FACTOR ANALYSES ON THE VALIDITY OF THE WECHSLER MEASURES OF COGNITIVE FUNCTIONING

A DISSERTATION

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIRMENTS

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

IN THE GRADUATE SCHOOL OF THE

TEXAS WOMAN'S UNIVERSITY

SCHOOL OF SOCIAL WORK, PSYCHOLOGY AND PHILOSOPHY

COLLEGE OF ARTS AND SCIENCES

BY

BRIANA EILEEN PAULMAN, B.A.

DENTON, TEXAS

AUGUST 2023

ACKNOWLEDGMENTS

I want to first thank my parents, Ronald and Joanne Paulman, for their love and unwavering support throughout my life. Thank you for igniting my passion for psychology, which has gotten me to where I am today, and for always believing in me. I love you both, and I am who I am today because of you. I cannot talk about my family without acknowledging my brother, Brendan Paulman, who has always encouraged me along the way. I also want to thank my rock, my biggest supporter, and my best friend and husband, Zach Keller. I could not have gotten through graduate school without you standing by my side. I cannot wait to start the next chapter of our lives together and see where the next adventure takes us (and Finley, of course). Next, I want to thank my dissertation committee, Dr. Wendi Johnson, Dr. Lisa Rosen, and Dr. Joel Schneider. I sincerely thank each of you for your guidance and wisdom throughout this process. Finally, I want to thank Dr. Denise Maricle. Thank you for believing in me and helping my reach my goals. During my time in graduate school, your courses, lessons, advice, and support have had the most profound impact on me that I know I will carry on far beyond graduate school. I am so grateful to have had your mentorship throughout my time in graduate school. I honestly could not have gotten through it without you; thank you.

ABSTRACT

BRIANA EILEEN PAULMAN

FACTOR ANALYSES ON THE VALIDITY OF THE WECHSLER MEASURES OF COGNITIVE FUNCTIONING

AUGUST 2023

Intelligence has been and currently is one of the most widely studied phenomena in the field of psychology. There is a lack of consensus in the field on the definition and conceptualization of intelligence with researchers developing numerous theories to better understand it. Corollary to the debate over the concept of intelligence itself, methods of reliably measuring intelligence are also of concern. Given the widespread use of intellectual assessments, it is imperative that practitioners utilizing such assessments are well-versed in the instruments, including what cognitive abilities are measured and how to interpret the results. The Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V), the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV), and the Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition (WPPSI-IV) are three related cognitive assessment measures developed to assess general intelligence and a combination of several broad abilities: verbal comprehension (*Gc*), visual-spatial processing (*Gv*), fluid reasoning (*Gf*), short-term working memory (*Gwm*), and processing speed (*Gs*). The Wechsler instruments were developed with theories and cognitive models from several domains, including cognitive psychology, neuropsychology, and cognitive abilities such as memory, attention, and executive functioning. However, it is debated whether the factor structure of the Wechsler instruments measures the cognitive abilities as outlined by the test publishers. The primary purpose of the present study was to examine the factor structure of the Wechsler cognitive assessment instruments against two models of cognitive abilities using data from the Wechsler standardization samples: Cattell-Horn-Carroll (CHC) theory of intelligence and the functional-CHC model of cognitive ability. The data analyses included multiple exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) of the Wechsler standardization samples for each instrument. The present study attempted to provide a more comprehensive understanding of general intelligence and cognitive abilities in children, adolescents, and adults. Results of the present study provided support for some of the hypothesized broad abilities posited by both the F-CHC model and CHC theory, particularly *Gc, Gr/Gf, Gcm/Gwm, Gv,* and *Gs*. Results also provided partial support for some of the hypothesized narrow abilities. However, the Wechsler instruments utilized do not appear robust enough to truly investigate the validity of the full F-CHC model and all the narrow abilities. This limitation and others are discussed as well as implications and future research directions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Cognitive assessments are administered every day in school, clinical, or private settings to assist in determining intellectual functioning and cognitive abilities in order to support diagnostic decisions. Intelligence, a construct that is predominant in the field, has a lengthy and well-researched history, including the theory and development of cognitive assessments utilized by many to this day (Wasserman, 2018). Given the widespread use of cognitive assessments, it is critical that practitioners utilizing such assessments are well-versed in the instruments, including which cognitive abilities are measured and how to interpret the results. An example of such instruments are the Wechsler cognitive assessment measures: The Wechsler Intelligence Scale for Children (WISC), introduced in 1949, the Wechsler Adult Intelligence Scale (WAIS), introduced in 1955, and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), introduced in 1967. The Wechsler cognitive assessment instruments are one of the most commonly used assessment tools by practitioners across settings. This chapter presents a review of the literature regarding the history of intelligence testing, intelligence theory, and contemporary theoretical perspectives. The Wechsler cognitive assessment instruments and the various revisions are examined, followed by an outline of the rationale and significance of the study.

Literature Review

The Wechsler cognitive assessment instruments have a long history of revisions designed to match current theories and best practice. While the Cattell-Horn-Carroll (CHC) theory, the present dominant theory of intelligence, was considered in the development of the current Wechsler instruments (i.e., WAIS-IV, WPPSI-IV, and WISC-V; Wechsler, 2008, 2012, 2014), other theoretical constructs and models were also utilized. When examining the underlying structure of the Wechsler instruments, the history of intellectual assessment and theory must also be examined.

Prior to the development of formal cognitive assessments based on theoretical conceptualizations of neurocognitive functioning, physicians and psychologists attempted to study human cognitive abilities using a variety of sensory and motor measures. It was posited that, by examining these basic human abilities, one could understand the construct of intelligence (Wasserman, 2018). This approach is exemplified in the work of Franz Joseph Gall and Johann Gaspar Spurzheim and their study of phrenology. Finding this conceptualization of intelligence limited, the study of intellectual functioning shifted to the study of human behavior and theories regarding latent cognitive traits or abilities. At the beginning of the 20th century, Charles Edward Spearman's discovery of a general intelligence factor laid the groundwork for the beginning of theoretical testing. At the turn of the 20th century, Alfred Binet, referred to as the father of intellectual assessment, in conjunction with Théodore Simon, began developing the first intellectual assessments. Binet and Simon were commissioned by the Minister of Public Instruction in France to develop standardized tests for making diagnostic and educational decisions regarding disabilities in children and adolescents (Wasserman, 2018). In 1905, the Binet-Simon Intelligence Scale was published, the first of its kind specifically created for children and adolescents (Wasserman, 2018). The Binet-Simon attempted to measure multiple cognitive abilities that accumulated to form a global estimate of intelligence.

Lewis Madison Terman revised the Binet-Simon Intelligence Scale by conducting several standardization studies utilizing large groups of children and adolescents with varying cognitive abilities, ultimately leading to the creation of the Stanford-Binet Intelligence Scale (SB) in 1916

(Wasserman, 2018). The SB was also the first instrument to introduce the term *intellectual quotient*, defined as an individual's mental age divided by chronological age, multiplied by 100 (Wasserman, 2018).

Large-scale intelligence testing began during World War I when there was a need for identifying men suited for service and where to place men in specific jobs. This need led to the creation of Examination Alpha and Examination Beta to examine the cognitive abilities of army recruits. Examination Alpha served as a largely verbal exam for those that spoke English while Examination Beta was created as a nonverbal exam (Wasserman, 2018). These exams influenced the current battery used by the United States Army today as well as contemporary cognitive assessment measures.

During this time, David Wechsler was a military test examiner administering Examinations Alpha and Beta. Wechsler subsequently developed his own ideas about assessment and intelligence, leading him to develop the Wechsler-Bellevue Intelligence Scale (WBIS; Wechsler, 1939b) designed for adults. The Wechsler-Bellevue provided practitioners with three measures: a measure of general intelligence (the Full-Scale Intelligence Quotient), a verbal intelligence quotient, and a performance intelligence quotient. The Wechsler-Bellevue was revised multiple times to provide a cognitive assessment instrument for varying ages (i.e., young children, children and adolescents, and adults). The current revision of the Wechsler scales includes the Wechsler Intelligence Scale for Children (WISC), Fifth Edition (Wechsler, 2014), the Wechsler Preschool and Primary Scale of Intelligence (WPPSI), Fourth Edition (Wechsler, 2012), and the Wechsler Adult Intelligence Scale (WAIS), Fourth Edition (Wechsler, 2008). The Wechsler cognitive assessment instruments are not the only tests available to measure cognitive functioning; there are a wide variety of instruments including the Woodcock-Johnson: Tests of

Cognitive Abilities, Fourth Edition (McGrew et al., 2014; Woodcock & Mather, 2014), the Kaufman Assessment Battery for Children, Second Edition, Normative Update (Kaufman & Kaufman, 2018), the Stanford Binet Intelligence Scales, Fifth Edition (Roid, 2003), the Differential Ability Scales, Second Edition (Elliot, 2007) and the Cognitive Assessment System, Second Edition (Naglieri et al., 2014).

Origins of Intelligence Theory

Modern intelligence theory has many origins; hence, there is still a lack of consensus among researchers on what constitutes intelligence. Each iteration considers previous theories while utilizing new research with each revision. Historically, there are philosophical, evolutionary, and sociological perspectives to intelligence theory. For example, intelligence research from an evolutionary perspective poses the question of whether intelligence enhances an individual's chance of survival (Wasserman, 2018). From its earliest days, the definition of intelligence has not been agreed upon. A predominate disagreement was seen (and is still a source of discussion and controversy) in the debate over ideas about whether or not a general factor of intelligence exists or if intelligence is comprised of distinct, separable factors of cognitive functions. Charles Edward Spearman (1904) posited there exists an overarching general intelligence factor derived from intercorrelations on individual test scores. This is the concept of "g" seen in his conceptualization of intellectual functioning as well as in current theoretical models of cognition. Spearman's theory had several detractors. Louis Leon Thurstone (1938) utilized factor analytic techniques to develop primary factors that he felt comprised intelligence. Thurstone posited eight primary factors and eschewed the idea of an overall or general factor in intelligence. However, Thurstone eventually accepted Spearman's concept of a general factor, although he still believed one single score of intelligence was not an adequate representation and recommended examining strengths and weaknesses of an individual instead (Thurstone, 1945; Wasserman, 2018). Philip Ewart Vernon (1950) developed the first hierarchical model of intelligence, proposing that the overarching, higher-order factor of intelligence (i.e., *g*) consisted of two lower-order factors, verbal/educational and spatial/mechanical. This dichotomous split is similar to the original Wechsler instruments.

Contemporary Perspectives

Following Vernon's model of a dichotomous split and using Thurstone's conceptualizations, Raymond Bernard Cattell proposed his own theory of intelligence, labeled the theory of fluid (*Gf*) and crystalized (*Gc*) intelligence or Gf-Gc theory. Cattell's (1943, 1944) theory postulated that fluid and crystalized intelligence are highly correlated. Over the next decades since his initial conceptualization, Cattell conducted several studies to validate his theory. He was later joined by John Leonard Horn, a doctoral student of his, who expanded upon Cattell's theory to include several additional broad factors of cognitive functioning and demonstrated that these factors have separate developmental trajectories. These factors included: fluid intelligence (*Gf*), crystalized intelligence (*Gc*), short-term acquisition and retrieval (SAR or *Gsm*), visual intelligence (*Gv*), auditory intelligence (*Ga*), long-term storage and retrieval (TSR or *Glr*), cognitive processing speed (*Gs*), correct decision speed (CDS), and quantitative knowledge (*Gq*; Cattell & Horn, 1978).

John Bissell Carroll (1993) did not agree with Cattell and Horn's Gf-Gc theory, instead developing his own theory based on analysis of over 450 ability matrices. Carroll's (1993) model consisted of three strata, aptly named Carroll's three-stratum model of human cognitive abilities. Within this model, the top stratum consisted of a general factor of intelligence subsuming broad (i.e., stratum II) and narrow (i.e., stratum III) abilities beneath it. Carroll's model provided a way for researchers to empirically organize decades of research in order to classify and understand intelligence.

Hitherto, Gf-Gc theory and Carroll's three-stratum theory were the predominant theories of intelligence. In fact, cognitive assessments, such as the Woodcock-Johnson Psychoeducational Battery-Revised (WJ-R; Woodcock & Johnson, 1989), began to incorporate the theory into the test's development. In the case of the WJ-R, Gf-Gc theory was utilized in its development as the primary organizational and interpretive premise for the test. However, there still existed contention and debate on what comprised intelligence; how many constructs does intelligence consist of and does there exist a general intelligence factor? In response to this debate, Kevin McGrew proposed the Cattell-Horn-Carroll (CHC) theory of cognitive abilities as a combination of Gf-Gc theory and Carroll's three-stratum theory. CHC theory posits that there is a hierarchical structure of cognitive abilities consisting of narrow and broad abilities (Schneider & McGrew, 2018). CHC theory has had several revisions since its conception based on emerging research throughout the last two decades. While continuously evolving, CHC theory remains the predominant psychometric theory of intelligence in the field. One proposed revision is Woodcock et al.'s (2017) and Woodcock et al.'s (2018) reconceptualization of CHC theory into the functional-CHC theory (F-CHC). The F-CHC was created in an effort to have a more functional theory for the average practitioner that is simpler to interpret and consistent with current neurocognitive research.

The Wechsler Instruments

The Wechsler instruments were considered one of the principal measures for adult and child cognitive assessment since the emergence and popularization of intelligence testing. Since the original conception of a cognitive assessment instrument from David Wechsler, the Wechsler instruments have undergone several revisions. The present study investigates the factor structure of the most recent iterations of the Wechsler cognitive instruments, the WAIS-IV (Wechsler, 2008), WPPSI-IV (Wechsler, 2012), and WISC-V (Wechsler, 2014). The present study examined the evolution of the Wechsler instruments in an effort to understand how the conceptualization has changed from the first iteration to the present.

The first assessment developed by Wechsler was the WBIS in 1939 (Wechsler, 1939b) as an instrument designed specifically for use with adults. The WBIS included a dichotomous split of verbal and performance IQ, which combined to form an overall Full-Scale Intelligence Quotient (FSIQ). Wechsler developed the WBIS based on subtests from the SB, the Army Examinations, and other existing measures at the time (Groth-Marnat, 2009). The WBIS has undergone several revisions, including developing measures specific for adults (WAIS), children and adolescents (WISC), and young children (WPPSI).

In 1949, Wechsler adapted the WBIS to develop a cognitive assessment for children and adolescents, the WISC (Wechsler, 1949). The WISC underwent several revisions including the WISC-R (Wechsler, 1974), WISC-III (Wechsler, 1991), WISC-IV (Wechsler, 2003), and the current WISC-V (Wechsler, 2014). As with each revision of the WBIS, the WISC consisted of a verbal and performance split with an overall FSIQ. The WISC-III first deviated from this structure to include additional cognitive abilities (Keith & Reynolds, 2010). The WISC-IV eliminated the dichotomous split and included four factors. The WISC-V followed suit with the WPPSI-IV to include a five-factor model represented through the Verbal Comprehension Index (VCI), Visual Spatial Index (VSI), Fluid Reasoning Index (FRI), Working Memory Index (WMI), and Processing Speed Index (PSI), in addition to the FSIQ (Wechsler, 2014).

In 1955, Wechsler adapted the WBIS to create a cognitive assessment measure specifically for adults, the WAIS (Wechsler, 1955). The WAIS was revised several times, including the WAIS-R in 1981, the WAIS-III in 1997, and the WAIS-IV in 2008. The WAIS-V is currently in development. The WAIS has moved from a two-factor structure consisting of a verbal and performance split to a four-factor model through factor analytic and validation studies. The four factors of the current WAIS-IV are exemplified by the VCI, Perceptual Organization Index (POI), WMI, and a PSI, in addition to the FSIQ (Wechsler, 2008).

In 1967, the WBIS was adapted to examine cognitive abilities of young children, aptly called the WPPSI (Wechsler, 1967). The first edition of the WPPSI fulfilled the need for preschool assessment. The WPPSI was revised in 1989 (WPPSI-R; Wechsler, 1989), 2002 (WPPSI-III; Wechsler 2002), and 2012 (WPPSI-IV; Wechsler 2012). The WPPSI-IV was the first Wechsler instrument to feature a five-factor model based on contemporary research on intelligence, neurocognitive research, and working memory models (Raiford & Coalson, 2014). In similar fashion to the WAIS-IV, the fourth iteration of the WPPSI eliminated the verbal and performance IQ in favor of five factors represented through the VCI, VSI, FRI, WMI, and PSI (Wechsler, 2012). The WPPSI-IV also features two separate age groups based on distinct cognitive abilities found within those age ranges.

While CHC theory may be the contemporary theory for many intelligence tests, it was not the only influence in revisions of the Wechsler instruments. Each revision of the Wechsler instruments took into consideration functional theories and structural theories, including theories from cognitive psychology, neuropsychology, and specific cognitive abilities such as memory, attention, and executive functioning, to create a comprehensive theoretical basis (Kaufman et al., 2016). The FSIQ may be considered a measure of overall intelligence, or *g*, as Spearman theorized, within the theoretical framework of the Wechsler cognitive instruments. Additional information regarding the Wechsler cognitive assessment instruments, including their composition, theoretical development, and psychometric specification are discussed in Chapter 2 and Chapter 3.

Purpose and Significance of the Present Study

The field of school psychology encompasses research and theory from many disciplines of psychology as practitioners are held responsible for supporting children and adolescents' academic, behavioral, social, and emotional needs. This includes understanding and interpreting cognitive abilities of children and adolescents using cognitive assessment. When practitioners understand a child or adolescent's cognitive profile, including strengths and weaknesses, they are able to provide more accurate diagnoses and interventions that are best suited for that individual. Oftentimes in order to provide a comprehensive evaluation of an individual, a cognitive assessment is administered. Therefore, it is critical that the particular cognitive assessment instrument used is structurally valid and well-researched. Further, practitioners must be aware of the theory that comprises the development of their chosen intelligence instrument in order to appropriately interpret results and to be assured that the instrument is measuring the cognitive factors it claims to measure.

The publishers of the Wechsler instruments have provided factor analytic support for the underlying factor structure of the instruments. The WISC-V and the WPPSI-IV both contain five factors characteristic of five broad abilities of the CHC theory: *Gc* represented by the VCI; *Gv* represented by the VSI; *Gf* represented by the FRI; immediate short-term memory and working memory (*Gwm*) as measured by the WMI; and *Gs* exemplified by the PSI. The WAIS-IV contains four factors characteristic of five broad abilities of the CHC theory represented by four factor indices including the VCI (*Gc*), POI (*Gf* and *Gv*), WMI (*Gwm*), and PSI (*Gs*). Although the publishers of the Wechsler instruments provide factor analytic data for the structure of the measure, researchers have found that the methodology provided is cause for psychometric concern (Benson et al., 2010; Canivez et al., 2016; Canivez, Watkins, & Dombrowski, 2017; Canivez, Dombrowski, & Watkins, 2017; Dombrowski et al., 2015). In particular, there is little psychometric support for a fifth factor of *Gf* as the WISC-V and WPPSI-V have done in their most recent iterations. Further, there are little details provided on the confirmatory factor analysis (CFA) methods employed by the publishers of the WAIS-IV, WISC-V, and WPPSI-IV, specifically the lack reasoning for certain model identification methods (Canivez et al., 2016; Canivez, Watkins, & Dombrowski, 2017). These methodological issues are further discussed in Chapter 5.

The present study examined the factor structure of the Wechsler cognitive assessment instruments in the standardization sample using the F-CHC model of cognitive ability and the CHC theory of cognitive abilities. The data analyses utilized included exploratory factor analysis (EFA) and CFA of the standardization samples of the WAIS-IV, WPPSI-IV, and WISC-V. EFA is an exploratory multivariate statistical method that examines the factor structure without a preconceived hypothesis, while CFA is a similar method utilizing a preconceived hypothesis. The following research questions and hypotheses were proposed.

Research Questions

1. To what degree does the proposed F-CHC model explain the latent structure of cognitive abilities as measured by data from the Wechsler standardization samples?
    a. Research Hypothesis 1: The data will support the proposed F-CHC model.

b. Research Hypothesis 2: In comparison to other models, the proposed F-CHC model will be the best fitting model for the data from the Wechsler standardization samples.

2. If the above-mentioned model does not adequately represent the Wechsler standardization samples, is there another factor structure that best fits the data?

   a. Research Hypothesis 3: The results from this analysis will not support the proposed factor structure that is identified by the test publishers of the Wechsler instruments.

CHAPTER II

REVIEW OF THE LITERATURE

The following chapter provides a review of the relevant literature to the present study. The history of intelligence testing, theoretical perspectives of cognitive abilities, and evolution of the Wechsler instruments will be examined. The Wechsler cognitive assessment instruments began with a two-factor structure (i.e., WBIS) and evolved into a four (i.e., WAIS) and five-factor (i.e., WPPSI; WISC) structure that is utilized today and closely aligns with CHC theory in addition to other neurocognitive theories and memory models.

Development of Intellectual Assessment

While intelligence testing is arguably the most studied phenomenon in psychology, there is still contention and controversy around the assessment of intelligence (Wasserman, 2018). The first known attempt to predict cognitive abilities based on measurements of the skull was developed in the early 1800s by physicians Franz Joseph Gall and Johann Gaspar Spurzheim in their study of phrenology. These physicians posited that the mind consisted of several faculties (i.e., mental functions) located in particular regions of the brain (Wasserman, 2018). While a novel concept, phrenology was and is a pseudoscience. Intelligence was further defined in the mid-to-late 1800s by Herbert Spencer as an organism's ability to adapt through interactions with their environment, followed by Charles Darwin who argued this adaptive behavior is similarly seen in humans and animals. Consequently, intelligence was viewed as advantageous from an evolutionary perspective.

Anthropometrics

The origins of intelligence testing began with the merging of theory and practice. Francis Galton was a pioneer in anthropometry, creating the first laboratory designed to study the measurement of man. Galton’s study of anthropometry proved instrumental in the development of cognitive assessment measures. Galton did not claim to measure intelligence; rather, he based the facility of intelligence on physical and sensory dimensions that could be measured (Wasserman, 2018). Further, Galton is credited with utilizing test batteries for the collection of data, the use of control groups in research, and various statistical methods, including regression and correlation (Wasserman, 2018). While Galton’s initiatives in his laboratory may be considered the beginning of intelligence testing, Charles Edward Spearman’s finding of a general intellectual factor in 1904 spearheaded the movement for theory development. Concurrently in the late 1800s, several researchers had begun large-scale efforts to develop intellectual tests in the United States and across the world; James McKeen Cattell studied at Columbia University and Alfred Binet was conducting his research in France. It was not until Spearman’s discovery of a way to quantify general intelligence did intellectual testing research move past sensory and motor measures.

Cattell studied under several psychologists, working in laboratories across the United States and Europe. In Wilhelm Wundt’s laboratory in Germany, Cattell shied away from Wundt’s “experimenter introspection” method and discovered his interest in measuring individual differences based on observable behaviors in the laboratory setting (Wasserman, 2018). After discovering this realization, Cattell transitioned into Galton’s anthropometry laboratory in 1887. Cattell and Galton combined each of their initiatives and conducted tests in the laboratory utilizing sensory measures, reaction time measures, and newly created measures (e.g., memory for digits, word association, rapid color naming, letter cancellation, logical memory). This combination of higher-order, lower-order, and anthropometric measures created a unique battery designed to quantitatively assess human abilities (Wasserman, 2018). However, subsequent researchers have discredited Cattell and Galton's work in anthropometrics due to a lack of theory and unproven tests (Wasserman, 2018). Nevertheless, Cattell is credited as a pioneer in testing and measurement.

Contemporary Intelligence Testing

Alfred Binet was pivotal in intelligence testing as the developer of the first working measure of intelligence (Wasserman, 2018). In 1890, Binet published several papers recounting experimental studies he conducted with his two daughters. Initial studies involved observations and tasks designed to measure cognition and personality; throughout his daughters' mid-adolescence, he expanded his studies to include measures of language, attention, memory, and reasoning (Wasserman, 2018). In 1895, Binet collaborated with Victor Henri to outline an intelligence test consisting of ten measurable abilities: memory, imagination, imagery, comprehension, attention, suggestibility, aesthetic sentiment, moral sentiment, muscular strength/will-power, and motor ability (Wasserman, 2018). Binet and Henri speculated that more complex abilities better measured intelligence than the previously thought sensory and motor abilities. In 1899, after Henri's departure, Binet began collaborating with Théodore Simon, which led to the development of intelligence tests. In 1904, Binet led an educational advocacy group and became a member of the Bourgeois Commission, a commission established by the Minister of Public Instruction in France to examine current public education laws and their application to students with disabilities (Wasserman, 2018). Binet took this opportunity as a member of the commission to move forward in his efforts in creating standardized tests for making diagnoses and educational decisions. In 1905, the revolutionary Binet-Simon Intelligence Scale was published; this scale consisted of 30 items assessing a variety of cognitive processes resulting in a global estimate of intelligence.

Throughout the years, the Binet-Simon Intelligence Scale underwent several revisions. Lewis Madison Terman developed one such revision that was hailed as the best adaption of the original scale. Terman, a professor out of Stanford University, conducted several standardization studies utilizing large groups of children and adolescents with varying cognitive abilities. These studies ultimately led to the 1916 SB (Wasserman, 2018). Terman further contributed to the SB by introducing the term intellectual quotient, defined as an individual's mental age divided by chronological age, multiplied by 100 (Wasserman, 2018). The SB was said to measure general intelligence, with IQ as an estimate of Spearman's *g.* The SB underwent even further revision throughout the years until its fifth revision in 2004 (SB-5; Roid, 2003).

Large-Scale Assessment and World War I

World War I (1914-1918) was the catalyst for the development of large-scale assessments with adult populations. At the onset of war with Germany in 1917, the United States saw an influx of draftees. However, no criteria existed for identifying and excluding men who were not suitable for service. Additional problems included the fact that there was no way of identifying potential officers from the draftees, and no procedure for assigning men to specific jobs in the military (Wasserman, 2018). Consequently, Robert M. Yerkes, the president of the American Psychological Association (APA) at that time, assembled a committee to assist the government in their endeavors. This committee of psychologists, including Terman and Arthur Otis, developed two measures: Examination Alpha and Examination Beta. Examination Alpha, or Army Alpha, was designed for use with individuals who were fluent, literate, English-language speakers while Examination Beta, or Army Beta, was designed as a largely nonverbal measure for those did not speak English or were illiterate (Wasserman, 2018). The Army Alpha and Army

Beta have been revised throughout the years ultimately resulting in the Armed Services Vocational Aptitude Battery, the current test of the United States military.

During this time, David Wechsler registered for the draft, despite his objections to the war. Rather than serving in active combat, Wechsler became a mental test examiner where he was trained on the Alpha and Beta Examinations in addition to other tests. While serving as an examiner in the army, Wechsler developed his own core ideas about assessment based on his experience with the existing measures (Wasserman, 2018), quickly becoming a pivotal figure in intellectual assessment. Regarding intelligence, Wechsler created a definition that would become one of the best-known definitions of intelligence (Wasserman, 2018). Wechsler defined intelligence as, "the aggregate or global capacity of the individual to act purposefully, to think rationally and to deal effectively with his environment" (Wechsler, 1939a, p. 3). Wechsler's definition combined Spearman's concept of *g* and Binet's ideas of adaption to the environment and dynamics of intelligence (Wasserman, 2018). Wechsler saw a need for adult intelligence tests due to the poor normative sample of the SB, and the inadequacy of the Army tests to make clinical decisions (Wasserman, 2018). The Bellevue Intelligence Scale, soon after known as the Wechsler-Bellevue, was Wechsler's first iteration of his own intelligence test; it was a conglomeration of subtests from existing instruments that Wechsler deemed psychometrically sound and clinically necessary.

The Bellevue took a similar format to the Army Alpha and Army Beta tests with a verbal and performance dichotomous split (Wasserman, 2018). The Bellevue and all subsequent Wechsler instruments contained a deviation FSIQ, explained as the distance from the normative mean expressed in standardized units, in addition to the Verbal Intelligence Quotient (VIQ) and Performance Intelligence Quotient (PIQ). Over the decades, Wechsler developed several instruments, including the WISC (Wechsler, 1949, 1974, 1991, 2003, 2014), WAIS (Wechsler, 1955, 1981, 1997, 2008), and the WPPSI (Wechsler, 1967, 1989, 2002, 2012).

Origins of Intelligence Theory

Francis Galton developed one of the earliest laboratories intended to examine human cognition. Galton developed several instruments using scientific methodology to examine intelligence, which he believed consisted of one general ability. At the time, these instruments allowed objective measurements of observed, physical variables and latent, cognitive variables; these instruments mostly measured sensorimotor abilities as they are known today (Tulsky et al., 2003). While Galton posited that testing can be used to examine intelligence, he never explicitly used the term intelligence (Kent, 2017). Galton had a large influence on James McKeen Cattell's work, who developed 10 tasks intended to measure intelligence. Cattell viewed intelligence as a function of the ability to process information in that those who have the ability to do so more quickly will ultimately learn and retain more information. While Cattell's viewpoint has proven to be inaccurate, the efficiency and ability to process information has been included in contemporary theories of intelligence (i.e., processing speed; Tulsky et al., 2003).

Cattell and Galton laid the framework for measuring intelligence as psychologists attempted to define the construct of intelligence. Spearman's (1904) paper examined Galton's hypothesis, stating that a general intellectual ability is involved in all tasks requiring intellectual effort (Kent, 2017). Spearman also is credited for his use of statistical procedures, such as factor analysis, in examining the structure of intelligence and its' underlying dimensions (Wasserman, 2018). Spearman was interested in the relationship between various kinds of intelligence, physical performance, such as measures of sight and touch, and academic performance (Kent, 2017; Tulsky et al., 2003). In 1904, Spearman published his findings of a factor of general intelligence (i.e., *g*), summarized as the following, "All branches of intellectual activity have in common one fundamental function (or group of functions), whereas the remaining or specific elements of the activity seem in every case to be wholly different from that in all the others" (p. 284). In this excerpt, Spearman was referring to *g* when he discussed the common fundamental factor. Spearman advanced his concept of *g* by developing a two-factor theory consisting of general intelligence, a shared factor across tasks, and specific factors unique to each task (Wasserman, 2018). These unique factors each share some level of general intelligence. Interestingly, Spearman (1927) made the connection that intelligence was largely hereditary.

While Spearman was developing his two-factor theory of intelligence, Edward Lee Thorndike developed his own theory, positing that there existed varying subtypes of intelligence (Wasserman, 2018). These subtypes include abstract or verbal intelligence, practical intelligence, and social intelligence. He argued that these separate factors could never be combined into one general factor of intelligence, as Spearman theorized.

Louis Leon Thurstone also disagreed with Spearman's concept of a single intellectual factor, speculating that, while interconnected, there were distinct factors. Thurstone is credited for the advancement of statistics with his development of multiple factor analysis. Moreover, he is known for his theory of primary mental abilities, a model of cognitive abilities derived from factor analysis. Through several administrations of an assessment battery, Thurstone (1938) originally landed on seven primary mental abilities: *Gc*, word fluency, number facility, memory, visualizing, perceptual speed, and induction (Kent, 2017; Wasserman, 2018). While Thurstone initially challenged Spearman's concept of *g*, he later accepted the concept of a general factor. However, he still believed one single score (i.e., IQ) was not an inadequate measure of cognitive abilities and examining cognitive profiles of strengths and weaknesses would provide a better understanding (Wasserman, 2018). Of note, Cattell (1941) pointed out that *g* could still be a derivative of Thurstone's primary mental abilities.

Philip Ewart Vernon (1950) developed the first hierarchical model of intelligence wherein *g* was the overarching factor with two lower-order factors below it: verbal/educational (v:ed) and spatial/mechanical (k:m; Wasserman, 2018). Even further, Vernon proposed there were narrow and specific factors below the verbal ability and spatial ability factors. While this may have been considered an oversimplification, this dichotomous split is similar to the verbal-performance (i.e., verbal and nonverbal) split seen in the Wechsler instruments.

Finally, sociological theories of intelligence focused on the societal influences on the development of intelligence through language, attitudes, and behaviors (Cianciolo & Sternberg, 2004). Sociological theories were largely driven by psychologist Lee Vygotsky. Vygotsky (1978) posited that people utilize psychological tools (e.g., language, thinking styles, images) to improve the mental capabilities of other individuals. Vygotsky (1978) developed the concept of zone of proximal development wherein the zone was characterized by the difference in capabilities when an individual is assisted and unassisted. This viewpoint on cognitive development with an emphasis on social influences differs from previous biological considerations. While Vygotsky's concepts do not directly relate to a theory of intelligence, his work influenced intelligence theory by integrating social factors and cognitive development (Cianciolo & Sternberg, 2004).

