studies, the first and second factors were identified as G*c* and G*wm*, respectively. The same subtests loaded under each factor, though not in the same order. For example, Oral Vocabulary had the strongest loading on Spurgin's first factor, while that position was filled by Picture Vocabulary in the present study. Spurgin identified G*a* as the third factor. While it contained three of the same subtests as G*a* in the present study, Spurgin's results included Visual-Auditory Learning and Concept Formation, whereas these subtests loaded under the Perceptual Reasoning factor in the present study. Spurgin posited that these subtests both require auditory processing and reasoning skills, thus explaining their associations with both G*a* and Perceptual Reasoning. Notably, G*a* was the last factor in the present study. According to Floyd, McGrew, Barry, Rafael, and Rogers (2009), G*a* appears to be more closely associated with general intelligence as age increases, which may help to explain why it accounted for more variance in Spurgin's study of the adolescent group. It may also follow that the auditory components of Visual-

Auditory Learning and Concept Formation (which incorporates verbal feedback throughout the test) play a more critical role in performance among adolescents, though more research is needed to examine this possibility. Furthermore, some researchers argue that cognitive abilities become more integrated in adolescence (Breit, Brunner, & Preckel, 2020; Tucker-Drob, 2009); this may at least partially explain how subtests typically associated with G*f* are also affiliated with G*a* among adolescents.

Spurgin's fourth factor was determined to represent Attention; the subtests generally aligned with the fourth factor of the present study, which was deemed G*s*. Spurgin's Attention factor included the Number Series subtest, which was not included in the final analysis of the present study. The G*s* factor also included Retrieval Fluency, which was removed from Spurgin's final analysis. Spurgin explained that each of these tasks require sustaining or dividing attention. Finally, Spurgin's fifth factor, Cognitive Reasoning, closely aligned with the present study's third factor, Perceptual Reasoning. Both appeared to represent a combination of G*f*, G*v*, and G*lr*. The primary difference is that the present study included Concept Formation here; this is more consistent with factorial classifications provided in the *Technical Manual*. Of course, these differences may have also been affected by the sequential removal of different subtests in each study. Table 25 highlights differences between the present study and Spurgin's findings.

Table 25

Hypothesized Factors and Subtests Found by Spurgin (2018) Versus the Current Study

Factor	Spurgin's Hypothesized Broad Ability Factor	Subtests	Current Hypothesized Broad Ability Factor	Subtests
I	G*c*	Oral Vocabulary General Information Picture Vocabulary Oral Comprehension	G*c*	Picture Vocabulary General Information Oral Vocabulary Oral Comprehension
II	G*wm*	Verbal Attention Memory for Words Object-Number Sequencing Sentence Repetition Understanding Directions Nonword Repetition	G*wm*	Sentence Repetition Nonword Repetition Verbal Attention Memory for Words Understanding Directions Object-Number Sequencing
III	G*a*	Phonological Processing Segmentation Sound Blending Sound Awareness Visual-Auditory Learning Concept Formation	Perceptual Reasoning	Visualization Visual-Auditory Learning Analysis-Synthesis Picture Recognition Concept Formation Story Recall
IV	Attention	Letter-Pattern Matching Pair Cancellation Number-Pattern Matching Number Series	G*s*	Letter-Pattern Matching Number-Pattern Matching Pair Cancellation Retrieval Fluency
V	Cognitive Reasoning	Visualization Picture Recognition Story Recall Analysis-Synthesis	G*a*	Phonological Processing Segmentation Sound Blending

Implications

Findings from the present study as well as the aforementioned studies imply that the WJ IV COG and WJ IV OL are likely over-factored. Though some broad abilities found by these studies were consistent with those reported in the *Technical Manual* (e.g., G*c* emerged clearly across studies), many factors were not clearly defined by subsequent research, and several subtests either cross-loaded or did not appear to load saliently onto any broad ability factors. This information may be useful for practitioners seeking to measure specific broad abilities, as some factor scores may need to be interpreted cautiously or interpreted in conjunction with another broad ability factor in order to have a more accurate understanding of a student's skillset. This research also informs practitioners that the WJ IV appears to primarily provide a measure of *g* and may best be interpreted as broadly as is possible. Practitioners should also be mindful that not all subtests are cleanly aligned with a single factor, and some (e.g., Numbers Reversed) may prove more useful in providing qualitative information rather than being used to quantify a student's abilities.

Though it is one of the premiere instruments used by school psychologists to assess student abilities, the WJ IV is not infallible, and additional means of assessment may be necessary to complement the measure and confirm a student's level of functioning. This of course is not limited to the WJ, but to all forms of intellectual assessment, which is why multiple data sources are typically required to make accurate diagnoses and recommendations. Knowing the limitations of an instrument is imperative for practitioners to make informed decisions. It is equally as important to appreciate that theories of intelligence are ever-changing. Therefore, finding ways to measure increasingly complex constructs is an incredibly challenging and often imperfect task.

