1-1-2010

The Relationship Between the Wechsler Intelligence Scale for Children-Fourth Edition and the Woodcock-Johnson III Tests of Cognitive Abilities in a Clinically Referred Pediatric Population

Lindsay Anne Shaw
Nova Southeastern University, lindsay.a.shaw@gmail.com

THE RELATIONSHIP BETWEEN THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN-FOURTH EDITION AND THE WOODCOCK-JOHNSON III TESTS OF COGNTIVE ABILITIES IN A CLINICALLY REFERRED PEDIATRIC POPULATION

By

Lindsay Anne Shaw, M.S.

A Dissertation Presented to the Center for Psychological Studies

of Nova Southeastern University

in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy

NOVA SOUTHEASTERN UNIVERSITY

2010

This dissertation was submitted by Lindsay A. Shaw, M.S. under the direction of the Chairperson of the dissertation committee listed below. It was submitted to the School of Psychological Studies and approved in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Clinical Psychology at Nova Southeastern University.

Approved:

____________________	____________________________________
Date of Defense	Charles Golden, Ph.D., Chairperson
	____________________________________
	Sarah Valley-Gray, Psy.D.
	____________________________________
	Christian DeLucia, Ph.D.
____________________	____________________________________
Date of Final Approval	Charles Golden, Ph.D., Chairperson

Acknowledgements

I would like to acknowledge the members of this dissertation committee, Drs. Charles Golden, Sarah Valley-Gray, and Christian DeLucia, for their guidance, patience, and support on this project. Specifically, I would like to thank Dr. Golden for his guidance and direction from the inception of this research idea to the completion of this document. I would especially like to thank Dr. Valley-Gray for her continued optimism and encouragement and I would like to thank Dr. DeLucia for his important contributions to the research design of this project. I am also highly indebted to my committee for their crucial insight and flexibility in completing this project.

I would also like to acknowledge my classmates and friends who willingly offered their knowledge, humor, and camaraderie during completion of this dissertation. Specifically, Lindsay Hines, Maritza Figueroa, and Howard Buddin were instrumental in providing understanding and encouragement, and likewise contributed greatly to my determination in completing this project. Furthermore, I would like to extend my deepest gratitude to my mentors, including Dr. Golden, Dr. Jessica Foley, and Dr. Judith Migoya, for facilitating my growth and development, and helping me achieve my full potential, both personally and professionally.

I am particularly grateful to my family for their unconditional generosity and love throughout my graduate school endeavors. They willingly offered their support, providing me with a place to call home, and I am forever grateful for their support.

Table of Contents

List of Tables

THE RELATIONSHIP BETWEEN THE WECHSLER INTELLIGENCE SCALE FOR CHILDREN- FOURTH EDITION AND THE WOODCOCK-JOHNSON III TESTS OF COGNTIVE ABILITIES IN A CLINICALLY REFERRED PEDIATRIC POPULATION

By

Lindsay Anne Shaw, M.S.

Nova Southeastern University

Abstract

This research involves an investigation of the construct validity of the Wechsler Intelligence Scale for Children- Fourth Edition (WISC-IV) when compared to the Woodcock-Johnson III Tests of Cognitive Abilities (WJ III COG) to provide evidence for the utility of using the WISC-IV in assessing cognitive abilities according to the Cattell-Horn-Carroll (CHC) theory. The study was conducted using archival data consisting of 92 children and adolescents between the ages of 6 years and 16 years, 11 months referred for a comprehensive neuropsychological evaluation at a university-affiliated assessment center. Data for all participants were collected following administration of a battery of measures as part of a neuropsychological evaluation, with tests administered in no particular order. The mean age of children was 9.82 years (*SD* = 2.81) with a mean grade level of 3.95 (*SD* = 2.63). Ten hypotheses were investigated specifically to examine the comparability of the general intellectual functioning scores for each battery among a sample of children with neuropsychiatric disorders, as well as to examine the convergent and discriminate validity of the WISC-IV index scores. The first hypothesis utilized a paired samples *t*-test and found that the WISC-IV Full Scale IQ score was significantly below that of the WJ III COG General Intellectual Ability-Extended score. For the remaining hypotheses, Pearson product-moment correlations revealed large correlations between the WISC-IV and WJ III COG convergent constructs of general intellectual functioning, comprehension-knowledge, fluid reasoning, working memory, and processing speed. For correlations between divergent constructs, the WISC-IV Verbal Comprehension Index and the WJ III COG Visual-Spatial Thinking (*Gv*) factor demonstrated a large correlation. Both the WISC-IV Processing Speed Index and Working Memory Index correlated moderately with the WJ III COG *Gv* factor, while the WISC-IV Perceptual Reasoning Index correlated moderately with the WJ III COG Auditory Processing factor. Fisher's r to Z transformation was used to assess for significant differences between the observed correlations and stipulated values determined. Results indicated that correlations between the global IQ, fluid reasoning, and short-term memory composite scores of the two measures were significantly greater than that found for the WISC-III and WJ III COG, while the relationship between the verbal ability and processing speed composite scores were consistent with past findings. Correlations between divergent constructs revealed a reliable pattern of significantly greater relationships than was found for research concerning the WISC-III and WJ III COG. Primarily, results of this study provided evidence that the substantive changes made to the WISC-IV have improved the ability to interpret the Full Scale IQ score as a measure of general intelligence similar to that obtained by the WJ III COG. However, the global IQ scores between the two measures cannot be assumed to be equivalent among children with neuropsychiatric disorders. Results also suggested that the WISC-IV appears to provide improved measurement of the CHC broad abilities of fluid reasoning (*Gf*) and short-term memory (*Gsm*). Correlations between divergent constructs provided evidence for relationships between cognitive abilities suggested to be significantly related to academic achievement. This study concluded that research findings for the WISC-III cannot be applied conclusively to the

WISC-IV and that the substantive changes made to the WISC-IV have improved the ability to interpret the battery under the CHC framework. However, findings underscore the importance of examining performance across second-order factors that may contribute to differences in general intelligence, as well as remaining aware of differences in narrow ability constructs measured, task demands, or shared variance between subtests when making interpretations of test performance.

Chapter I: Statement of the Problem

In 1949, David Wechsler published the first version of the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949), developed as a downward extension of his original test, the Wechsler-Bellevue Intelligence Scale (Wechsler, 1939) (Kaufman, Flanagan, Alfonso, & Mascolo, 2006). While the original Wechsler measure primarily sought to classify individuals based upon global aspects of intelligence, subtests included did not align with an explicit theory of intelligence (Coalson & Weiss, 2002). Nonetheless, consistencies have been demonstrated to exist between the Wechsler scales and other measures of intelligence because of the inclusion of significant areas of cognitive ability: verbal comprehension, perceptual organization, quantitative reasoning, memory, and processing speed (Carroll, 1993, 1997; Horn, 1991).

Over the last 60 years, the Wechsler scales have undergone numerous revisions, reflecting the evolution of intellectual assessment, and allowing for the clinical (e.g., improved norms) and practical (e.g., simplified administration procedures) utility of the tests across a wide range of settings and purposes (Wechsler, 2003; Wechsler, 2004). The Wechsler scales have demonstrated continued diagnostic applicability, utilized for the purposes of identification of mental retardation and learning disabilities, placement in specialized programs, determining clinical intervention, and for neuropsychological evaluation (Beres, Kaufman, & Perlman, 2000). It was previously thought that intelligence tests were not useful in neuropsychological evaluations because they were reported to lack specificity regarding underlying brain impairments, limiting conclusions regarding brain function. Likewise, their ability to accurately predict functional outcomes was questionable (Yeates & Donders, 2005). However, research has provided evidence for the relationship between neurological substrates and performance on the

Wechsler scales, possibly allowing for the prediction of performance based on brain damage (Gläscher et al., 2009; Riccio & Hynd, 2000). Moreover, such tests are practically relevant for validating recommendations in special education and clinical practice, and provide useful hypotheses regarding a child's cognitive strengths and weaknesses (Yeates & Donders). Among the various cognitive ability measures available, the WISC has been purported to be the most frequently used measure of intelligence among child neuropsychologists (Camara, Nathan, & Peunte, 2000).

The usefulness of intelligence testing is also an area of debate among school psychologists. According to Pfeiffer, Reddy, Kletzel, Schmeizer, and Boyer (2000), concerns have continued regarding the use of the WISC with minority students or among children for whom English is not their primary language. In their survey of nationally certified school psychologists regarding the perceived usefulness of the Wechsler Intelligence Test for Children-Third Edition (WISC-III, Wechsler, 1991), concerns were reported regarding the applicability of the measure in directing psychoeducational interventions and strategies for instruction. Regardless of such concerns and other perceived weaknesses of the Wechsler scale, such as outdated visual stimuli, lack of utility in re-evaluations, and high verbal content, Pfeiffer et al. found that 70% of the school psychologists surveyed rated the WISC-III factor scores and profile analysis as the most practically useful feature of the measure. Moreover, it characterized the WISC-III as playing a useful role in diagnosis and educational placement.