Contemporary Theoretical Perspectives

Gf-Gc Theory of Intelligence

In the 1930s, Raymond Bernard Cattell attempted to answer his doubts of a general ability factor as Spearman conceived (Cattell, 1971). Cattell developed a theory of intelligence based on *Gf* and *Gc* intelligence, which he presented at the 1941 APA National Convention. Cattell's Gf-Gc theory was influenced by Thurstone's theory of intelligence, characterized as distinct, interconnected, abilities, without a general factor (i.e., *g*) as Spearman conceptualized. Cattell's (1943) original definition of fluid ability referred to the "ability to discriminate and perceive relations between any fundaments, new or old" (p. 178). This ability was said to develop through adolescence followed by a slow decline in adulthood. Later definitions of *Gf* included inductive and deductive reasoning abilities where individuals must adapt and learn from their environment rather than relying on prior knowledge (Flanagan & Dixon, 2013; Wasserman, 2018). Cattell (1963, 1971) posited that *Gf* intelligence dictated the acquisition of knowledge; individuals with higher levels of *Gf* intelligence will have fewer limitations on their ability to learn information (Schneider & McGrew, 2012). *Gf* intelligence is also susceptible to neurological insults such as adverse events and aging (Kent, 2017). Conversely, Cattell (1943) defined *Gc* ability as consisting of "discriminatory habits long established in a particular field, originally through the operation of fluid ability, but no longer requiring insightful perception for their successful operation" (p.178). While *Gf* intelligence peaks in adolescence to young adulthood, *Gc* intelligence is predominant in adulthood. *Gc* intelligence refers to the ability to acquire knowledge and the accessibility of knowledge, largely influenced by culture and one's development of *Gf* intelligence (Cattell, 1943; Wasserman, 2018).

Twenty years elapsed before Cattell revisited his Gf-Gc theory. In 1963, Cattell published an update to his Gf-Gc theory after conducting studies to validate his original theory. Specifically, Cattell examined the data utilizing factor analytic studies of 277 seventh- and eighth-grade students. The measures used in the study consisted of Thurstone's primary abilities (i.e., verbal, spatial, reasoning, number, and fluency), four subtests from the Institute of

Personality and Ability Testing (IPAT) Culture Fair Intelligence Test, Scale 2a (i.e., Perceptual Series, Perceptual Classification, Matrices, and Topology), and the High School Personality Questionnaire (HSPQ; Cattell, 1963). The guiding hypothesis of the *Gf-Gc* theory was that certain abilities would load separately onto either the *Gc* ability factor or the *Gf* ability factor; Cattell's research corroborated his hypotheses. The results of this study supported the concept of two general abilities: a general *Gf* ability factor correlated with a *Gc* ability factor.

John Leonard Horn, a doctoral student of Cattell, further expanded on Cattell's model in his dissertation to include several other broad abilities. Horn demonstrated that *Gf* and *Gc* intelligence have separate developmental trajectories. Together, Cattell and Horn expanded the model to include five factors (i.e., visualization, short-term memory, long-term storage and retrieval capacity, and speed of processing; Horn & Cattell, 1966). In 1968, Cattell and Horn developed nine main factors based on multiple factor analytic studies. These nine factors were based on intellectual (*Gf* and *Gc* intelligence) and personality factors (i.e., social desirability and carelessness; Kent, 2017). Cattell (1987) would later use the term "investment" to explain the strong correlation between *Gf* and *Gc*; for individuals with lower *Gf*, the amount of effort in learning does not return as much as those with higher *Gf* (Schneider & McGrew, 2012). Throughout the next few decades, Horn, Cattell, and several other researchers expanded the model to include additional factors. In 1991, Gf-Gc was expanded to include the following additional abilities: *Gf*, *Gc*, SAR or *Gsm*, *Gv*, *Ga*, TSR or *Glr*, *Gs*, *CDS*, and *Gq* (Schneider & McGrew, 2012).

The Three-Stratum Theory of Cognitive Abilities

John Bissell Carroll (1993) disagreed with the two-factor model developed by Cattell and Horn. Instead, Carroll (1993) proposed a hierarchical model of cognitive abilities based on three strata in his seminal work entitled "Human Cognitive Abilities: A Survey of Factor-Analytic Studies," where he stated the Gf-Gc model:

> …appears to offer the most well-founded and reasonable approach to an acceptable theory of the structure of cognitive abilities. The major revision I would make about it is that it appears not to provide a third-order g factor to account for correlations among the broad second-order factors. (p. 62)

Through a comprehensive analysis of over 450 ability matrices, Carroll managed to develop an empirically based model of cognitive abilities utilizing research dating back to Spearman's theory of *g* (Kent, 2017; Schneider & McGrew, 2012). Within Carroll's (1993) model, the highest stratum, stratum III, contains the general factor, *g,* subsuming all broad and narrow abilities below it. Stratum II contains eight factors that are considered broad abilities. Carroll's eight factors, presented in relation to their association with *g* are as follows: *Gf*, *Gc*, *Gy* or *Gsm*, *Gv*, *Ga*, retrieval ability (*Gr*), *Gs*, and reaction time/decision speed (*Gt*; Carroll, 1993; Schneider & McGrew, 2012). Finally, stratum I contains over 70 narrow, specific abilities (Flanagan & Dixon, 2013; Wasserman, 2018). Carroll's three-stratum theory provided a way to organize decades of research and helped further the understanding of intelligence theory.

While the Gf-Gc model and Carroll's three-stratum theory have many similarities, there are several major differences. Carroll's model consists of *g* at stratum III while the Gf-Gc model does not contain an overarching intellectual ability factor (Flanagan & Dixon, 2013). Second, Carroll subsumes *Gq* under *Gf*, while the Gf-Gc theory contains *Gq* and leaves *Gq* as its own broad ability. Third, Gf-Gc includes a separate broad ability for reading and writing (*Grw*), while Carroll subsumes these abilities under *Gc*. Finally, both Cattell-Horn and Carroll differ in their interpretation of narrow memory abilities (Flanagan & Dixon, 2013). The WJ-R (Woodcock &

Johnson, 1989) became the first intelligence battery to utilize the *Gf-Gc* theory, bridging the gap between theory and practice.

The CHC Theory of Cognitive Abilities

Prior to the development of the CHC theory, Carroll's three-stratum model (1993) and Cattell-Horn's Gf-Gc model (1963) remained the two main theories of intelligence with dispute ongoing in the field on how intelligence should best be conceptualized (Horn & Blankson, 2012). Carroll's three-stratum theory was integrated with Horn and Cattell's Gf-Gc theory to create the CHC theoretical framework, first proposed in 1997 by Kevin McGrew (McGrew, 1997). This newly proposed CHC theory was further refined by McGrew, Flanagan, and colleagues and eventually accepted by Horn and Carroll.

The CHC theory posits that cognitive abilities have a hierarchical structure. The base of the hierarchical model of CHC theory consists of specific, directly measurable abilities related to a specific task (i.e., ability to repeat sentences; Schneider & McGrew, 2018). Narrow, broad, and general abilities are latent variables taken from observations among specific abilities (Schneider & McGrew, 2018). More specifically, narrow abilities consist of highly correlated abilities explained by a particular theoretical concept. For example, the ability to repeat sentences may be correlated with the ability to repeat digits; these specific abilities may be explained by an individual's auditory short-term memory capacity (Schneider & McGrew, 2018). Broad abilities (i.e., working memory capacity) are a compilation of highly correlated narrow abilities (i.e., auditory and visual short-term memory capacity, attentional control). Broad abilities within CHC theory begin with a capital G (i.e., general), followed by specific lowercase letters for that particular ability (i.e., *Gwm* for working memory). The introduction of the CHC theory of intelligence included nine broad factors: *Gf, Gc, Gv, Glr, Gsm, Gs, Ga, Gq,* and reading and writing (*Grw*; Flanagan & McGrew, 1997).

Prior to its conceptualization, factor analytic studies of existing intelligence tests indicated the existence of the CHC theory. In fact, upon reexamination of the factor structure, Carroll found that the 1989 WJ-R (Woodcock & Johnson, 1989) aligned closely with the CHC theory of intelligence (Carroll, 2003). This analysis paved the way for future versions of WJ batteries to utilize the CHC structure in their test development (Schneider & McGrew, 2018). While the WJ-R was originally based on Gf-Gc theory, the underlying structure was consistent with CHC theory (Keith & Reynolds, 2010). Ten years later, the authors and publishers of the Woodcock-Johnson III Tests of Cognitive Abilities (WJ-III; McGrew & Woodcock, 2001; Woodcock et al., 2001) and the SB-5 (Roid, 2003), had the opportunity to meet with Horn and Carroll to discuss their theories in regard to the new intelligence instruments (Newton & McGrew, 2010). Horn and Carroll ultimately agreed to align their theories with the recently developed CHC theory of cognitive abilities (Newton & McGrew, 2010). Since its introduction, the CHC theory has been researched and validated throughout the literature, mostly through factor analytic studies. The WJ-III was the first instrument to be entirely based on CHC theory. Further, studies have shown evidence of the CHC theory through factor analytic studies of assessment instruments (McGrew & Woodcock, 2001).

While the WJ instruments were the first to be developed with CHC theory in mind, other intelligence instruments have followed suit in their subsequent revisions. The Kaufman Assessment Battery for Children, Second Edition (KABC-II; Kaufman & Kaufman, 2018) provides both the CHC theory of intelligence and Lurian theory for clinicians to interpret the results. Though the Differential Ability Scale, Second Edition (DAS-II; Elliot, 2007) was developed based on several different theories of intelligence, including *Gf-Gc* theory, it is still consistent with several CHC broad abilities (Keith & Reynolds, 2010). Finally, factor analytic studies supported a CHC-based five-factor model in the SB-5 (Roid, 2003). This evidence, in addition to numerous research studies on the development, trajectory, and relationships to neuropsychology constructs, has led to the CHC theory being the primary psychometric theory of intelligence. However, while several broad abilities of CHC theory are represented (i.e., fluid reasoning, crystalized knowledge, visual-spatial processing, and short-term memory), other abilities, such as auditory processing and long-term storage and retrieval, are often underrepresented in intelligence assessments (Flanagan et al., 2013). The WJ IV (Schrank et al., 2014) batteries, consisting of the Tests of Cognitive Abilities, the Tests of Achievement, and the Tests of Oral Language, is the only measure to assess for all nine of the original CHC factors.

The incorporation of CHC theory into the Wechsler instruments was more gradual. The first revision to include theory aligning with CHC theory was seen in the revision of the WISC-III to the WISC-IV. Further integration was seen in the development of the WISC-V (Schneider & McGrew, 2018). While the Wechsler developers integrated CHC theory throughout the last two revisions, the Wechsler instruments still utilized other theories in addition to revisions based on clinical utility, as historically seen in the development of the original Wechsler instruments.

In 2012, Joel Schneider and Kevin McGrew proposed a reconceptualized version of CHC theory. The basis for this reconceptualization was the identification of 16 broad abilities and the need to organize these abilities between functional and conceptual groupings in an information processing model (Schneider & McGrew, 2012). The three domains within this proposed model include acquired knowledge, sensory-motor domain-specific abilities, and domain-independent general capacities. At the top of Schneider and McGrew's (2012) CHC map is acquired knowledge, consisting of *Gc*, domain-specific knowledge (*Gkn*), *Grw*, and *Gq*. Within the sensory-motor domain, the proposed sensory-linked abilities include *Gv*, *Ga*, olfactory abilities (*Go*), tactile abilities (*Gh*), kinesthetic abilities (*Gk*), and psychomotor abilities (*Gp*), which determine perceptual processing abilities in an individual (Schneider & McGrew, 2012). Within the domain-independent abilities lies parameters of cognitive efficiency, including *Gsm, Glr,* psychomotor speed (*Gps*)*, Gs,* and decision speed (*Gt*), with *Gf* the sole domain-free ability. These abilities are not associated with any specific sensory functions.

In 2017, Woodcock et al. proposed a revision to the CHC theory that reconceptualized the structure of intelligence. The purpose behind this revision was to provide practitioners with a practical, functional theoretical model. This newly proposed theory, the F-CHC includes three theoretical domains: acquired knowledge, thinking abilities, and cognitive efficiency. Further, Woodcock et al. (2017) and Woodcock et al. (2018) reduced the number of narrow abilities that comprise each broad ability to two factors. This reduction in narrow abilities did not result in a loss of data or precision required for clinical interpretation. As this is an emerging theory, further validation support is needed to corroborate the F-CHC Model.

In 2018, Schneider and McGrew proposed yet another revision to the CHC theory. Prior to their update, other revisions had been made including deletion, consolidation, and reorganization of abilities (Schneider & McGrew, 2018). For example, *Gsm* was renamed as working memory capacity (*Gwm*) to reflect new research and theoretical developments (Schrank et al., 2014), in addition to other logistical reclassifications. Perhaps the most dramatic shift from Schneider and McGrew's 2012 version of CHC to their 2018 version is the division of *Glr* into long-term storage and learning (*Gl*) and retrieval fluency (*Gr*). In this new conceptualization, *Gl* is characterized as the effort required to store new information into long-term memory, and *Gr* is the speed and ease at which said information can be recovered from long-term memory (Schneider & McGrew, 2018). Additionally, emotional intelligence (*Gei*), defined as the ability to understand social and emotional behaviors, was hypothesized to be a broad ability similar to those of the current CHC abilities. *Gf* is suspected to overlap with other broad abilities of CHC theory (i.e., *Gc, Gq,* and *Gv*) and is based on a process component and a content component (Schneider & McGrew, 2018).

Another radical update by Schneider and McGrew (2018) to the CHC theory was the revision of *Gs*. This revision suggests that there is an overall general speediness that is comprised of four broad abilities: broad psychomotor speed (*Gps*), broad decision speed (*Gt*), broad cognitive speed (*Gs*), and *Gr* (Schneider & McGrew, 2018). These broad abilities are then further comprised of narrow abilities. Regarding *Ga*, the authors proposed that there should be a distinction between verbal and nonverbal abilities. Within the *Gv* domain, Schneider and McGrew (2018) suggested the addition of large-scale spatial navigation abilities. Finally, domains previously considered less essential (i.e., *Gh*, *Gk*, and *Gt*) have been elevated based on recent research and theory on embodied cognition. Clearly, the CHC theory of intelligence has proven to be a complex conceptualization of intelligence that is continuously updated based on current research. Further, CHC theory has since become the contemporary means of interpreting results of assessment instruments in the field of psychological science. Thus, practitioners should be well-familiarized with the intricacies of CHC when utilizing and interpreting data of intelligence instruments.

The Evolution of the Wechsler Instruments

David Wechsler studied under several prominent psychologists such as Cattell, Thorndike, and Spearman, who had their own theories of intelligence and testing which

Wechsler attempted to combine into a single approach. This proved challenging as Spearman and Thorndike debated the existence of a general intellectual factor (Tulsky et al., 2003). Wechsler's experience with Army Alpha and Army Beta testing inspired his own conceptualization of intelligence and testing; he speculated that intelligence could be measured by various verbal and nonverbal tasks. Wechsler adopted Spearman's concept of *g* while still aligning with Thorndike's views of specific abilities of intelligence to create the Wechsler-Bellevue Intelligence Scale, Form I (Tulsky et al., 2003). The Wechsler-Bellevue featured the FSIQ as a reflection of general intelligence. Wechsler discussed the limitations of the FSIQ as follows:

> It thus appears that the entity or quantity which we are able to measure by intelligence tests is not a simple quantity. Certainly it is not something which can be expressed by one single factor alone, say "g", whether you define it in its most general terms as mental energy, the ability to reduce relations or merely as the intellectual factor. Intelligence is this and yet something more. (Wechsler, 1939a, p. 11)

Wechsler determined one should examine two broad abilities, verbal and performance ability, in addition to the FSIQ. He took this opportunity to focus more on specific factors that comprise general intelligence rather than just focusing on general intelligence. Wechsler's subsequent instruments included subtests designed to address various cognitive abilities, such as verbal and nonverbal reasoning, and practical intelligence (Tulsky et al., 2003). Therefore, Wechsler was able to develop an instrument that combined Spearman and Thorndike's separate theories into one cohesive theory of intelligence, positing that intelligence consists of global and specific abilities.

However, Wechsler's original scales were not based on theory per se; rather, he based his tests on practical and clinical applications (Lichtenberger & Kaufman, 2012). Nevertheless, Wechsler's concept of intelligence is consistent with the contemporary CHC theory of intelligence, which would be developed almost 50 years later (Kaufman et al., 2016). While CHC theory became the contemporary theory for many current intelligence tests, it was not the only influence in revisions of the Wechsler instruments. The Wechsler revisions took into consideration functional theories in addition to structural theories, including theories from cognitive psychology, neuropsychology, and specific cognitive abilities such as memory, attention, and executive functioning, in an attempt to create a comprehensive theoretical basis (Kaufman et al., 2016; Wechsler, 2008, 2012, 2014).

For example, the WPPSI-IV and WISC-V utilize working memory models in their development: the multi-component model and the embedded-process model. The multi-component model (Baddeley, 2000; Wechsler, 2012, 2014) stipulates two memory storage systems that temporarily store or manipulate information: the phonological loop for verbal information and the visual-spatial domain for visual and spatial information. The embedded-process model (Cowan, 1999; Wechsler, 2012, 2014) posits that information in working memory is developed when a subset of long-term memories are activated due to stimuli from the environment bringing these memories to attention.

The Wechsler-Bellevue Intelligence Scale

After the creation of the Binet intelligence scales, Wechsler noted that there was not an assessment instrument specifically designed for adults, which led him to question the validity of the Binet scales for use with adults. These limitations led to Wechsler's development of the first adult intelligence test, which he created during his time at the Bellevue Psychiatric Hospital in New York as the chief psychologist (Holdnack, 2019). The Wechsler-Bellevue Intelligence

Scale, Form I (WBIS-I) was developed in 1939 (Wechsler, 1939b), followed by Form II (WBIS-II) in 1946.

The WBIS was based on a two-part structure including the Verbal Intelligence Quotient (VIQ) and the Performance Intelligence Quotient (PIQ); combined, these formed the Full-Scale Intelligence Quotient (FSIQ; Weiss et al., 2016). Although Wechsler divided his subtests into verbal and performance domains, this does not necessarily mean Wechsler viewed intelligence as a two-factor structure. Wechsler (1958) explained that the purpose of grouping subtests into verbal and performance domain:

> …does not imply that these are the only abilities involved in the tests…The subtests are different measures of intelligence, not measures of different kinds of intelligence, and the dichotomy of Verbal and Performance areas is only one of several ways to in which the tests could be grouped. (p. 64)

Wechsler developed the WBIS based on subtests from the 1937 SB (i.e., Comprehension, Arithmetic, Digit Span, Similarities, and Vocabulary) and the Army Exam (i.e., Picture Arrangement), Koh's Block Design (i.e., Block Design), Army Alpha (i.e., Information and Comprehension), Army Beta (i.e., Digit Symbol-Coding), Healy Picture Completion (i.e., Picture Completion), and Pinther-Paterson Test (i.e., Object Assembly; Groth-Marnat, 2009). The full WBIS included 11 subtests, and Wechsler decided to remove the organization of items by mental age in favor of a point-scale where credit was given for correct items and summed together for each subtest total raw score, then converted to standard scores (Holdnack, 2019). Wechsler's use of standard scores was more psychometrically-sound than Terman's mental age divided by chronological age (Lichtenberger & Kaufman, 2012). While innovative, the WBIS had numerous technical issues including subtest reliability, sample size, and representativeness.

Wechsler Adult Intelligence Scale

The original WAIS was developed in 1955 as an adaptation of the WBIS-I designed for adults. While the WBIS included norms from a small sample of adults, the WAIS was able to expand those norms from a nationally stratified sample based on census data (Wechsler, 2008). With the revision of the WBIS to the first edition of the WAIS, there was an increase in the number of items for nine out of the 11 subtests (Lichtenberger & Kaufman, 2012; Wechsler, 1955). Subsequent revisions included the WAIS-R in 1981, the WAIS-III in 1997, and the WAIS-IV in 2008. The WAIS (1955) and WAIS-R (1981) included six subtests in the Verbal Scale and five subtests in the Performance Scale (Lichtenberger & Kaufman, 2012). While no new subtests were introduced, the WAIS-R included new items, dropped items, and made changes in administration and scoring (Wechsler, 1981).

The revision of the WAIS-R to the WAIS-III included updated norms, modified items with higher floors and ceilings, extended age ranges, the creation of index and factor scores, validation with other cognitive and achievement measures at that time, and extensive reliability and validity research (Groth-Marnat, 2009). While the WAIS-III maintained the VIQ, PIQ, and FSIQ, the addition of three new subtests allowed practitioners to calculate four index scores. Further, the WAIS-III moved from a two-factor model to a four-factor model, following suit of the WISC-III (Weiss et al., 2016). The four index scores included the VCI, POI, WMI, and PSI.

In 2008, the WAIS-IV (Wechsler, 2008) was adapted to fit the four-factor model designed for the WISC-IV. The WAIS-IV eliminated the dichotomous verbal and performance IQ in favor of four index scores (i.e., VCI, WMI, Perceptual Reasoning Index (PRI), and PSI) in addition to the FSIQ (Groth-Marnat, 2009; Wechsler, 2008). The four index scores were carried over from the WAIS-III with the name change of Perceptual Organization to Perceptual

Reasoning to align with the indexes of the WISC-IV (Lichtenberger & Kaufman, 2012). While the overall verbal and performance IQs were eliminated, the four indexes of the WAIS-IV still allow a verbal (i.e., VCI and WMI) and nonverbal (i.e., PRI and PSI) model. The WAIS-IV also updated the FSIQ which is now calculated from the sum of 10 subtests within the four indexes. This differs from the WAIS-III where the FSIQ was comprised of 11 subtests. Further, only eight of the 10 subtests within the WAIS-IV FSIQ are similar to those of the WAIS-III (Lichtenberger & Kaufman, 2012). While the WAIS-IV eliminated the Verbal and Performance IQ, the General Ability Index (GAI) was added as a new global index score. The GAI is derived from scaled scores on the VCI and PRI subtests, excluding the WMI and PSI.

Wechsler Preschool and Primary Scale of Intelligence

The original WPPSI was developed in 1967 as an adaptation of the WBIS-II designed for young children. The first edition of the WPPSI was developed for young children between the ages of 4 years, 0 months and 6 years, 6 months to fulfil the increased need for preschool assessment (Wechsler, 1967). The WPPSI was later revised in 1989 (WPPSI-R; Wechsler, 1989) and 2002 (WPPSI-III; Wechsler 2002). The 1989 revision (WPPSI-R) expanded the age range to 3 years, 0 months and 7 years, 3 months while retaining the same subtests from the original WPPSI and adding one more subtest (Raiford & Coalson, 2014; Wechsler, 1989). Further, both easier and more challenging items were created to extend the floors and ceilings of certain subtests. The third edition of the WPPSI (Wechsler, 2002) divided the age range into two age groups (i.e., 2 years, 6 months to 3 years, 11 months and 4 years, 0 months to 7 years, 3 months) with each age group consisting of different batteries. Psychometric advances also guided the replacement and addition of certain subtests. The younger age band included three verbal subtests, two performance subtests, and four composite scores (i.e., VIQ, PIQ, FSIQ, and GLC).

The older age band included five verbal subtests, five performance subtests, two Processing Speed subtests, and five composite scores (i.e., VIQ, PIQ, PSQ, FSIQ, and GLC; Wechsler, 2012).

In 2012, the WPPSI-IV (Wechsler, 2012) was the first Wechsler instrument to introduce a five-factor model by developing new subtests to create a FRI (Weiss et al., 2016). The WPPSI-IV eliminated the verbal and performance IQ model in favor of this five-factor model. The WPPSI-IV (Wechsler, 2012) older age band includes five primary index scores: VCI, VSI, FRI, WMI, and PSI, while the younger age band only includes the VCI, VSI, and WMI. A FSIQ, Nonverbal Index, GAI, Cognitive Proficiency Index, and Vocabulary Acquisition Index are also included. Key revisions of the WPPSI-III to the WPPSI-IV include updated theoretical foundations that incorporated contemporary research on intelligence, neurocognitive research, and models of working memory (Raiford & Coalson, 2014; Wechsler, 2014). The developers also increased the level of developmental appropriateness of the subtests and user friendliness. As with each revision of a Wechsler instrument, the psychometric properties were improved to include updated norms, increase evidence of validity and reliability, improve the floors and ceilings of subtests, and reduce item bias (Raiford & Coalson, 2014; Wechsler, 2014). Finally, true to form, these revisions were made to enhance clinical utility, including new comparison score methods in order to aid interpretation.

Wechsler Intelligence Scale for Children

The original WISC was developed in 1949 as an adaptation of the WBIS-II designed for children and adolescents. The first edition of the WISC followed the two-factor structure of the WBIS and progressed throughout each new revision; the WISC-R had a three-factor structure, and the WISC-III had a four-factor structure (Keith et al., 2006). The original WISC included 11 subtests from the WBIS adapted for children and added one more subtest created specifically for the WISC. The original WISC was revised in 1974 to include a more representative sample (WISC-R; Wechsler, 1974). The WISC-R retained the same 12 subtests from the first edition but included additional ages, shifting from ages 5 to 15 years to ages 6 to 16 years. Additionally, the WISC-R maintained the VIQ, PIQ, and FSIQ.

The WISC-III (Wechsler, 1991) was the first deviation in structure to acknowledge new abilities in addition to the verbal and performance abilities (Keith & Reynolds, 2010). These new index scores included the VCI, POI, Freedom from Distractibility Index (FDI; i.e., memory), and PSI to represent more specific domains of cognitive functioning (Wechsler, 2014; Weiss et al., 2016). While these four factors were offered, they were purely supplemental to the traditional VIQ, PIQ, and FSIQ. Additionally, while the WISC-III retained the same subtests from the WISC-R, an additional subtest was introduced. A cross-battery factor analytic study of the WISC-III and WJ-III showed that domains measured by the WISC-III could also be considered to measure similar constructs of the WJ-III (Keith & Reynolds, 2010).

The WISC-IV (Wechsler, 2003) revisions were considered the largest departure from the original two-factor structure. In 2003, with the release of the WISC-IV, the developers made several large changes; they completely removed the VIQ and PIQ. This newly formed four-factor model of the WISC-IV was intended to be the primary measure that replaced the VIQ and PIQ. The four factors of the WISC-IV included the VCI, PSI, WMI, PRI (Keith et al., 2006). In creating the WISC-IV, developers changed the FDI to the WMI to reflect more accurately what the factor measures (Keith et al., 2006). In a similar fashion, the POI was changed to the PRI as newly added subtests assessed the construct of *Gf*, and the PRI reflected this change (Keith et al., 2006; Weiss et al., 2016). Additionally, five new subtests were added, three subtests were dropped (i.e., Mazes, Object Assembly, and Picture Arrangement), and three subtests were now considered supplemental (i.e., Arithmetic, Information, and Picture Completion; Keith et al., 2006). The revision from the WISC-III to the WISC-IV represented an attempt to incorporate modern CHC theory. At this point, the WISC-IV and WAIS-IV were not developed around a specific theory. Conversely, other intelligence measures at that time such as the Kaufman Assessment Battery for Children (KABC; Kaufman & Kaufman, 2004), Stanford-Binet-5 (SB-5; Roid, 2003) and the Woodcock-Johnson Tests of Cognitive Abilities, Third Edition (WJ-III; McGrew & Woodcock, 2001; Woodcock et al., 2001) were developed in alignment with the CHC theory of intelligence.

In 2013, researchers concluded that a five-factor solution was a better fit for the data of the WISC-IV and WAIS-IV than the original four-factor solution (Wechsler, 2014; Weiss et al., 2013a, 2013b; Weiss et al., 2016). The pivotal Weiss and colleagues (2013a) validation study for the WISC-V five-factor model reorganized the 15 WISC-IV subtests to include an additional fifth factor, *Gf*, consisting of Arithmetic, Matrix Reasoning, and Picture Concepts. Factor analytic studies provided support for the separation of the original PRI into two composites: FRI and VSI. The final five-factor solution included a VCI, VSI, FRI, WMI, and PSI (Weiss et al., 2016). Each index is measured by two subtests with most indexes providing optional secondary subtests. The FSIQ of the WISC-V consists of seven subtests total pulled from each of the indexes. The FSIQ is a well-established measure of general intelligence utilizing decades of research from Spearman, Wechsler, Thurstone, Vernon, Carroll, Horn, and Cattell (Weiss et al., 2016). It is also considered the most reliable score and best predictor of *g* (Wahlstrom et al., 2018). This five-factor solution the Wechsler developers created for the WISC-V and WPPSI-IV has significant overlap with the CHC theory of intelligence.

Wechsler Indexes

Although the latest updates of the Wechsler instruments to a four- and five-factor model were guided by research in neuropsychology and cognition, research on the CHC theory of intelligence has produced similar models (Weiss et al., 2016). Therefore, the Wechsler instruments conceptualization of intelligence is similar to the contemporary CHC theory of intelligence.

The VCI of the Wechsler instruments is similar to the *Gc* domain of the CHC theory. The VCI is derived from two subtests within the VCI, and the two subtests differ across instruments (Wahlstrom et al., 2018). The VCI is intended to measure "verbal concept formation, verbal reasoning, verbal comprehension and expression, acquired knowledge, and practical knowledge and judgement" (Wahlstrom et al., 2018, p. 250). *Gc* subtests are intended to examine the breadth and depth of factual knowledge and the ability to utilize one's own experiences with verbal concepts in a novel way (Wahlstrom et al., 2018). The VSI is similar to the *Gv* domain of the CHC theory. The VSI is intended to measure the "ability to evaluate visual details, understand spatial relations among objects, and construct geometric designs using a model" (Wahlstrom et al., 2018, p. 251). The FRI is the equivalent to the *Gf* domain of the CHC theory. The FRI is intended to measure one's ability to use inductive and deductive reasoning to solve problems. The younger age band of the WPPSI-IV does not contain a FRI due to the level of comprehension required to solve the tasks.

The WMI is similar to the *Gwm* domain of CHC theory. The WMI is intended to examine "the ability to consciously register, maintain, and manipulate auditory and visual information" which requires "paying attention and focusing, keeping the information in conscious awareness, mentally processing the information in a manner that conforms to the task demands, and then providing a result" (Wahlstrom et al., 2018, p. 253). Proactive interference is utilized for *Gwm* subtests; "proactive interference involves repeated exposure to stimuli across items, such that prior exposure interferes with memory for the present item" (Wahlstrom et al., 2018, p. 253). The PSI is equivalent to the *Gs* domain of the CHC theory. The PSI involves the use of reasoning in identifying solutions to problems. As with the FRI, the younger age band of the WPPSI-VI does not contain a PSI due to younger children having difficulty understanding the task of working quickly.

Models Proposed for the Present Study

Factor Structure of the Wechsler Instruments

The Wechsler instruments examined in the present study were developed with theories and cognitive models from several domains. While the Wechsler instruments utilized some broad abilities as described by CHC theory, other theories and models were derived from cognitive psychology, neuropsychology, and specific cognitive abilities such as memory, attention, and executive functioning. The Wechsler instruments currently assess general intelligence (i.e., FSIQ) and a combination of several broad abilities: *Gc, Gv, Gf, Gwm,* and *Gs.* The WAIS-IV utilizes a four-factor structure while the WPPSI-IV and the WISC-V utilize a five-factor structure of cognitive abilities.

The CHC Theory of Cognitive Abilities

The prominent theory of intelligence with the most empirically supported research base is the CHC theory of cognitive abilities. The CHC theory is supported through numerous factor analytic studies of cognitive assessments and has even guided the development of revisions of said cognitive assessments. CHC theory is an integration of the Cattell-Horn Gf-Gc model and Carroll's three-stratum theory of cognitive abilities. The contemporary CHC theory consists of an overall intellectual ability (i.e., *g*) subsuming 16 broad abilities and further subsuming 80 narrow cognitive abilities (Schneider & McGrew, 2012). While this model is utilized by many cognitive assessments today, the Wechsler instruments are developed based on a conglomeration of theories, including CHC theory.

The F-CHC Model of Cognitive Abilities

Richard W. Woodcock et al. (2017) and Woodcock et al. (2018) proposed a new, practical approach to the original CHC theory by grouping cognitive abilities into three broad domains: acquired knowledge, thinking abilities, and cognitive efficiency. Within the acquired knowledge domain lies *Gc,* broad reading (*Grw-R*), broad writing (*Grw-W*), broad mathematics (*Gq*), and psychomotor abilities (*Gp*). The thinking abilities domain includes *Gv*, *Ga*, *Glm*, and *Gr*. Finally, within the cognitive efficiency domain is conscious memory (*Gcm*) and *Gs.* The F-CHC nomenclature also reconceptualized the narrow abilities included in each domain to propose that only two narrow abilities are in each domain. As this model is a recent proposal, there is limited validation research. A dissertation by Nadine Ndip (2020) attempted to validate the F-CHC model using the WJ IV (Schrank et al., 2014) batteries: The Tests of Cognitive Abilities, the Tests of Achievement, and the Tests of Oral Language. Through factor analytic studies, Ndip (2020) found that the factor structure of the WJ IV batteries deviated from what was originally presented in the Technical Manual and supported some of the broad and narrow abilities proposed by the F-CHC model. Ndip (2020) concluded that the F-CHC model of cognitive abilities best fit the data from the WJ IV standardization sample. The present study sought to add to the validation of the F-CHC model utilizing the Wechsler cognitive assessment instruments.

CHAPTER III

METHOD

The following chapter outlines the methodology for the present study, including participants, procedures, and materials. Detailed information regarding the psychometric properties of the measures and statistical procedures used in the analyses are provided.

The normative data set for the Wechsler instruments was requested and received from the publisher. The cognitive measures of the Wechsler instruments (i.e., WAIS-IV, WPPSI-IV, and WISC-V) were similarly normed; therefore, the procedures, validity, and reliability are outlined together. These measures will be discussed as a whole with individual differences addressed in the respective sections.

Participants

The publisher used a stratified sampling plan to ensure that the normative sample included representative proportions of children, adolescents, and adults across selected demographic factors. Consequently, participants in the normative sample comprised a demographically diverse group consisting of various ages, sexes, races/ethnicities, parent education level, and geographic regions.