Methodological Concerns, Limitations, and Conclusions

Several methodological concerns regarding the design of the current study should be noted, as they may have implications for the results and discussion. First, there were a number of decisions made when conducting the EFA, such as the extraction method, the type of rotation, and the most appropriate means of determining the number of factors to extract. There are specific advantages and disadvantages with each choice. For example, oblique rotations are useful for correlated data; however, they can be complex to interpret if data is excessively correlated (Beavers et al., 2013). While there were numerous examples of EFAs in the literature to provide guidance on the best procedures, researchers often provided different recommendations regarding elements such as sample size, acceptable strength of items to include in factors, and the rationales behind various decisions which must be made when conducting an EFA. One argument is that the availability of choices allowed for a more flexible analysis, as the data could be explored in a myriad of ways. However, this also presented a limitation; deciphering the various EFA methodologies leads to a more subjective analysis, which subsequently makes the results dependent on researchers making the most appropriate decisions for the data (Henson & Roberts, 2006). Current and relevant literature on the different options to conduct an EFA was consulted to assist in clarifying this process and ensuring that the data was analyzed properly.

The WJ IV *Technical Manual* (McGrew et al., 2014) provides correlation matrices for ages 3 through 90 and older; however, a narrow age group was selected to be the focus of this study. The 9 to 13 matrix was chosen due to the frequency of psychoeducational assessments of children in this age range; however, this limits the applicability of the current study to other age groups. Therefore, broad implications for the population should not be made based on the results of this study, and results should be interpreted with caution among other age groups.

An additional methodological concern of the current study is the fact that all data was gathered from the WJ IV *Technical Manual* (McGrew et al., 2014). Although this is helpful in that the manual provides a data set with a considerable number of participants from varied backgrounds, the present study is wholly dependent on the information in the manual being accurately gathered and reported. Each statistical decision prior to the EFA will have been made by McGrew et al. (2014), leaving little control over some aspects of the data. While much information on data collection and statistical methods is present in the manual, it is likely that some of the sequence of events used to clean and present the data was not reported. For example, the authors included relatively little information about their use of imputation. Uncertainty regarding which variables include imputed data added an additional limitation for this study.

Future Directions

Results of this study were consistent with the test publishers' findings to a degree, as some factors mirrored those reported by McGrew et al. (2014). A CFA would be useful to further solidify these findings and examine structural differences. A CFA may also prove beneficial in clarifying G*f* as its own construct and informing future iterations of CHC theory. As Spurgin's (2018) research demonstrated, new broad ability factors may need to be introduced which address various executive functions such as attentional control or higher-order cognitive reasoning, or in the case of the present study, perceptual reasoning.

Additionally, access to the full normative dataset would alleviate concerns about clustering children at diverse development stages into a single group (e.g., 9 to 13 years old). This could especially yield important information for practitioners, such as the structural profile among neurotypical 9-year-olds will likely differ from those of neurotypical 13-year-olds, for example. Along the same lines, there is a need for data regarding the performance of various clinical populations in order to better inform diagnoses.

Chapter Summary

This study's results were consistent with the findings of other researchers who examined the WJ IV — the WJ IV COG and WJ IV OL appear to be over-factored, and the tests primarily provide a measure of *g*. This is also consistent with research on previous versions of the WJ (Dombrowski, 2014; Dombrowski & Watkins, 2013). The factor structure found by the current study generally aligned with test publishers – G*c*, G*s*, and G*wm* remained mostly unchanged with the exception of subtests which were removed during analyses. However, this study also found blended CHC factors and subtests which either cross-loaded or failed to saliently load onto any factor. This was not entirely unsurprising given that many tasks have a wide array of narrow abilities.

Results of this study support a five-factor solution as opposed to the seven factors reported by test publishers. The CHC factors that clearly emerged included: comprehension-knowledge (G*c*), short-term working memory (G*wm*), cognitive processing speed (G*s*), auditory processing (G*a*), and visual processing (G*v*). Subtests typically reported as measures of G*f* and G*lr* clustered together into a single factor hypothesized to represent perceptual reasoning. These findings have several implications for practitioners, chiefly that the WJ IV is an infallible instrument that does a better overall job of assessing general intelligence. Based on this study, scores pertaining to G*f* and G*lr* should especially be interpreted with caution.

This study has its own limitations related to data availability and the need for professional judgment to inform certain statistical decisions (for example, determining factor retention). Additional research is needed to replicate these results as well as confirm the factor structure of the WJ IV COG and WJ IV OL. Furthermore, there remains a paucity of research on the WJ IV, leaving many opportunities to continue exploring the utility of the instrument and informing its evolution.

References

Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Psychology*, *3*, 77–85.