The viability of utilizing measures of intelligence for the evaluation of child psychopathology and developmental disabilities has been substantiated (see Matson, Andrasik, & Matson, 2008, for the uses of intelligence testing when evaluating childhood pathologies), with diverse and intricate methods available for assessing the complicated construct of cognitive functioning (Sattler, 2008). The Wechsler measure, in its many revised forms, has historically demonstrated considerable popularity as part of a comprehensive psychological assessment (Oakland & Hu, 1992; Prifitera, Weiss, Saklofaske, & Rolfhus, 2005; Zhu, & Weiss, 2005). Nonetheless, criticism has been made regarding the continued failure of the Wechsler scales to incorporate contemporary cognitive theory and research (e.g., Braden, 1995; Little, 1992; Kamphaus, 1993; Shaw, Swerdlik, & Laurent, 1993; Thorndike, 1997; Witt & Gresham, 1985), calling into question the substantive foundation of the scales.

The WISC, now in its fourth revision, has undergone numerous changes since the first publication with regards to the test's content and structure. Each successive revision of the measure has allowed for updated norms, provided more contemporary and less biased testing materials, improved the psychometric properties of the test, and has clarified the factor structure in support of a four-factor solution (Prifitera, Weiss, et al., 2005). However, criticism has been made regarding the failure to incorporate new developments in cognitive theory across revisions (e.g., Little, 1992). Of importance, the Wechsler Intelligence Test for Children- Fourth Edition (WISC-IV, Wechsler, 2003) has been reported to integrate current research regarding cognitive functions and learning (Shaughnessy, 2006).

Following suit with regards to revisions made to other measures of intelligence, the WISC-IV more closely aligns with the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (Flanagan & Kaufman, 2004), which is considered to be one of the most widely accepted theoretical models of cognitive ability (Keith, Kranzler, & Flanagan, 2001; McGrew, 1997). The Cattell-Horn-Carroll theory is a contemporary framework that integrates two similar cognitive models: Carroll's three-tier model of human cognitive abilities and the Horn-Cattell *Gf*-*Gc* model (McGrew, 1997), where *Gf* and *Gc* refer, respectively, to "fluid" and "crystallized" intelligence. This hierarchical theory, which provides a theoretical taxonomy for understanding the cognitive constructs measured by major intelligence test batteries, classifies intellectual abilities within a three-tiered structure, integrating general abilities (*g*; stratum III), ten broad abilities (stratum II) [crystallized intelligence (*Gc*), fluid reasoning (*Gf*), quantitative knowledge (*Gq*), short-term memory (*Gsm*), visual processing (*Gv*), auditory processing (*Ga*), long-term retrieval (*Glr*), processing speed (*Gs*), reaction time (*Gt*), and reading/writing (*Grw*)], and 73 narrow abilities (stratum I). The WISC-IV manual provides evidence of test validity through factor analytic research and comparative studies with other Wechsler-based measures of cognitive ability, while other research has examined the within- and cross-battery factor-structure of the Wechsler scales. However, the WISC-IV's correlation with measures more aligned with current CHC theory is unclear. Correlational studies provided in the manual were conducted mainly with Wechsler-based measures and limited independent research has been conducted, failing to allow for understanding of the correlational patterns between the WISC-IV and more diverse cognitive tests.

A shortcoming in the current research is the lack of studies evaluating the relationship between the WISC-IV and other current measures of cognitive functioning based upon CHC theory. The Woodcock-Johnson Tests of Cognitive Abilities- Third Edition (WJ III COG; Woodcock, McGrew, & Mather, 2001) is one such primary measure. The instrument has strong construct validation that allows for CHC based evaluations, with recent revisions based on the Cattell-Horn-Carroll theory (Taub & McGrew, 2004). Research concerning the correlations between the WISC-III and WJ III COG completed at the time that the WJ III COG was standardized has provided considerable empirical evidence regarding the convergent and divergent relationships of both measures. Although the WISC-III and the WISC-IV demonstrate significant correlations, comparisons made between the WISC-III and WJ III COG are not applicable to the understanding of the validity of the WISC-IV when considering the substantial changes made to the content and structure of the Wechsler scale. As such, research is needed to examine the correlations between the WISC-IV and the WJ III COG global intellectual scores and composite factor scores.

The CHC-based Cross Battery Assessment (XBA) approach (Flanagan & McGrew, 1997; Flanagan, McGrew, & Ortiz, 2000; Flanagan, Ortiz, & Alfonso, 2007; McGrew & Flanagan, 1998) was developed to provide researchers and clinicians with a comprehensive theory to interpret performance on intellectual test batteries. The Cattell-Horn-Carroll framework, supplemented by tests from other batteries, allows for a more thorough analysis of CHC abilities. This approach is based on a series of joint confirmatory factor analyses examining the classification of individual intelligence tests at both the broad and narrow ability stratums (Flanagan et al., 2000; McGrew, 1997; McGrew & Flanagan, 1998; McGrew & Woodcock, 2001). This approach not only assists in interpretations made of the WISC-IV in light of its lack of a specified formal theory (Keith, Fine, Taub, Reynolds, & Kranzler 2006), but it can also be a powerful tool for understanding a student's intellectual abilities by providing a common language by which to describe cognitive abilities (e.g., Alfonso, Flanagan, & Radwan, 2005; Flanagan & McGrew, 1997; Flanagan et al., 2007). However, it remains to be seen whether the composite index and factor scores for the WISC-IV and WJ III COG characterized by the same construct actually measure the same abilities. Examining correlations between these two measures will provide evidence for the comparability of the broad abilities measured, allowing for the tests to be used interchangeably.

Another shortcoming of the current research concerns the failure to determine the statistical difference between the mean scores for the WISC-IV and the WJ III COG. While the relationship between the mean full scale scores for earlier versions of the WISC and Woodcock-Johnson has been explored, there have been no current studies to determine if any statistical difference exists between the most recently published versions of these tests. Research is needed to determine if there are any statistical differences between the WISC-IV Full Scale Intelligence Quotient (FSIQ) and the WJ III COG General Intellectual Ability (GIA) scores.

Given the regular application of the WISC-IV in determining eligibility for exceptional student education, professional standards mandate evidence regarding the test's psychometric robustness (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). Traditionally, the identification of children with a specific learning disability (SLD) has been based upon an ability-achievement discrepancy model. With the revision of the Individuals with Disabilities Education Improvement Act in 2004 (IDEA; 2004), identification of children with learning disabilities has moved away from requiring evidence of a significant discrepancy between intellectual functioning and academic performance. Instead, there is a focus upon the evaluation of intraindividual differences in cognitive functioning and/or achievement relative to intellectual development. This alternative discrepancy method involves examining an individual's pattern of cognitive or academic strengths and weaknesses to determine the impact on academic performance (Mather & Gregg, 2006). While some states continue to enforce the use of a single criterion for the identification of a learning disorder (Holdnack & Weiss, 2006), arguments have been made for the limited utility of global intelligence scores or using a discrepancy model when determining eligibility for SLD (Fletcher et al., 2001; Kavale & Forness, 1995; Mather & Gregg, 2006; Vellutino, 2001; Warner, Dede, Garvan, & Conway, 2002).

Despite these concerns, general intelligence test scores continue to be used in the identification of individuals with learning disabilities, mental retardation, giftedness, or low achievement (Saklofske, Prifitera, Weiss, Rolfhus, & Zhu, 2005). As such, it is important to understand if different measurements of cognitive ability assess similar constructs or result in different mean IQ scores, as this could have implications for those being evaluated for special education placement (Brown & Morgan, 1991; Naglieri, Salter, & Rojahn, 2005) and could adversely impact an individual's functioning across multiple life domains (Silver et al., 2008).

It is especially important to examine the comparability of the Full Scale scores across various measures of cognitive functioning when considering the changes made to the overall factor structure of the WISC-IV. The WISC-III Freedom from Distractibility Index (FDI) was renamed the Working Memory Index (WMI) with the revision of the WISC-IV. The subtests that comprised the index were modified, with the Arithmetic subtest being moved to supplemental status, reducing the emphasis on school achievement. Moreover, an additional task of working memory (Letter-Number Sequencing) was added to the index. The WISC-III Perceptual Organization Index (POI) was renamed the Perceptual Reasoning Index (PRI), with the core subtests measuring distinct processes involved in fluid reasoning (*Gf*) and visual processing

(*Gv*). Within the Verbal Comprehension Index, the Information subtest was moved to supplemental status to reduce the influence of abilities regarding general factual knowledge.