The WAIS-IV normative data set consists of 2,200 participants divided into 13 age groups: 16:0–17:11, 18:0–19:11, 20:0–24:11, 25:0–29:11, 30:0–34:11, 35:0–44:11, 45:0–54:11, 55:0–64:11, 65:0–69.11, 70:0–74:11, 75:0–79:11, 80:0–84:11, and 85:0–90:11. The nine younger age groups each consist of 200 participants while the four older age groups each consist of 100 participants. The normative sample consists of an equal number of male and female participants in each age group with the exception for the age groups within the 65:0–90:11 range. This age group included more women than men which is consistent with census data.

The WPPSI-IV data set consists of 1,700 participants divided into 9 age groups: 2:6–2:11, 3:0–3:5, 3:6–3:11, 4:0–4:5, 4:6–4:11, 5:0–5:5, 5:6–5:11, 6:0–6:11, and 7:0–7:11. Each age group consists of 200 children with the exception for the 7:0–7:11 age group, which consists of 100 children. Each age group contains an equal number of male and female participants.

The WISC-V data set consists of 2,200 participants divided into 11 age groups: 6:0–6:11, 7:0–7:11, 8:0–8:11, 9:0–9:11, 10:0–10:11, 11:0–11:11, 12:0–12:11, 13:0–13:11, 14:0–14:11, 15:0–15:11, and 16:0–16:11. Each age group consists of an equal number of male and female participants.

Procedures

An archival data set previously obtained through the WAIS-IV, WPPSI-IV, and WISC-V norming studies was used for the analyses. Descriptions of the procedures used to collect the WAIS-IV, WPPSI-IV, and WISC-V are provided in the instrument's respective technical manuals (Wechsler, 2008, 2012, 2014) and briefly summarized here. Data for the WAIS-IV were collected in October 2005. Data for the WPPSI-IV were collected between December 2010 and May 2012. Data for the WISC-V were collected between April 2013 and March 2014. Norming study participants were identified by trained recruiters and independent examiners based on inclusion criteria of the standardization sample who also fit the sampling plan matrix (Wechsler, 2008, 2012, 2014). Examiners and the participant, or the participant's parent or guardian, were paid for their participation. Data were gathered by examiners with extensive experience testing children. Potential examiners submitted information regarding educational, professional, and assessment experience in addition to certification and licensing status information. Examiners reviewed the test kit materials and a summary of common administration and scoring errors prior to the first testing session. For the WAIS-IV, examiners were required to complete a training quiz consisting of questions about administration and scoring rules. Any errors were required to be reviewed and corrected, and selected examiners scored 100% correct on the training quiz. Following the first testing session, examiners submitted the results and awaited feedback prior to testing additional participants. Any scoring and administration errors were discussed with the examiner. Periodic updates regarding frequent scoring or administration errors were sent to the examiners (Wechsler, 2008, 2012, 2014).

Data collected by the examiners were then scored by trained personnel who were required to have a bachelor's degree at minimum and attend a training program conducted by the research team. Due to the subjective nature of the *Gc* subtests, two scorers were assigned with any discrepancy resolved by a third scorer, the resolver (Wechsler, 2008, 2012, 2014). All scorers received feedback on errors and received additional training as needed. Members of the research team were resolvers for a minimum of 10 cases per scorer to ensure accuracy over the course of the standardization process. The average agreement between scorers for nonverbal subtests was greater than 98%, and the average agreement for the subtests requiring verbatim response was greater than 95% (Wechsler, 2008, 2012, 2014). Scorers were instructed not to discuss scoring with one another to prevent scoring drift. Following data collection, the research team examined verbatim responses for scoring drift that may have occurred.

Measurement Instrumentation

The data set for the present study consisted of subtests from the Wechsler cognitive assessment instruments, the WAIS-IV (Wechsler, 2008), WPPSI-IV (Wechsler, 2012), and WISC-V (Wechsler, 2014).

Reliability

The Wechsler instruments were developed according to the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA] et al., 2014). Reliability (*r*) refers to a test score's accuracy, consistency, and stability across various situations (Wechsler, 2008, 2012, 2014). Each test score is an approximation of an individual's hypothetical true test score, and the difference between the actual test score and the hypothetical true test score is defined as measurement error. A test is said to be reliable if it has a relatively small amount of measurement error and produces similar results within the administration and on separate occasions (Wechsler, 2008, 2012, 2014). Further, a Pearson correlation coefficient of at least .80 is considered highly reliable.

The Wechsler instruments provide several types of reliability analyses, including test-retest reliability, split-half reliability, internal consistency, and interscorer agreement. Evidence of internal consistency was obtained utilizing the split-half method. The split-half reliability coefficient refers to the correlation between total scores of two half-tests (Field, 2013) that is then corrected by using the Spearman-Brown formula for the full test (Wechsler, 2008, 2012, 2014). Split-half coefficients were not used for the *Gs* and *Glr* (i.e., Coding, Symbol Search, Cancellation, Naming Speed Literacy, Naming Speed Quantity, Immediate Symbol Translation, Delayed Symbol Translation, Bug Search, and Animal Coding); therefore, test-retest coefficients were used to estimate reliability for these subtests. The reliability coefficient for all other subtests is found by taking the correlation between an individual's scores on the first and second testing session then corrected for the variability of the norming sample. Internal consistency reliability coefficients were calculated for subtests, process, and composite scores by age group and overall sample. Average reliability coefficients of the overall normative sample were calculated using Fisher's *z* transformation. The standard error of measurement (SE) was calculated and expressed in terms of units of standard score. The SE is inversely related to reliability; as reliability increases, SE decreases. Evidence of test-retest reliability for the WAIS-IV was obtained by administering the WAIS-IV twice within a range of 8-82 days with a mean of 22 days. The WAIS-IV provides stability coefficients for all ages as well as for four separate age groups (16–29, 30–54, 55–69, and 70–90; Wechsler, 2008). Evidence of test-retest reliability for the WPPSI-IV was obtained by administering the WPPSI-IV twice within a range of 7-48 days with a mean of 23 days. The WPPSI-IV provides stability coefficients for all ages as well as for three separate age groups (2:6–3:11, 4:0–5:5, 5:6–7:7; Wechsler, 2012). Evidence of test-retest reliability for the WISC-V was obtained by administering the WISC-V twice within a range of 9-82 days with a mean of 26 days. The WISC-V provides stability coefficients for all ages as well as for five separate age groups (6–7, 8–9, 10–11, 12–13, 14–16; Wechsler, 2014).

During the norming process of the Wechsler instruments, most protocols were scored by two independent scorers. For the WAIS-IV objective subtests, there was a high interscorer agreement ranging from .98 to .99. Subjective subtest interscorer reliabilities were also high at .93 for Similarities, .95 for Vocabulary, .97 for Information, and .91 for Comprehension. For the WPPSI-IV objective subtests, there was a high interscorer agreement ranging from .98 to .99. Subjective subtest interscorer reliabilities were also high at .99 for Similarities, .96 for Vocabulary, .96 for Information, .97 for Comprehension, and .99 for Picture Naming. For the WISC-V objective subtests, there was a high interscorer agreement ranging from .98 to .99. Subjective subtest interscorer reliabilities were also high at .98 for Similarities, .97 for Vocabulary, .99 for Information, and .97 for Comprehension. The aforementioned evidence of reliability provides strong support for the precision of WAIS-IV, WPPSI-IV and WISC-V scores and their use for valid interpretation (Wechsler, 2008, 2012, 2014).

Validity

The Wechsler instruments provide several types of validity analyses, including content validity, convergent and discriminant validity, and construct validity. Content validity refers to the relationship between the test's content and the intended construct that is measured. Content validity was evaluated by completing comprehensive literature reviews and advisory panel reviews of items and subtests at each stage of research and development. Response process validity was also utilized to provide support that the participant utilizes the expected cognitive process when completing a task. Evidence of this is provided by psychometric analyses and theoretical sources. Confirmatory factor analytic research and mean score comparisons between special and matched control groups were utilized to provide evidence of construct validity (Wechsler, 2008, 2012, 2014). Examining the internal structure provides support by evaluating interrelations among items, subtests, and composite scores. Convergent validity refers to associations between closely related constructs while discriminant validity refers to associations between distantly related constructs, both of which are important when determining the constructs a certain test is measuring and providing evidence for the accuracy of inferences of the test scores. For the WAIS-IV (Wechsler, 2008), evidence of convergent and discriminant validity was obtained by examining nonclinical samples of the WAIS-IV in relation to the WAIS-III, WISC-IV, WMS-III, CMS, WIAT-II, Brown ADD, D-KEFS, CVLT-II, and RBANS. For the WPPSI-IV (Wechsler, 2012), evidence of convergent and discriminant validity was obtained by examining nonclinical samples of the WPPSI-IV in relation to the WPPSI-III, Bayley-III, WISC-IV, DAS-II, NNAT2, NEPSY-II, WIAT-III, and BASC-2 Parent Rating

Scales. Finally, for the WISC-V (Wechsler, 2014), evidence of convergent and discriminant validity was obtained by examining nonclinical samples of the WISC-V in relation to the WISC-IV, WPPSI-IV, WAIS-IV, KABC-II, KTEA-3, WIAT-III, Vineland-II, and BASC-2 Parent Rating Scales.

Instruments

The WAIS-IV (Wechsler, 2008) battery includes 15 subtests measuring global intellectual functioning, broad cognitive abilities, and narrow cognitive abilities. The WAIS-IV consists of core and supplemental subtests. The first 10 subtests yield the FSIQ, which includes three subtests each *Gc* and *Gf*, and two subtests each of *Gwm* and *Gs*. Another cognitive composite yielded by the subtests of the WAIS-IV is the GAI. The GAI includes three subtests each of *Gc* and *Gf*.

The subtests that comprise the WPPSI-IV (Wechsler, 2012) battery are determined by age bands provided by the publisher. The WPPSI-IV battery for ages 2:6 to 3:11 years consists of 7 subtests measuring global intellectual functioning, broad cognitive abilities, and narrow cognitive abilities. The FSIQ is comprised of 5 subtests, including two each of *Gc* and *Gv*, and one of *Gwm*. Other cognitive composites yielded by the subtests of this age band of the WPPSI-IV are the Vocabulary Acquisition Index (VAI), Nonverbal Index (NVI), and GAI. The VAI includes two subtests of *Gc*. The NVI consists of two subtests, one *Gv* and one *Gwm*. The GAI includes two subtests each of *Gc* and *Gv*. The WPPSI-IV battery for ages 4:0 to 7:7 years consists of 13 subtests measuring global intellectual functioning, broad cognitive abilities, and narrow cognitive abilities. The FSIQ is comprised of 6 subtests, including two of *Gc*, and one each of *Gv*, *Gf*, *Gs*, and *Gwm*. Other cognitive composites yielded by the subtests of this age band of the WPPSI-IV are the VAI, NVI, GAI, and Cognitive Proficiency Index (CPI). The VAI includes two subtests of *Gc*. The NVI consists of two subtests of *Gf*, and one subtest each of *Gv*, *Gwm*, and *Gs*. The GAI includes two subtests of *Gc*, and one subtest each of *Gv* and *Gf*. The CPI consists of two subtests, one each of *Gs* and Gwm.

The WISC-V (Wechsler, 2014) battery includes 21 subtests measuring global intellectual functioning, broad cognitive abilities, and narrow cognitive abilities. The WISC-V consists of primary, secondary, and complementary subtests. The first seven subtests yield the FSIQ, which includes one subtest each of *Gv*, *Gwm*, and *Gs*, and two subtests each of *Gc* and *Gf*. Other cognitive composites yielded by the subtests of the WISC-V are the GAI, NVI, Auditory Working Memory Index (AWMI), CPI, and Quantitative Reasoning Index (QRI). The GAI includes two subtests each of *Gc* and *Gf*, and one subtest of *Gv*. The NVI consists of two subtests each of *Gv* and *Gf*, and one subtest each of *Gw*m and *Gs*. The AWMI includes two subtests of *Gwm*. The CPI includes two subtests each of *Gwm* and *Gs*. The QRI consists of two subtests of *Gf*.

The following are descriptions and psychometric properties of the subtests from the WAIS-IV, WPPSI-IV, and WISC-V batteries that were used in the present study. Average reliability coefficients are presented for the overall normative sample from both age groups of the WPPSI-IV, as well as the overall normative sample for the WAIS-IV and WISC-V. The following information was provided from the *Technical and Interpretive Manuals* of the WAIS-IV (Wechsler, 2008), WPPSI-IV (Wechsler, 2012), and WISC-V (Wechsler, 2014).

Block Design

This subtest is available on the WAIS-IV, WPPSI-IV, and WISC-V. This subtest is considered a measure of *Gv*; it assesses the narrow abilities of visualization and spatial relations. On the Block Design subtest, an individual is asked to view a model and/or picture and use two-color blocks to re-create the design within a specified time limit. Earlier items on the WPPSI-IV utilize one-color blocks in addition to two-color blocks. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.87), WPPSI-IV (0.85), and WISC-V (0.84).

Similarities

This subtest is available on the WAIS-IV, WPPSI-IV, and WISC-V. This subtest is considered a measure of *Gc*; it assesses the narrow abilities of lexical knowledge and language development. On the Similarities subtest, an individual is read two words that represent common objects or concepts and asked to describe how they are similar. Earlier items on the WPPSI-IV consist of pictures where an individual is asked to select the response option that is from the same category as two other depicted objects. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.87), WPPSI-IV (0.93), and WISC-V (0.87).

Matrix Reasoning

This subtest is available on the WAIS-IV, WPPSI-IV, and WISC-V. This subtest is considered a measure of *Gf*; it assesses the narrow ability of inductive reasoning. On the Matrix Reasoning subtest, an individual is asked to view an incomplete matrix or series and select the response option that completes the matrix or series. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.90), WPPSI-IV (0.90), and WISC-V (0.87).

Digit Span

This subtest is available on the WAIS-IV and WISC-V. This subtest is considered a measure of *Gwm*; it assesses the narrow abilities of memory span and working memory. On the Digit Span subtest, an individual is asked to recall numbers read by the examiner in the same order, reverse order, and ascending order. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.93) and WISC-V (0.91).

Coding

This subtest is available on the WAIS-IV and WISC-V. This subtest is considered a measure of *Gs*; it assesses the narrow ability of rate-of-test-taking and perceptual speed. On the Coding subtest, an individual is asked to use a key to copy symbols that correspond with simple geometric shapes or numbers within a specified time limit. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.86) and WISC-V (0.82).

Vocabulary

This subtest is available on the WAIS-IV, WPPSI-IV, and WISC-V. This subtest is considered a measure of *Gc*; it assesses the narrow ability of lexical knowledge. On the Vocabulary subtest, an individual is asked to either name depicted objects in pictures or define words read aloud by the examiner. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.94), WPPSI-IV (0.89), and WISC-V (0.87).

Figure Weights

This subtest is available on the WAIS-IV and WISC-V. This subtest is considered a measure of *Gf*; it assesses the narrow ability of *Gq*. On the Figure Weights subtest, an individual is asked to view a scale with missing weight(s) and select the response option that keeps the scale balanced within a specified time limit. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.90) and WISC-V (0.94).

Visual Puzzles

This subtest is available on the WAIS-IV and WISC-V. This subtest is considered a measure of *Gv*; it assesses the narrow abilities of spatial relations and visualization. On the Visual Puzzles subtest, an individual is asked to view a completed puzzle and select the three response options that, when combined, reconstructs the puzzle within a specified time limit. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.89) and WISC-V (0.89).

Symbol Search

This subtest is available on the WAIS-IV and WISC-V. This subtest is considered a measure of *Gs*; it assesses the narrow ability of perceptual speed. On the Symbol Search subtest, an individual is asked to scan search groups and indicate whether target symbols are present within a specified time limit. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.81) and WISC-V (0.81).

Information

This subtest is available on the WAIS-IV, WPPSI-IV, and WISC-V. This subtest is considered a measure of *Gc*; it assesses the narrow ability of general verbal information and retention and retrieval of learned information. On the Information subtest, an individual is asked to answer questions about a broad range of general knowledge topics. Earlier items on the WPPSI-IV contain pictures where an individual is asked to select the response option that best answers a question about a general knowledge topic. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.93), WPPSI-IV (0.89), and WISC-V (0.86).

Letter-Number Sequencing

This subtest is available on the WAIS-IV and WISC-V. This subtest is considered a measure of *Gwm*; it assesses the narrow ability of working memory capacity. On the Letter-Number Sequencing subtest, an individual is asked to recall sequences of numbers and letters read by the examiner in ascending numerical order and then in alphabetical order. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.88) and WISC-V (0.86).

Cancellation

This subtest is available on the WAIS-IV, WPPSI-IV, and WISC-V. This subtest is considered a measure of *Gs*; it assesses the narrow ability of perceptual speed. On the Cancellation subtest, an individual is asked to scan two arrangements of objects, one random and one structured, and mark target objects within a specified time limit. The WAIS-IV contains only a structured arrangement consisting of shapes. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.78), WPPSI-IV (0.76), and WISC-V (0.82).

Comprehension

This subtest is available on the WAIS-IV, WPPSI-IV, and WISC-V. This subtest is considered a measure of *Gc*; it assesses the narrow ability of general verbal information. On the Comprehension subtest, an individual is asked to answer questions based on his or her understanding of general principles and social situations. Earlier items on the WPPSI-IV consist of pictures where an individual is asked to select the response option that represents the best response to a general principle or social situation. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.87), WPPSI-IV (0.91), and WISC-V (0.83).

Arithmetic

This subtest is available on the WAIS-IV and WISC-V. This subtest is considered a measure of *Gq*; it assesses the narrow abilities of math achievement, *Gwm*, and *Gf*. On the Arithmetic subtest, an individual is asked to mentally solve arithmetic problems within a specified time limit. The mean reliability coefficient demonstrated strong reliability for the WAIS-IV (0.88) and WISC-V (0.90).

Picture Completion

This subtest is available on the WAIS-IV. This subtest is considered a measure of *Gf*; it assesses the narrow ability of contextual reasoning. On the Picture Completion subtest, an individual is asked to view a picture with an important part missing and identify the missing part within a specified time limit. The mean reliability coefficient demonstrated strong reliability (0.84).

Picture Span

This subtest is available on the WISC-V. This subtest is considered a measure of *Gwm*; it assesses the narrow ability of memory span. On the Picture Span subtest, an individual is asked to view a stimulus page with one or more pictures for a specified time and then select the picture(s) in sequential order from options on a response page. The mean reliability coefficient demonstrated strong reliability (0.85).

Naming Speed Literacy

This subtest is available on the WISC-V. This subtest is considered a measure of *Gs*; it assesses the narrow abilities of long-term storage and retrieval, speed of lexical access, and naming facility. On the Naming Speed Literacy subtest, an individual is asked to name objects of various size and color, letters, and numbers as quickly as possible. The mean reliability coefficient demonstrated strong reliability (0.86).

Naming Speed Quantity

This subtest is available on the WISC-V. This subtest is considered a measure of *Gs*; it assesses the narrow abilities of long-term storage and retrieval, speed of lexical access, and naming facility. On the Naming Speed Quantity subtest, an individual is asked to name the quantity of squares inside a series of boxes as quickly as possible. The mean reliability coefficient demonstrated strong reliability (0.83).

Immediate Symbol Translation

This subtest is available on the WISC-V. This subtest is considered a measure of *Glr*; it assesses the narrow ability of associative memory. On the Immediate Symbol Translation subtest, an individual is asked to learn visual-verbal pairs and then translate symbol strings into phrases or sentences. The mean reliability coefficient demonstrated strong reliability (0.88).

Delayed Symbol Translation

This subtest is available on the WISC-V. This subtest is considered a measure of *Glr*; it assesses the narrow ability of associative memory. On the Delayed Symbol Translation subtest, an individual is asked to translate symbols into words, phrases, or sentences using recalled visual-verbal pairs from Immediate Symbol Translation. The mean reliability coefficient demonstrated strong reliability (0.87).

Recognition Symbol Translation

This subtest is available on the WISC-V. This subtest is considered a measure of *Glr*; it assesses the narrow ability of associative memory. On the Recognition Symbol Translation subtest, an individual is asked to view a symbol and select the correct translation from response options the examiner reads aloud using recalled verbal-visual pairs from Immediate Symbol Translation. The mean reliability coefficient demonstrated strong reliability (0.82).

Picture Concepts

This subtest is available on the WPPSI-IV and WISC-V. This subtest is considered a measure of *Gf*; it assesses the narrow abilities of inductive reasoning and general verbal information. On the Picture Concepts subtest, an individual is asked to view two or three rows of pictures and select one picture from each row to form a group with a common characteristic. The mean reliability coefficient demonstrated strong reliability for the WPPSI-IV (0.89) and WISC-V (0.83).

Receptive Vocabulary

This subtest is available on the WPPSI-IV. This subtest is considered a measure of *Gc*; it assesses the narrow ability of lexical knowledge. On the Receptive Vocabulary subtest, an individual is asked to select the response option that best represents the word the examiner reads aloud. The mean reliability coefficient demonstrated strong reliability (0.90).

Picture Memory

This subtest is available on the WPPSI-IV. This subtest is considered a measure of *Gwm*; it assesses the narrow ability of working memory. On the Picture Memory subtest, an individual is asked to view a stimulus page of one or more pictures for a specified time and then select he picture(s) from options on a response page. The mean reliability coefficient demonstrated strong reliability (0.91).

Bug Search

This subtest is available on the WPPSI-IV. This subtest is considered a measure of *Gs*; it assesses the narrow ability of perceptual speed. On the Bug Search subtest, an individual is asked to mark the bug in the search group that matches the target bug within a specified time limit. The mean reliability coefficient demonstrated strong reliability (0.83).

Zoo Locations

This subtest is available on the WPPSI-IV. This subtest is considered a measure of *Gwm*; it assesses the narrow ability of visual working memory. On the Zoo Locations subtest, an individual is asked to view one or more animal cards placed on a zoo layout for a specified time and then place each card in the previously viewed locations. The mean reliability coefficient demonstrated strong reliability (0.86).

Object Assembly

This subtest is available on the WPPSI-IV. This subtest is considered a measure of *Gv*; it assesses the narrow ability of visualization. On the Object Assembly subtest, an individual is asked to assemble the pieces of a puzzle to create a representation of an identified object within a specified time limit. The mean reliability coefficient demonstrated strong reliability (0.85).

Animal Coding

This subtest is available on the WPPSI-IV. This subtest is considered a measure of *Gs*; it assesses the narrow ability of perceptual speed. On the Animal Coding subtest, an individual is asked to mark shapes that correspond to pictured animals using a key within a specified time limit. The mean reliability coefficient demonstrated strong reliability (0.75).

Picture Naming

This subtest is available on the WPPSI-IV. This subtest is considered a measure of *Gc*; it assesses the narrow ability of general verbal information. On the Picture Naming subtest, an individual is asked to name depicted objects. The mean reliability coefficient demonstrated strong reliability (0.88).

Data Analysis

The present study evaluated the factor structure of the WAIS-IV, WISC-V, and WPPSI-IV using the F-CHC model of cognitive abilities proposed by Woodcock et al. (2017) and Woodcock et al. (2018). Findings from this study may provide a better understanding of the latent structure of these measures with children, adolescents, and adults. The following research questions were proposed:

1. To what degree does the proposed F-CHC model explain the latent structure of cognitive abilities as measured by data from the Wechsler standardization samples?
    a. Research Hypothesis 1: The data will support the proposed F-CHC model.
    b. Research Hypothesis 2: In comparison to other models, the proposed F-CHC model will be the best fitting model for the data from the Wechsler standardization samples.
2. If the above-mentioned model does not adequately represent the Wechsler standardization samples, is there another factor structure that best fits the data?
    a. Research Hypothesis 3: The results from this analysis will not support the proposed factor structure that is identified by the test publishers of the Wechsler instruments.

Statistical Analyses

Factor analysis is a multivariate statistical procedure used to determine relationships between observed variables by examining Pearson correlations. Observed variables that are highly correlated with each other can be said to represent underlying factors. These factors are not directly measurable, rather, they are latent variables (Meyers et al., 2016). Latent variables are not observable (e.g., intelligence), but they are related to factors that are observable and can be measured (e.g., test scores). Latent variables are also called hypothetical constructs or factors, while observed variables are often referred to as indicators (Kline, 2011). Factor analysis is a broad term that includes exploratory factor analysis, principal components analysis, confirmatory factor analysis, and structural equation modeling (Meyers et al., 2016). EFA is an inductive method that begins with a set of variables to determine the underlying factor structure. Conversely, CFA begins with a predetermined factor structure where the researcher decides which sets of variables are correlated with one another and are associated with a particular factor (Meyers et al., 2016). Structural equation modeling (SEM) is a method designed to determine the direction of the latent variables and their associated observed variables (Meyers et al., 2016). Typically, EFA and CFA are complimentary; EFA is first employed to examine the factor structure followed by CFA to examine if the said factor structure is acceptable.

The present study analyzed the factor structure of the Wechsler instruments. Specifically, this study sought to provide evidence for the hypothesized factor structure of the F-CHC model. As part of the analyses, the following psychometric issues were addressed.

Descriptive Data

The full data set consists of 6,100 participants. The *Technical and Interpretive Manuals* of the Wechsler instruments indicate that the normative data was stratified based on age, sex, race/ethnicity, self or parent education level, and geographic region. Detailed demographic characteristics are provided in each *Technical and Interpretive Manual* of the Wechsler instruments (Wechsler, 2008, 2012, 2014). The data set that was utilized for this study includes un-imputed standardization data requested from the publishers of the Wechsler instruments. Additionally, the Wechsler instruments *Technical and Interpretive Manual Supplement* provides intercorrelation matrices for various age ranges specific to each instrument to exhibit the correlation amongst subtests (Wechsler, 2008, 2012, 2014). There are 11 age groups in the WISC-V, 13 age groups in the WAIS-IV, and 9 age groups in the WPPSI-IV.

Sample Size

Sample size must be carefully considered when conducting an EFA or CFA. Consensus is scant on appropriate sample size recommendations in the literature. However, the reliability of a factor analysis is dependent on the size of the sample (Field, 2013). Numerous guidelines have been outlined for calculating sample size requirements, such as: minimum sample size of 100 or 200, 5 or 10 observations per parameter, or 10 cases per variable (Wolf et al., 2013). One most common rule is to have at least 10-15 participants per variable measured; however, this rule does not have a clear empirical basis (Field, 2013). Instead, it is recommended to consider the overall sample rather than a case-to-variables ratio. While sample size calculations are more nuanced than a simple rule of thumb, general recommendations are a sample size of 300 as good and 1000 as excellent (Comrey & Lee, 1992; Tabachnick & Fidell, 2012). Regardless of whether case-to-variable ratios or overall sample are utilized, the present study meets the minimum requirements for reliable analyses. In the present analysis, 49 subtests function as observed variables in the model, indicating a minimum sample size of 490 participants. The overall sample for the present study consisted of 6,100 participants; however, even considering each measure on its own, the sample size for the study met the minimum requirement.

Addressing the Issue of Missing Data

The Wechsler standardization samples were examined for missing data values prior to proceeding with the analysis. On the WAIS-IV, a missing values analysis indicated that 18% of values were missing from three of the subtests: Letter-Number Sequencing, Figure Weights, and Cancellation. However, these three subtests do not contribute to the overall Full-Scale Intelligence Quotient (FSIQ). Further, these subtests were not administered to the older age group of which the WAIS-IV standardization sample is comprised (i.e., 70:0 to 90:11 age group). Therefore, the researcher utilized pairwise computation of the correlation matrix to handle the missing data. Pairwise computation is accomplished by computing covariance based on all cases with non-missing values for a specific pair of variables (Marsh, 1998). This technique is available in RStudio, the programming language used for the analyses. For the WISC-V, a missing values analysis indicated that 1.86% of values were missing from two of the subtests: Letter-Number Sequencing and Digit Span. Therefore, due to the minimal amount of missing data, the researcher determined that it was unnecessary to outline a particular procedure for handling the missing values. The sample for the present analysis was deemed adequate. Finally, for the WPPSI-IV, no data were missing, indicating the sample was suitable for analysis.

Exploratory Factor Analysis

Prior to the development of an assessment, the developers typically hypothesize the latent factors they expect to find on the assessment. This may include utilizing current theoretical models, such as CHC theory, as a way to conceptualize the assessment instrument and said latent factors. However, an EFA is truly exploratory in nature. EFA allows the developer to generate theories based on the latent constructs.

EFA is a complex statistical procedure involving many components that can be broken down into steps. The first step involves identifying suitability of the data set for the factor analysis. The researcher examined the correlation matrix for correlation coefficients exceeding 0.30, the minimal factor loading acceptable for the analysis (Williams et al., 2010). Additionally, the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy and Bartlett's Test of Sphericity were conducted. A KMO index of 0.50 is considered adequate for continuation of factor analysis. Finally, a significant ($p < .05$) Bartlett's Test of Sphericity was necessary (Williams et al., 2010). Once the researcher conducted these tests and examined the adequacy of the data, it was then determined to extract the factors using the principal component analysis (PCA) method. PCA is one of the most commonly used extraction methods (Williams et al., 2010). The researcher utilized PCA to determine the number of factors to extract. Researchers typically use the Eigenvalue criterion stating factors with an eigenvalue greater than 1 can be extracted (Williams et al., 2010). Visually, scree plots are beneficial. Within a scree plot, the researcher drew a line through the smaller eigenvalues until there was a break in the line; any values above the break indicate the number of factors to retain (Williams et al., 2010). In the present study, the researcher utilized parallel analysis and the Scree Test as factor extraction criteria.

Researchers must also consider rotations within the matrix in determining whether a variable relates to multiple factors. Orthogonal varimax/quartimax and oblique oblimin/promax are two such methods for addressing this. Orthogonal varimax results in uncorrelated factor structures while oblique oblimin results in correlated factors (Williams et al., 2010). The researcher determined the rotation method to use prior to conducting the analysis. In the present study, the researcher used an oblique, oblimin rotation. This rotation is often used when data are hypothesized to be correlated with one another and allows for correlated factors (Kline, 2016). Finally, the researcher interpreted the factors based on the variables that comprise said factors (Williams et al., 2010). Most importantly, the labels for these factors should be meaningful as a reflection of the original theoretical concept. The researcher attempted to do this by using contemporary theoretical perspectives.

Confirmatory Factor Analysis/Structural Equation Modeling

CFA is a type of Structural Equation Modeling (SEM) intended to evaluate the proposed factor structure (Wechsler, 2014). SEM is a broad term that also describes other techniques such as causal modeling, covariance structure analysis, covariance structure modeling, and analysis of covariance structure (Kline, 2011). SEM, as it is intended, involves researchers defining relationships a priori based on theory and then testing to see if those relationships accurately reflect the data (Weston & Gore, 2006). In the general case, a researcher will subset the complete dataset into two parts. The first part will be explored with EFA models in order to generate empirically informed hypotheses. These hypotheses will then be applied in the form of CFA on the second subset of data. This is done so that the data used to generate the hypotheses is separate from, and uninformed by, the data used to test the hypotheses. For the purpose of this study, the data was not split into parts. The reasoning for this choice was that the hypotheses, with respect to which CFA structures were implemented, were generated not by EFA of the existing data, but rather from previously published F-CHC theories. In this way, by construction, the data used to generate the hypotheses is not present in the data used for the CFA, so the researcher made no distinction with respect to what data was used for EFA and CFA models.

SEM consists of the measurement model, the correlation between observed variables and latent constructs, and the structural model, the relationship between latent constructs or covariances (Weston & Gore, 2006). There are five steps in the SEM process: Model specification, model identification, model estimation, model evaluation/testing, and model modification (Lomax, 2013; Weston & Gore, 2006). The model specification step refers to the initial theoretical selection in which the researcher determined the relationships among observed and latent variables. Once variables were selected and relationships predetermined, model identification helped the researcher consider whether the data contained enough information for the parameters. That is, the researcher determined whether the model is over-, under-, or just-identified by determining the number of degrees of freedom (Weston & Gore, 2006). In the present study, several of the models were just-identified. Next, the researcher determined the parameters of the model (i.e., model estimation) using ordinary least-squares to find the minimum residual. Model testing was then conducted to examine the fit of the data to the theoretical model (Lomax, 2013). Fit was determined using the root mean square error of approximation (RMSEA) and the comparative fit index (CFI; Lomax, 2013; Weston & Gore, 2006). Finally, researchers typically make model modifications if the data do not fit the initial theoretical model identified during the model specification stage (Lomax, 2013). Modification can be done by using statistical strategy called a specification search to determine what changes need to be made for a better fit (Weston & Gore, 2006). However, researchers must be cognizant that modifications make substantive sense and are not merely done to find a better fit (Lomax, 2013). In the present study, the researcher made modifications to models that were under-identified.

In the analytic plan, numerous CFAs were conducted on models consisting of one, two, three, four, five, and six factors. The present study utilized models consistent with current theoretical perspective (i.e., CHC theory), perspectives in the literature from previous factor analytic research of the Wechsler instruments (Wechsler, 2008, 2012, 2014), and the F-CHC model of cognitive abilities (i.e., the hypothesized model of interest). The researcher analyzed each model for goodness of fit and determined which model best fits the data. It was hypothesized that the F-CHC model of cognitive abilities would be the best fitting model.

Conclusion

This chapter outlined the methodology utilized in the research study. Information was provided on the psychometric properties of the Wechsler instruments, including the reliability and validity of each instrument. Additionally, the subtests that comprise each measure are explained including the hypothesized related broad and narrow abilities. This chapter also outlined the statistical procedures employed in the data analysis. The hypothesized models, the F-CHC model and CHC theory, were utilized in examining the factor structure of the Wechsler cognitive assessment instruments.