Beavers, A., Lounsbury, J. W., Richards, J. K., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research & Evaluation*, *18*(5/6), 1–13.

Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative review. *Journal of Applied Psychology*, *96*, 881–906. doi:10.1037/a0023222

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:10.1076/jcen.24.3.383.981

Breit, M., Brunner, M., & Preckel, F. (2020). General intelligence and specific cognitive abilities in adolescence: Tests of age differentiation, ability differentiation, and their interaction in two large samples. *Developmental Psychology*, *56*(2), 364–384. doi:10.1037/dev0000876.supp

Brown, J. D. (2009). Statistics Corner. Questions and answers about language testing statistics: Principal components analysis and exploratory factor analysis—Definitions, differences, and choices. *Shiken: JALT Testing & Evaluation SIG Newsletter*, *13*(1), 26–30.

Brown, T. A. (2014). *Confirmatory factor analysis for applied research (*2nd ed.*)*. New York, NY: Guilford.

Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory component analysis. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99–136). Washington, DC: APA.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2016). Factor structure of the Wechsler Intelligence Scale for Children-Fifth Edition: Exploratory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment*, *28*(8), 975–986.

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

Carroll, J. B. (2005). The three-stratum theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 69–76). New York, NY: Guilford.

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

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

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing [Abstract]. *Psychology Bulletin*, *38*, 592.

Cormier, D. C., McGrew, K. S., Bulut, O., & Funamoto, A. (2017). Revisiting the relations between the WJ-IV measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during the school-age years. *Journal of Psychoeducational Assessment*, *35*(8), 731–754.

Cornell, E. L., & Coxe, W. C. (1934). *A performance ability scale*. New York, NY: World Book.

Cox, T. F., & Cox, M. A. (2000). *Multidimensional scaling*. Boca Raton, FL: CRC press.

Darwin, C., Duthie, J. F., & Hopkins, W. (1859). *On the origin of species by means of natural selection: Or, the preservation of favoured races in the struggle for life.* London, England: John Murray, Albemarle Street.

Dombrowski, S. C. (2014). Investigating the structure of the WJ-III Cognitive in early school age through two exploratory bifactor analysis procedures. *Journal of Psychoeducational Assessment*, *32*(6), 483–494.

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2017). Exploratory and hierarchical factor analysis of the WJ IV Cognitive at school age. *Psychological Assessment*, *29*, 394–407.

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). Heirarchical exploratory factor analyses of the Woodcock-Johnson IV full test battery: Implications for CHC application in school psychology. *School Psychology Quarterly*, *33*(2), 235–250.

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.). [Measurement instrument]. San Antonio, TX: Pearson.

Evans, V. (1999). *Higher-order factor analysis: An introductory primer.* Annual meeting of the Southwest Educational Research Association; San Antonio, TX.

Flanagan, D. P., & McDonough, E. M. (2018). *Contemporary intellectual assessment: Theories, tests, and issues*. New York, NY: Guilford.

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*(2), 249–265.

Gathercole, S. E., & Alloway, T. P. (2008). *Working memory and learning: A practical guide for teachers*. Los Angeles, CA: Sage.

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

Gustafsson, J. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, *8*, 179–203.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling.* Cary, NC: SAS Institute, Inc.

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*(3), 393–416.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrica*, *30*(2), 179–185.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. Mcgrew, J. K. Werder, & R. W. Woodcock (Eds.), *Woodcock-Johnson technical manual* (pp. 197–232). Chicago, IL: Riverside.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*(1), 141–151.

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

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children interpretive manual.* Circle Pines, MN: American Guidance Service.

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

Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it.* Beverly Hills, CA: Sage.

Kline, R. B. (2011). *Principles and practice of structural equation modelling.* New York, NY: Guilford.

Kohs, S. C. (1923). *Intelligence measurement: A psychological and statistical study based upon the block design tests*. New York, NY: Macmillan.

Kwate, N. O. A. (2001). Intelligence or misorientation? Eurocentrism in the WICS-III. *Journal of Black Psychology*, *27*(2), 221–238.

Lawley, D. N., & Maxwell, A. E. (1971). *Factor analysis as a statistical method.* London: Butterworth.

Leiter, R. G. (1936). *The Leiter International Performance Scale.* [Measurement instrument]. Honolulu, HI: University of Hawaii Press.

McCredie, H. (2018). Making intelligence testing more scientific: Charles Spearman (1863–1945). *Assessment & Development Matters*, *10*(1), 28–33.

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a current 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.

McGrew. K. S. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment* (2nd ed., pp. 136–181). New York, NY: Guilford.

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

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

McGrew, K. S., & Woodcock, R. W. (2001). *Technical Manual: Woodcock Johnson III*. Rolling Meadows, IL: Riverside.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage.