While the constructs of working memory and processing speed did not contribute as heavily to the calculation of the WISC-III FSIQ, the WISC-IV four factor structure allows for a more equal weighting of performance on verbal comprehension, perceptual reasoning, working memory, and processing speed in the construct of overall intelligence. The WISC-IV evidences lower composite scores than the WISC-III, indicating that prior research concerning the WISC-III cannot be generalized to the WISC-IV. Moreover, the changes made have resulted in a Full Scale score that is more representative of the CHC broad abilities measured by the battery (Kaufman et al., 2006). Research by Flanagan and Kaufman (2004) has examined the content validity of new and revised intelligence test batteries, including the WISC-IV, based on CHC theory, whereas Keith et al. (2006) used confirmatory factor analysis to determine if the WISC-IV structure is better described by CHC theory. Consistent findings were shown suggesting that the WISC-IV provides measurement of the CHC broad abilities of crystallized ability (*Gc*), fluid reasoning (*Gf*), visual processing (*Gv*), short-term memory (*Gsm*), and processing speed (*Gs*).

Research suggests that overall intelligence scores should be equivalent to the extent that they measure *g* or that they may be more comparable based on the extent to which they measure similar content (i.e., broad and narrow abilities). Differences between intelligences scores may be found based on the extent to which the specific abilities measured are more closely related to academic performance. Students with learning disabilities may obtain lower scores on intelligence tests that place greater emphasis on measuring cognitive processes found to be weaker in LD samples (e.g., phonological awareness, rate, memory, and perceptual speed)

(Mather & Wendling, 2005). Research has shown that the WISC Full Scale IQ score is typically significantly higher than the Woodcock-Johnson global score in samples of children with learning difficulties. These discrepancies exist because the Woodcock-Johnson battery includes tasks that more discretely measure weaker cognitive processes associated with learning difficulties, specifically, auditory processing (*Ga*) and long-term retrieval (*Glr*). Though the WISC-IV FSIQ now more accurately measures CHC broad abilities, the FSIQ is likely less influenced by abilities related to school achievement than its predecessor. It remains to be seen what effect the changes made to the structure of the Wechsler scales have had on the comparability of the WISC-IV FSIQ and WJ III COG GIA scores. Research is needed to address this concern by determining if the WISC-IV results in lower, higher, or equivalent mean Full Scale IQ scores when compared to the WJ III COG.

A third shortcoming of the current research involves the populations studied. Research concerning the correlations between the WISC-III and the WJ III COG was conducted by comparing how children without learning difficulties performed across both measures. While these comparisons helped to provide evidence for the convergent and discriminant validity of both measures, it failed to provide evidence for the comparability of the measures for children referred for academic difficulties. Prior research examining the relationship between the Wechsler Scales and the Woodcock-Johnson batteries has been conducted in samples of children with learning and behavior difficulties (Bracken, Prasse, & Breen, 1984; Phelps, Rosso, & Falasco, 1984; Thompson & Brassard, 1984; Ysseldyke, Shinn, & Epps, 1981). However, these studies utilized much earlier versions of the tests, limiting the generalizability of the results.

The technical manuals for both the WISC-IV and the WJ III assessment instruments provide results for special populations, providing important information regarding each test's specificity and the clinical utility for diagnostic assessment (Hebben, 2004). However, the generalizability of these results is limited for a number of reasons. Studies for each measure were completed at different times and with different samples. Moreover, the studies completed during the standardization of the WJ III COG were limited, and included only one sample of students with either Attention-Deficit/Hyperactivity Disorder (ADHD) or a learning disability who were administered both the WJ III COG and the WISC-III. Limitations also exist because of the nature of the samples used for the WISC-IV special group studies. Sample sizes were generally small and participants were not randomly selected. Also, data were derived from independent clinical settings, suggesting that different criteria and procedures were used for diagnosis (Hebben). Specifically, though, the WISC-IV studies do not provide comparisons of group performance across a number of different intelligence measures (outside the realm of the Psychological Corporation). Also, the correlations between the WISC-IV and the WJ III COG are unclear regarding the extent to which they measure similar cognitive processes. The revisions to the WISC suggest that the WISC-IV would prove to be more correlated with the WJ III COG, even among samples of children identified with neuropsychiatric impairments. This is due to the fact that the previous versions of the WISC and WJ have demonstrated adequate convergent validity and the improved structure of the WISC-IV now aligns the battery with more current cognitive theory. However, such comparison studies have yet to be conducted.

In sum, there are several shortcomings in the research surrounding the validity of the WISC-IV. Numerous revisions were made during the development of the WISC-IV, resulting in a measure that is significantly different in content and structure from the WISC-III. Therefore, validity research regarding the WISC-III does not provide an adequate understanding of the current test's convergent and discriminate validity. Additionally, because the changes made more closely align the WISC-IV with CHC theory, the generalizability of findings concerning the WISC-III is questionable. Finally, research is lacking regarding the comparison of the WISC-IV and other cognitive measures utilized among referred groups of children. Given these limitations, research is needed to explore the relationship between the WISC-IV and the WJ III COG in a clinically referred population. More specifically, this study will examine the relationships between the WISC-IV and WJ III COG Full Scale scores and Index/Composite scores, as well as help to determine if any significant differences exist between the global scores within a clinical sample of children.

Chapter II: Review of the Literature

In order to understand the relevance of the proposed investigation, it is necessary to present an examination of: (a) theories underlying the conceptualization and measurement of IQ; (b) the development of the Cattell-Horn-Carroll (CHC) theory of intelligence; and (c) the stages of construct validity used to establish the test validity of the Wechsler scales, including examination of its construct validity from the perceptive of the Cattell-Horn-Carroll theory.

Conceptualization and Measurement of IQ

Intelligence, as defined in the psychometric sense, is the general reasoning capacity used in various problem-solving tasks (Kline, 1991) and results from variations in brain structure following the interaction of genetic and non-genetic factors (Draganski, Gaser, Busch, Schuierer, Bogdahn, & May, 2004; Thompson et al., 2001). Theories regarding the nature of intelligence have evolved across time (Carroll, 1993, 1997; Gardner, 1983; Horn & Noll, 1997; Neisser et al., 1996; Spearman, 1932; Thurstone, 1938) and have offered tremendous variation in the conceptualization and assessment of cognitive functioning (Das, Naglieri, & Kirby, 1994; Flannagan & McGrew, 1997; Sattler, 2001; Sternberg & Berg, 1986).

Considerable differences have arisen over time concerning those aspects included in the measurement and definition of intelligence. Some theorists, such as Jean Piaget, conceptualized intelligence using developmentally based experiences. Piaget's theory (1972) holds that new information is assimilated into and accommodated by existing cognitive structures, with this framework being applied to the intellectual development of all children. More recent investigations have explored the anatomical and physiological brain substrates related to intelligence. Those who have attempted to clarify intelligence according to more explicit theories have sought to provide a framework by which one can take into account individual differences in comprehension, adaptation, learning, and problem-solving, which can vary across domains and according to settings and standards of performance even within a single individual (Neisser et. al, 1996).

According to Sattler (2001), early theories of intelligence focused more on examining sensorimotor functions thought to contribute to mental functioning. The possibility of studying mental ability through experimental and practical means did not formally arise until the work of J. M. Cattell in 1890, whereas the focus on examining cognitive functions as typically seen in modern intelligence tests did not take place until the turn of the century following the work of Carl Wernicke and Theodor Ziehen.

Intelligence tests were originally developed independent from theory. The first intelligence test, developed by Binet and Simon in 1905, attempted to assess degrees of individual mental ability (Sattler, 2001). Although the Binet-Simon scale was not theory driven and lacked indicators of performance (Thorndike, 1997), it was the first of its kind to incorporate administration standards and items ranked according to level of difficulty while also acknowledging age-based cognitive development in order to measure higher mental processes associated with intelligence (Sattler, 2001). In 1916, the test was revised to include a ratio-based intelligent quotient that, despite criticism, allowed for the comparison of intellectual functioning among individuals (Thorndike, 1997). Since its conception, the psychometric approach to intelligence testing has evolved considerably, with certain tests even created to measure specific constructs, such as verbal and nonverbal intelligence (Neisser et al., 1996). More recently developed tests emphasize empirically based theories of cognitive functioning. With this in mind, intelligence is thought to best be measured by instruments that take into account the multiple and fairly independent factors that contribute to the phenomenon (Strauss, Sherman, & Spreen, 2006).

Spearman (1904) brought the idea of psychometric testing to life by developing an approach to understanding intelligence using factor analytic methods. Spearman formulated a two factor theory of intelligence. He proposed a general factor, *g*, to represent what all tests have in common and, therefore, to reflect how a person would perform across batteries of intelligence tests, regardless of the domains included (Thorndike, 1997). Spearman hypothesized that positive correlations found between diverse measures were accounted for by *g*, with varying amounts of *g* represented within each measure, such that those with higher *g* loadings were more highly correlated (Brody, 1999). Spearman later included smaller specific factors (*s*) required for particular cognitive tasks (Thorndike, 1997), which accounted for the overlapping variance between tests beyond the sole influence of *g* (Wasserman & Tulsky, 2005). Many factor analytic studies have found support for *g*, with all modern intelligence tests purported to measure a general intelligence factor that accounts for the largest proportion of variance in an intelligence test battery (Kamphaus, 2005).