CHAPTER IV

RESULTS

The primary goal of the present study was to examine the factor structure of general intellectual ability and broad and narrow cognitive abilities using traditional and contemporary models of cognitive abilities. This chapter outlines the results of the analyses including preliminary data analysis, exploratory factor analyses, and confirmatory factory analyses for each instrument. The programming language RStudio was utilized for the analyses.

For each instrument, data were first analyzed as a whole to examine the hypothesized combinations of broad abilities that comprise each instrument. Data were then examined by broad abilities of intelligence (e.g., *Gc, Gr, Gv, Gcm,* and *Gs*) to investigate the hypothesized combinations of narrow abilities that comprise each measure. Correlation matrices were first generated and analyzed prior to conducting the EFA. Correlation matrices were evaluated with Bartlett's Test of Sphericity (Bartlett, 1950) and the KMO Measure of Sampling Adequacy (Kaiser & Rice, 1974). These evaluations allowed the researcher to determine whether the data was suitable for factor analysis. Once deemed suitable for analysis, EFA was then used to examine various models of the data utilizing factor extraction methods such as Scree Plot and parallel analysis. Factors were extracted via an oblique, oblimin rotation. This rotation is often used when data are hypothesized to be correlated with one another and allows for correlated factors (Kline, 2016). Standardized factor loadings, or coefficients, were computed in order to directly compare the variables. Standardized coefficients are typically between 0 and 1; however, there are instances where coefficients may be greater than 1. This does not indicate an issue with the analysis, rather, there may exist multicollinearity between the variables (Schreiber, 2008). Multicollinearity exists when two or more predictor variables are strongly correlated with one another, which can distort the interpretation of the results (Meyers et al., 2016). In the present study, certain coefficients for the latent variables exceeded 1, indicating possible multicollinearity of the latent variables. However, this is to be expected given the nature of the latent variables of interest (i.e., hypothesized constructs of intelligence).

After exploring the factor structure of the various instruments, the researcher then employed CFA to further examine the factor structure based on a theoretical concept (i.e., F-CHC and CHC). The CFI and the RMSEA were utilized as fit measures. The CFI is an incremental fit index that examines relative improvement in fit (Kline, 2013). The CFI includes values ranging from 0 to 1, where higher values indicate better fit (Cangur & Ercan, 2015). A fit above .95 is deemed acceptable. The RMSEA is an absolute fit index that takes into account model complexity. RMSEA values range from 0 to 1, where lower values indicate better fit (Cangur & Ercan, 2015), and a value of 0 indicates best fit (Kline, 2013). A fit value smaller than .05 indicates reasonable fit. The results will be presented by instrument: the WAIS-IV, WISC-V, and WPPSI-IV.

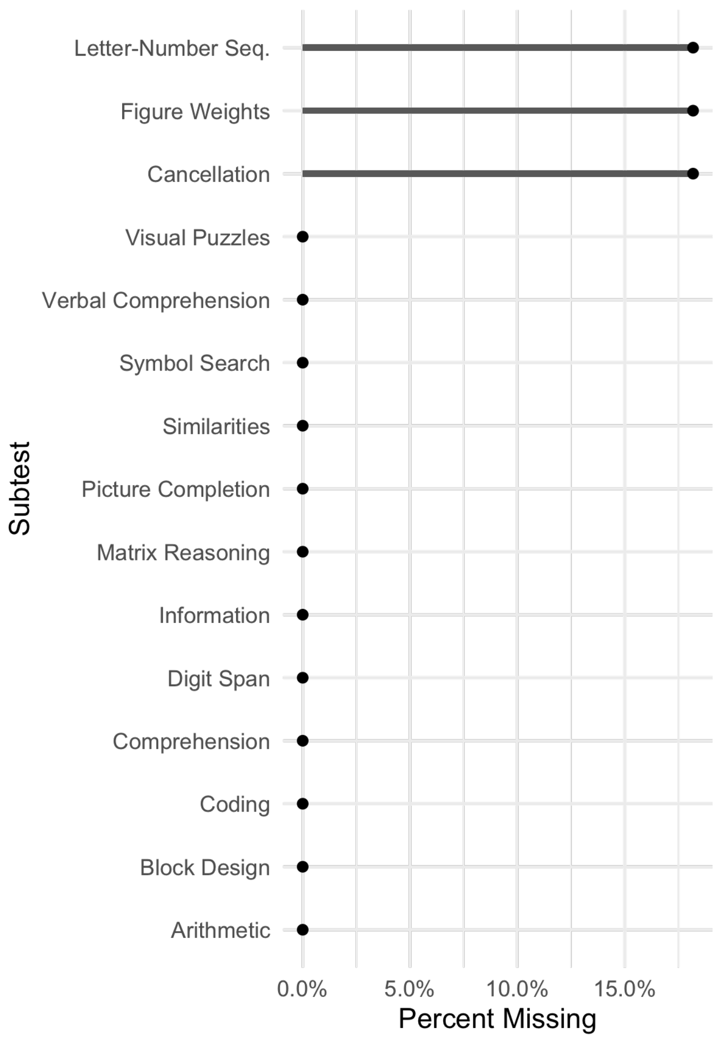
WAIS-IV

Preliminary Data Analysis

The analyses in this section are based on the standardization sample from the WAIS-IV norming study and consisted of 2,200 participants. The preliminary analysis examined the 15 subtests from the WAIS-IV. A missing values analysis indicated that 18% of values were missing from three of the subtests: Letter-Number Sequencing, Figure Weights, and Cancellation (see Figure 1). However, these three subtests do not contribute to the overall FSIQ. Further, these subtests were not administered to the older age group of which the WAIS-IV standardization sample is comprised (i.e., 70:0 to 90:11 age group). Therefore, the researcher utilized pairwise computation of the correlation matrix to handle the missing data. Pairwise computation is accomplished by computing covariance based on all cases with non-missing values for a specific pair of variables (Marsh, 1998). This technique is available in RStudio.

Figure 1

Percentage of Missing Values in Each Subtest in the WAIS-IV Dataset



Primary Data Analysis

The WAIS-IV subtests were used to create a correlation matrix (see Table 1).

Table 1

Correlation Matrix of the Hypothesized WAIS-IV Subtests

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Similarities	1														
2 Vocabulary	.73	1													
3 Comprehension	.71	.74	1												
4 Information	.63	.72	.64	1											
5 Symbol Search	.34	.34	.31	.33	1										
6 Cancellation	.22	.24	.21	.22	.46	1									
7 Coding	.38	.39	.38	.33	.64	.42	1								
8 Visual Puzzles	.44	.43	.43	.43	.37	.31	.35	1							
9 Picture Completion	.42	.39	.39	.40	.40	.33	.35	.48	1						
10 Matrix Reasoning	.49	.50	.48	.47	.38	.26	.44	.52	.41	1					
11 Block Design	.48	.45	.45	.44	.40	.33	.38	.64	.49	.54	1				
12 Figure Weights	.53	.53	.53	.51	.34	.28	.36	.57	.41	.57	.55	1			
13 Arithmetic	.53	.57	.53	.56	.36	.30	.43	.48	.36	.52	.50	.61	1		
14 Letter-Number Seq.	.44	.48	.46	.42	.37	.29	.37	.40	.37	.44	.41	.48	.55	1	
15 Digit Span	.47	.50	.47	.42	.40	.34	.44	.40	.38	.47	.44	.50	.60	.68	1

Exploratory Factor Analysis of the WAIS-IV

Bartlett's Test of Sphericity revealed that the correlation matrix was not random (χ^2 = 18373.07, *df* = 105, *p* < .001). The KMO Measure of Sampling Adequacy was .94, indicating that the data is appropriate for analysis (Kline, 2013). The WAIS-IV factor structure was analyzed to determine the number of factors to be extracted. Parallel analysis and examination of the visual scree plot (see Figure 2) both suggested a maximum of four factors. The EFA supported four factors labeled as *Gc*, *Gcm*, and *Gs*. There was a fourth factor consisting of subtests based on the broad abilities of *Gr* and *Gv* to create a combined perceptual reasoning grouping (*Gv-Gr*). The factor labels were chosen to be consistent with the F-CHC model. The five-factor model revealed similar broad ability factors as the four-factor model, but the five-factor model also revealed weak factor loadings onto a fifth factor with no clear definition of the factor. Therefore, while the analysis allowed for a fifth factor, there were no significant loadings on the fifth factor, suggesting a four-factor solution is best fitting. Refer to Tables 2 and 3 for the factor loadings of the WAIS-IV subtests in the four-factor and five-factor model.

Figure 2

Visual Scree Plot of the Hypothesized WAIS-IV Subtests

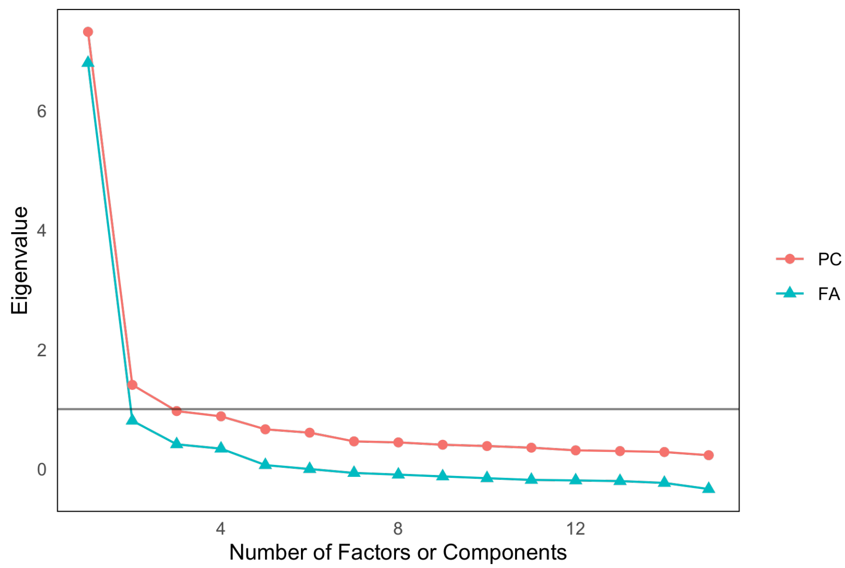


Table 2

EFA Standardized Factor Loadings for the WAIS-IV Four-Factor Solution

	Comprehension Knowledge	Perceptual Reasoning	Conscious Memory	Cognitive Processing Speed
Vocabulary	**.92**	−.07	.02	.02
Comprehension	**.81**	.02	.02	−.01
Similarities	**.78**	.06	−.02	.03
Information	**.75**	.08	−.02	.00
Visual Puzzles	−.03	**.84**	−.02	.00
Block Design	.01	**.74**	.01	.07
Figure Weights	.19	**.50**	.23	−.08
Matrix Reasoning	.18	**.41**	.15	.07
Picture Completion	.10	**.41**	−.01	.21
Digit Span	−.03	−.03	**.87**	.05
Letter-Number Seq.	.01	.03	**.76**	.00
Arithmetic	.25	.19	**.43**	−.01
Symbol Search	−.01	.03	−.03	**.83**
Coding	.09	−.04	.07	**.71**
Cancellation	−.12	.15	.10	**.47**

Table 3

EFA Standardized Factor Loadings for the WAIS-IV Five-Factor Solution

	Comprehension Knowledge	Perceptual Reasoning	Conscious Memory	Cognitive Processing Speed	Factor 5
Vocabulary	**.93**	−.08	.02	.01	.00
Comprehension	**.81**	.01	.01	.00	−.01
Similarities	**.80**	.05	−.02	.01	.02
Information	**.76**	.08	−.02	−.02	.01
Visual Puzzles	−.02	**.83**	−.02	−.01	.03
Block Design	.03	**.74**	.01	.04	.06
Figure Weights	.11	**.57**	.15	.00	−.19
Matrix Reasoning	.10	**.49**	.07	.15	−.17
Picture Completion	.18	**.38**	.06	.08	.26
Digit Span	−.02	−.03	**.87**	.03	.00
Letter-Number Seq.	.03	.02	**.77**	−.04	.03
Arithmetic	.17	.25	**.37**	.07	−.19
Symbol Search	.01	.07	.05	**.67**	.18
Coding	.02	-.02	−.01	**.89**	−.08
Cancellation	−.08	.13	.17	**.37**	.22

Confirmatory Factor Analysis of the WAIS-IV

In the CFA, various models of the WAIS-IV were analyzed for model fit. The one-factor model (*g*) revealed poor fit indices (CFI = .802, RMSEA = .135). Fit indices improved with the addition of each factor. Specifically, the two-factor model exhibited better fit than the one-factor (*Gc* + *Gr;* CFI = .885, RMSEA = .104), and the three-factor model exhibited better fit than the two-factor model (*Gc* + *Gr* + *Gs,* CFI = .928, RMSEA = .083).

The four-factor model revealed reasonable fit indices (CFI = .952, RMSEA = .068). As hypothesized by F-CHC, in the four-factor structure of the WAIS-IV, Similarities, Information, Comprehension, and Vocabulary form a broad factor of *Gc*. Also hypothesized by the F-CHC model, Letter-Number Sequencing and Digit Span form a broad factor of *Gcm*, while Symbol Search, Coding, and Cancellation form a broad factor of *Gs*. However, Matrix Reasoning, Visual Puzzles, Picture Completion, Block Design, Figure Weights, and Arithmetic form a broad factor labeled as perceptual reasoning by the researcher. This factor is inconsistent with both the F-CHC and CHC models.

The five-factor model also revealed reasonable fit indices (CFI = .974, RMSEA = .051). Additionally, the five-factor model also revealed several cross-loadings where two subtests loaded onto two or more of the factors. In this case, Arithmetic loaded on *Gc, Gr,* and *Gcm*. Picture Completion is another subtest that cross-loaded on *Gr* and *Gv*. As hypothesized by the F-CHC model, in the five-factor structure of the WAIS-IV, Similarities, Information, Comprehension, Vocabulary, and Arithmetic form a broad factor of *Gc*. Letter-Number Sequencing, Digit Span, and Arithmetic form a broad factor of *Gcm*, while Symbol Search, Coding, and Cancellation form a broad factor of *Gs*. Matrix Reasoning, Picture Completion, Figure Weights, and Arithmetic now form a broad factor of *Gr*, and Visual Puzzles, Block Design, and Picture Completion form a broad factor of *Gv*.

A six-factor model of intelligence was attempted; however, this only revealed incremental improvement of the model fit. Further, this would require one subtest to form a broad factor, which according to current research (Flanagan et al., 2013), is not recommended. Given the principle of parsimony, the five-factor model, as hypothesized by the F-CHC model, appears to best fit the standardization sample for the WAIS-IV. Refer to Figure 3 and 4 for the diagrams of the four-factor and five-factor structure.

Figure 3

Four-Factor Model of WAIS-IV

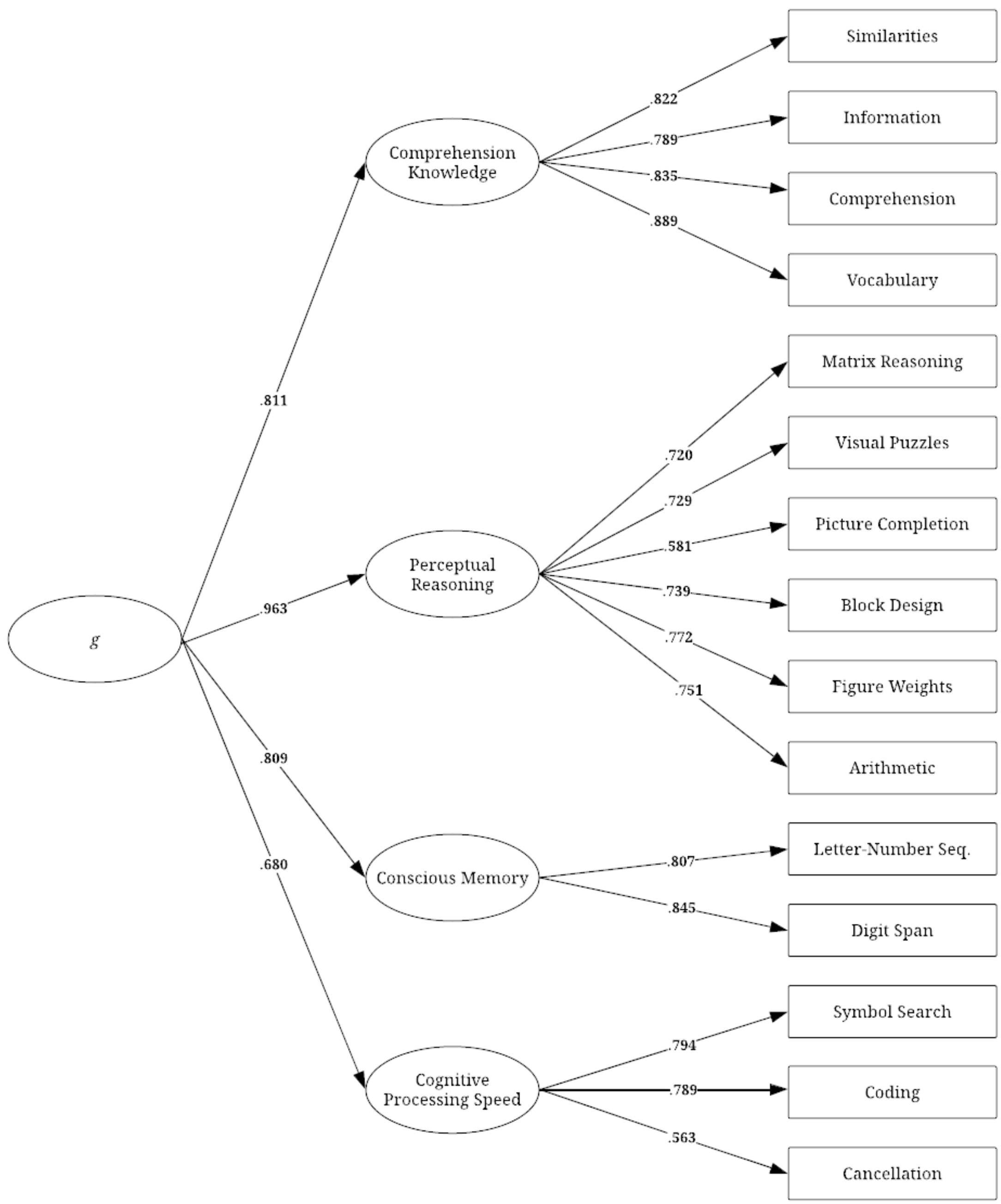
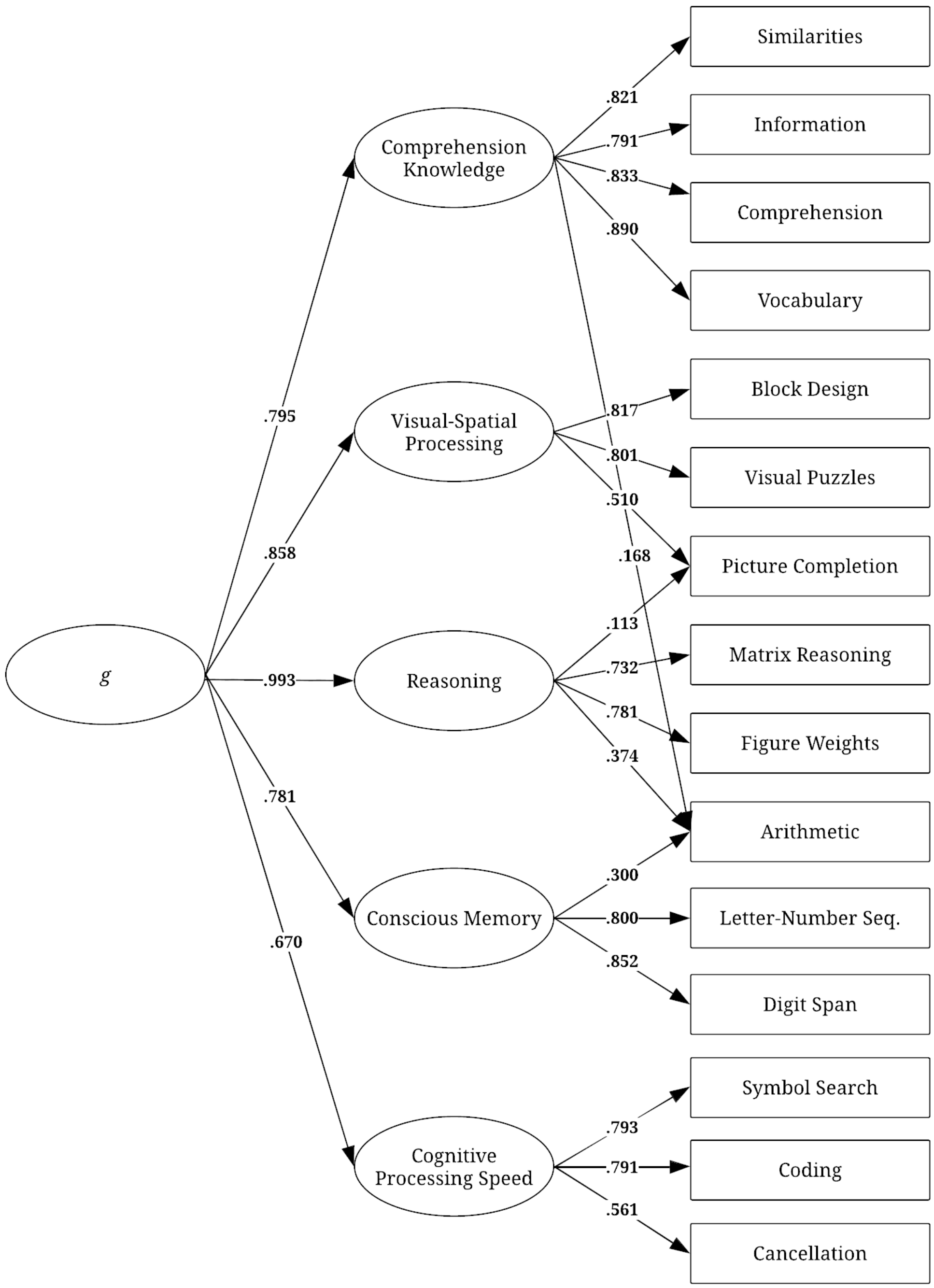


Figure 4

Five-Factor Model of WAIS-IV



The following analyses were conducted to further examine the factor structure of the WAIS-IV. Specifically, factor grouping from the CFA (*Gc, Gv, Gcm, Gr,* and *Gs*) were investigated. Multiple EFAs and CFAs were conducted to investigate whether the narrow abilities, as hypothesized by the F-CHC model, could be extracted.

Comprehension Knowledge

In the F-CHC model, Similarities, Vocabulary, Comprehension, and Information are hypothesized to be measures of the construct of *Gc*. These subtests were used to create a correlation matrix (see Table 4).

Table 4

Correlation Matrix of the Hypothesized WAIS-IV Gc Subtests

	Similarities	Vocabulary	Comprehension	Information
Similarities	1.00			
Vocabulary	.73	1.00		
Comprehension	.71	.74	1.00	
Information	.63	.72	.64	1.00

Exploratory Factor Analysis of the WAIS-IV Gc. Bartlett's Test of Sphericity revealed that the correlation matrix was not random ($\chi^2 = 5500.23$, $df = 6$, $p < .001$). The KMO Measure of Sampling Adequacy was .84, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis and examination of the visual scree plot (see Figure 5) both suggested a maximum of two factors. Based on the two factors determined by the parallel analysis and visual scree plot, a one-factor and three-factor solution were also examined in the EFA. The EFA revealed that the factors loaded onto two distinct factors. As hypothesized by the F-CHC model, the EFA supported a verbal ability narrow factor consisting of Similarities, Vocabulary, and

Comprehension, in addition to a factual knowledge narrow factor consisting of Information. Refer to Table 5 for the factor loadings of the *Gc* subtests.

Figure 5

Visual Scree Plot of the Hypothesized WAIS-IV Gc Subtests

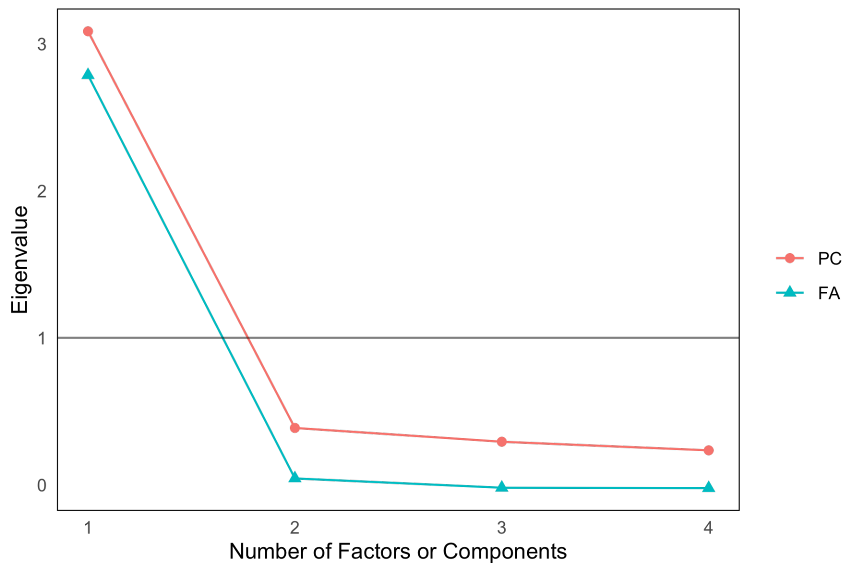


Table 5

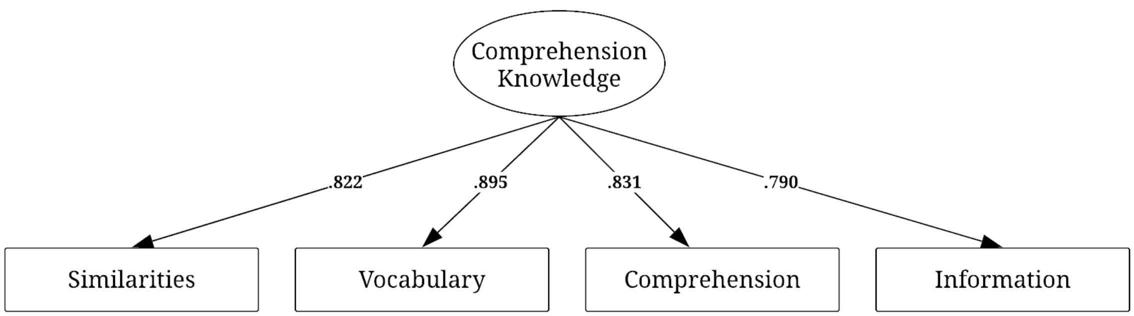
EFA Standardized Factor Loadings for the WAIS-IV Gc Two-Factor Solution

	Verbal Ability	Factual Knowledge
Similarities	**.87**	−.05
Vocabulary	**.77**	.13
Comprehension	**.87**	−.03
Information	.01	**.97**

Confirmatory Factor Analysis of the WAIS-IV Gc. In the CFA, various models of *Gc* were analyzed for model fit. The one-factor model of *Gc* revealed reasonable fit indices (CFI = .994, RMSEA = .084). In the two-factor model of *Gc*, as hypothesized by the F-CHC model, one of the approximate fit indices was considered poor (RMSEA = .120), whereas the CFI was deemed reasonable at .994. In the two-factor structure of *Gc,* Similarities, Vocabulary, and Comprehension form a narrow factor labeled as verbal ability whereas Information forms a narrow factor labeled as factual knowledge. However, the one-factor model of *Gc* proved to be the best fitting model. Refer to Figure 6 for the one-factor structure.

Figure 6

One-Factor Model of WAIS-IV Gc



Cognitive Processing Speed

In the F-CHC model, Symbol Search, Cancellation, and Coding are hypothesized to be measures of the construct of *Gs*. These subtests were used to create a correlation matrix (see Table 6).

Table 6

Correlation Matrix of the Hypothesized WAIS-IV Gs Subtests

	Symbol Search	Coding	Cancellation
Symbol Search	1.00		
Coding	.64	1.00	
Cancellation	.46	.42	1.00

Exploratory Factor Analysis of the WAIS-IV Gs. Bartlett's Test of Sphericity revealed that the correlation matrix was not random ($\chi^2 = 1762.129$, $df = 3$, $p < .001$). The KMO Measure of Sampling Adequacy was .66, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis suggested a maximum of one factor while examination of the visual scree plot (see Figure 7) suggested a maximum of two factors. However, given the limited number of indicators (e.g., subtests), a two-factor model is not statistically relevant. Therefore, a one-factor EFA was conducted and provided support for the broad factor of *Gs* as hypothesized by the F-CHC model. Refer to Table 7 for the factor loadings of the *Gs* subtests for the one-factor model.

Figure 7

Visual Scree Plot of the Hypothesized WAIS-IV Gs Subtests

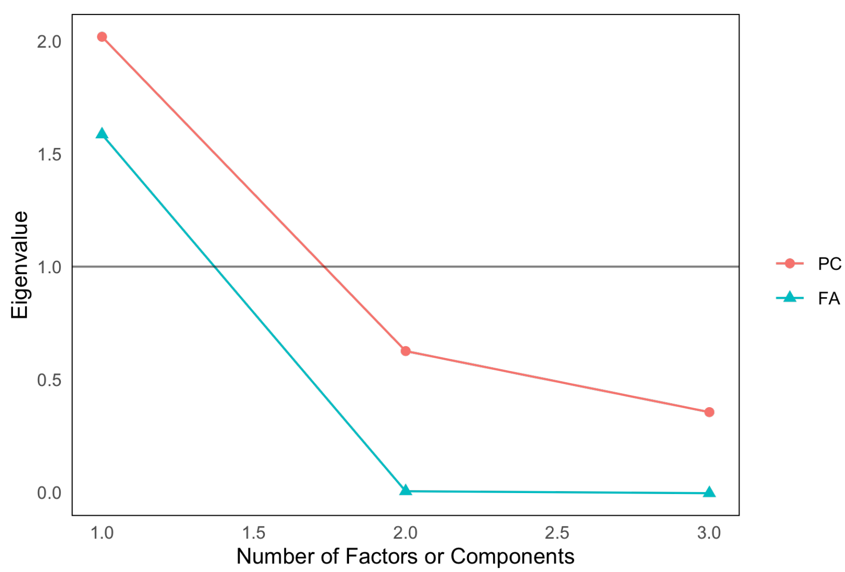


Table 7

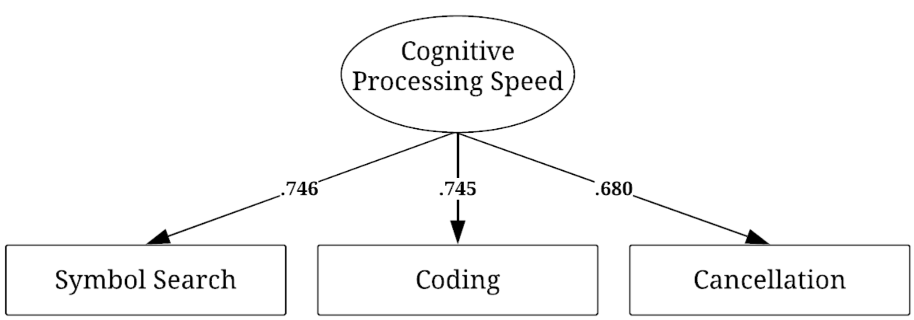
EFA Standardized Factor Loadings for the WAIS-IV Gs One-Factor Solution

	Cognitive Processing Speed
Symbol Search	**.83**
Coding	**.78**
Cancellation	**.55**

Confirmatory Factor Analysis of the WAIS-IV Gs. The researcher specified a constraint on the factor loadings in the CFA due to the limited data from the hypothesized subtests. In this case, the indicator loadings were set equal to one another in order to create a just-identified model. The results suggest perfect fit for a one-factor model of *Gs* (CFI = 1.000, RMSEA = .000), indicating support for the broad factor of *Gs*. Refer to Figure 8 below for the one-factor structure.

Figure 8

One-Factor Model of WAIS-IV Gs



Visual-Spatial Processing

In the F-CHC model, Visual Puzzles and Block Design are hypothesized to be measures of the construct of *Gv.* These subtests were used to create a correlation matrix (see Table 8).

Table 8

Correlation Matrix of the Hypothesized WAIS-IV Gv Subtests

	Visual Puzzles	Block Design
Visual Puzzles	1.00	
Block Design	.64	1.00

Given the limited number of indictors for this particular factor (e.g., two subtests), methods such as EFA and CFA are not statistically relevant. In particular, there is not enough information to fit the number of parameters required by a factor analytic approach. Therefore, this does not allow the researcher to accurately examine the narrow abilities of visual-spatial processing as hypothesized by the F-CHC model or CHC theory. Therefore, the researcher did not pursue the EFA and CFA for the broad ability of *Gv.* While narrow abilities could not be examined, the broad ability of *Gv* was previously supported in the CFA for the WAIS-IV instrument as a whole. This evidence for the broad ability of *Gv* is consistent with the F-CHC model, CHC theory, and that posited by the Wechsler publishers.

Reasoning

In the F-CHC model, Matrix Reasoning, Figure Weights, Picture Completion, and Arithmetic are hypothesized to be measures of the construct of *Gr*. These subtests were used to create a correlation matrix (see Table 9).