Mosier, C. I. (1943). On the reliability of a weighted composite. *Psychometrika*, *8*, 161–168.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System*. [Measurement instrument] Itasca, IL: Riverside.

Osborne, J. W. (2014). *Best practices in exploratory factor analysis.* Scotts Valley, CA: CreateSpace Independent Publishing.

Otis, A. S. (1918). *Otis Group Intelligence Scale: Manual of directions for primary and advanced examinations*. Yonkers-on-Hudson, NY: World Book Company.

Pintner, R., & Paterson, D. G. (1917). A comparison of deaf and hearing children in visual memory for digits. *Journal of Experimental Psychology*, *2*(1), 76–88.

Pressey, S. L., & Pressey, L. W. (1918). A group point scale for measuring general intelligence with first results from 1,100 school children. *Journal of Applied Psychology*, *2*, 250–269.

Raubenheimer, J. (2004). An item selection procedure to maximize scale reliability and validity. *Journal of Industrial Psychology*, *30*(4), 59–64.

Reynolds, C. R., & Kamphaus, R. W. (2003). *RIAS*: *Reynolds Intellectual Assessment Scales*. Lutz, FL: Psychological Assessment Resources, Inc.

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology*, *51*(4), 535–555.

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

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*(1), 53.

Schneider, W. J., & McGrew, K. S. (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: Guilford.

Schneider, W. J., & McGrew, K. S. (2018). The Cattell–Horn–Carroll theory of cognitive abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 73–163). New York, NY: Guilford.

Schrank, F. A. (2010). Woodcock-Johnson III Tests of Cognitive Abilities. In A. Davis (Ed.), *Handbook of pediatric neuropsychology* (pp. 415–434). New York, NY: Springer.

Schrank, F. A., Decker, S. L., & Garruto, J. M. (2016). *Essentials of WJ IV cognitive abilities assessment.* Hoboken, NJ: Wiley.

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

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

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

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

Schrank, F. A., & Wendling, B. J. (2018). The Woodcock–Johnson IV: Tests of Cognitive Abilities, Tests of Oral Language, Tests of Achievement. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (p. 383–451). New York, NY: Guilford.

Scott, C. A. (1913). General intelligence or "school brightness." *Journal of Educational Psychology*, *4*, 509–524.

Seguin, E. (1856). Origin of treatment and training of idiots. *American Journal of Education*, *2,* 145–152.

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

Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, *15*, 72–101.

Spurgin, A. R. (2018). *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* (Doctoral dissertation). Retrieved from ProQuest Dissertations & Theses Global.(Order No. 13804283).

Sternberg, R. J. (Ed.). (1986). *Advances in the psychology of human intelligence: Volume* Hillsdale, NJ: Lawrence Erlbaum Associates.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (4th ed.). Hillsdale, NS: Erlbaum.

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*(2), 689–697.

Suhr, D. D. (2006, March). *Exploratory or confirmatory factor analysis*. Paper presented at the Proceedings of the 31st Annual SAS Users Group International Conference, Cary, NC.

Users Group International Conference. Cary, NC: SAS Institute Inc.

Terman, L. M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale*. New York, NY: Houghton Mifflin.

Thorndike, E. L. (1919). Tests of intelligence, reliability, significance, susceptibility to special training and adaptation to the general nature of the task. *School and Society*, *9*, 189–195.

Thurstone, L. L. (1934). The vectors of mind. *Psychological Review*, *41*(1), 1–32.

Tucker-Drob, E. M. (2009). Differentiation of cognitive abilities across the life span. *Developmental Psychology*, *45*, 1097–1118. doi:10.1037/a0015864

Urbina, S. (2004). *Essentials of psychological testing*. Hoboken, NJ: Wiley.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321–327.

Vernon, P. E. (1961). Intelligence and attainment tests. *British Journal of Educational Studies*, *10*(1), 97–98.

Vogt, W. P. (1999). *Dictionary of statistics & methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Wasserman, J. D., & Tulsky, D. S. (2005). A history of intelligence assessment. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3–22). New York, NY: Guilford.

Wechsler, D. (1939). *The measurement of adult intelligence.* New York, NY: Williams & Wilkins.

Wechsler, D. (1946). *Wechsler-Bellevue Intelligence Scale, form II. Manual for administering and scoring the test.* New York, NY: Psychological Corporation.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children (WISC).* [Measurement instrument]. New York, NY: Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale-Restandardized (WAIS).* [Measurement instrument]. New York, NY: Psychological Corporation.

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

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

Wechsler, D. (2014b). *Wechsler Intelligence Scale for Children–Fifth edition technical and interpretive manual*. San Antonio, TX: NCS Pearson.

Woodcock, R. W. (1958). An experimental prognostic test for remedial readers. *Journal of Educational Psychology*, *49*, 23–27.

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). Woodcock-Johnson III. Rolling Meadows, IL: Riverside.

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