Another factor-analytic based theory included that of Primary Mental Abilities (PMA) (Thurstone, 1938), which posited that *g* resulted from the relationship between a number of primary factors (Thorndike, 1997). In contrast to Spearman's unitary theory of intelligence, this theory offered insight into understanding how intelligence could be measured through distinct cognitive factors (Flanagan et al., 2000). Using new methods of factor analysis, Thurstone's research suggested that intelligence was comprised of seven independent factors, or primary abilities, with intraindividual profiles shown among samples of individuals with similar overall IQ scores. Later research suggested that both *g* and more specific abilities contributed to an individual's overall IQ score. Wechsler (1975) considered intelligence to arise not only from cognitive factors, but also from ecological factors (e.g. planning and goal awareness, enthusiasm, impulsiveness, anxiety, and persistence), describing intelligence as the capacity of an individual to understand the world and meet its demands. Moreover, he believed that intelligence was not localizable to any one particular area of the brain (Kamphaus, 2005).

Wechsler's original Wechsler-Bellevue test was based on the conceptualization of intelligence as both as a global and specific entity. That is, qualitatively distinct abilities contribute to the individual's behavior as a whole, producing intelligent behavior that reflects *g*. In classifying individuals based upon their overall level of cognitive functioning, Wechsler chose tests that he believed were the most clinically useful and ecologically valid (Coalson & Weiss, 2002; Kamphaus, 2005). His conceptualization of intelligence as a global construct that can be measured by distinct abilities is consistent with current research. Likewise, his hypothesis that performance on intelligence tests is tied to specific test content and that "what we measure with tests is not what tests measure" (Wechsler, 1975, p. 139), reflects the notion that assessing an individual's intelligence involves more than just obtaining the individual's intelligence test scores (Coalson & Weiss, 2002).

Cattell-Horn-Carroll (CHC) Theory

Moving beyond a unitary model of *g,* the theory of fluid and crystallized intelligence (*Gf-Gc* theory), as originally proposed by Raymond Cattell (1941, 1943, 1957), offered significant understanding regarding the dichotomous nature of intelligence. Though not immediately contributing to the development of assessment instruments, the expanded Cattell-Horn *Gf-Gc* theory (Horn & Noll, 1997) was later integrated with John Carroll's three-stratum theory (1993, 1998) to become the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (McGrew, 1997). The CHC theory subsequently became the basis for the WJ III. The Cattell-Horn-Carroll theory is purported to be the most comprehensive, empirically supported, and psychometrically-based framework for use in conceptualizing the structure of human cognitive abilities (McGrew, 2005).

Gf-Gc theory was created following Thurstone's Primary Mental Abilities theory, which argued against a general factor underlying all intelligent behavior (Horn & Noll, 1997). Cattell expanded on Spearman's concept of general intelligence (*g*) by identifying fluid (*Gf*) and crystallized (*Gc*) intelligence as two other types of general abilities (Phelps, McGrew, Knopik, & Ford, 2005). *Gf* is purported to include the more biologically and neurologically influenced factors of nonverbal, mental efficiency and adaptive and new learning capabilities, while *Gc* includes knowledge and information, individual abilities influenced by acculturation and supported by fluid intelligence (McGrew & Flanagan, 1998; Sattler, 2001). This dichotomous theory was later expanded by Cattell's student, John Horn (1965, 1968), who recognized additional broad cognitive abilities, including visual perception or processing (*Gv*), short-term memory (*Gsm*), long-term storage and retrieval (*Glr*), speed of processing (*Gs*), and auditory processing (*Ga*). The theory was again expanded to include a total of 10 broad abilities with the addition of quantitative ability (*Gq*) (Horn, 1988, 1989) and facility with reading and writing (*Grw*) (Woodcock, 1998) (McGrew & Woodcock, 2001). The expanded *Gf-Gc* theory provided a basis for analysis of development and neurological functioning by offering a description of the structural organization of broad abilities among primary abilities and the variables with which abilities correlate (Horn & Null, 1997).

In 1993, Carroll conducted a comprehensive analysis of independent-source structural research on human cognitive abilities to unify the study of cognitive abilities. This comprehensive review included a re-factor-analysis of data from 461 of the major psychometric post-1925 data sets, including many of the studies investigating aspects of the *Gf-Gc* theory. Using exploratory factor analysis, Carroll proposed that human cognitive abilities could be represented in a hierarchical structure, with *g* (general intelligence) existing as an overall, general ability (Stratum III). Carroll proposed that Stratum II included eight broad cognitive abilities, while Stratum I consisted of 69 specific, or narrow, abilities that were grouped accordingly into the broad categories (McGrew & Woodcock, 2001; Sanders, McIntosh, Dunham, Rothlisberg, & Finch, 2007).

Carroll's theory was intended to extend or replace those theories regarding the structure of cognitive abilities already in use (Carroll, 1997). While differences exist between the expanded Cattell-Horn *Gf-Gc* (fluid-crystallized intelligence) theory and Carroll's three-stratum theory, particularly regarding the inclusion of a higher-order *g* factor, the overarching similarities provided the impetus for McGrew's (1997) proposal for an integrated Carroll and Horn-Cattell *Gf-Gc* framework. McGrew's hierarchical model identified *g* (general intelligence) at the apex (stratum III), 10 broad abilities at the second stratum, and numerous narrow abilities pertaining to each of the broad abilities (stratum I). This model was presented to provide a means by which to classify intelligence tests according to the two factor theory of fluid (*Gf)* and crystallized (*Gc*) intelligence and to allow clinicians to use a *Gf-Gc* cross-battery approach to assessment. A description of the factors included in the structure of CHC theory is outlined in Table 1.

Table 1
CHC Broad Ability Factors

Factor	Description Abilities
Fluid Reasoning (*Gf*)	The ability to reason and/or problem-solve given novel or unfamiliar information.
Crystallized Intelligence (*Gc*)	Knowledge acquired through verbal communication, and/or factual information
Short-Term Memory (*Gsm*)	The ability to hold information in immediate memory and manipulate it for a task.
Quantitative Knowledge (*Gq*)	Ability to reason using numbers and applying numerical concepts.
Visual-Spatial Reasoning (*Gv*)	Ability to organize and synthesize visual stimuli.
Long-Term Retrieval (*Glr*)	Ability to store information in memory and retrieve it at a later time.
Auditory Processing (*Ga*)	Ability to organize and synthesize information that is presented auditorily.
Processing Speed (*Gs*)	The ability to perform automatic cognitive tasks and maintain focused attention under the influence of time pressure.
Reading and Writing Ability (*Grw*)	Ability to decode and synthesize information and apply this information in written form.
Decision/Reaction time (*Gt*)	Ability to react quickly or to quickly make decisions.

Note. Adapted from *Examiner's Manual. Woodcock-Johnson Tests of Cognitive Abilities* (pp. 19-20) by N. Mather and R. W. Woodcock, 2001, Itasca, IL: Riverside Publishing. Copyright 2001 by Riverside Publishing; "The three-stratum theory of cognitive abilities" by J.B. Carroll, 1997. In D. P. Flanagan, J. L. Genshaft, and P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91), New York: Guilford. Copyright 1997 by Guilford.

Since its inception, the Cattell-Horn-Carroll model has been refined and was subsequently used as the basis for development of the WJ III COG in providing measurement of the broad Cattell-Horn-Carroll factors (McGrew & Flanagan, 1998; Shrank & Flanagan, 2003). Factor analytic research has contributed to the development of the Cross-Battery Assessment (XBA) approach (Flanagan & McGrew, 1997; Flanagan et al., 2000; Flanagan et al., 2007; McGrew & Flanagan, 1998), which has provided a systematic method for interpretation of intelligence tests with other assessment measures based upon CHC theory (Flanagan et al., 2000). The CHC model is "…a well reasoned… psychometric taxonomic framework…that can improve research vis-à-vis the use of a common nomenclature," (McGrew, 2009a, p. 4). It is supported by structural factor-analytic evidence and developmental, neurocognitive, and heritability evidence (see Horn & Noll, 1997). However, this model is limited in that it is only partially empirically tested and not based upon a series of comprehensive empirical confirmatory factor analysis comparison studies (McGrew). Despite these limitations, research has indicated that interpretations of the Wechsler scales can be benefited by linking the analysis to CHC theory.

Psychometric Validity

Validity refers to "the degree to which a test measures what it is intended to measure" (Garrett, 1937, p. 324). While this definition was at one time considered sufficient to judge a test's validity, the validity of a test's use and interpretation is now established by the degree to which it is supported by empirical evidence and theory (Woodcock et al., 2001). A test's construct validity is supported through substantial, structural, and external validity evidence (Benson, 1998; Cronbach, 1971; Cronbach & Meehl, 1955; Loevinger, 1957; Messick, 1989).