Table 9

Correlation Matrix of the Hypothesized WAIS-IV Gr Subtests

	Picture Completion	Matrix Reasoning	Figure Weights	Arithmetic
Picture Completion	1.00			
Matrix Reasoning	.41	1.00		
Figure Weights	.41	.57	1.00	
Arithmetic	.36	.52	.61	1.00

Exploratory Factor Analysis of the WAIS-IV Gr. Bartlett's Test of Sphericity revealed that the correlation matrix was not random ($\chi^2 = 2572.29$, $df = 6$, $p < .001$). The KMO Measure of Sampling Adequacy was .77, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis and examination of the visual scree plot (see Figure 9) both suggested a maximum of two factors. Based on the factors determined by the parallel analysis and visual scree plot, a one-factor and three-factor solution were also examined in the EFA. The EFA revealed that the factors loaded onto two distinct factors. As hypothesized by the F-CHC model, the EFA supported an inductive/deductive reasoning narrow factor consisting of Figure Weights, Matrix Reasoning, and Arithmetic in addition to a contextual reasoning narrow factor consisting of Picture Completion. Refer to Table 10 for the factor loadings of the *Gr* subtests for the two-factor model.

Figure 9

Visual Scree Plot of the Hypothesized WAIS-IV Gr Subtests

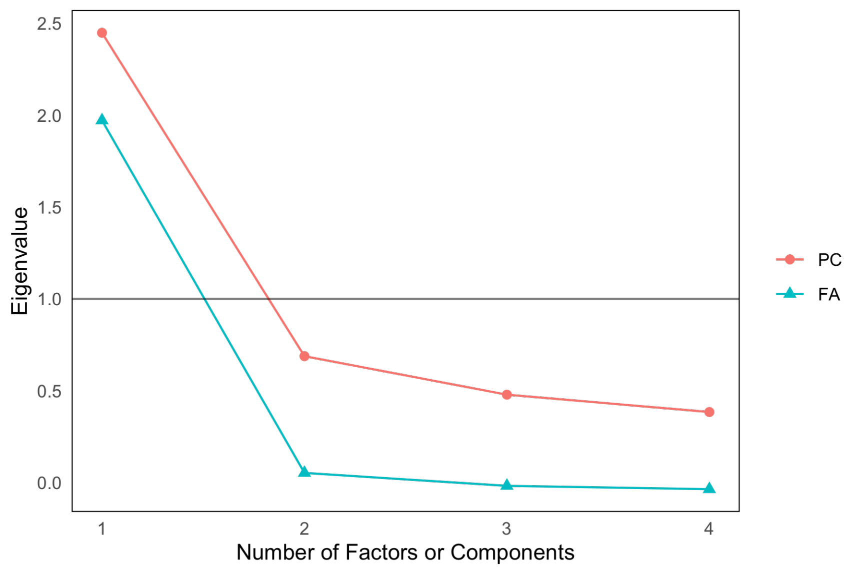


Table 10

EFA Standardized Factor Loadings for the WAIS-IV Gr Two-Factor Solution

	Inductive/Deductive Reasoning	Contextual Reasoning
Figure Weights	**.81**	−.01
Arithmetic	**.77**	−.03
Matrix Reasoning	**.66**	.07
Picture Completion	.00	**.99**

Confirmatory Factor Analysis of the WAIS-IV Gr. In the CFA, various models of *Gr* were analyzed for model fit. The one-factor model of *Gr* revealed reasonable fit indices (CFI = .995, RMSEA = .052). The two-factor model of *Gr*, as hypothesized by the F-CHC model, also revealed reasonable fit indices (CFI = .995, RMSEA = .077). In the two-factor structure of *Gr,* Matrix Reasoning, Arithmetic, and Figure Weights form a narrow factor labeled as inductive/deductive reasoning whereas Picture Completion forms a narrow factor labeled as contextual reasoning. However, the one-factor model of *Gr* proved to be the best fitting model. Refer to Figure 10 and 11 for the diagrams of the one-factor and two-factor structure.

Figure 10

One-Factor Model of WAIS-IV Gr

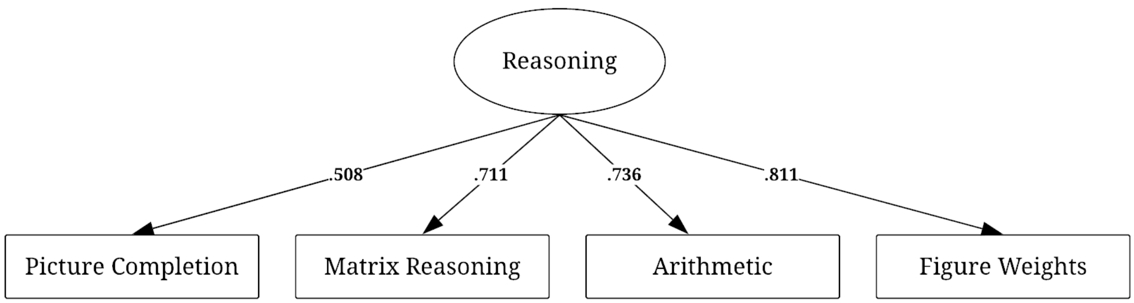
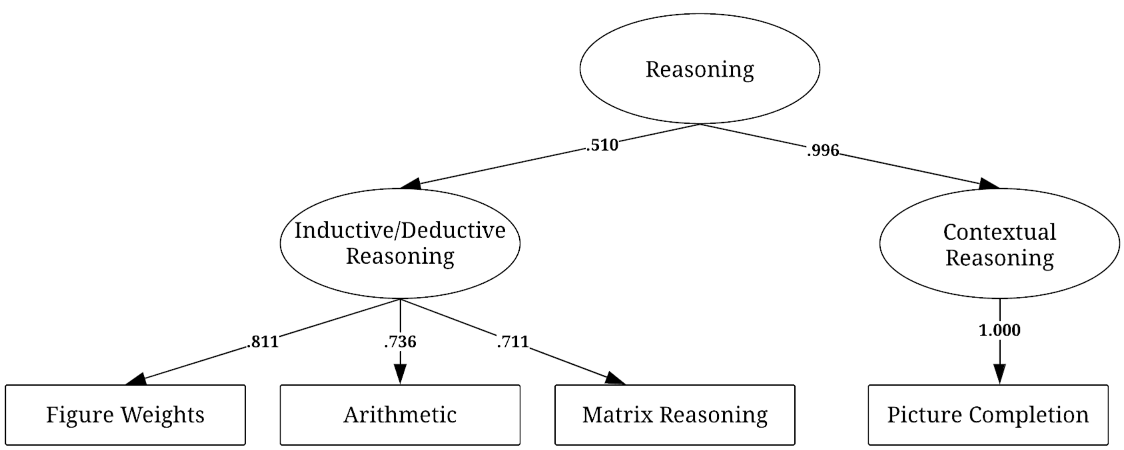


Figure 11

Two-Factor Model of WAIS-IV Gr



Conscious Memory

In the F-CHC model, Letter-Number Sequencing and Digit Span are hypothesized to be measures of the construct of *Gcm*. These subtests were used to create a correlation matrix (see Table 11).

Table 11

Correlation Matrix of the Hypothesized WAIS-IV Gcm Subtests

	Letter-Number Sequencing	Digit Span
Letter-Number Sequencing	1.00	
Digit Span	.68	1.00

Given the limited number of indictors for this particular factor (e.g., two subtests), statistical methods such as EFA and CFA are not feasible. In particular, there is not enough information to accurately examine narrow abilities of F-CHC and CHC theory. Furthermore, there is a foregone conclusion that the broad factor of *Gcm* will be supported given the subtests that comprise this factor. Therefore, the researcher discontinued with the EFA and CFA for the broad ability of *Gcm.*

Conclusion

The factor analytic structure of the WAIS-IV was examined. Evidence was found for both a four-factor and five-factor model, with stronger support for a five-factor model. This five-factor model is more consistent with the F-CHC model and CHC theory rather than the factor structure posited by the Wechsler publishers. Examination of narrow abilities was attempted for each broad factor.

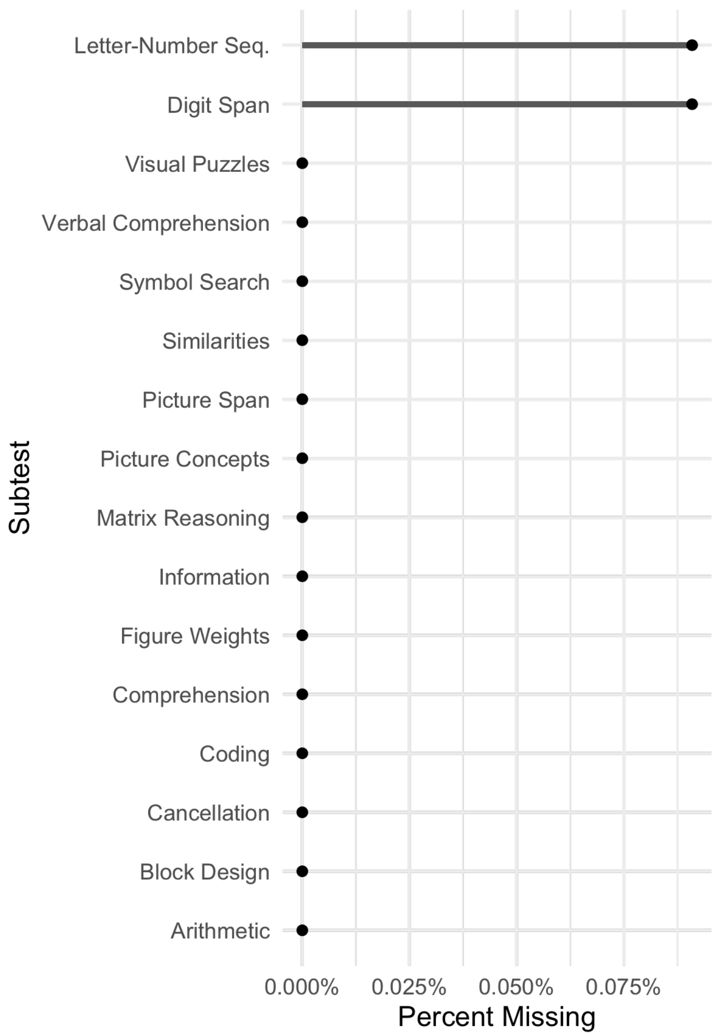
WISC-V

Preliminary Data Analysis

The analyses in this section are based on the standardization sample from the WISC-V norming study and consisted of 2,200 participants. The preliminary analysis examined the 16 subtests from the WISC-V. Complementary subtests (i.e., Naming Speed Literacy, Naming Speed Quality, Immediate Symbol Translation, Delayed Symbol Translation, and Recognition Symbol Translation) were not utilized in the analyses as these are not intelligence subtests and cannot be substituted for primary or secondary subtests (Canivez et al., 2016). Further, complimentary subtests were also not utilized in the CFAs reported in the WISC-V *Technical and Interpretive Manual* (Wechsler, 2014); therefore, they are not included in the present analysis. A missing values analysis indicated that 1.86% of values were missing from two of the subtests: Letter-Number Sequencing and Digit Span (see Figure 12). Therefore, due to the minimal amount of missing data, the researcher determined that it was unnecessary to outline a particular procedure for handling the missing values. The sample for the present analysis was deemed adequate.

Figure 12

Percentage of Missing Values in Each Subtest in the WISC-V Dataset



Primary Data Analysis

The WISC-V subtests were used to create a correlation matrix (see Table 12).

Table 12

Correlation Matrix of the Hypothesized WISC-V Subtests

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1 Similarities	1															
2 Vocabulary	.68	1														
3 Comprehension	.59	.60	1													
4 Information	.65	.71	.56	1												
5 Figure Weights	.46	.49	.38	.48	1											
6 Arithmetic	.54	.53	.46	.55	.50	1										
7 Matrix Reasoning	.45	.45	.39	.45	.47	.45	1									
8 Picture Concepts	.39	.42	.35	.40	.33	.37	.35	1								
9 Visual Puzzles	.48	.51	.40	.48	.50	.46	.47	.39	1							
10 Block Design	.46	.47	.38	.47	.47	.46	.47	.34	.60	1						
11 Digit Span	.47	.46	.42	.46	.43	.55	.44	.34	.42	.42	1					
12 Picture Span	.39	.38	.36	.36	.35	.43	.38	.30	.36	.35	.51	1				
13 Letter-Num Seq.	.48	.49	.43	.47	.40	.54	.43	.33	.39	.38	.65	.49	1			
14 Symbol Search	.28	.25	.26	.29	.23	.32	.29	.23	.28	.34	.32	.27	.28	1		
15 Cancellation	.11	.10	.14	.13	.11	.15	.13	.11	.13	.19	.11	.09	.11	.33	1	
16 Coding	.23	.21	.24	.20	.19	.31	.24	.19	.20	.31	.28	.25	.29	.58	.30	1

Exploratory Factor Analysis of the WISC-V

Bartlett's Test of Sphericity revealed that the correlation matrix was not random (χ^2 = 15604.29, *df* = 120, *p* < .001). The KMO Measure of Sampling Adequacy was .94, indicating that the data is appropriate for analysis (Kline, 2013). The WISC-V factor structure was analyzed to determine the number of factors to be extracted. Parallel analysis suggested a maximum of four factors while examination of the visual scree plot suggested four to five factors (see Figure 13). The EFA supported four factors labeled as *Gc*, *Gcm*, and *Gs*, and a fourth factor consisting of subtests based on the broad abilities of *Gr* and *Gv* to create a combined *Gv*-*Gr* factor. The five-factor model revealed similar broad factors for the first four factors. The five-factor model also produced a fifth factor with a single subtest, Cancellation, with no clear definition. Refer to Tables 13 and 14 below for the four-factor and five-factor solutions for the EFA.

Figure 13

Visual Scree Plot of the Hypothesized WISC-V Subtests

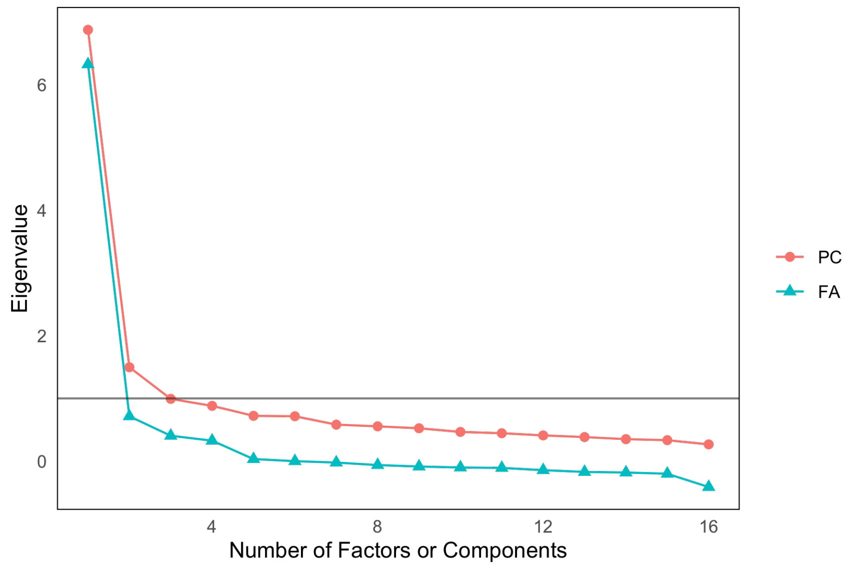


Table 13

EFA Standardized Factor Loadings for the WISC-V Four-Factor Solution

	Comprehension Knowledge	Conscious Memory	Perceptual Reasoning	Cognitive Processing Speed
Vocabulary	**.87**	−.03	.02	−.03
Information	**.79**	−.02	.05	.00
Similarities	**.77**	.04	.01	.02
Comprehension	**.70**	.06	−.08	.07
Picture Concepts	.24	.10	**.22**	.03
Digit Span	−.05	**.81**	.06	.01
Letter-Number Seq.	.08	**.77**	−.06	.01
Picture Span	.00	**.55**	.09	.04
Arithmetic	.24	**.37**	.17	.07
Visual Puzzles	.03	.00	**.78**	−.04
Block Design	−.01	−.01	**.71**	.12
Figure Weights	.12	.16	**.49**	−.06
Matrix Reasoning	.07	.21	**.43**	.03
Coding	−.02	.06	−.05	**.75**
Symbol Search	.03	−.02	.05	**.74**
Cancellation	.02	−.12	.08	**.43**

Table 14

EFA Standardized Factor Loadings for the WISC-V Five-Factor Solution

	Comprehension Knowledge	Conscious Memory	Perceptual Reasoning	Cognitive Processing Speed	Factor 5
Vocabulary	**.88**	−.03	.02	−.02	−.01
Information	**.79**	.01	.05	−.01	.02
Similarities	**.77**	.03	.01	.03	−.01
Comprehension	**.70**	.06	−.08	.06	.02
Picture Concepts	.24	.10	.23	.02	.01
Digit Span	−.06	**.81**	.06	.01	.00
Letter-Number Seq.	.08	**.78**	−.07	.00	.01
Picture Span	.00	**.54**	.09	.06	−.02
Arithmetic	.24	**.37**	.17	.07	.01
Visual Puzzles	.03	.00	**.78**	−.04	.00
Block Design	−.01	−.02	**.71**	.11	.02
Figure Weights	.12	.16	**.48**	−.06	.01
Matrix Reasoning	.07	.21	**.43**	.03	.01
Coding	−.01	.02	−.03	**.82**	−.02
Symbol Search	.03	−.01	.08	**.67**	.05
Cancellation	.00	.00	.00	.00	**.99**

Confirmatory Factor Analysis of the WISC-V

In the CFA, various models of the WISC-V were analyzed for model fit. The one-factor model (*g*) exhibited poor fit indices (CFI = .843, RMSEA = .103). Fit indices improved with the addition of each factor. Specifically, the two-factor model exhibited better fit than the one-factor

(*Gc* + *Gr;* CFI = .892, RMSEA = .086), and the three-factor model exhibited better fit than the two-factor model (*Gc* + *Gr* + *Gs,* CFI = .946, RMSEA = .061).

The four-factor model revealed reasonable fit indices (CFI = .970, RMSEA = .046). As hypothesized by the F-CHC model, in the four-factor structure of the WISC-V, Similarities, Information, Comprehension, and Vocabulary form a broad factor of *Gc*. Picture Span, Letter-Number Sequencing, and Digit Span form a broad factor of *Gcm*, while Symbol Search, Coding, and Cancellation form a broad factor of *Gs*. However, Matrix Reasoning, Visual Puzzles, Picture Concepts, Block Design, Figure Weights, and Arithmetic formed a broad factor labeled as perceptual reasoning by the researcher that was inconsistent with the F-CHC or CHC models.

The five-factor model also revealed reasonable fit indices (CFI = .984, RMSEA = .034). The five-factor model also revealed several cross-loadings where two subtests loaded onto two or more of the factors. Similarly to the WAIS-IV, Arithmetic loaded on *Gc, Gr,* and *Gsm.*. Picture Concepts is another subtest that cross-loaded on *Gr* and *Gc.* As hypothesized by the F-CHC model, in the five-factor structure of the WISC-V, Similarities, Information, Comprehension, Vocabulary, Arithmetic, and Picture Concepts form a broad factor of *Gc*. Picture Span, Letter-Number Sequencing, Digit Span, and Arithmetic form a broad factor of *Gcm*, while Symbol Search, Coding, and Cancellation form a broad factor of *Gs*. In the five-factor model, Matrix Reasoning, Picture Concepts, Figure Weights, and Arithmetic now form a broad factor of *Gr*, and Visual Puzzles and Block Design form a broad factor of *Gv*. These results are also consistent with the results of the WAIS-IV CFA analyses and with the F-CHC model.

A six-factor model of intelligence was attempted; however, this only revealed incremental improvement of the model fit. Further, this would require one subtest to form a broad factor, which according to current research (Flanagan et al., 2013), is not recommended. Given the principle of parsimony, the five-factor model, as hypothesized by the F-CHC model, appears to best fit the standardization sample for the WISC-V. Refer to the Figure 14 and 15 for the diagrams of the four-factor and five-factor structure.

Figure 14

Four-Factor Model of WISC-V

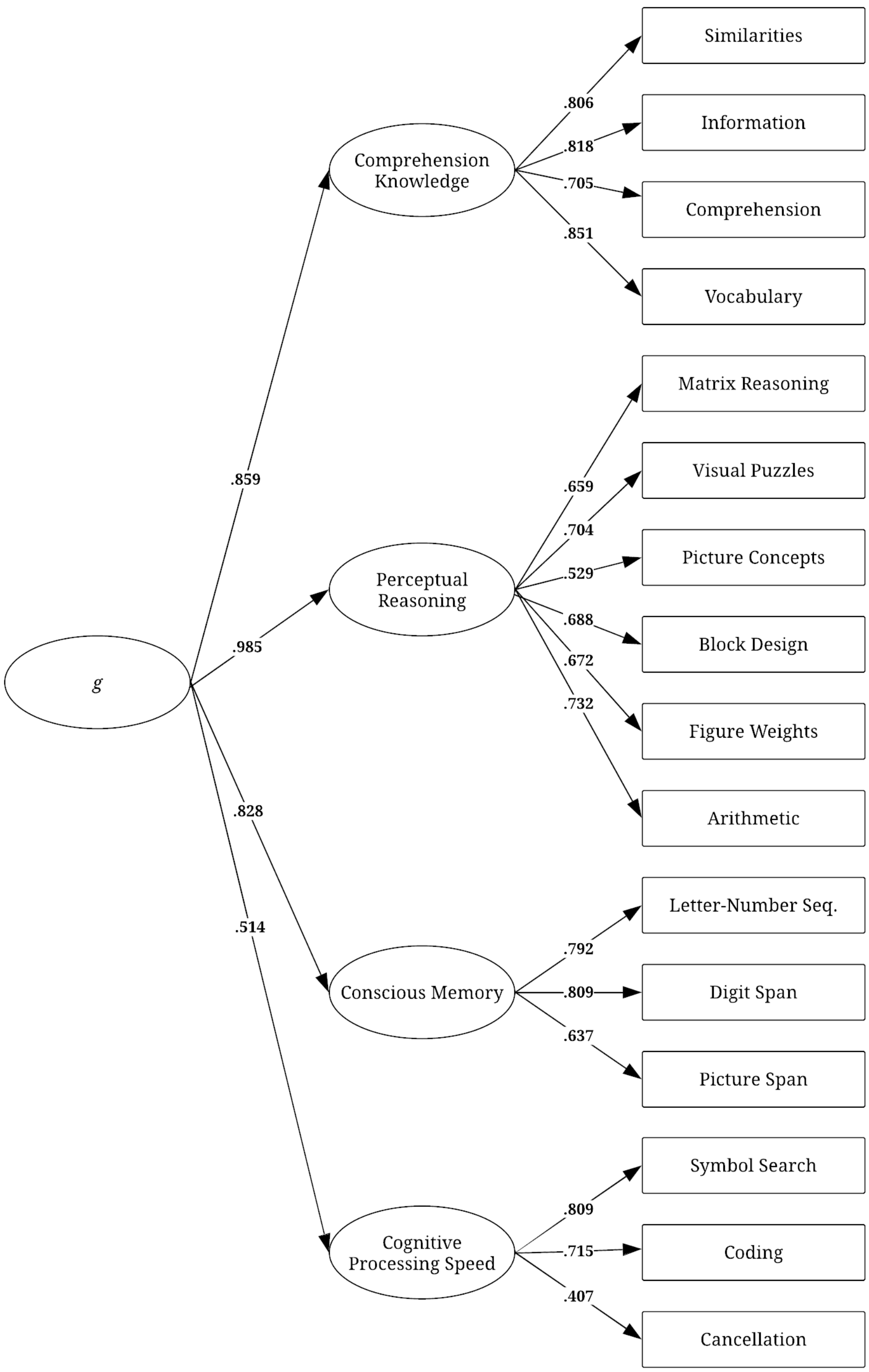
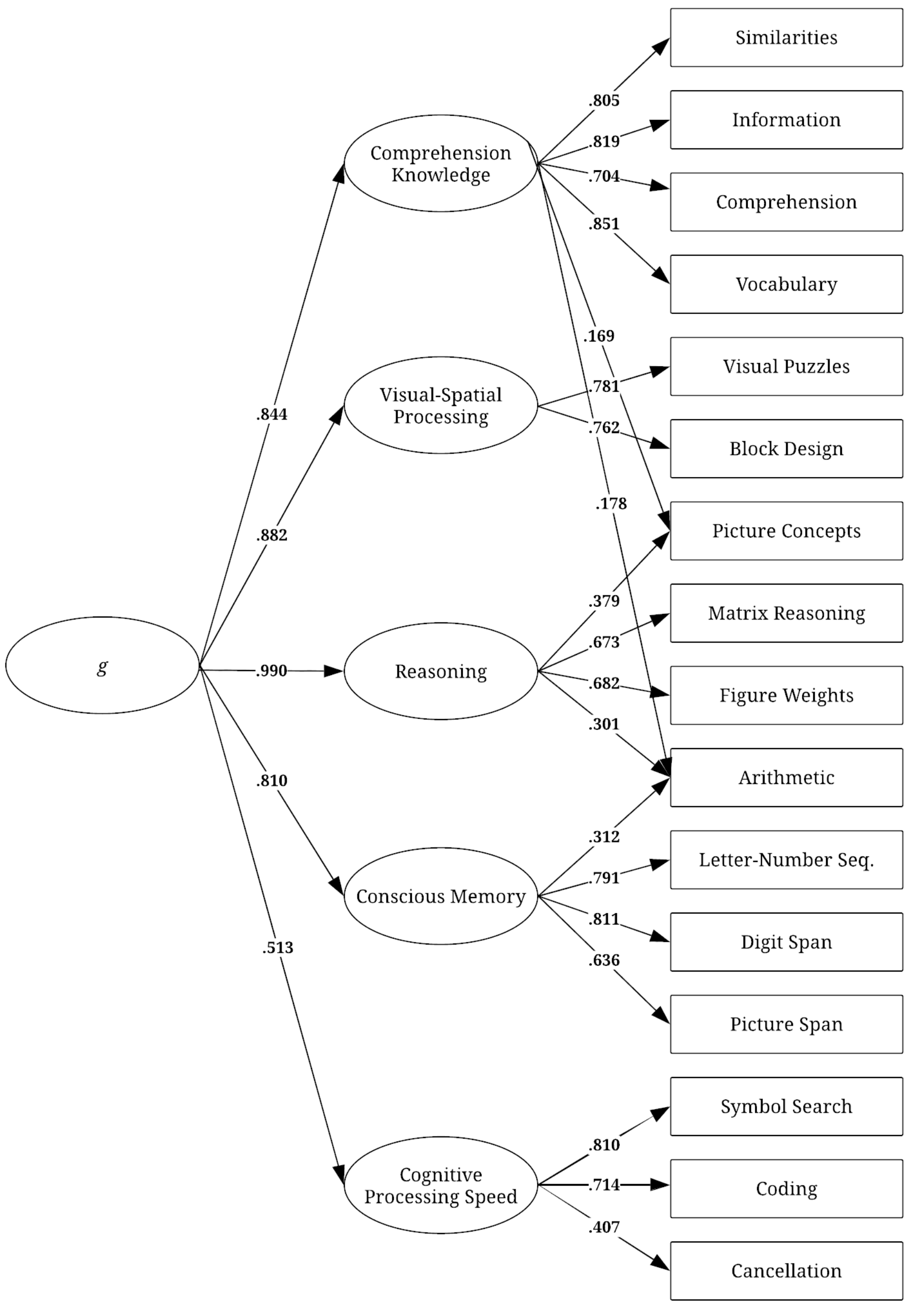


Figure 15

Five-Factor Model of WISC-V



The following analyses were conducted to further examine the factor structure of the WISC-V. Specifically, factor groupings from the CFA (*Gc, Gv, Gcm, Gr,* and *Gs*) were investigated. Multiple EFAs and CFAs were conducted to investigate whether the narrow abilities, as hypothesized by the F-CHC model, could be extracted.

Comprehension Knowledge

In the F-CHC model, Similarities, Vocabulary, Comprehension, and Information are hypothesized to be measures of the construct of *Gc*. These subtests were used to create a correlation matrix (see Table 15).

Table 15

Correlation Matrix of the Hypothesized WISC-V Gc Subtests

	Similarities	Vocabulary	Comprehension	Information
Similarities	1.00			
Vocabulary	.68	1.00		
Comprehension	.59	.60	1.00	
Information	.65	.71	.56	1.00

Exploratory Factor Analysis of the WISC-V Gc. Bartlett's Test of Sphericity revealed that the correlation matrix was not random ($\chi^2 = 4422.94$, $df = 6$, $p < .001$). The KMO Measure of Sampling Adequacy was .83, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis and examination of the visual scree plot (see Figure 16) both suggested a maximum of two factors. Based on the two factors determined by the parallel analysis and visual scree plot, a one-factor and three-factor solution were also examined in the EFA. The EFA revealed that the factors loaded onto two distinct factors. As hypothesized by the F-CHC model, the EFA supported a verbal ability narrow factor consisting of Similarities, Vocabulary, and

Information in addition to a factual knowledge narrow factor consisting of Comprehension. Refer to Table 16 for the factor loadings of the *Gc* subtests.

Figure 16

Visual Scree Plot of the Hypothesized WISC-V Gc Subtests

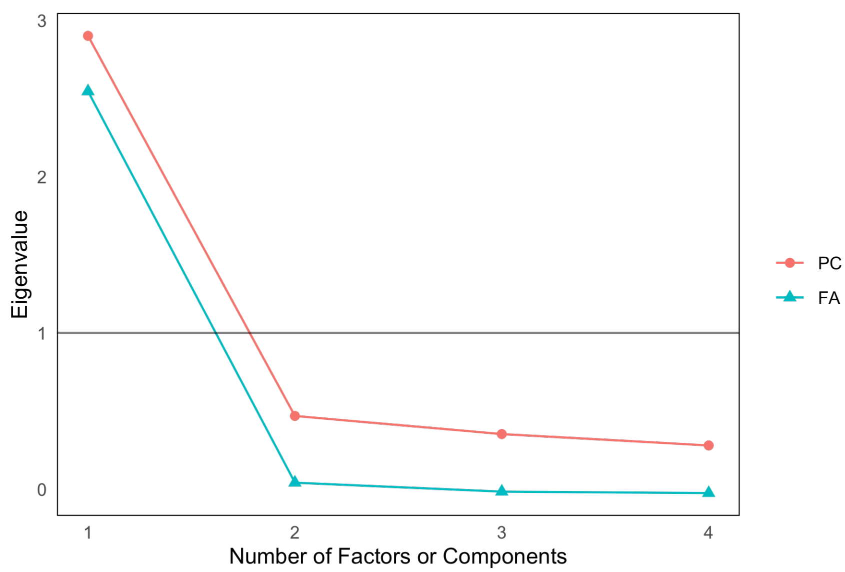


Table 16

EFA Standardized Factor Loadings for the WISC-V Gc Two-Factor Solution

	Verbal Ability	Factual Knowledge
Similarities	**.72**	.09
Vocabulary	**.87**	−.01
Information	**.85**	−.04
Comprehension	.00	**.99**

Confirmatory Factor Analysis of the WISC-V Gc. In the CFA, various models of *Gc* were analyzed for model fit. The one-factor model of *Gc* revealed reasonable fit indices (CFI = .996, RMSEA = .064). The two-factor model of *Gc*, as hypothesized by the F-CHC model, also revealed reasonable fit indices (CFI = .996; RMSEA = .093). In the two-factor structure of *Gc,* Similarities, Vocabulary, and Comprehension form a narrow factor labeled as verbal ability whereas Information forms a narrow factor labeled as factual knowledge. The two-factor structure also appeared to result in a Heywood case (Cooperman & Waller, 2021). A Heywood case is a factor loading value that is near impossible; in this case, that factor loading value was above 1.000. However, the one-factor model of *Gc* proved to be the best fitting model. Refer to Figures 17 and 18 for the diagrams of the one- and two-factor structure.

Figure 17

One-Factor Model of WISC-V Gc

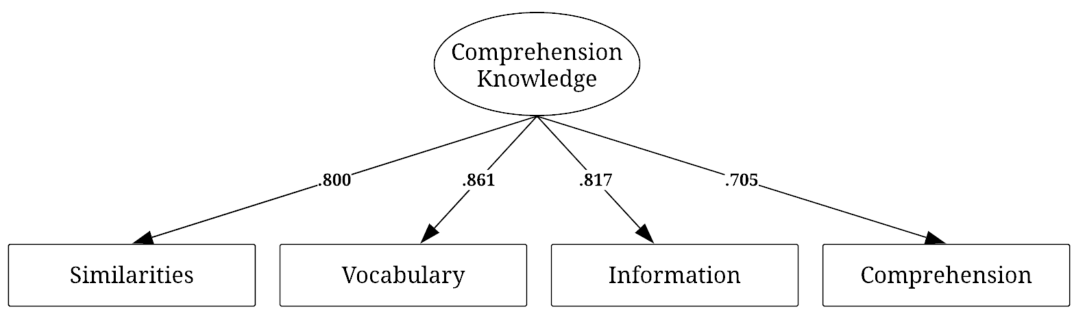
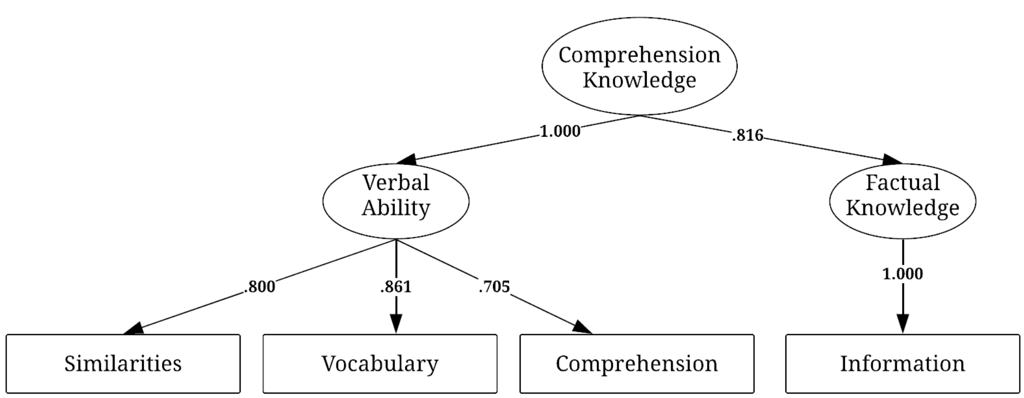


Figure 18

Two-Factor Model of WISC-V Gc



Cognitive Processing Speed

In the F-CHC model, Symbol Search, Cancellation, and Coding are hypothesized to be measures of the construct of *Gs*. These subtests were used to create a correlation matrix (see Table 17).

Table 17

Correlation Matrix of the Hypothesized WISC-V Gs Subtests

	Symbol Search	Cancellation	Coding
Symbol Search	1.00		
Cancellation	.33	1.00	
Coding	.58	.30	1.00

Exploratory Factor Analysis of the WISC-V Gs. Bartlett's Test of Sphericity revealed that the correlation matrix was not random (χ^2 = 1190.414, df = 3, p < .001). The KMO Measure of Sampling Adequacy was .61, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis suggested a maximum of one factor while examination of the visual scree plot (see Figure 19) suggested a maximum of two factors. However, given the limited number of indicators (e.g., subtests), a two-factor model is not statistically relevant. Therefore, a one-factor EFA was conducted and provided support for the broad factor of *Gs* as hypothesized by the F-CHC model. Refer to Table 18 for the factor loadings of the *Gs* subtests.

Figure 19

Visual Scree Plot of the Hypothesized WISC-V Gs Subtests

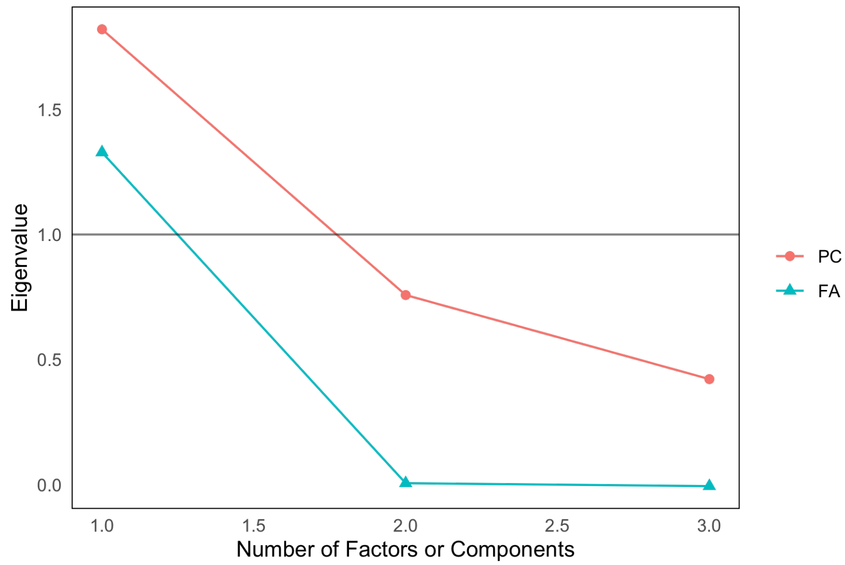


Table 18

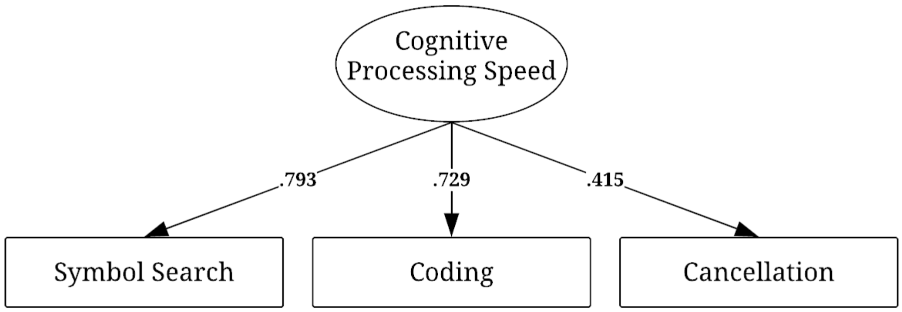
EFA Standardized Factor Loadings for the WISC-V Gs One-Factor Solution

	Cognitive Processing Speed
Symbol Search	**.78**
Coding	**.74**
Cancellation	**.42**

Confirmatory Factor Analysis of the WISC-V Gs. The researcher specified a constraint on the factor loadings in the CFA due to the limited data from the hypothesized subtests. In this case, the indicator loadings were set equal to one another in order to create a just-identified model. The results suggest perfect fit for a one-factor model of *Gs* (CFI = 1.000, RMSEA = .000), indicating support for the broad factor of *Gs*. Refer to Figure 20 for the one-factor structure.

Figure 20

One-Factor Model of WISC-V Gs



Visual-Spatial Processing

In the F-CHC model, Visual Puzzles and Block Design are hypothesized to be measures of the construct of *Gv*. These subtests were used to create a correlation matrix (see Table 19).

Table 19

Correlation Matrix of the Hypothesized WISC-V Gv Subtests

	Visual Puzzles	Block Design
Visual Puzzles	1.00	
Block Design	.60	1.00

Given the limited number of indictors for this particular factor (e.g., two subtests), methods such as EFA and CFA are not statistically relevant. In particular, there is not enough information to fit the number of parameters required by a factor analytic approach. Therefore, this does not allow the researcher to accurately examine the narrow abilities of visual-spatial processing as hypothesized by the F-CHC model or CHC theory. Therefore, the researcher did not pursue the EFA and CFA for the broad ability of *Gv.* While narrow abilities could not be examined, the broad ability of *Gv* was previously supported in the CFA for the WISC-V instrument as a whole. This evidence for the broad ability of *Gv* is consistent with the F-CHC model, CHC theory, and that posited by the Wechsler publishers.

Conscious Memory

In the F-CHC model, Letter-Number Sequencing, Picture Span, and Digit Span are hypothesized to be measures of the construct of *Gcm*. These subtests were used to create a correlation matrix (see Table 20).

Table 20

Correlation Matrix of the Hypothesized WISC-V Gcm Subtests

	Digit Span	Picture Span	Letter-Number Seq.
Digit Span	1.00		
Picture Span	.51	1.00	
Letter-Number Seq.	.65	.49	1.00

Exploratory Factor Analysis of the WISC-V Gcm. Bartlett's Test of Sphericity revealed that the correlation matrix was not random ($\chi^2 = 2002.788$, $df = 3$, $p < .001$). The KMO Measure of Sampling Adequacy was .68, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis suggested a maximum of one factor while examination of the visual scree plot (see Figure 21) suggested a maximum of two factors. However, given the limited number of indicators (e.g., subtests), a two-factor model is not statistically relevant. Therefore, a one-factor EFA was conducted and provided support for the broad factor of *Gcm* as hypothesized by the F-CHC model. Refer to Table 21 for the factor loadings of the *Gcm* subtests.

Figure 21

Visual Scree Plot of the Hypothesized WISC-V Gcm Subtests

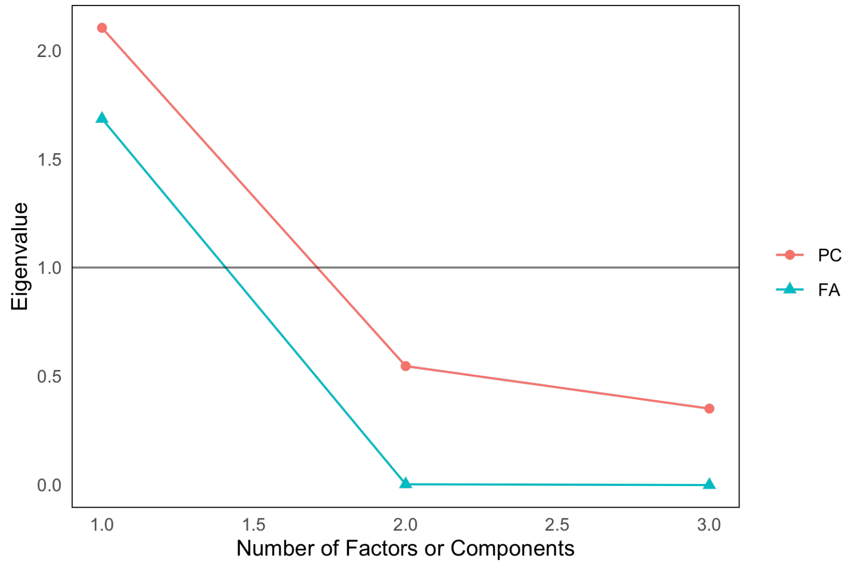


Table 21

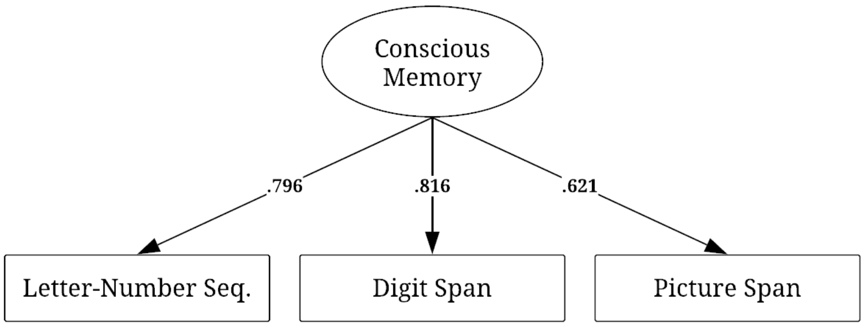
EFA Standardized Factor Loadings for the WISC-V Gcm One-Factor Solution

	Conscious Memory
Digit Span	**.81**
Letter-Number Sequencing	**.80**
Picture Span	**.62**

Confirmatory Factor Analysis of the WISC-V Gcm. The researcher specified a constraint on the factor loadings in the CFA due to the limited data from the hypothesized subtests. In this case, the indicator loadings were set equal to one another in order to create a just-identified model. The results suggest perfect fit for a one-factor model of *Gs* (CFI = 1.000, RMSEA = .000), indicating support for the broad factor of *Gs*. Refer to Figure 22 for the one-factor structure.

Figure 22

One-Factor Model of WISC-V Gcm



Reasoning

In the F-CHC model, Figure Weights, Arithmetic, Matrix Reasoning, and Picture Concepts are hypothesized to be measures of the construct of *Gr*. These subtests were used to create a correlation matrix (see Table 22).

Table 22

Correlation Matrix of the Hypothesized WISC-V Gr Subtests

	Figure Weights	Arithmetic	Matrix Reasoning	Picture Concepts
Figure Weights	1.00			
Arithmetic	.50	1.00		
Matrix Reasoning	.47	.45	1.00	
Picture Concepts	.33	.37	.35	1.00

Exploratory Factor Analysis of the WISC-V Gr. Bartlett's Test of Sphericity revealed that the correlation matrix was not random (χ^2 = 1830.218, *df* = 6, *p* < .001). The KMO Measure of Sampling Adequacy was .76, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis suggested a maximum of one factor while examination of the visual scree plot (see Figure 23) suggested a maximum of two factors. Based on the two factors determined by the parallel analysis and visual scree plot, a one-factor and three-factor solution were also examined in the EFA. The EFA revealed that the factors loaded onto two distinct factors. As hypothesized by the F-CHC model, the EFA supported an inductive/deductive reasoning narrow factor consisting of Figure Weights, Arithmetic, and Matric Reasoning in addition to a contextual reasoning narrow factor consisting of Picture Concepts. It is hypothesized that Picture Concepts is a contextual reasoning narrow ability based on F-CHC theory due to the tasks of this subtest. Further, this subtest cross-loaded on *Gc*, which involves understanding and comprehension of the images that the individual is viewing. Refer to Tables 23 and 24 for the factor loadings of the *Gr* subtests.

Figure 23

Visual Scree Plot of the Hypothesized WISC-V Gr Subtests

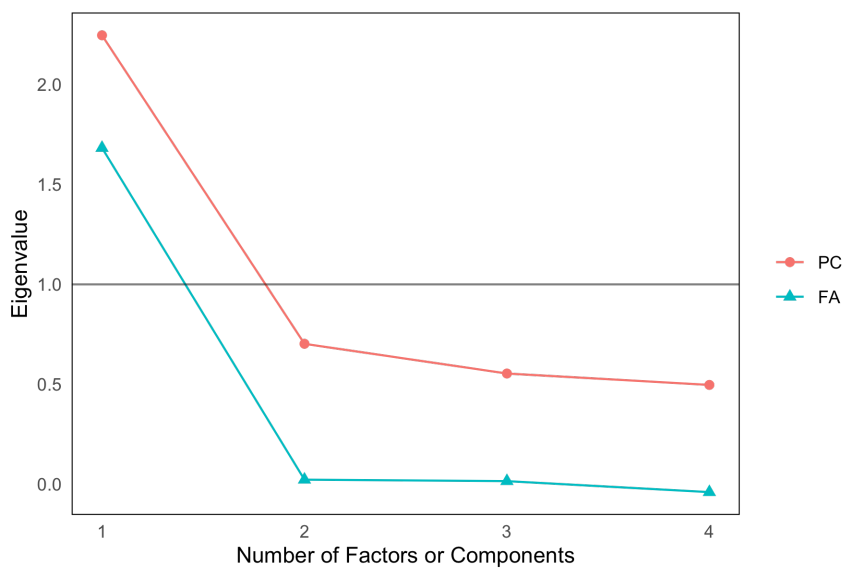


Table 23

EFA Standardized Factor Loadings for the WISC-V Gr One-Factor Solution

	Reasoning
Figure Weights	**.69**
Arithmetic	**.71**
Matrix Reasoning	**.66**
Picture Concepts	**.51**

Table 24

EFA Standardized Factor Loadings for the WISC-V Gr Two-Factor Solution

	Inductive/Deductive Reasoning	Contextual Reasoning
Figure Weights	**.74**	−.04
Arithmetic	**.68**	.03
Matrix Reasoning	**.63**	.03
Picture Concepts	.00	**.99**

Confirmatory Factor Analysis of the WISC-V Gr. In the CFA, various models of *Gr* were analyzed for model fit. The one-factor model of *Gr* revealed reasonable fit indices (CFI = .997, RMSEA = .035). The two-factor model of *Gr*, as hypothesized by the F-CHC model, also revealed reasonable fit indices (CFI = .996; RMSEA = .054). In the two-factor structure of *Gr,* Figure Weights, Arithmetic, and Matrix Reasoning form a narrow factor labeled as inductive/deductive reasoning whereas Picture Concepts forms a narrow factor labeled as contextual reasoning. However, the one-factor model of *Gr* proved to be the best fitting model. Refer to Figure 24 and 25 for the diagrams of the one- and two-factor structure.

Figure 24

One-Factor Model of WISC-V Gr

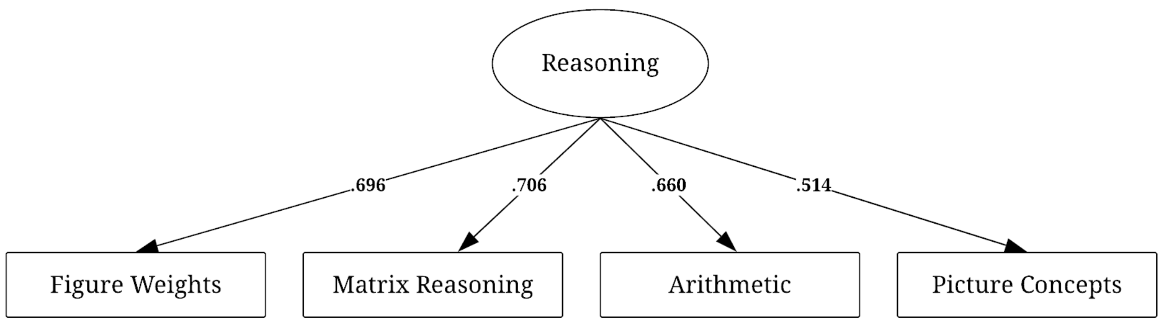
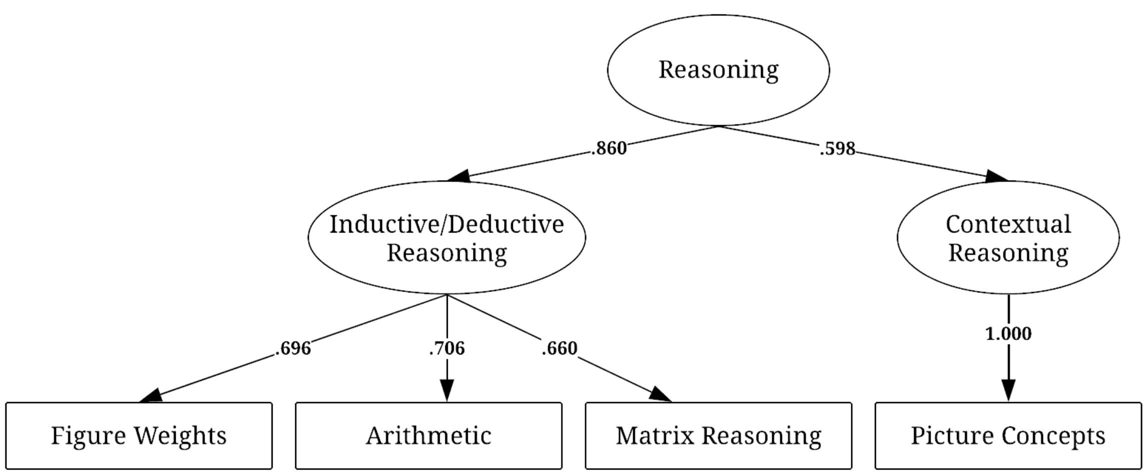


Figure 25

Two-Factor Model of WISC-V Gr



Conclusion

The factor analytic structure of the WISC-V was examined. Evidence was found for both a four-factor and five-factor model, with stronger support for a five-factor model. This five-factor model is consistent with the F-CHC model, CHC theory, and the factor structure posited by the Wechsler publishers. Examination of narrow abilities was attempted for each broad factor.

WPPSI-IV

Preliminary Data Analysis

The analyses in this section are based on the standardization sample from the WPPSI-IV norming study. The preliminary analysis examined the 15 subtests from the WPPSI-IV. For the present study, the researcher removed 600 participants from the standardization sample from the younger age band (i.e., 2:6 to 3:11). These participants were removed from the analysis as the younger age band was not administered eight subtests, resulting in a higher percentage of missing data (see Figure 26). The remaining sample included 1,100 participants and resulted in no missing data (see Figure 27), indicating the sample is suitable for analysis.

Figure 26

Percentage of Missing Values in Each Subtest in the entire WPPSI-IV Dataset

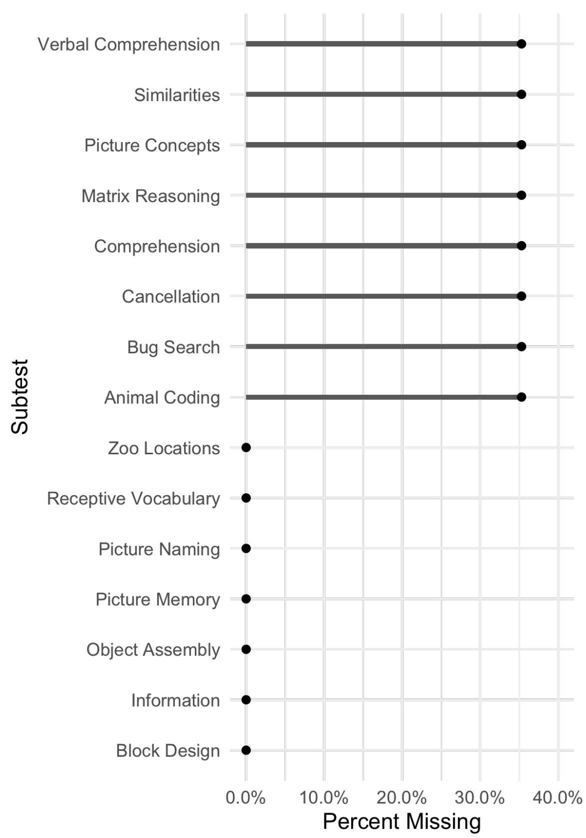
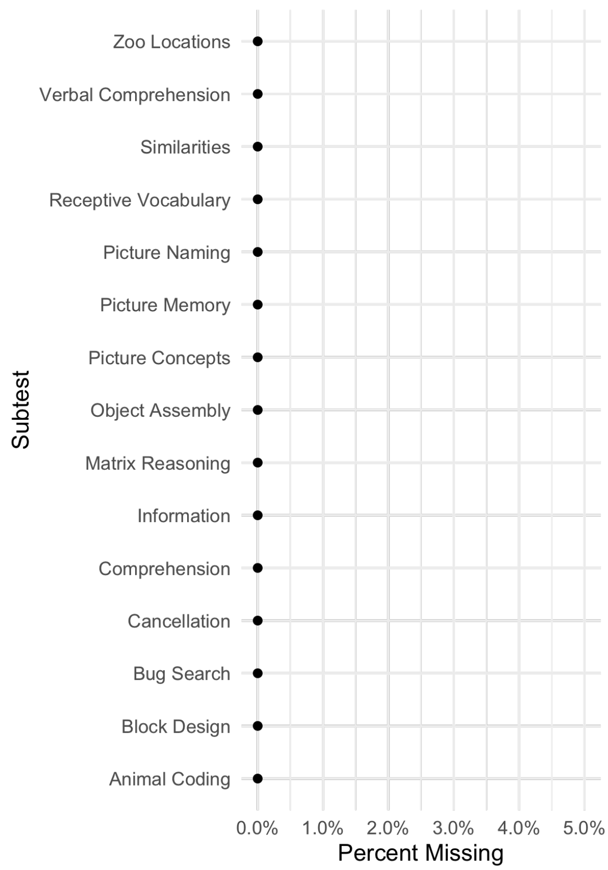


Figure 27

Percentage of Missing Values in Each Subtest in the older age band of the WPPSI-IV Dataset



Primary Data Analysis

The WPPSI-IV subtests were used to create a correlation matrix (see Table 25).

Table 25

Correlation Matrix of the Hypothesized WPPSI-IV Subtests

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1 Similarities	1														
2 Vocabulary	.67	1													
3 Comprehension	.64	.65	1												
4 Information	.64	.60	.59	1											
5 Receptive Vocabulary	.53	.51	.51	.57	1										
6 Picture Naming	.59	.55	.53	.66	.58	1									
7 Object Assembly	.45	.39	.40	.42	.40	.41	1								
8 Block Design	.44	.43	.40	.45	.43	.41	.47	1							
9 Zoo Locations	.30	.32	.31	.32	.33	.30	.32	.36	1						
10 Picture Memory	.43	.42	.39	.43	.40	.36	.37	.38	.39	1					
11 Matrix Reasoning	.46	.44	.40	.45	.45	.43	.44	.49	.39	.41	1				
12 Picture Concepts	.43	.45	.39	.40	.38	.37	.32	.39	.28	.39	.44	1			
13 Bug Search	.36	.36	.34	.37	.36	.31	.36	.39	.38	.39	.39	.33	1		
14 Cancellation	.27	.30	.31	.32	.30	.27	.30	.30	.25	.30	.28	.27	.46	1	
15 Animal Coding	.33	.34	.28	.34	.29	.31	.31	.33	.32	.36	.34	.30	.54	.43	1

Exploratory Factor Analysis of the WPPSI-IV

Bartlett's Test of Sphericity revealed that the correlation matrix was not random (χ^2 = 7211.42, *df* = 105, *p* < .001). The KMO Measure of Sampling Adequacy was .94, indicating that the data is appropriate for analysis (Kline, 2013). The WPPSI-IV factor structure was analyzed to determine the number of factors to be extracted. Parallel analysis suggested a maximum of four factors while examination of the visual scree plot suggested four to five factors (see Figure 28). The EFA supported four to five broad factors. In the four-factor model, a strong broad factor consisted of *Gs* tasks. A second factor appeared to include tasks of verbal ability, a narrow ability of *Gc*. The second narrow ability of factual knowledge under *Gc* appears to be represented by a third broad factor and included the tasks of Picture Naming, Information, and Receptive Vocabulary. This divide of *Gc* into distinct broad factors suggests that the tasks are measuring different cognitive constructs than originally hypothesized by the publishers. The fourth factor consisted of a compilation of subtests from *Gv, Gcm,* and *Gr* with no clear definition as to what the factor was measuring, although it could be hypothesized that a visual component is what the tasks share or have in common.

In the five-factor structure for the WPPSI-IV, the broad factors also consisted of *Gc* and *Gs*, with an additional factor of *Gv-Gr*. Picture Memory, Picture Naming, and Receptive Vocabulary did not load strongly on any one factor. Zoo Locations appeared to create its own factor, which might be labeled as an immediate recall factor. Refer to Tables 26 and 27 below for the four-factor and five-factor solutions for the EFA.

Figure 28

Visual Scree Plot of the Hypothesized WPPSI-IV Subtests

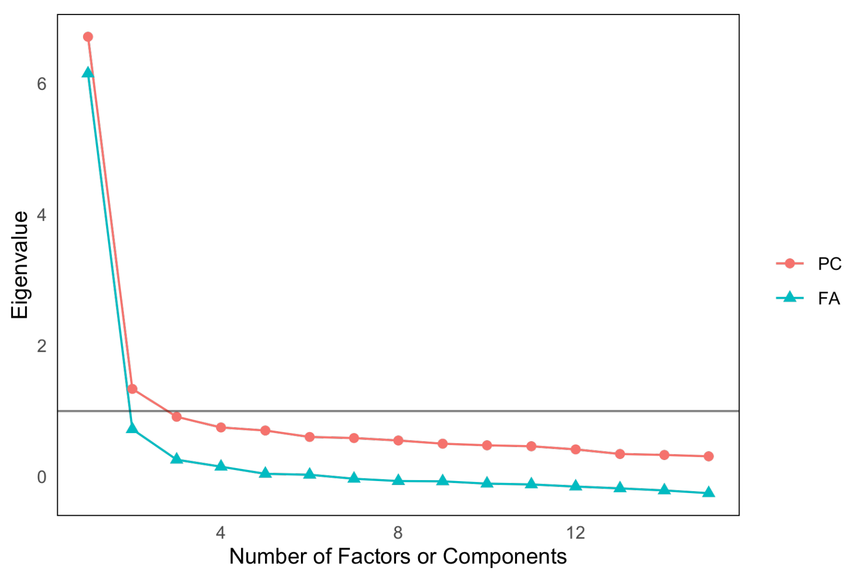


Table 26

EFA Standardized Factor Loadings for the WPPSI-IV Four-Factor Solution

	Verbal Ability	Factor 2	Cognitive Processing Speed	Factual Knowledge
Vocabulary	**.85**	−.02	.02	−.03
Comprehension	**.69**	−.01	.02	.10
Similarities	**.66**	.07	-.01	.15
Picture Concepts	.36	**.32**	.05	−.10
Block Design	−.02	**.69**	.00	.07
Matrix Reasoning	.09	**.61**	.00	.03
Object Assembly	−.05	**.56**	.04	.17
Zoo Locations	.00	**.51**	.15	−.07
Picture Memory	.29	**.37**	.15	−.16
Animal Coding	.00	−.05	**.74**	.01
Bug Search	−.02	.08	**.72**	−.02
Cancellation	.01	−.05	**.61**	.06
Picture Naming	.09	.07	.06	**.69**
Information	.32	.09	.09	**.42**
Receptive Vocabulary	.14	.22	.07	**.41**

Table 27

EFA Standardized Factor Loadings for the WPPSI-IV Five-Factor Solution

	Comprehension Knowledge	Cognitive Processing Speed	Perceptual Reasoning	Immediate Recall	Factor 5
Vocabulary	**.83**	.03	−.02	.02	−.06
Comprehension	**.74**	.02	−.06	.08	.06
Similarities	**.70**	.01	.12	−.02	.07
Information	**.46**	.09	.13	.03	.29
Picture Memory	.27	.15	.18	.21	−.13
Animal Coding	.00	**.75**	−.04	−.01	−.01
Bug Search	−.03	**.72**	.03	.05	−.02
Cancellation	.01	**.63**	−.01	−.05	.03
Block Design	−.05	.03	**.70**	.05	.02
Matrix Reasoning	.09	.03	**.58**	.07	−.03
Object Assembly	−.03	.07	**.54**	.07	.12
Picture Concepts	.34	.06	**.39**	−.05	−.19
Zoo Locations	.00	.00	.01	**.81**	−.01
Picture Naming	.34	.06	.14	.03	.44
Receptive Vocabulary	.27	.07	.21	.08	.29

Confirmatory Factor Analysis of the WPPSI-IV

In the CFA, various models of the WPPSI-IV were analyzed for model fit. The one-factor model (*g*) exhibited poor fit indices (CFI = .879, RMSEA = .093). Fit indices improved with the addition of each factor. Specifically, the two-factor model exhibited better fit than the one-factor (*Gc* + *Gr;* CFI = .942, RMSEA = .066), and the three-factor model exhibited better fit than the two-factor model (*Gc* + *Gr* + *Gs,* CFI = .972, RMSEA = .046).

The four-factor model revealed reasonable fit indices (CFI = .974, RMSEA = .044). As hypothesized by the F-CHC model, in the four-factor structure of the WPPSI-IV, Similarities, Comprehension, Vocabulary, Receptive Vocabulary, Information, and Picture Naming form a broad factor of *Gc*. Also hypothesized by the F-CHC model, Animal Coding, Bug Search, and Cancellation appear to form a broad factor of *Gs*. Zoo Locations and Picture Memory formed a broad factor of *Gsm*. However, in the four-factor model, Block Design, Object Assembly, Matrix Reasoning, and Picture Concepts formed a broad factor labeled as perceptual reasoning by the researcher, different than their hypothesized role in the F-CHC and CHC models.

The five-factor model also revealed reasonable fit indices (CFI = .976, RMSEA = .043). As hypothesized by the F-CHC model, in the five-factor structure of the WPPSI-IV, Similarities, Comprehension, Vocabulary, Receptive Vocabulary, Information, and Picture Naming form a broad factor of *Gc*. Also hypothesized by the F-CHC model, Animal Coding, Bug Search, and Cancellation appear to form a broad factor of *Gs*. Block Design and Object Assembly formed a broad factor of *Gv*, while Matrix Reasoning and Picture Concepts formed a broad factor of *Gr*. Finally, Zoo Locations and Picture Memory form a broad factor of *Gcm*. The five-factor model corresponds the closest with the F-CHC model and appears to be the best fit for the standardization sample for the WPPSI-IV. Refer to Figure 29 and 30 for the diagrams of the four-factor and five-factor CFA structures.