According to Flanagan (2000), "Substantive validity" refers to the use of an underlying theoretical domain to create and provide operational definitions of measured constructs. Structural validity relates to the degree to which a test measures the constructs it purports to measure, which is reflected by a test's internal structure. Both substantive and structural validity should be established before determining a test's external validity. External validity is revealed when a structurally valid measure is shown to demonstrate expected convergent and divergent relations with substantively and structurally strong external measures.

While the WISC-IV is purported to align more closely with the Cattell-Horn-Carroll (CHC) theory than it's predecessors, it is important to examine research exploring (a) how well previous and current versions of the Wechsler scales have mapped onto CHC theory, (b) the within- and cross-battery structural validity of the Wechsler scales, (c) the construct validity of the Wechsler scales, and (d) WISC-III and WISC-IV differences. Data in this regard will strengthen the validity of interpretations made from the Wechsler measure and provide support for the convergent and discriminate relationships between the WISC-IV and the WJ III COG.

Substantive Validity

A test's construct validity is enhanced when an identified theory is used as the foundation for the formulation of test items intended to measure defined theoretical constructs. Intelligence measures with adequate content validity, or construct representation, can be identified by the extent to which a large number of different tests provide measurement of a unique aspect of the major components of the theoretical domain of constructs (Flanagan et al, 2000; Loevinger, 1957). While Watkins, Wilson, Kotz, Carbone, and Babula (2006) found evidence of the substantive validity of the WISC-IV four-factor structure among students referred for special education services, previous versions of the Wechsler scales have been criticized for their lack of adherence to a strong theory of intelligence, resulting in weak substantive validity, particularly when viewed from the perspective of CHC theory (Carroll, 1993; Flanagan, 2000; Flanagan et al., 2000; Wasserman & Bracken, 2003).

According to Macmann and Barnett (1994), the Wechsler scales were faulted for not reflecting the complexity of intelligence or providing an adequate sampling of relevant variables. This is attributed to the fact that the instruments reflect a general factor model, suggesting that the tests are not fit for testing cognitive theory because of their lack of a strong substantive foundation. The authors also suggested that the verbal and performance constructs were not discriminately different. While the WISC-IV incorporates recent theory regarding the nature of intelligence, the structure of the Wechsler scales provides measurement of an underlying "informal theory," which includes four first-order abilities that are manifestations of a global general intelligence (Keith & Witta, 1997).

Previous research has attempted to evaluate the content validity of the WISC outside of the perspective of the Cattell-Horn-Carroll theory. This research has provided insight into the validity of inferences that can be drawn from the measure, suggesting that interpretations can be made based on the Guttman model of intelligence (Cohen, Fiorello, & Farley, 2006). Also, interpretations of working memory can be made based on Baddeley's theory of working memory (Leffard, Miller, Bernstein, DeMann, Mangis, & McCoy, 2006). However, because the CHC theory is considered to be an excellent means of evaluating the content validity of intelligence tests (Flanagan, 2000), an extensive body of literature is available concerning how well the Wechsler scales map onto the CHC model. Research examining test classifications according to

CHC theory considers a broad construct to be strongly represented when it is measured by three or more narrow abilities. From this viewpoint, research has suggested that previous Wechsler measures have only provided strong measurement of crystallized knowledge (*Gc*) and visual processing (*Gv*) (Flanagan, 2000; Flanagan et al., 2000; McGrew & Flanagan, 1998). While the Wechsler Adult Intelligence Scale- Third Edition (WAIS-III; Wechsler, 1997) is also reported to demonstrate adequate representation (defined by the measurement of two or more narrow abilities representing a broad ability) of processing speed (*Gs*) and short-term memory (*Gsm*), the WISC-III has been shown to only provide adequate measurement of *Gs*.

More recent research (Flanagan & Kaufman, 2004; Kaufman et al, 2006) has suggested that revisions made to the WISC-IV more closely align the measure with CHC theory. According to Keith, Fine, Taub, Reynolds, and Kranzler (2006), although the WISC-IV is not entirely consistent with CHC theory, the measure draws on the theory in its organization and structure. In comparing the WISC-IV model with that of one based on CHC theory, results of higher order confirmatory factor analysis (CFA) using the WISC-IV standardization data revealed that the WISC-IV appears to measure crystallized ability (*Gc*), visual processing (*Gv*), fluid reasoning (*Gf*), short-term memory (*Gsm*), and processing speed (*Gs*). Keith and colleagues concluded that *Gc* is strongly represented by administration of the core verbal subtests of Vocabulary, Similarities, and Comprehension, whereas, *Gsm* is adequately measured by the core Working Memory subtests of the battery (i.e., Digit Span and Letter-Number Sequencing).

Other broad abilities, however, are only adequately represented with the inclusion of supplemental subtests. For example, it was suggested that the WISC-IV Picture Concepts and Matrix Reasoning subtests provide adequate representation of *Gf* (fluid reasoning). However this construct is more strongly represented when Arithmetic is included as a measure of fluid reasoning. Both the Coding and Symbol Search subtests provide adequate representation of processing speed (*Gs*), whereas the addition of Cancellation more strongly measures this construct. According to Keith et al., (2006), the Block Design subtest appears to be the only core test to measure visual processing (*Gv*). However, the addition of the Picture Completion subtest provides adequate measurement of this construct, and strong representation is found when the Symbol Search subtest is included as a measure of visual processing. Overall, Keith and colleagues suggest that the WISC-IV provides greater measurement of the CHC constructs when supplemental subtests are included. Also, construct representation is increased when additional constructs measured by a particular subtest are considered. Taken together, these results suggest that the WISC-IV provides greater measurement of the Cattell-Horn-Carroll model than its predecessors.

Consistent across all Wechsler scales is the lack of measurement of auditory processing (*Ga*) and long-term retrieval (*Glr*) (Flanagan, 2000; Flanagan & Kaufman, 2004; Flanagan et al, 2000). However, according to Flanagan and Kaufman, the substantive validity of *Ga* and *Glr* can be improved if interpretations of the WISC-IV are made congruently with joint findings from the Children's Memory Scale (CMS; Cohen, 1997), which provides measurement of *Glr*, and the Wechsler Individual Achievement Test–Second Edition (WIAT-II; The Psychological Corporation, 2002), which provides measurement of *Ga* to some extent.

Moreover, research has suggested that the validity of interpretations made with the Wechsler scales can be improved using the XBA approach to assessment. Studies examining the relationship between CHC-related cognitive abilities and reading achievement (Evans, Floyd, McGrew, & Leforgee, 2002; Flanagan, 2000; Floyd, Keith, Taub, & McGrew, 2007; Vanderwood, McGrew, Flanagan, & Keith, 2002) have suggested that while the *g* factor underlying a cross-battery assessment approach accounts for a significant proportion of variance in reading achievement, specific CHC cognitive abilities [e.g. auditory processing (*Ga*), crystallized intelligence (*Gc*), processing speed (*Gs*), long-term retrieval *(Glr)*] can be used to better explain and understand academic achievement.

Research has also examined the relationship between CHC-related cognitive abilities and mathematics achievement (Floyd, Evans, & McGrew, 2003; Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001; Proctor, Floyd, & Shaver, 2005). Findings indicate that while general intelligence is associated with math achievement, the broad cognitive ability factors of Comprehension–Knowledge (*Gc*), Fluid Reasoning (*Gf*), and Processing Speed (*Gs*) display consistent relationships with measures of math calculation skills and math reasoning. Likewise, long-term retrieval (*Glr*) abilities may be important to early mathematics calculation skill development. Moreover, the narrow cognitive ability of working memory may display a strong relationship with mathematics achievement, whereas auditory processing (*Ga*) abilities may demonstrate a moderate relationship with math calculation skills. These findings suggest that the weak substantive foundation of the Wechsler scales and investigations of specific cognitive deficits associated with specific learning disabilities can be improved through a cross-battery approach to assessment.

Overall, research has suggested that the WISC-IV provides greater representation of the Cattell-Horn-Carroll theory constructs, reflecting a stronger substantive foundation. As such, it is likely that the measure may provide a better explanation of achievement without the support added via a cross-battery approach. Furthermore, these results lend support for the convergent relationships between the WISC-IV and the WJ III COG.

Structural Validity

The structural stage of validity involves studying only the observed variables for a measure to determine if they are internally consistent. Evaluating the internal consistency of a test determines if the test measures the constructs it purports to measure (Benson, 1998). Evidence for the structural validity of the Wechsler Scales has been demonstrated with research examining the internal structure of the scales.

Within-battery structural research. Despite criticisms regarding the lack of application to contemporary theory and research, the Wechsler scales have maintained an unrivaled position of dominance within the field of psychological assessment (Flanagan & Kaufman, 2004). According to Flanagan (2000), within-battery factor analytic studies have lent support for the initial two-factor solution involving a verbal and nonverbal factor and later versions [e.g. Wechsler Intelligence Scale for Children – Revised (WISC-R)] comprised of three factors: Verbal Comprehension, Perceptual Organization, and Memory/Freedom from Distractibility. Research by Keith and Witta (1997) has provided support for the four-factor model of the WISC-III, which has also been cross-validated in a variety of samples (e.g. Donders & Warschausky, 1996; Grice, Krohn, & Logerquist, 1999; Konold, Kush, & Canivez, 1997; Tupa, Wright, & Fristad, 1997).