Figure 29

Four-Factor Model of WPPSI-IV

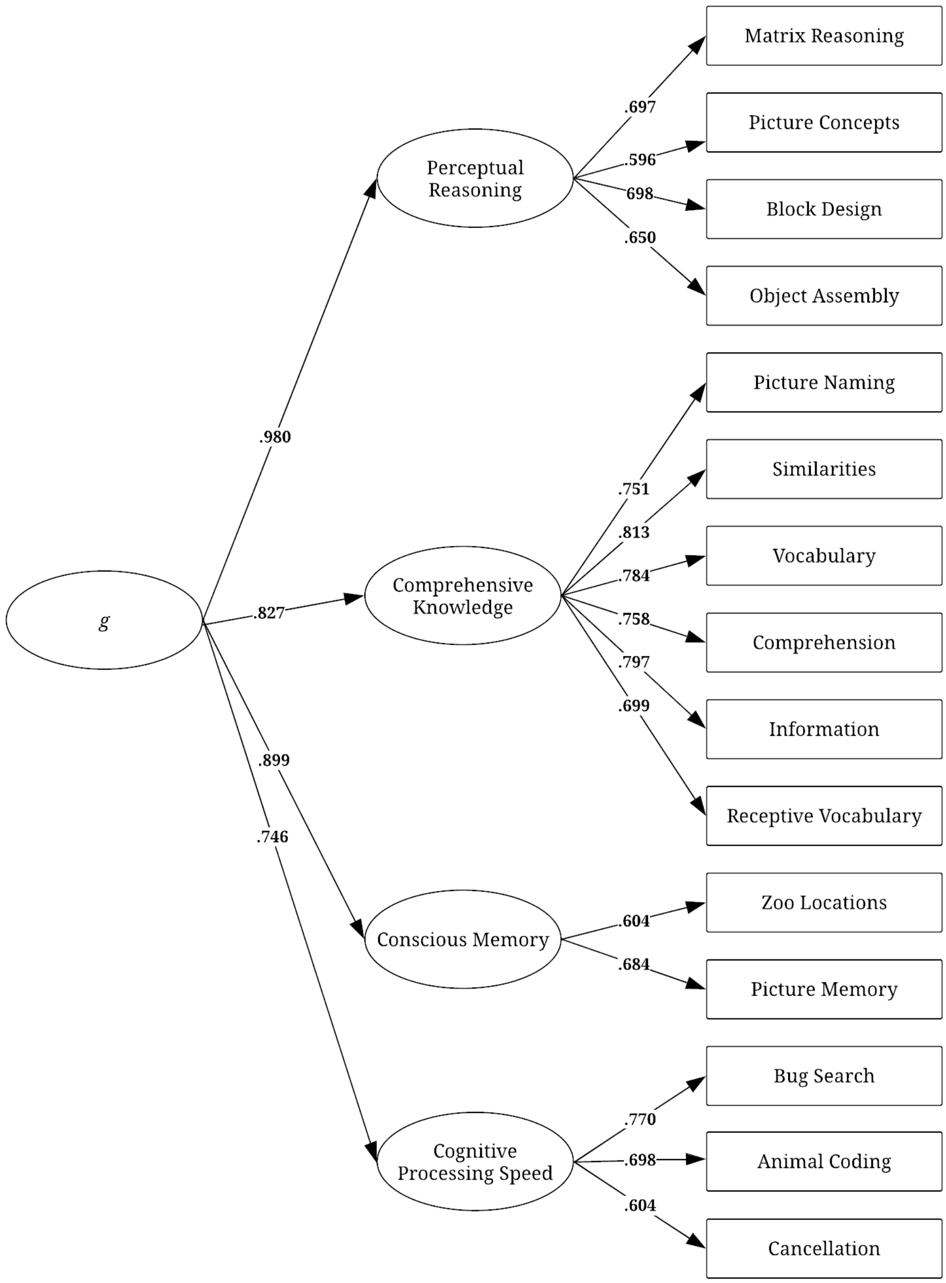
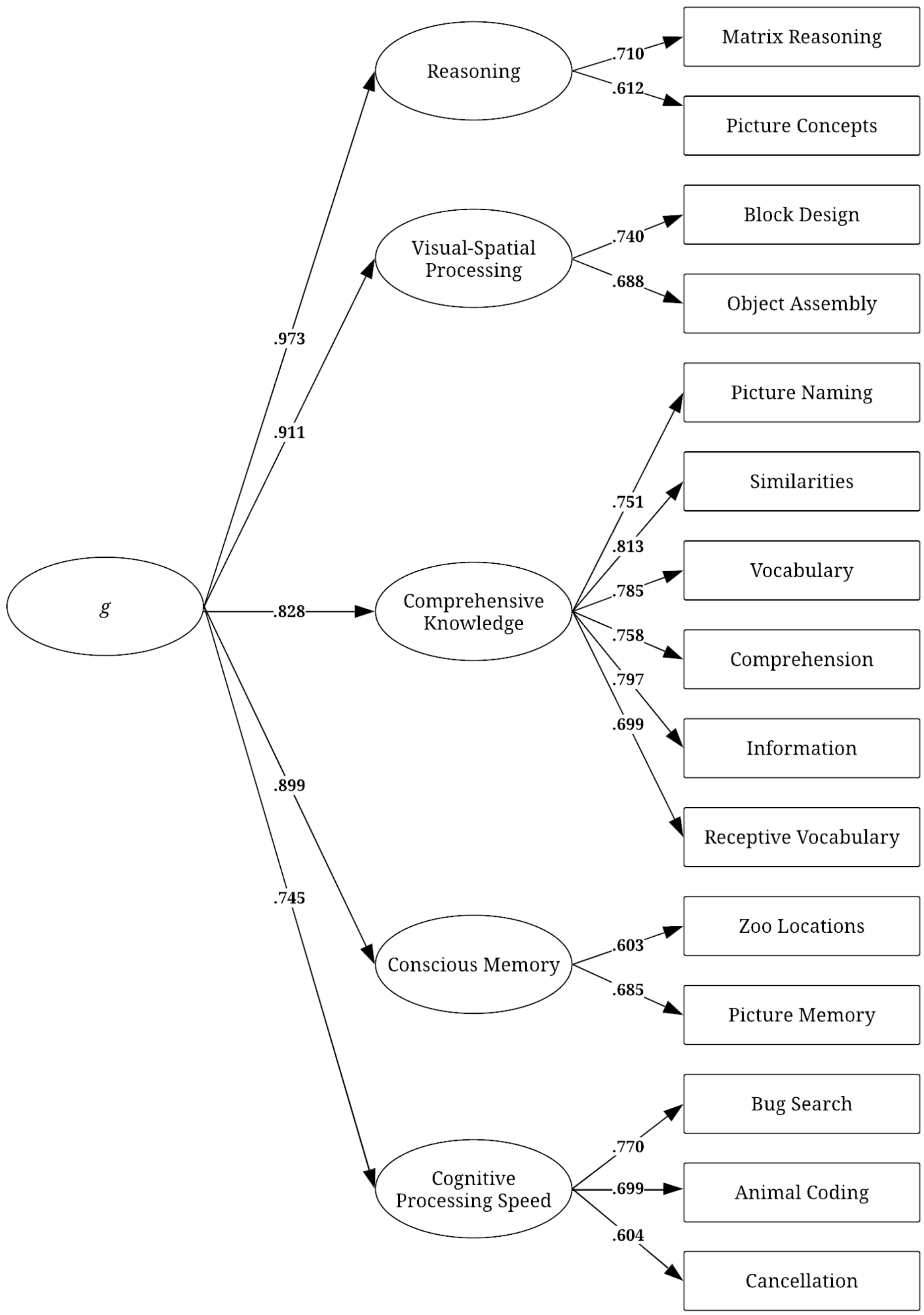


Figure 30

Five-Factor Model of WPPSI-IV



The following analyses were conducted to further examine the factor structure of the WPPSI-IV. Specifically, factor grouping from the CFA (*Gc, Gv, Gcm, Gr,* and *Gs*) were investigated. Multiple EFAs and CFAs were conducted to investigate whether the narrow abilities, as hypothesized by the F-CHC model, could be extracted.

Comprehension Knowledge

In the F-CHC model, Similarities, Vocabulary, Comprehension, Information, Receptive Vocabulary, and Picture Naming are hypothesized to be measures of the construct of *Gc*. These subtests were used to create a correlation matrix (see Table 28).

Table 28

Correlation Matrix of the Hypothesized WPPSI-IV Gc Subtests

	Similarities	Vocabulary	Comprehension	Information	Receptive Vocab.	Picture Naming
Similarities	1.00					
Vocabulary	.67	1.00				
Comprehension	.64	.65	1.00			
Information	.64	.60	.59	1.00		
Receptive Vocab.	.53	.51	.51	.57	1.00	
Picture Naming	.58	.55	.53	.66	.58	1.00

Exploratory Factor Analysis of the WPPSI-IV Gc. Bartlett's Test of Sphericity revealed that the correlation matrix was not random ($\chi^2 = 3510.57$, $df = 15$, $p < .001$). The KMO Measure of Sampling Adequacy was .90, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis and examination of the visual scree plot (see Figure 31) both suggested a maximum of two factors. Based on the two factors determined by the parallel analysis and visual scree plot, a one-factor and three-factor solution were also examined in the EFA. The EFA revealed that the factors loaded onto two distinct factors. As hypothesized by the

F-CHC model, the EFA supported a verbal ability narrow factor consisting of Similarities, Vocabulary, and Comprehension in addition to a factual knowledge narrow factor consisting of Picture Naming, Receptive Vocabulary, and Information. Refer to Table 29 for the factor loadings of the *Gc* subtests.

Figure 31

Visual Scree Plot of the Hypothesized WPPSI-IV Gc Subtests

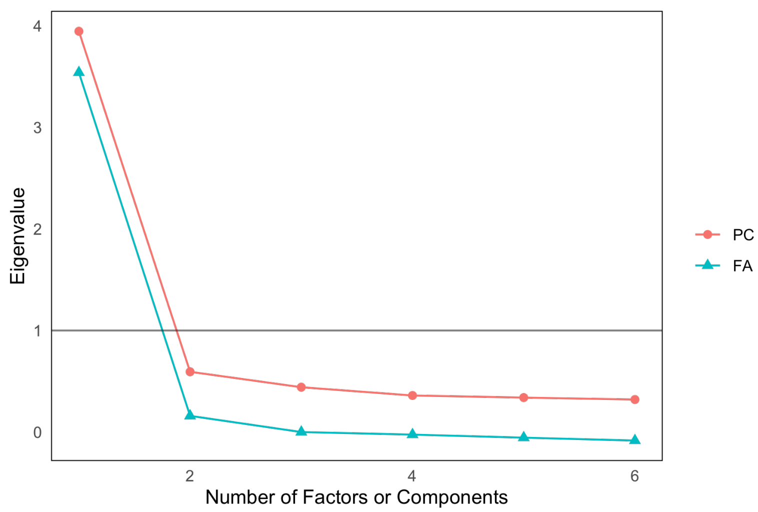


Table 29

EFA Standardized Factor Loadings for the WPPSI-IV Gc Two-Factor Solution

	Verbal Ability	Factual Knowledge
Vocabulary	**.87**	−.06
Comprehension	**.77**	.02
Similarities	**.70**	.14
Picture Naming	−.04	**.85**
Receptive Vocabulary	.03	**.70**
Information	.27	**.56**

Confirmatory Factor Analysis of the WPPSI-IV Gc. In the CFA, various models of *Gc* were analyzed for model fit. The one-factor model of *Gc* revealed reasonable fit indices (CFI = .974, RMSEA = .096). The two-factor model of *Gc*, as hypothesized by the F-CHC model, also revealed reasonable fit indices (CFI = .995, RMSEA = .045). Finally, a three-factor model of *Gc* revealed better fit indices (CFI = .999, RMSEA = .026). As hypothesized by the F-CHC model, in the two-factor structure Similarities, Vocabulary, and Comprehension form a narrow factor labeled as verbal ability, whereas Information, Picture Naming, and Receptive Vocabulary form a narrow factor labeled as factual knowledge. However, the three-factor structure appears to fit the data better than the two-factor structure. In the three-factor structure of *Gc,* Similarities, Vocabulary, and Comprehension form a narrow factor labeled as verbal ability, Information forms the narrow ability of factual knowledge, and Picture Naming and Receptive Vocabulary form a narrow factor labeled as lexical access. Refer to Figures 32, 33, and 34 for the diagrams of the one-, two-, and three-factor structure.

Figure 32

One-Factor Model of WPPSI-IV Gc

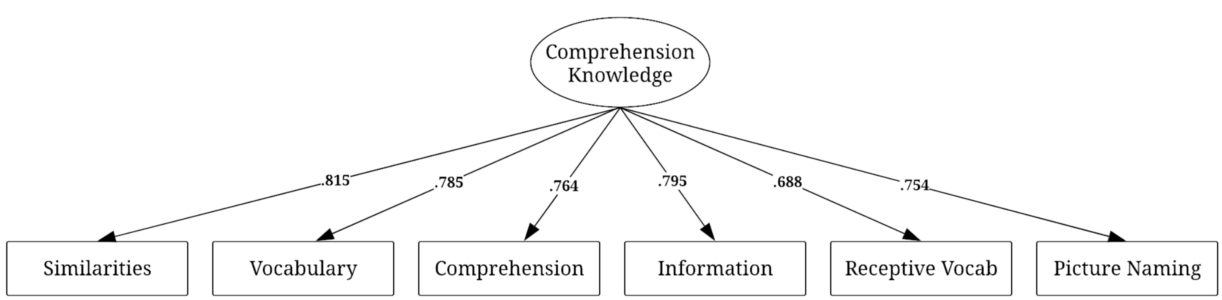


Figure 33

Two-Factor Model of WPPSI-IV Gc

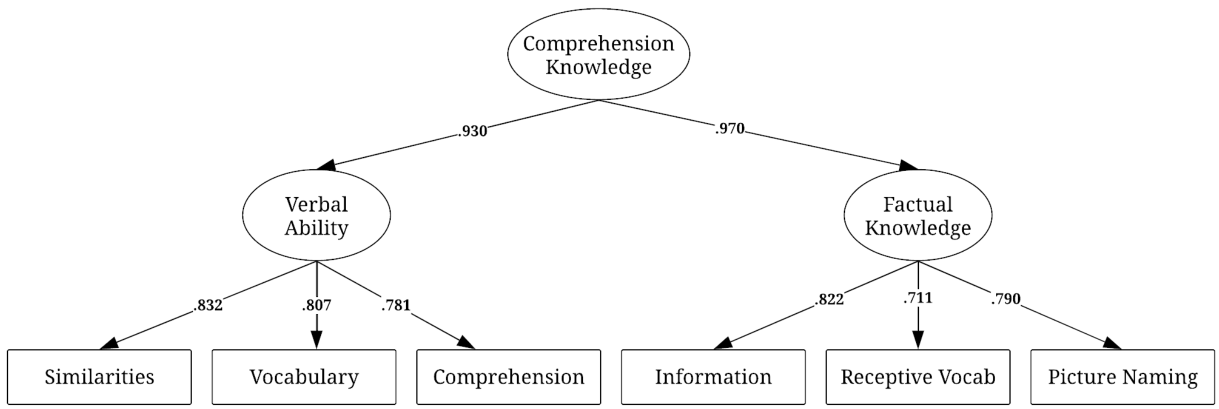
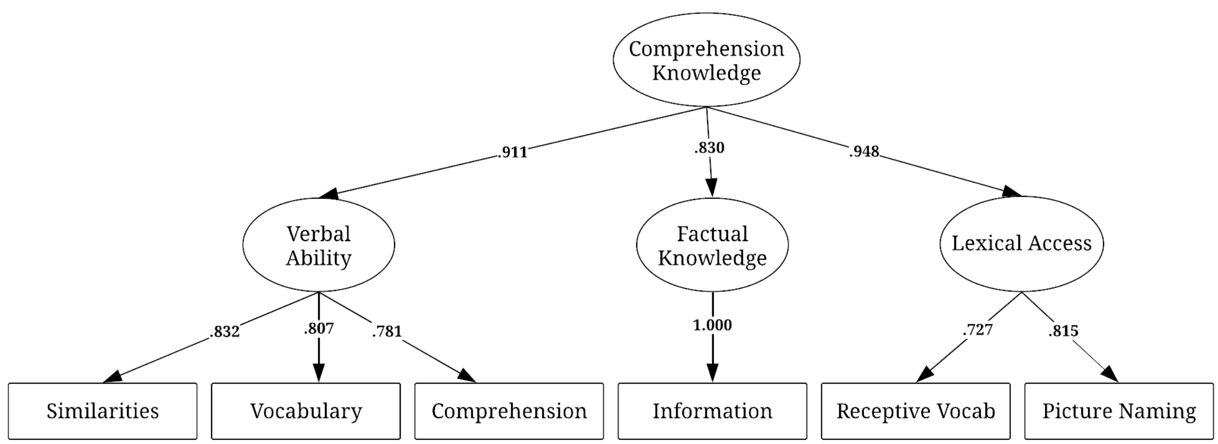


Figure 34

Three-Factor Model of WPPSI-IV Gc



Cognitive Processing Speed

In the F-CHC model, Bug Search, Cancellation, and Animal Coding are hypothesized to be measures of the construct of *Gs*. These subtests were used to create a correlation matrix (see Table 30).

Table 30

Correlation Matrix of the Hypothesized WPPSI-IV Gs Subtests

	Bug Search	Cancellation	Animal Coding
Bug Search	1.00		
Cancellation	.46	1.00	
Animal Coding	.54	.43	1.00

Exploratory Factor Analysis of the WPPSI-IV Gs. Bartlett's Test of Sphericity revealed that the correlation matrix was not random ($\chi^2 = 701.7826$, $df = 3$, $p < .001$). The KMO Measure of Sampling Adequacy was .67, indicating that the data is appropriate for analysis (Kline, 2013). Parallel analysis suggested a maximum of one factor while examination of the visual scree plot (see Figure 35) suggested a maximum of two factors. However, given the limited number of indicators (e.g., subtests), a two-factor model is not statistically relevant. Therefore, a one-factor EFA was conducted and provided support for the broad factor of *Gs* as hypothesized by the F-CHC model. Refer to Table 31 for the factor loadings of the *Gs* subtests.

Figure 35

Visual Scree Plot of the Hypothesized WPPSI-IV Gs Subtests

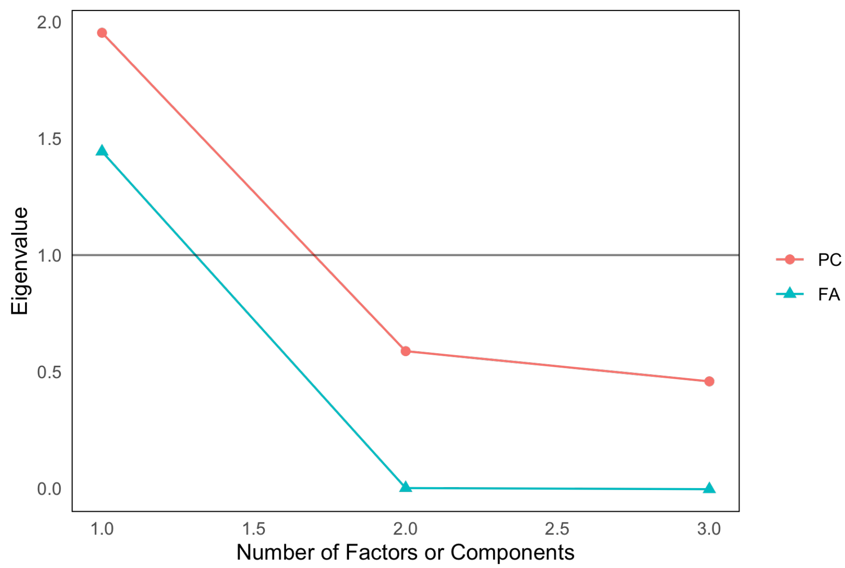


Table 31

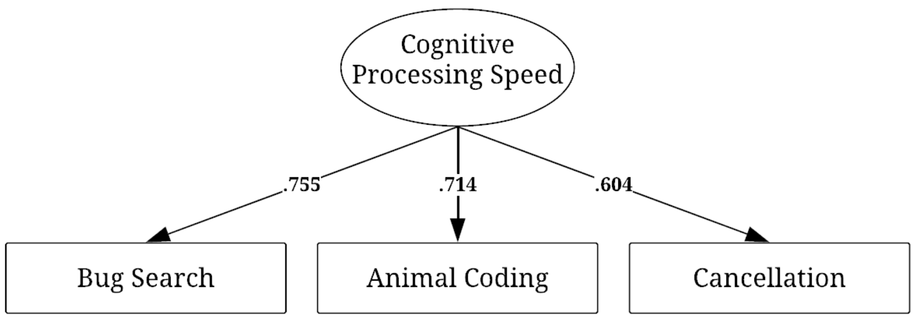
EFA Standardized Factor Loadings for the WPPSI-IV Gs One-Factor Solution

	Cognitive Processing Speed
Bug Search	**.75**
Animal Coding	**.72**
Cancellation	**.60**

Confirmatory Factor Analysis of the WPPSI-IV Gs. The researcher specified a constraint on the factor loadings in the CFA due to the limited data from the hypothesized subtests. In this case, the indicator loadings were set equal to one another in order to create a just-identified model. The results suggest perfect fit for a one-factor model of *Gs* (CFI = 1.000, RMSEA = .000), indicating support for the broad factor of *Gs*. Refer to Figure 36 for the one- and two-factor structure.

Figure 36

One-Factor Model of WPPSI-IV Gs



Visual-Spatial Processing

In the F-CHC model, Object Assembly and Block Design are hypothesized to be measures of the construct of *Gv*. These subtests were used to create a correlation matrix (see Table 32).

Table 32

Correlation Matrix of the Hypothesized WPPSI-IV Gv Subtests

	Object Assembly	Block Design
Object Assembly	1.00	
Block Design	.51	1.00

Given the limited number of indictors for this particular factor (e.g., two subtests), methods such as EFA and CFA are not statistically relevant. In particular, there is not enough information to fit the number of parameters required by a factor analytic approach. Therefore, this does not allow the researcher to accurately examine the narrow abilities of *Gv* as hypothesized by the F-CHC model or CHC theory. Therefore, the researcher did not pursue the EFA and CFA for the broad ability of *Gv*. While narrow abilities could not be examined, the broad ability of *Gv* was previously supported in the CFA for the WPPSI-IV instrument as a whole. This evidence for the broad ability of *Gv* is consistent with the F-CHC model, CHC theory, and that posited by the Wechsler publishers.

Conscious Memory

In the F-CHC model, Zoo Locations and Picture Memory are hypothesized to be measures of the construct of *Gcm*. These subtests were used to create a correlation matrix (see Table 33).

Table 33

Correlation Matrix of the Hypothesized WPPSI-IV Gcm Subtests

	Zoo Locations	Picture Memory
Zoo Locations	1.00	
Picture Memory	.43	1.00

Given the limited number of indictors for this particular factor (e.g., two subtests), methods such as EFA and CFA are not statistically relevant. In particular, there is not enough information to fit the number of parameters required by a factor analytic approach. Therefore, this does not allow the researcher to accurately examine the narrow abilities of *Gcm* as hypothesized by the F-CHC model or CHC theory. Therefore, the researcher did not pursue the EFA and CFA for the broad ability of *Gcm.* While narrow abilities could not be examined, the broad ability of *Gcm* was previously supported in the CFA for the WPPSI-IV instrument as a whole. This evidence for the broad ability of *Gcm* is consistent with the F-CHC model, CHC theory, and that posited by the Wechsler publishers.

Reasoning

In the F-CHC model, Matrix Reasoning and Picture Concepts are hypothesized to be measures of the construct of *Gr*. These subtests were used to create a correlation matrix (see Table 34).

Table 34

Correlation Matrix of the Hypothesized WPPSI-IV Gr Subtests

	Matrix Reasoning	Picture Concepts
Matrix Reasoning	1.00	
Picture Concepts	.44	1.00

Given the limited number of indictors for this particular factor (e.g., two subtests), methods such as EFA and CFA are not statistically relevant. In particular, there is not enough information to fit the number of parameters required by a factor analytic approach. Therefore, this does not allow the researcher to accurately examine the narrow abilities of *Gr* as hypothesized by the F-CHC model or CHC theory. Therefore, the researcher did not pursue the EFA and CFA for the broad ability of *Gr.* While narrow abilities could not be examined, the broad ability of *Gr* was previously supported in the CFA for the WAIS-IV instrument as a whole. This evidence for the broad ability of *Gr* is consistent with the F-CHC model, CHC theory, and that posited by the Wechsler publishers.

Conclusion

The factor analytic structure of the WPPSI-IV was examined. Evidence was found for both a four-factor and five-factor model, with stronger support for a five-factor model. This five-factor model is consistent with the F-CHC model, CHC theory, and the factor structure posited by the Wechsler publishers. Examination of narrow abilities was attempted for each broad factor.

Final Conclusions

This chapter presented and described the results of the analyses. Preliminary analyses for each of the instruments revealed minimal missing data for the WISC-V and no missing data for the WPPSI-IV. For the WAIS-IV, the missing values analysis indicated that 18% of values were missing from three of the subtests. Therefore, the researcher utilized pairwise computation of the correlation matrix to handle the missing data. The primary analyses conducted included multiple exploratory and confirmatory factor analyses on the 15 subtests from the WAIS-IV, 16 subtests from the WISC-V, and 15 subtests from the WPPSI-IV. First, the hypothesized groupings of broad abilities were examined followed by the hypothesized groupings of narrow abilities for each instrument. In the EFA, factors were extracted using principal axis factoring with an oblique, oblimin rotation. In the CFA, fit indices included CFI and RMSEA. The results of the present analysis suggest that the structure of the Wechsler cognitive assessment instruments may differ from what is presented by the publishers and contemporary theories of intelligence (i.e., F-CHC and CHC). The results provide partial support for the proposed factor structure of the F-CHC model. However, the Wechsler instruments utilized do not appear robust enough to truly investigate the validity of the full F-CHC model and all the narrow abilities. Specifically, there is a limited number of indicators (e.g., subtests) for each instrument to appropriately examine narrow abilities. These limitations and implications are further discussed in Chapter 5.

CHAPTER V

DISCUSSION

The present study sought to examine the factor structure of the Wechsler cognitive assessment instruments (WAIS-IV, WPPSI-IV, and WISC-V). Specifically, EFA was first employed to explore the factor structure with no restraint, followed by CFA to determine whether the CHC theory or the F-CHC model best explained the factor structure of the instruments compared to that outlined by the publishers. The F-CHC model and CHC theory were both examined as these are the prominent theories of intelligence in the field. The publishers of the Wechsler cognitive assessment instruments utilized functional and structural theories, including theories from cognitive psychology, neuropsychology, and specific cognitive abilities such as memory, attention, and executive functioning (Kaufman et al., 2016; Wechsler, 2008, 2012, 2014) in the creation of the cognitive instruments. Upon the discovery of the CHC theory, the publishers applied this theory post-hoc to the instruments and posited that the index scores were related to cognitive abilities of the CHC theory. The WAIS-IV (Wechsler, 2008) claims to measure *Gc*, *Gwm*, *Gs*, and *Gv-Gr*. The WISC-V (Wechsler, 2014) and WPPSI-IV (Wechsler, 2012) claim to measure *Gc*, *Gf*, *Gv*, *Gwm*, and *Gs*. The results of the present study provided some support for the factor structure of the F-CHC model. However, the Wechsler instruments utilized do not appear robust enough to truly investigate the validity of the full F-CHC model and all the narrow abilities. This chapter provides a review of the results from the statistical analyses and a discussion of the limitations of the study, implications for practice, and directions for future research.

Review of Statistical Analyses

Statistical analyses were conducted in RStudio. Preliminary analyses for each of the instruments revealed minimal missing data for the WISC-V and no missing data for the WPPSI-IV. For the WAIS-IV, the missing values analysis indicated that 18% of values were missing from three of the subtests. Therefore, the researcher utilized pairwise computation of the correlation matrix to handle the missing data. Bivariate correlations were conducted to examine the similarities between the observed variables. The primary analyses conducted included multiple exploratory and confirmatory factor analyses on the 15 subtests from the WAIS-IV, 16 subtests from the WISC-V, and 15 subtests from the WPPSI-IV. First, the hypothesized groupings of broad abilities were examined followed by the hypothesized groupings of narrow abilities for each instrument. The CHC theory of intelligence (Schneider & McGrew, 2012, 2018) was examined as it is the most well-known theory of intelligence, and because of its post-hoc application to the Wechsler instruments. The F-CHC model of cognitive abilities was also examined as it is an emerging, application-based, reconceptualization of CHC theory based on current neurocognitive research (Woodcock et al., 2017; Woodcock et al., 2018). The F-CHC model was included in an attempt to further validate the model and provide support for its hypothesized broad and narrow cognitive abilities.

In regard to the researcher's hypotheses, the results of the present analysis partially supported the third hypothesis, suggesting that the structure of the Wechsler cognitive assessment instruments may differ from what is presented by the publishers and contemporary theories of intelligence (i.e., F-CHC and CHC). However, many of the analyses provided support for broad abilities largely consistent with that of the publishers, the F-CHC model, and the CHC theory. The results also provide partial support for the researcher's first and second hypotheses.

The data from the Wechsler standardization samples for each instrument appear to partially support the F-CHC model. In particular, the broad abilities are supported by the model. However, there does not appear to be enough data from the Wechsler instruments to provide support for all of the narrow abilities as hypothesized by the F-CHC model and/or the CHC theory. The present study is one of the first studies to examine the narrow abilities of the Wechsler cognitive assessment instruments. This could be due to the limited subtest availability of the Wechsler instruments, and the common perspective that at least two subtests must comprise a narrow ability (Flanagan et al., 2013). These results are also consistent with previous research examining the factor structure of the WAIS-IV, WISC-V, and WPPSI-IV, further described below.

It must be noted that minimal research exists on narrow abilities that comprise the Wechsler instruments. Prior validation studies examined broad abilities in their factor analyses, but to date no studies have examined narrow abilities. This could be due to the lack of available subtests required to adequately examine and interpret narrow abilities, which is a general consensus of two subtests per narrow ability (Flanagan et al., 2013). This appears to be a limitation of the Wechsler measures themselves, not a product of the contemporary theories lacking evidence of validity (i.e., F-CHC and CHC). That being said, David Wechsler did not create these instruments with a theory in mind, per se. Wechsler's original cognitive instruments were a conglomeration of his experiences with Army Alpha and Army Beta testing, his mentor's, Spearman and Thorndike, work, and practical and clinical applications (Lichtenberger & Kaufman, 2012; Tulsky et al., 2003). Revisions over the years took into consideration functional theories in addition to structural theories, including theories from cognitive psychology, neuropsychology, and specific cognitive abilities such as memory, attention, and executive functioning in an attempt to create a comprehensive theoretical basis (Kaufman et al., 2016; Wechsler, 2008, 2012, 2014). Therefore, as Wechsler's instruments were not developed with a theory that posits the existence of narrow abilities in mind, it is not surprising that subtests do not always precisely measure narrow abilities.

Discussion of Findings

WAIS-IV

The WAIS-IV has been well-researched in validation studies. Benson and colleagues (2010) utilized the WAIS-IV standardization sample in their investigation of the factor structure of the WAIS-IV. In their confirmatory factor analysis, the authors analyzed the WAIS-IV factor structure against CHC theory. However, the authors noted that the subtests insufficiently measured narrow abilities as at least two subtests are needed to adequately do so. Therefore, the authors state that this exclusion of narrow cognitive abilities from the model analysis may result in a discrepancy between the data set and the hypothesized model. The authors concluded that a CHC-based structure provided a better interpretation than that of the publishers with support for the factors of *Gc, Gwm, Gs, Gf,* and *Gv*. The results of the analysis also revealed an index of *Gq* consisting of Figure Weights and Arithmetic, not included in the final model due to subtest limitations. It was also suggested that Cancellation not be included as a measure of *Gs* due to weak loadings. An additional study by Weiss and colleagues (2013a) also used the WAIS-IV standardization sample. Results of the Weiss et al. (2013a) study found that both the four-factor (*Gc, Gv, Gs, Gwm*) and five-factor (*Gc, Gv, Gc, Gwm, Gf*) models provided good fit statistics. However, the five-factor model split perceptual reasoning into separate *Gf* and *Gv* factors, also evidenced by previous research (Benson et al., 2010). Additionally, Weiss and colleagues

(2013a) found evidence of a *Gq* factor that consisted of Arithmetic and Figure Weights, also evidenced by Benson et al. (2010).

The publishers of the WAIS-IV (Wechsler, 2008) provide factor analytic support in the *Technical and Interpretive Manual* for a four-factor structure consisting of indices equivalent to broad abilities of the CHC theory (i.e., *Gc*, *Gwm*, and *Gs*), in addition to a PRI. In the present analysis, exploratory analysis of the WAIS-IV revealed a four-factor structure best fit the data, as outlined by the publishers. A five-factor structure revealed a weak fifth factor with no strong loadings and no clear definition of the factor. Confirmatory analyses were conducted on a one-, two-, three-, four-, five-, and six-factor model as proposed by the F-CHC model and the CHC theory. The WAIS-IV does not adequately measure *Ga* or *Glm*; therefore, the researcher did not attempt to examine additional factor structures. Both the four-factor and five-factor model of the CFA provided reasonable fit indices, with the five-factor model slightly outperforming the four-factor. The five-factor model was consistent with both the F-CHC model and CHC theory as it included the broad abilities of *Gc, Gr/Gf, Gcm/Gwm, Gv,* and *Gs.* In the five-factor model, the factor labeled as perceptual reasoning split into separate *Gv* and *Gr* factors. This split allowed the *Gv* subtests, Block Design and Visual Puzzles, to produce stronger factor loadings on the *Gv* factor than they previously had on the perceptual reasoning factor.

The results of the analyses revealed that the alignment of subtests may differ from that proposed by the publishers. Arithmetic appeared to load stronger on *Gr* than *Gcm*; however, this is not surprising given the difficulties Arithmetic has provided other researchers in prior validation studies (Benson et al., 2010). Arithmetic was also found to cross-load on *Gc.* This provides evidence for the impurity of this particular subtest. Arithmetic involves an individual mentally solving arithmetic problems within a specified time limit. This requires the individual to comprehend the problem (e.g., *Gc*), use reasoning skills to solve the problem (e.g., *Gr*), all while doing so in their mind (e.g., *Gcm*). Picture Completion also did not load strongly on *Gr,* as it was previously thought of as a perceptual reasoning subtest when perhaps it fits better elsewhere. Picture Completion is a subtest that requires an individual to view a picture with an important part missing and identify the missing part within a specified time limit. This subtest cross-loaded on *Gv,* likely given the visual component of the task. Cancellation also does not load strongly on *Gs,* indicating it may not be a strong indicator of *Gs*.

The broad factor of *Gc* preserved the same subtests as posited by the *Technical and Interpretive Manual* (Wechsler, 2008) of the WAIS-IV. Specifically, Vocabulary, Similarities, Information, and Comprehension loaded on *Gc.* The researcher examined *Gc* from a two-factor structure as hypothesized by the F-CHC model, including a factual knowledge and verbal ability factor. However, in the CFA, a one-factor solution provided the most support for the broad ability of *Gc*, consistent with both the F-CHC model and the CHC theory.