Studies examining the structural validity of the WISC-IV within the normative sample (Watkins, 2006) and within referred samples (Bodin, Pardini, Burns, & Stevens, 2009; Watkins et al., 2006) have suggested that WISC-IV interpretations should not neglect the strong influence of the general factor (*g*). This is because it accounts for more variance in any of the first-order factors, and in each of the 10 core subtests than any first-order factor. Though these findings provide evidence for the multilevel four-factor structure of the WISC-IV, they call into question the predictive validity of the first-order (broad) factor scores. However, research has argued against focusing on a global FSIQ (Fiorello, Hale, McGrath, Ryan, & Quinn, 2001; Hale, Fiorello, Kavanagh, Holdnack, & Aloe, 2007) and has provided evidence that a multifactoral model better represents the WISC-IV standardization, particularly the clinical group data over and above a single factor model (e.g. Fiorello et al.; Hale et al., 2001; Keith et al., 2006; Wechsler, 2003).

Chen and Zhu (2008) examined the consistency of the WISC-IV factorial structure across genders using the WISC-IV standardization data. Results suggested that the structure did not vary across groups, providing support for the overall factor pattern of the measure. In contrast, research by Keith et al. (2006) has led to inconsistent results regarding the validity of the WISC-IV factor structure. Results of their higher order, multi-sample CFA indicated that the WISC-IV measures the same constructs across the 11-year age range of the test. However, they found that a five-factor CHC model provided a better fit to the WISC-IV standardization data when comparing the scoring structure of the WISC-IV against a theory-derived model more closely aligned with CHC theory.

Overall, research has supported the various Wechsler factor structures, with general improvement shown for each new structure pattern. However, concern has been raised regarding factor-based interpretations because of continued changes regarding the factorial nature of the batteries (Flanagan, 2000). Concern has also been raised regarding the strength of interpretations made from the perspective of the WISC-IV's four-factor structure because test performance may be better explained by the general factor or from the perspective of the CHC theory. However, the multilevel structure of the WISC-IV has received broad support, likely as a result of its attempt to align more closely with current theoretical foundations, contributing to the internal validity of the measure.

Cross-battery structural research. Cross-battery factor analysis incorporates tests from more than one intelligence measure, allowing tests from the different individual batteries to load on the theoretical factors specified by another theory (Flanagan et al., 2000). CHC theory has been widely used in factor analytic studies (Elliott, 1997; Keith et al., 2001; Reynolds, Keith, Fine, Fisher, & Low, 2007; Sanders et al., 2007) because it allows for understanding the constructs measured by intelligence tests. Research has provided evidence of the CHC classifications of previous Wechsler measures (Flanagan et al., 2000; McGrew & Flanagan, 1998). While this has lent support for the construct validity of the Wechsler scales, there is a lack of current cross-battery investigations examining how the WISC-IV measures distinct CHC broad and narrow abilities.

A study by Phelps, McGrew, Knopik, and Ford, (2005) examined the validity of the WISC-III and WJ III COG broad and narrow Cattell-Horn-Carroll ability classifications. Consistent with previous cross- and within-battery research, results revealed that, in contrast to the WJ III COG, the Wechsler scale did not load on the broad constructs of *Ga* (auditory processing), *Gf* (fluid reasoning), or *Glr* (long-term retrieval). However, results were also consistent with previous research (Flanagan et al., 2000; McGrew & Flanagan, 1998) demonstrating that the batteries are similar in the broad constructs measured, including: *Gc* (crystallized intelligence), *Gq* (quantitative knowledge), *Gs* (processing speed), *Gsm* (short-term memory), and *Gv* (visual processing). Although changes have been made to the core subtests compromising the WISC-IV, these results suggest that the WISC-IV and WJ III will continue to demonstrate a greater concordance of broad ability constructs.

The Phelps et al. (2005) study was also consistent with previous research (Flanagan et al., 2000; McGrew & Flanagan, 1998; McGrew & Woodcock, 2001) suggesting that the Wechsler and Woodcock-Johnson measures demonstrate some consistency in the CHC narrow ability constructs measure, including: crystallized intelligence [language development (LD), lexical knowledge (VL), and general information (KO)] and visual processing [spatial relations (SR) and visualization (VZ)]. Other results of the Phelps et al. study call into question the correlations that may exist between the WISC-IV and the WJ III COG. While the WISC-III contained a greater proportion of *Gv* (visual processing) tests when compared to the WJ III, the WISC-III Picture Completion and Picture Arrangement subtests demonstrated low *Gv* (visual processing) loadings (.37). However, since these subtests are no longer included as core subtests on the WISC-IV, it is unclear how well the Wechsler Scale provides measurement of *Gv.*

Results of the Phelps et al. (2005) study also suggest that the WISC-IV Working Memory Index shows improved measurement of *Gsm* (short-term memory). This is in contrast to the Freedom from Distractibility Index from previous versions of the Wechsler scale, which has not been found to be consistent with any CHC broad ability (Flanagan, 2000), demonstrating improvement in the scale's factor structure. Phelps et al. found only one strong measure of *Gsm* (Digit Span) from the WISC-III, in contrast to the WJ III, which had four strong measures of *Gsm*. The WISC-IV has made changes to the structure of the Working Memory Index with the inclusion of an adapted version of the WAIS-III subtest, Letter-Number Sequencing, a strong measure of short-term memory (Flanagan et al., 2000; Flanagan & Ortiz, 2001; McGrew & Flanagan, 1998). This suggests that the WISC-IV shows improved measurement of the CHC broad ability construct of short-term memory (*Gsm*) because it now includes measurement of the two narrow abilities of memory span (Digit Span) and working memory (Letter-Number Sequencing), providing evidence for the convergent validity between the WISC-IV and the WJ III COG.

External Validity

Construct validity is evidenced through research assessing the convergent and discriminant validity of independent measures. This is needed to justify novel trait measures or validate test interpretation (Campbell & Fiske, 1959). Convergent validity is supported if stronger relationships are shown between independent measures that are intended to measure the same construct than with variables that do not measure the same construct or that share a similar method. Moreover, tests can be invalidated when they do not demonstrate appropriate discriminant validity, such that too high correlations are shown with other independent measures that purport to measure a different construct.

Correlations with other measures. Research evaluating the external validity of previous editions of the Wechsler measures with differing ability measures has generally provided positive external validity results (Kaufman & Kaufman, 1993; Raskin, Bloom, Klee, & Reese, 1978; Reynolds & Kamphaus, 2003). Similar results have been shown regarding the correlations of the WISC-III with differing ability measures and within different populations (Bell, Rucker, Finch, & Alexander, 2002; Canivez, Neitzel, & Martin, 2005; DiCerbo & Barona, 2000;

Dumont, Cruse, Price, & Whelley, 1996; Kaufman & Kaufman, 2004; Law & Faison, 1996; Prewett & Matavich, 1994; Vo, Weisenberger, Becker, & Jacob-Timm, 1999; Wechsler, 1991; Zimmerman & Woo-Sam, 1997). However, the significant differences between the WISC-III and WISC-IV make one unable to generalize findings from one version to another.

Correlational research involving the WISC-IV has continued to demonstrate evidence for positive external validity. The WISC-IV manual provides the correlations between the Full Scale IQ and composite scores with other measures of intelligence. While the WISC-IV FSIQ was shown to correlate significantly with the WISC-III, Wechsler Preschool and Primary Scale of Intelligence- Third Edition (WPPSI-III), and WAIS-III FSIQs (.89, respectively) (Wechsler, 2004), correlations were largely examined between measures guided by Wechsler theory, limiting the validity of interpretations made regarding the test scores.

Further research has explored the correlations between the WISC-IV and other independent measures within clinical samples. Edwards and Paulin (2007) examined the convergent relations between the WISC-IV and the Reynolds Intellectual Assessment Scale (RIAS; Reynolds & Kamphaus, 2003) in a sample of 48 elementary school children referred for psychoeducational testing. Results revealed a strong correlation of .90 between the WISC-IV FSIQ and the RIAS Composite Intelligence Index. Moreover, results provided support for relationships between composites scores measuring similar constructs of comprehension-knowledge and fluid reasoning. While the RIAS does not measure other constructs included on the WISC-IV, namely processing speed and working memory, conclusions regarding the external validity of these constructs and the FSIQ score are limited.

Comparisons between the WISC-IV Full Scale IQ score and the Kaufman Assessment Battery for Children, Second Edition (KABC-II) Fluid-Crystallized Index (FCI) have demonstrated adjusted correlation means of .89 across age-groups of the standardization samples (Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005). The KABC-II is based on a dual theoretical foundation. In addition to following the CHC approach, as measured by the FCI, this battery also provides measurement of the Luria neuropsychological model, yielding a global score referred to as the Mental Processing Index (MPI). Adjusted correlation means across age-groups between the WISC-IV FSIQ and MPI were .88, providing support for the convergent validity of the Wechsler battery outside of CHC theory. Similar correlations between the WISC-IV FSIQ and the KABC-II FCI and MPI were also found in a sample of 30 Taos Indian Pueblo children (Fletcher-Janzen, 2003), further suggesting that the tests are measuring the same major abilities.

Unlike the Edwards and Paulin (2007) study, support for the convergent and discriminate validity of the WISC-IV has been provided through its correlations with the KABC-II composite scores. The WISC-IV was shown to have correlations ranging from .66 to .85 with constructs measuring sequential/*Gsm* (short-term memory), simultaneous/*Gv* (visual-spatial reasoning), planning/*Gf* (fluid reasoning), and knowledge/*Gc* (crystallized intelligence) abilities (Kaufman et al., 2005). Nonetheless, results are limited regarding the validity of the WISC-IV Processing Speed Index as a measure of processing speed because the KABC-II does not measure this construct. Overall, research examining the correlations between the WISC-IV and external measures has provided support for the validity of inferences drawn from the tests comprising the

Wechsler FSIQ score. However, results have not provided sufficient support for understanding and interpreting inferences drawn for the composite scores.

Correlations between the Wechsler Intelligence Scale for Children and the Woodcock-Johnson. Although there are no published studies investigating the relationship between the WISC-IV and the WJ III COG, research evaluating earlier versions of the measures has provided evidence for the convergent and discriminant validity of the broad scores. Research that has focused on examining the concurrent validity between the Wechsler Intelligence Scale for Children-Revised (WISC-R) and the Woodcock-Johnson Tests of Cognitive Ability (WJTCA) (Woodcock and Johnson, 1977) has generally shown positive correlations between the WJTCA Broad Cognitive score and the WISC-R Full Scale IQ score (i.e., .79).

Similar findings have also been demonstrated in studies involving referred samples (McGrew, 1983; Sanville & Cummings, 1983). Thompson and Brassard (1984) examined the performance of three groups, including typical children and children diagnosed with mild-to-moderate, and severe learning disabilities (LD), on the WJTCA and WISC-R. Results revealed strong correlations between the WJTCA and WISC-R Full Scale IQ scores in each group (correlations of .86, .74, and .93 for the normal, mild-to-moderate LD, and severe LD groups, respectively). Reeve, Hall, and Zakreski (1979) compared the WJTCA and the WISC-R in a sample of children with reading and/or mathematics disorders. Results revealed a correlation of .79 between the Full Scale IQ scores, with 60 percent of the total measured variance common to the two instruments. Removal of outlier cases increased the correlation between the Full Scale scores to .89. Results also revealed differing correlations across genders, but only a small sample of eight girls was included in the study.

Other studies (i.e., Phelps et al., 1984) have reported smaller correlations between the WISC-R and WJCTA Full Scale scores. Ysseldyke, Shinn, and Epps (1981) reported a correlation of .67 between WISC-R Full Scale IQ score and the WJCTA Broad Cognitive Scale score among 50 fourth grade students with learning disabilities. The relatively small correlation between the mean scores is suggested to have been due to the restricted age range used in the study. Bracken, Prasse, and Breen (1984) administered the WJTCA and the WISC-R to 142 children referred for psychoeducational evaluation, with 104 children identified as having a learning disability and 39 students retained in a regular classroom after evaluation placement decisions. Results demonstrated a correlation of .63 among children identified as learning-disabled and a correlation of .72 among children remaining in regular education. The authors of this study interpreted the differences in results as suggesting that the two tests measure differing abilities to differing degrees.

Research suggests that at least part of the relationship between the WISC and the WJ is related to the degree to which the subtests included measure a general intelligence factor (*g*). The magnitude of the correlations found between the WISC-R and WJCTA Full Scale scores demonstrates that they are similar measures of general intelligence, suggesting some overlap of the abilities measured by each instrument. The variability of correlations across samples indicates that the composition of the two instruments differs to a degree, either because subtests included demonstrate less of a relationship to general intelligence or due to the differences between the broad cognitive constructs and specific abilities measured.

Other studies (Reeve et al., 1979; Sanville & Cummings, 1983; Thompson & Brassard, 1984; Ysseldyke et al., 1981) have demonstrated differences between the mean composite scores of previous versions of the WISC and WJ instruments. More recently, research has indicated that the WISC-IV demonstrates variability in global scores when compared to other instruments (Edwards & Paulin, 2007; Fletcher-Janzen, 2003). This highlights the need to not only consider sample characteristics and the expected cognitive profile of particular populations, but also how different assessment instruments could subsequently impact the determination of diagnoses and services received.

While the Wechsler Intelligence Scale for Children-Revised (WISC-R) and the Woodcock-Johnson Psychoeducational Battery-Revised (WJ-R) were administered for a series of validity studies reported in the WJ-R technical manual (McGrew, Werder, & Woodcock, 1991), more recently, the correlations between the broad factor scores of the WISC-III and the WJ III COG were examined in a validity study included in the WJ III technical manual. The tests were administered to 150 students without learning difficulties, ranging from 8 to 12 years of age, as part of the Phelps Grades 3 through 5 Normal Sample validity study for the WJ III COG (McGrew & Woodcock, 2001). Results showed high correlations between the WJ III General Intellectual Ability (GIA)-Standard (Std) and the GIA-Extended (Ext) scores and the WISC-III FSIQ (.71 and .76, respectively), supporting the convergent validity of the broad constructs being measured within each battery.

Though high Full Scale score correlations were found in the Phelps study (McGrew & Woodcock, 2001), only moderate correlations were demonstrated between the WJ III Thinking Ability-Std and –Ext scores and the WISC-III FSIQ (57 and .58, respectively). Since the Thinking Ability composite score provides measurement of only four broad categories [Long-term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Auditory Processing (*Ga*), Fluid Reasoning

(*Gf*)], this likely suggests some overlap in visual-spatial thinking and fluid reasoning abilities across both instruments. However, results also reflect the differentiation of abilities measured by each battery, providing support for the differences in mean scores likely to be found between the batteries.

In the Phelps study reported in the WJ III manual, correlations were examined between factors hypothesized to measure verbal knowledge and comprehension abilities (*Gc*), specifically, the WJ III Verbal Ability-Ext, Verbal Ability-Std, Comprehension-Knowledge (*Gc*) factor, and Knowledge factor scores and the WISC-III Verbal IQ (VIQ) and Verbal Comprehension Index (VCI). As would be expected, findings suggest high correlations between measures assessing *Gc* abilities; the correlations between the WJ III Verbal Ability-Ext, Verbal Ability-Std, Comprehension-Knowledge (*Gc*) factor, and Knowledge factor scores and the WISC-III VIQ were .79, .73, and.79, .76, respectively. Also, the WJ III COG factor scores showed correlations of .78, .71, .78, and .75, respectively, with the WISC-III VCI.

The Phelps study (McGrew & Woodcock, 2001) also compared broad cognitive factor scores corresponding with other broad *Gf-Gc* abilities to determine the degree of similarity between the WISC-III and WJ III COG. The WISC-III Processing Speed Index (PSI) and the WJ III COG Processing Speed (*Gs*) factor are hypothesized to measure the broad cognitive factor of processing speed (*Gs*). This is because of the extent to which they measure the ability to fluently and automatically perform cognitive tasks, particularly when focused attention and concentration are taxed due to time limits (Flanagan, McGrew, & Ortiz, 2000). The moderate correlation of.59 evidenced between these scales provides support for the measurement of *Gs* abilities on both measures, but also suggests that there is a differentiation of *Gs* abilities measured by each battery.

Correlations between factor scores hypothesized to measure short-term memory (*Gsm*) abilities, specifically, the WISC-III Freedom from Distractibility Index (FDI) and the WJ III COG Short-Term Memory factor, were obtained. Research has shown that the FDI included only one subtest (Digit Span) that provided measurement of short-term memory (Woodcock, 1990). The moderate correlation of .58 found between these broad factor scores demonstrates some overlap of the short-term memory abilities being measured by each battery.

The Phelps study further explored the relationship between the WISC-III Perceptual Organization Index (POI) and the WJ III COG Fluid Reasoning factor (*Gf*) as both are hypothesized to measure fluid reasoning (*Gf*) abilities. A correlation of .46 was found between the WJ III COG *Gf* factor and the WISC-III POI, providing limited support for the measurement of *Gf* abilities by the WISC-III POI. Previously, the construct of fluid reasoning was underrepresented on the WISC-III POI, which showed greater weighting for measures of visual processing (*Gv*). It was further shown to measure verbal knowledge and comprehension (*Gc*), making it a factorially complex measure and thereby limiting interpretations that could be made (Flanagan, 2000).