The broad factor of *Gs* also preserved the same subtests as posited by the *Technical and Interpretive Manual* (Wechsler, 2008). Coding, Symbol Search, and Cancellation loaded on *Gs.* In the CFA, the researcher had to specify a constraint on the factor loadings due to the limited data from the hypothesized subtests. In this case, the indicator loadings were set equal to one another in order to create a just-identified model. Therefore, the fit indices are not reliable sources of fit; however, this is a product of the dataset, not the hypothesized F-CHC or CHC model. In the CFA, a one-factor solution provided the most support for the broad ability of *Gs*, consistent with both the F-CHC model and the CHC theory.

The broad factor of *Gcm* maintained similar subtests as posted by the *Technical and Interpretive Manual* (Wechsler, 2008). While Digit Span and Letter-Number Sequencing loaded on *Gcm,* Arithmetic loaded more strongly on another factor (*Gr*). Due to a lack of indicators required for the analyses, EFA and CFA were not attempted to examine narrow abilities. Therefore, a one-factor solution provided the most support for the broad ability of *Gcm*, consistent with both the F-CHC model and the CHC theory.

Arithmetic appeared to migrate toward *Gr*, suggesting that this subtest may best measure reasoning abilities. However, Arithmetic also cross-loaded onto *Gc* and *Gcm*. This provides evidence of the impurity of the tasks on the Wechsler instruments, meaning they may measure numerous abilities and do not fit cleanly in any one factor. Contrary to the *Technical and Interpretive Manual* (Wechsler, 2008), Figure Weights, Matrix Reasoning, and Picture Completion also loaded on *Gr.* This provided support for separation of perceptual reasoning into two distinct factors, *Gr* and *Gv*. The researcher examined a two-factor structure of *Gr* as hypothesized by the F-CHC model, including an inductive/deductive reasoning and contextual reasoning factor. However, in the CFA, a one-factor solution provided the most support for the broad ability of *Gr*, consistent with both the F-CHC model and the CHC theory.

Finally, Block Design and Visual Puzzles loaded on a broad factor of *Gv,* contrary to the perceptual reasoning factor proposed by the publishers. Given the limited data from the hypothesized subtests, EFA and CFA were not attempted to examine narrow abilities. Therefore, a one-factor solution provided the most support for the broad ability of *Gv,* consistent with both the F-CHC model and the CHC theory.

Upon examination of the WAIS-IV in its entirety, it is clear that the instrument is not able to adequately measure narrow abilities as hypothesized by contemporary theories of intelligence (i.e., F-CHC model and CHC theory). Specifically, the dataset of the WAIS-IV does not contain enough indicators, or subtests, to adequately and statistically examine narrow abilities. The

WAIS-IV standardization sample provided factor analytic support for both the F-CHC and CHC broad abilities of *Gc, Gr/Gf, Gcm/Gwm, Gv,* and *Gs*. However, only two of the broad abilities, *Gc* and *Gr,* could be appropriately examined for narrow abilities. Even so, one-factor solutions provided the most support for said broad abilities. This lack of evidence for narrow abilities is largely a product of the tasks that comprise the WAIS-IV; the tasks do not adequately measure separate cognitive constructs. This appears to be a limitation of the instrument rather than a limitation of the F-CHC model or the CHC theory.

WISC-V

Several validation studies have been conducted on the WISC-V, particularly as it transitioned from a four-factor to a five-factor model with little psychometric support from the publishers. Dombrowski and colleagues (2015) utilized the WISC-V standardization sample in their investigation of the factor structure of the WISC-V. A bifactor factor analysis revealed that a three-factor model consisting of *Gf, Gwm,* and *Gs* most reasonably fit the data. Further, there was no evidence of a distinct *Gf* factor or *Gc* factor (Dombrowski et al., 2015). Dombrowski et al. (2015) concluded that the WSC-V primarily measures *g* due to the fact that it most accounts for the subtests' total and common variance. Canivez et al. (2016) and Canivez, Watkins, and Dombrowski (2017) also examined the factor structure utilizing the WISC-V standardization sample. Their preferred EFA model consisted of four factors, similar to the WISC-IV. Similar to Dombrowski et al. (2015), Canivez et al. (2016), and Canivez, Watkins, and Dombrowski (2017), found that *Gv* and *Gf* converged to form a *Gv-Gr* factor. Further, Picture Concepts proved problematic in that it did not contribute significantly to any one factor. Arithmetic resulted in similar issues where it did not load significantly on the *Gwm* factor. Canivez, Dombrowski, and Watkins (2017) also examined the factor structure of the WISC-V using the standardization sample divided into age groups (i.e., ages 6-8, 9-11, 12-14, and 15-16 years). Their findings were consistent with those of Canivez and colleagues in 2016 and 2017; there was no EFA evidence for a five-factor model as outlined by the publishers. Further, when examined by age group, Arithmetic proved problematic as it was associated with *Gwm* in the 6-8 and 9-11 age groups, *Gc* in the 12-14 age group, and no clear association was found for the 15-16 age group. Picture Concepts failed to load strongly on any one factor in the 6-8 and 9-11 age groups. Finally, Figure Weights and Matrix Reasoning emerged as weak representations of *Gv-Gr*, but they still did not split into a salient *Gf* factor (Canivez, Dombrowski, & Watkins, 2017).

The publishers of the WISC-V (Wechsler, 2014) provide factor analytic support in the *Technical and Interpretive Manual* for a five-factor structure consisting of indices equivalent to broad abilities of the CHC theory: *Gc, Gf, Gv, Gwm,* and *Gs*. In the present analysis, exploratory analysis of the WISC-V revealed a four-factor structure best fits the data, contrary to that posited by the publishers. A five-factor structure revealed a weak fifth factor with only one strong subtest loading: Cancellation. Confirmatory analyses were conducted on a one-, two-, three-, four-, five-, and six-factor model as proposed by the F-CHC model and the CHC theory. The WISC-V does not adequately measure *Ga* or *Glm*; therefore, the researcher did not attempt to examine additional factor structures. Both the four-factor and five-factor model of the CFA provided reasonable fit indices, with the five-factor model slightly outperforming the four-factor. The five-factor structure found in the present analysis was consistent with the publishers of the WISC-V, the F-CHC model, and the CHC theory as it included the broad abilities of *Gc, Gr/Gf, Gcm/Gwm, Gv,* and *Gs.*

The results of the analyses revealed that the alignment of subtests may differ from that proposed by the publishers. Arithmetic was found to cross-load on *Gc, Gr,* and *Gcm.* Similar to the WAIS-IV, this provides evidence for the impurity of this particular subtest. Arithmetic involves an individual mentally solving arithmetic problems within a specified time limit. This requires the individual to comprehend the problem (e.g., *Gc*), use reasoning skills to solve the problem (e.g., *Gr*), all while doing so in their mind (e.g., *Gcm*. Picture Concepts did not load strongly on *Gr,* as it was previously thought of as a reasoning subtest when perhaps it fits better elsewhere. In the EFA, Picture Concepts loaded just as strongly to *Gc* as it did *Gr.* In the CFA, Picture Concepts cross-loaded on both *Gc* and *Gr,* although less strongly on *Gc.* One explanation for this are the tasks that comprise this subtest itself. On the Picture Concepts subtest, an individual is asked to view two or three rows of pictures and select one picture from each row to form a group with a common characteristic. However, an individual must have prior knowledge of what the picture is in order to understand how the picture relates to the other pictures, using both inductive reasoning and general verbal information. Cancellation also does not load strongly on *Gs,* indicating it may not be a strong indicator of *Gs*.

The broad factor of *Gc* preserved the same subtests as posited by the *Technical and Interpretive Manual* (Wechsler, 2014) of the WISC-V. Specifically, Vocabulary, Similarities, Information, and Comprehension loaded on *Gc.* In the CFA, the researcher examined *Gc* from a two-factor structure as hypothesized by the F-CHC model, including a factual knowledge and verbal ability factor. While the two-factor solution revealed reasonable fit indices, a one-factor solution provided the most support for the broad ability of *Gc*, consistent with both the F-CHC model and the CHC theory.

The broad factor of *Gs* also preserved the same subtests as posited by the *Technical and Interpretive Manual* (Wechsler, 2014). Coding, Symbol Search, and Cancellation loaded on *Gs.* In the CFA, the researcher had to specify a constraint on the factor loadings due to the limited data from the hypothesized subtests. In this case, the indicator loadings were set equal to one another in order to create a just-identified model. Therefore, the fit indices are not reliable sources of fit; however, this is a product of the dataset, not the hypothesized F-CHC or CHC model. In the CFA, a one-factor solution provided the most support for the broad ability of *Gs*, consistent with both the F-CHC model and the CHC theory.

The broad factor of *Gcm* maintained similar subtests as posted by the *Technical and Interpretive Manual* (Wechsler, 2014). Digit Span, Letter-Number Sequencing, and Picture Span loaded on *Gcm.* Similar to *Gs,* given the limited data from the hypothesized subtests, the researcher had to specify a constraint. This does not allow for a true examination of fit. In the CFA, a one-factor solution provided the most support for the broad ability of *Gcm*, consistent with both the F-CHC model and the CHC theory.

The broad factor of *Gr* maintained similar subtests as posted by the *Technical and Interpretive Manual* (Wechsler, 2014). Figure Weights, Matrix Reasoning, Arithmetic, and Picture Concepts loaded on *Gr.* In the CFA, the researcher examined a two-factor structure of *Gr* as hypothesized by the F-CHC model, including an inductive/deductive reasoning and contextual reasoning factor. While the two-factor solution provided reasonable fit indices, a one-factor solution provided the most support for the broad ability of *Gr*, consistent with both the F-CHC model and the CHC theory.

Finally, Block Design and Visual Puzzles loaded on a broad factor of *Gv,* as posted by the *Technical and Interpretive Manual* (Wechsler, 2014). Due to a lack of indicators required for the analyses, EFA and CFA were not attempted to examine narrow abilities. Therefore, a one-factor solution provided the most support for the broad ability of *Gv*, consistent with both the F-CHC model and the CHC theory.

Similar to the WAIS-IV, it is clear that the WISC-V is not able to adequately measure narrow abilities as hypothesized by contemporary theories of intelligence (i.e., F-CHC model and CHC theory). Specifically, the dataset of the WISC- does not contain enough indicators, or subtests, to adequately and statistically examine narrow abilities. The WISC-V standardization sample provided factor analytic support for both the F-CHC and CHC broad abilities of *Gc, Gr/Gf, Gcm/Gwm, Gv,* and *Gs*. However, only two of the broad abilities, *Gc* and *Gr,* could be appropriately examined for narrow abilities. Even so, one-factor solutions provided the most support for said broad abilities. This lack of evidence for narrow abilities is largely a product of the tasks that comprise the WISC-V; the tasks do not adequately measure separate cognitive constructs. This appears to be a limitation of the instrument rather than a limitation of the F-CHC model or the CHC theory.

WPPSI-IV

The factor structure of the WPPSI-IV has also been researched. Watkins and Beaujean (2014) examined the factor structure of the WPPSI-IV using the standardization sample. In their bifactor factor analysis, the authors found that *g* accounted for more total and common variance, similar to Dombrowski et al. (2015) and the WISC-V. While the subtests that comprise *Gc, Gv,* and *Gs* had stronger factor loadings (i.e., > .30), the subtests of *Gwm* and *Gf* had weak factor loadings below .30.

The publishers of the WPPSI-IV (Wechsler, 2012) provide factor analytic support in the *Technical and Interpretive Manual* for a five-factor structure consisting of indices equivalent to broad abilities of the CHC theory: *Gc*, *Gf*, *Gv*, *Gwm*, and *Gs*. In the present analysis, exploratory analysis of the WPPSI-IV revealed factor loadings inconsistent with that of the publishers of the WPPSI-IV, the F-CHC model, and CHC theory. In the four-factor structure, a strong broad factor consisted of *Gs*. A second factor appeared to include tasks of verbal ability, a narrow ability of *Gc*. The second narrow ability of factual knowledge under *Gc* appears to be represented by a third broad factor and included the tasks of Picture Naming, Information, and Receptive Vocabulary. This divide of *Gc* into distinct broad factors suggests that the tasks are measuring different cognitive constructs. The fourth factor consisted of subtests from *Gv, Gcm,* and *Gr* with no clear definition as to what the factor was measuring, although it could be hypothesized that a visual component is what the tasks share or have in common. In the five-factor structure, the broad factors also consisted of *Gc* and *Gs,* with an additional factor of *Gv-Gr*. Additionally, Picture Memory, Picture Naming, and Receptive Vocabulary did not load strongly on any one factor. Zoo Locations appeared to create its own factor, which might be labeled as an immediate recall factor.

Confirmatory analyses were conducted on one-, two-, three-, four-, five-, and six-factor models as proposed by the F-CHC model and CHC theory. The WPPSI-IV does not adequately measure *Ga* or *Glm*; therefore, the researcher did not attempt to examine additional factor structures. Both the four-factor and five-factor model of the CFA provided reasonable fit indices. However, the four-factor model included a broad factor of perceptual reasoning, combining *Gr* and *Gv* subtests. The five-factor model was consistent with that of the publishers of the WPPSI-IV, the F-CHC model, and the CHC theory as it included the broad abilities of *Gc, Gcm/Gwm,* and *Gs.*

The broad factor of *Gc* preserved the same subtests as posited by the *Technical and Interpretive Manual* (Wechsler, 2012) of the WPPSI-IV. Specifically, Vocabulary, Similarities, Information, Comprehension, Picture Naming, and Receptive Vocabulary loaded on *Gc.* In the CFA, the researcher examined *Gc* from a two-factor structure as hypothesized by the F-CHC model, including a factual knowledge and verbal ability factor. A three-factor structure was also examined, including a factual knowledge, verbal ability, and lexical access factor. While the one- and two-factor solutions revealed reasonable fit indices, a three-factor solution provided the most support for the narrow abilities of factual knowledge and verbal ability, consistent with the F-CHC model, and the narrow ability of lexical access, consistent with CHC theory. This divide into three narrow abilities suggests that the subtests that comprise *Gc* actually measure different cognitive constructs. Vocabulary, Similarities, and Comprehension may measure the ability to verbalize one's knowledge based on words or concepts. Information is a task that explicitly measures previously learned factual knowledge. Receptive Vocabulary and Picture Naming appear to fall under a category of lexical access as the tasks involve an individual simply being able to access vocabulary words. This novel finding is pertinent when interpreting results of the WPPSI-IV; a student who performs poorly on tasks of factual knowledge may still have intact verbal ability and lexical access, indicating specific areas to target for intervention.

The broad factor of *Gs* also preserved the same subtests as posited by the *Technical and Interpretive Manual* (Wechsler, 2012). Animal Coding, Bug Search, and Cancellation loaded on *Gs.* Due to a lack of indicators required for the analyses, EFA and CFA were not attempted to examine narrow abilities. Therefore, a one-factor solution provided the most support for the broad ability of *Gs*, consistent with both the F-CHC model and the CHC theory.

The broad factor of *Gcm* maintained similar subtests as posted by the *Technical and Interpretive Manual* (Wechsler, 2012). Zoo Locations and Picture Memory loaded on *Gcm.* Similar to *Gs,* given the limited data from the hypothesized subtests, EFA and CFA were not attempted to examine narrow abilities. Therefore, a one-factor solution provided the most support for the broad ability of *Gcm*, consistent with both the F-CHC model and the CHC theory.

The broad factor of *Gr* maintained similar subtests as posted by the *Technical and Interpretive Manual* (Wechsler, 2014). Matrix Reasoning and Picture Concepts loaded on *Gr.* Due to a lack of indicators required for the analyses, EFA and CFA were not attempted to examine narrow abilities. Therefore, a one-factor solution provided the most support for the broad ability of *Gr*, consistent with both the F-CHC model and the CHC theory.

Finally, Block Design and Object Assembly loaded on a broad factor of *Gv,* as posted by the *Technical and Interpretive Manual* (Wechsler, 2012). Due to a lack of indicators required for the analyses, EFA and CFA were not attempted to examine narrow abilities. Therefore, one-factor solution provided the most support for the broad ability of *Gv*, consistent with both the F-CHC model and CHC theory.

Similar to the WAIS-IV and WISC-V, it is clear that the instrument is not able to adequately measure narrow abilities as hypothesized by contemporary theories of intelligence (i.e., F-CHC model and CHC theory). Specifically, the dataset of the WPPSI-IV does not contain enough indicators, or subtests, to adequately and statistically examine narrow abilities.

Implications for Practice

As practitioners who utilize cognitive assessment instruments, particularly the Wechsler instruments, it is imperative to understand the underlying theory and structure that comprises said instruments. Specifically, practitioners in the schools use these instruments to assist with educational eligibility decisions while practitioners in clinical or private practice settings are making clinical, diagnostic decisions. While the factor structure of the Wechsler cognitive assessment instruments is largely similar to that posited by the publishers, results should still be interpreted with caution. In particular, the constructs of *Gr* and *Gv* appear to converge into a *Gr-Gv* factor. This appears to be a product of the subtests of *Gr*, which do not load strongly on a separate factor. Practitioners should be cognizant when interpreting the factor of *Gr* and its subtests. Further, the tasks that comprise the VCI of the WPPSI-IV appear to measure separate narrow abilities of *Gc*. Given this information, practitioners should use cation when interpreting subtest performance. Particularly, a child who performs poorly on one or two subtests may have a weakness in only one specific narrow ability but could have strengths in others.

Additionally, the results of the present study have made it more abundantly clear that the Wechsler instruments only measure limited broad cognitive abilities, consistent with the F-CHC model and CHC theory. Broad abilities such as *Ga* and *Glm* are not included in the development of the Wechsler cognitive assessment instruments. While the WISC-V includes three subtests intended to measure aspects of *Glm,* the subtests are minimal and not adequate at fully representing this broad ability. Further, these subtests are not intelligence subtests and cannot be substituted for primary or secondary subtests (Canivez et al., 2016). This is important for practitioners should they wish to investigate *Ga* or *Glm* in their cognitive assessment measure. Specifically, school psychologists who utilize these instruments for examining specific learning disabilities should understand the limitations and the impact on evaluations.

While the publishers of the Wechsler cognitive assessment instruments do not claim to measure narrow abilities, it is important to understand the implications of this given the contemporary theories of intelligence that exist in the field of school psychology. These contemporary theories, the CHC theory of intelligence and the F-CHC model of cognitive abilities, both posit that broad cognitive abilities are comprised of narrow cognitive abilities. Cognitive instruments, such as the WJ IV (Schrank et al., 2014), were developed based on the CHC theory, and therefore, the instrument is able to adequately measure the narrow abilities these theories posit. As the Wechsler instruments were developed prior to CHC theory and the F-

CHC model, the publishers have attempted to interpret the instruments from a theoretical perspective post-hoc. While the broad abilities of these instruments are largely consistent with the F-CHC model and CHC theory (see results of the present analyses), the instruments do not adequately measure narrow abilities.

Limitations of the Present Study

While the researcher has attempted to address potential issues in the planning stages of the research study, limitations still exist. In the present statistical analyses, the WAIS-IV and WPPSI-IV contained 15 subtests while the WISC-V contained 16 subtests. While this allowed the researcher to examine the instruments as whole, the limited number of variables (i.e., subtests), proved challenging to appropriately examine each broad cognitive ability through an EFA and CFA/SEM model. This resulted in the researcher having to place constraints on the data, resulting in "perfect" fit indices, when this is likely not the case. As previously mentioned, the Wechsler cognitive assessment instruments also do not adequately measure narrow cognitive abilities as hypothesized by the CHC theory of intelligence and the F-CHC model of cognitive abilities. Benson and colleagues (2010) previously examined the WAIS-IV, noting that the subtests insufficiently measured narrow abilities as at least two subtests are needed to adequately do so. Therefore, the authors state that this exclusion of narrow cognitive abilities from the model analysis may result in a discrepancy between the data set and the hypothesized model. This lack of evidence for narrow abilities is largely a product of the tasks that comprise each instrument; the tasks do not adequately measure separate cognitive constructs. Further, the analysis provided evidence for the impurity of the tasks, meaning they may measure numerous constructs. Therefore, the inability to validate the F-CHC model is not a limitation of the model; rather, this appears to be a product of the instruments' lack of available variables (i.e., subtests).

Additionally, there are different methods of employing an EFA and CFA. For example, some previous studies have examined these instruments in a bifactor factor analytic model (Dombrowski et al., 2015), or examined the instruments by higher-order factors (i.e., *g*), while the present study chose to examine only broad and narrow abilities. The rotation method of the CFA and the fit indices utilized (i.e., CFI and RMSEA) can also impact the results; therefore, results may not be as generalizable depending on how the researcher examines the data.

Additionally, the data utilized for the present study were not collected and/or managed by the researcher of the present study. This may be disadvantageous as the researcher has no control over data collection and, therefore, cannot provide justification for any errors in the data. Further, the statistical analyses involved in the present study are complex to employ and interpret and are vulnerable to human error. The present study examined the CHC theory and the F-CHC model, although additional theories of cognitive abilities and intelligence exist. As the researcher was seeking to validate the F-CHC model, confirmation bias may occur. The researcher may misidentify or misinterpret these complex theories as they are applied to the current data, which can impact the fit indices.

Finally, the present study is an adaptation of Ndip's (2020) factor analytic dissertation validating the F-CHC model for the WJ IV instruments. While the WJ IV standardization sample consisted of all ages, the ages that comprise the Wechsler standardization samples will differ based on the cognitive measure. While analyses can be conducted to validate the F-CHC theory, cognitive abilities at each age range may differ and confound results.

Directions for Future Research

As the results of the present study provided partial evidentiary support for the F-CHC model, future research may wish to further validate this model. Further, future research may also seek to investigate specific age groups within each of the Wechsler instruments to examine the factor analytic data and support for either the F-CHC model or CHC theory. Finally, the results of the present analysis apply to the standardization sample of the Wechsler instruments; therefore, it is not clear how generalizable these findings are to other populations.

Conclusion

The present study sought to validate the F-CHC model of cognitive abilities utilizing the Wechsler cognitive assessment instruments, the WAIS-IV, WISC-V, and WPPSI-IV. Each instrument was investigated separately due to the differing age groups in the standardization samples. Exploratory and confirmatory analyses revealed broad factors consistent with broad abilities as hypothesized by the F-CHC model and CHC theory (i.e., *Gc, Gs, Gwm/Gcm, Gf/Gr,* and *Gv*). Upon further examination of the narrow abilities hypothesized by the F-CHC model and/or CHC theory, it was revealed that the Wechsler instruments do not adequately measure narrow abilities of either the F-CHC model or CHC theory. In analysis of the WAIS-IV and the WISC-V, only *Gc* and *Gr* could be appropriately examined for narrow abilities. Even so, one-factor solutions provided the most support for said broad abilities. This appears to be a limitation of the instrument rather than a limitation of the F-CHC model or CHC theory. In an analysis of the WPPSI-IV, *Gc* was the only broad ability that could be appropriately examined for narrow abilities. A three-factor solution consisting of factual knowledge, verbal ability, and lexical access was found to best fit the data. This divide into three narrow abilities suggests that the subtests that comprise *Gc* actually measure different cognitive constructs when provided with enough data (i.e., subtests) to adequately measure said cognitive constructs. Overall, the lack of evidence for narrow abilities is largely a product of the tasks that comprise each instrument; the tasks do not adequately measure separate, distinct cognitive constructs.

Limitations included a lack of available subtests to adequately examine narrow abilities of the F-CHC model and CHC theory. This resulted in the researcher having to place constraints on the data, resulting in “perfect” fit indices, when this is likely not the case. Additionally, the researcher utilized an archival dataset which may be disadvantageous as the researcher has no control over data collection and, therefore, cannot provide justification for any errors in the data. Practitioners using the Wechsler instruments must be cognizant of the limitations of the instruments and interpret with caution. Specifically, high stakes decisions, such as special education eligibility in the school setting or diagnoses in the clinical setting, should not be based on one measure. Further, practitioners cannot discuss narrow abilities within the Wechsler instruments in their interpretation of the results as narrow abilities are not supported by the instruments.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing.* AERA.

Baddeley, A. (2000). The episodic buffer: A new component of working memory. *Trends in Cognitive Sciences, 4*(11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2

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

Benson, N., Hulac, D. M., & Kranzler, J. H. (2010). Independent examination of the Wechsler Adult Intelligence Scale–Fourth Edition (WAIS-IV): What does the WAIS-IV measure? *Psychological Assessment, 22*(1), 121–130. https://doi.org/10.1037/a0017767

Cangur, S., & Ercan, I. (2015). Comparison of model fit indices used in structural equation modeling under multivariate normality. *Journal of Modern Applied Statistical Methods, 14*(1), 152–167. https://doi.org/10.22237/jmasm/1430453580

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. https://doi.org/10.1037/pas0000238

Canivez, G. L., Dombrowski, S. C., & Watkins, M. W. (2017). Factor structure of the WISC-V in four standardization age groups: Exploratory and hierarchical factor analyses with the 16 primary and secondary subtests. *Psychology in the Schools, 55*(7), 741–769. https://doi.org/10.1002/pits.22138

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children–Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29*(4), 458–472. https://doi.org/10.1037/pas0000358

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

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

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

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

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

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

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action.* Houghton Mifflin

Cattell, R. B. (1987). *Intelligence Its structure, growth, and action.* Elsevier Science.

Cattell, R. B. & Horn, J. L. (1978). A check on the theory of fluid and crystallized intelligence with description of new subtest designs. *Journal of Educational Measurement, 15*, 139–164.

Cianciolo, A. T., & Sternberg, R. J. (2004). *Intelligence: A brief history.* Blackwell Publishing.

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis* (2nd ed.). Lawrence Erlbaum Associates, Inc.

Cooperman, A. W., & Waller, N. G. (2021). Heywood you go away! Examining causes, effects, and treatments for Heywood cases in exploratory factor analysis. *Psychological Methods.* Advance online publication. https://doi.org/10.1037/met0000384

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

Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests. *Intelligence, 53,* 194–201. https://doi.org/10.1016/j.intell.2015.10.009

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

Field, A. (2013). *Discovering statistics using IBM SPSS statistics.* (4th ed.). SAGE Publications Inc.

Flanagan, D. P., & Dixon, S. G. (2013). The Cattell-Horn-Carroll theory of cognitive abilities. In C. R. Reynolds, K. J. Vannest & E. Fletcher-Janzen (Eds.), *Encyclopedia of special education: A reference for the education of children, adolescents, and adults disabilities and other exceptional individuals.* (4th ed.). John Wiley & Sons, Inc.

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

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Wiley.

Groth-Marnat, G. (2009). *Handbook of psychological assessment* (5th ed.). John Wiley & Sons, Inc.

Holdnack, J. A. (2019). The development, expansion, and future of the WAIS-IV as a cornerstone in comprehensive cognitive assessments. In G. Goldstein, D. N. Allen & J. DeLuca (Eds.), *Handbook of psychological assessment* (4th ed., pp 103-139). Academic Press.

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

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligences. *Journal of Educational Psychology, 57*(5), 253–270. https://doi.org/10.1037/h0023816

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement, 34*(1), 111–117.

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

Kaufman, A. S., & Kaufman, N. L. (2018). *Kaufman Assessment Battery for Children, Normative Update* (2nd ed.). Pearson

Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2016). *Intelligent testing with the WISC-V.* John Wiley & Sons, Inc.

Keith, T. Z., Fine, J. G., Taub, G. E., Reynolds, M. R., & Kranzler, J. H. (2006). Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children--Fourth Edition: What does it measure? *School Psychology Review, 35*(1), 108–127.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell-Horn-Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47*(7), 635–650. https://doi.org/10/1002/pits

Kent, P. (2017). Fluid intelligence: A brief history. *Applied Neuropsychology, 6*(3), 193–203. https://doi.org/10.1080/21622965.2017.1317480

Kline, R. B. (2011). *Principles and practice of structural equation modeling.* (3rd ed.). The Guilford Press.

Kline, R. B. (2013). Exploratory and confirmatory factor analysis. In Y. Petscher, C. Schatschneider, & D. L. Compton (Eds.), *Applied quantitative analysis in education and the social sciences* (pp. 171-205). Routledge.

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Lichtenberger, E. O., & Kaufman, A. S. (2012). *Essentials of WAIS-IV assessment.* (2nd ed.). John Wiley & Sons, Inc.

Lomax, R. G. (2013). Structural equation modeling. In Y. Petscher & C. Schatschneider (Eds.), *Applied quantitative analysis in the social sciences* (pp. 245-264). Routledge.

Marsh, H. W. (1998) Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling: A Multidisciplinary Journal, 5*(1), 22–36. http://doi.org/10.1080/10705519809540087

McGrew, K. S. (1997). Analysis of the major intelligence batteries according to a proposed comprehensive Gf-Gc framework. In D. P. Flanagan, J. L., Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151-179). The Guilford Press.

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

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

Meyers, L. S., Gamst, G., & Guarino, A. J. (2016). *Applied multivariate research: Design and interpretation*. (3rd ed.). SAGE Publications Inc.

Naglieri, J. A., Das, J. P., & Goldstein, S. (2014). *Cognitive Assessment System–Second Edition.* Pro-Ed.

Ndip, N. (2020). *Factor analyses of the Woodcock-Johnson IV Tests of Cognitive Abilities, Tests of Achievement, and Tests of Oral Language* [Unpublished doctoral dissertation]. Texas Woman's University.

Newton, J. H., & McGrew, K. (2010). Introduction to the special issue: Current research in Cattell-Horn-Carroll-based assessment. *Psychology in the Schools, 47*(7), 621–634. https://doi.org/10.1002/pits

Raiford, S. E., & Coalson, D. L. (2014). *Essentials of WPPSI-IV assessment*. John Wiley & Sons, Inc.

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

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). The Guilford Press.

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* (4th ed., pp. 73-163). The Guilford Press.

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

Schreiber, J. B. (2008). Core reporting practices in structural equation modeling. *Research in Social and Administrative Pharmacy, 4,* 93–97. https://doi.org/10.1016.j.sapharm.2007.04.003

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–293.

Spearman, C. (1927). *The abilities of man.* Macmillan.

Tabachnick, B. G., & Fidell, L. S. (2012). *Using multivariate statistics* (6th ed.). Pearson.

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

Thurstone, L. L. (1945). The prediction of choice. *Psychometrika, 10,* 237–253. https://doi.org/10.1007/BF02288891

Tulsky, D. S., Saklofske, D. H., & Ricker, J. (2003). Historical overview of intelligence and memory: Factors influencing the Wechsler scales. In D. S. Tulsky, D. H. Saklofske G. J. Chelune, R. K. Heaton, R. J. Ivnik, R. Bornstein, A. Prifitera, & M. F. Ledbetter (Eds.), *Clinical interpretation of the WAIS-III and WMS-II.* Academic Press.

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

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

Wahlstrom, D., Raiford, S. E., Breaux, K. C., Zhu, J., & Weiss, L. G. (2018). The Wechsler Preschool and Primary Scale of Intelligence–Fourth Edition, Wechsler Intelligence Scale of Children–Fifth Edition, and Wechsler Individual Achievement Test–Third Edition. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 3-55). The Guilford Press.

Wasserman, J. D. (2018). A history of intelligence assessment: The unfinished tapestry. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp 3-55). The Guilford Press.

Watkins, M. W., & Beaujean, A. A. (2014). Bifactor structure of the Wechsler Preschool and Primary Scale of Intelligence – Fourth Edition. *School Psychology Quarterly, 29*(1), 52–63. https://doi.org/1037/spq0000038

Wechsler, D. (1939a). *The measurement of adult intelligence.* Williams & Wilkins.

Wechsler, D. (1939b). *Wechsler-Bellevue Intelligence Scale.* Psychological Corporation.

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

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale – Restandardized.* Psychological Corporation.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (4th ed.). Williams & Wilkins Co. https://doi.org/10.1037/11167-000

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

Wechsler, D. (1974). *Wechsler Intelligence Scale for Children – Revised.* Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale – Revised.* Psychological Corporation.

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

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

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale* (3rd ed.). Psychological Corporation.

Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence (*3rd ed.). Pearson.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children* (4th ed.). Pearson, Inc.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale* (4th ed.). Pearson, Inc.

Wechsler, D. (2012). *Wechsler Preschool and Primary Scale of Intelligence* (4th ed.). NCS Pearson, Inc.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). NCS Pearson, Inc.

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013a). WAIS-IV and clinical validation of the four- and five-factor interpretive approaches. *Journal of Psychoeducational Assessment, 31*(2), 114–131. https://doi.org/10.1177/0734282913478030

Weiss, L. G., Keith, T. Z., Zhu, J., & Chen, H. (2013b). WISC-IV clinical validation of the four- and five-factor interpretive approaches. *Journal of Psychoeducational Assessment, 31*(2), 94–113. https://doi.org/10.1177/0734282913478032

Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (Eds.). (2016). *WISC-V assessment and interpretation: Scientist-practitioner perspectives.* Elsevier Academic Press.

Weston, R., & Gore, P. A., Jr. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist, 34*(5), 719–751. https://doi.org/10.1177/0011000006286345

Williams, B., Brown, T., & Onsman, A. (2010). Exploratory factor analysis: A five-step guide for novices. *Australian Journal of Paramedicine, 8*(3), 1–13. https://doi.org/10.33151/ajp.8.3.93

Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Samples size requirements for structural equation models: An evaluation of power, bias, and solution propriety. *Educational and Psychological Measurement, 76*(6), 913–934. https://doi.org/10.1177/0013164413495237

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

Woodcock, R. W., & Mather, N. (2014). *Woodcock-Johnson IV.* Riverside

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III.* Riverside.

Woodcock, R. W., Maricle, D. E., Miller, D. C., & McGill, R. J. (2018). Functional Cattell-Horn-Carroll nomenclature for practical applications. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp 901-911). The Guilford Press.

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