Discriminant validity results were also provided by the Phelps study, although relationships between divergent constructs may have been underestimated because the sample included restricted ranges of scores (McGrew & Woodcock, 2001). Negligible correlations (i.e., 10) were evidenced between the WJ III COG Visual-Spatial Thinking (*Gv*) factor and both the WISC-III Verbal Comprehension Index and Processing Speed Index. While the WISC-IV Block

Design and Picture Completion subtests provide some assessment of visual processing abilities, subtests measuring components of comprehension-knowledge and processing speed are thought to be unrelated to pure measures of visual-spatial thinking (Flanagan, 2001; Flanagan et al., 2007; Keith et al., 2006), providing support for the findings.

However, the relationship between the verbal and visual-spatial thinking composite scores is inconsistent with other research demonstrating a large (i.e., .50) correlation between a composite measure of verbal ability and the WJ III COG *Gv* factor among normal preschool children (see Ford, Teague, and Tusing Preschool Normal Sample, McGrew & Woodcock, 2001). While the Visual-Spatial Thinking (*Gv*) subtests included on the WJ III COG are not thought to include components of crystallized knowledge (*Gc*), verbal subtests are highly representative of general intelligence (*g*), suggesting some degree of relationship between the constructs. Furthermore, as both language development and visual-spatial processing abilities are related to math achievement (Flanagan et al, 2007), this would suggest that a greater degree of relationship exists than was found for the WISC-III.

Other findings would also suggest that the relationship between the verbal and visual-spatial thinking composites was underestimated. The WJ III COG Visual-Spatial Thinking (*Gv*) subtest of Spatial Relations has been suggested to contribute unique ability variance beyond that of the narrow abilities of visualization and spatial relations (Phelps, McGrew, Knopik, & Ford, 2005). Likewise, Anjum (2004) found a moderate correlation between the WJ III COG Spatial Relations subtest and the Word Definitions subtest from the Differential Ability Scales (DAS). Furthermore, while the WJ III COG Picture Recognition subtest involves short-term visual memory and the WISC-IV Vocabulary and Information subtests involve retrieval of stored declarative or semantic knowledge, to some extent, the constructs exhibit shared content variance, suggesting that the divergent constructs are more strongly related than was indicated by the Phelps study.

Though there was a negligible relationship between the processing speed and visual-spatial thinking composite scores, the WISC-IV Symbol Search subtest has been shown to load on the visual-spatial thinking (*Gv*) factor (Keith et al., 2006). This suggests that the WISC-IV and WJ III COG composite may show a different pattern of divergent validity. It is also likely that the Phelps study finding was truncated when considering that the WISC-IV Digit-Symbol Coding subtest also likely involves aspects of visual-spatial thinking (i.e., processing visual stimuli in order to copy simple geometric shapes). Furthermore, as both perceptual speed and visual-spatial processing abilities are related to math achievement (Flanagan et al., 2007), it would be expected that the divergent constructs would show a greater degree of relationship.

The Phelps Study (McGrew & Woodcock, 2001) also showed a negligible correlation between the WISC-III Perceptual Organization Index and the WJ III COG Auditory Processing (*Ga*) factor (.19), providing support for the discriminant validity of these factors across the batteries. However, the WISC-IV Perceptual Reasoning Index (PRI) now measures a greater proportion of fluid reasoning (*Gf*) abilities. Other research (see McIntosh and Dunham Grades 3 through 5 Normal Sample, McGrew & Woodcock, 2001) has demonstrated a moderate correlation (i.e., .41) between composite measures of fluid reasoning and auditory processing. As fluid reasoning abilities are highly related to general intelligence (*g*), it is likely that these divergent constructs would show some degree of relationship because of the high *g* loading of the WISC-IV PRI's underlying subtests.

Furthermore, it is likely that the relationship between the WISC-IV PRI and WJ III COG *Ga* factor may be greater when considering that both include subtests that reflect "thinking abilities". This is inferred from the composition of the WJ III COG Thinking Ability cluster, which includes the Cattell-Horn-Carroll (CHC) broad abilities of long-term retrieval (*Glr*), visual-spatial thinking (*Gv*), auditory processing (*Ga*), and fluid reasoning (*Gf*). Likewise, it is thought that a relationship exists between the divergent constructs when considering that both are significantly related to reading achievement (Flanagan et al., 2007).

The WISC-III Freedom from Distractibility Index, a measure of short-term memory, was demonstrated to show a negligible correlation of .17 with the WJ III COG Visual-Spatial Thinking (*Gv*) factor. While this provides evidence for the discriminant validity between measures of short-term memory and visual-spatial thinking across the batteries, the changes made to the structure of the WISC-IV short-term memory composite score is likely to result in different findings. Specifically, WISC-IV Letter-Number Sequencing subtest has been shown to load appropriately onto short-term memory and involves manipulation of auditory information in working memory. However, performance on this subtest may require visuospatial imaging to manipulate the sequence of letters and numbers in immediate awareness, suggesting that the WISC-IV Working Memory Index may show a greater degree of relationship with the WJ III COG *Gv* factor.

Likewise, it is thought that the Phelps study findings are underestimated when considering that tasks of visual-spatial thinking include working memory components. The WJ III COG Picture Recognition subtest requires that a presented picture be help in mind in order to identify the visual stimuli when comparing it to its stored representation. This working memory demand would suggest that it would be correlated with other subtests measuring short-term or working memory. Furthermore, other research (see Phelps and Ford Preschool Normal Sample Study, McGrew & Woodcock, 2001) without restricted ranges of scores demonstrated a moderate relationship (i.e., .47) between a measure of short-term memory (from the Stanford-Binet Intelligence Scale- Fourth Edition) and the WJ III *Gv* factor in preschool children.

Research examining the relationship between the WISC-III index scores and the WJ III cognitive factor scores contributed much to the literature regarding the convergent and discriminant validity of the WJ III COG, lending support for predictions regarding the convergent and discriminant validity of the WISC-IV. To date, this is the most comprehensive study to examine the relationships between the factor and composite scores for each measure, as well as to investigate the convergent and discriminant validity of more recent versions of the tests. However, the findings cannot be applied conclusively to interpret relationships between the WISC-IV and the WJ III COG because of the significant changes that have been made to the overall structure of the WISC. Furthermore, though there is some concern regarding the pattern of divergent validity found for the WISC-III, it is likely that the WISC-IV indexes will demonstrate a relative degree of divergence from the WJ III COG. Research is needed to understand the pattern of divergent validity between the WISC-IV and WJ II COG.

While similarities remain between the WISC-III and WISC-IV, their structures are considerably different with regards to the composition of the individual factor composites and the Full Scale score (Sattler & Dumont, 2004). The revisions made to the WISC provide improved measurements of fluid reasoning, working memory, and processing speed. These revisions were influenced by research demonstrating the importance of such constructs as components of cognitive functioning (Wechsler, 2004) and place greater emphasis on important neuropsychological constructs that are of prime interest in evaluating children (Baron, 2005). Understanding the changes made in regards to the four factors of the battery and the overall composition of the Full Scale score highlights the errors that arise in attempting to make interpretations of the WISC-IV based on research using the WISC-III, as well as lends support for exploration of the convergent and discriminant relationships between the WISC-IV and the WJ III COG.

Differences between the WISC-III and WISC-IV structure. While the name of the Verbal Comprehension factor was retained from the WISC-III, the composition was changed. The WISC-III VCI included four core subtests (Information, Similarities, Vocabulary, and Comprehension). On the WISC-IV VCI, the Information subtest was moved to supplemental status, while the remaining subtests were retained as core subtests (Sattler & Dumont, 2004). This modification created an index with less emphasis on the knowledge of facts (Kaufman et al., 2006). This suggests that the WISC-IV VCI is an even stronger measure of language development, lending support for strong convergent relationships between the WISC-IV and WJ III COG *Gc* factor scores.

The WISC-III Perceptual Organization Index (POI) included four core subtests: Picture Arrangement, Block Design, Picture Completion, and Object Assembly. For the WISC-IV, this factor was renamed the Perceptual Reasoning Index (PRI) to more accurately reflect the constructs that comprise the index: fluid reasoning (*Gf)* and visual processing (*Gv*). Block Design was the only subtest from the WISC-III Perceptual Organization Index that was retained as a core subtest of the PRI. Picture Completion was moved to supplemental status and two new subtests adapted from other Wechsler measures, Picture Concepts and Matrix Reasoning, were added to the PRI (Sattler & Dumont, 2004). Although still considered to be a factorially complex index, the addition of the Matrix Reasoning and Picture Concepts subtests broadened the scale to assess the construct of fluid reasoning (Keith et al., 2006), suggesting a greater correlation with the WJ III COG Fluid Reasoning factor (*Gf*) than has previously been found.

The WISC-III Freedom from Distractibility Index contained two core subtests: Arithmetic and Digit Span. For the WISC-IV, this factor was renamed the Working Memory Index (WMI). The Arithmetic subtest was moved to supplemental status on the WISC-IV, reducing the influence of math achievement (Kaufman et al., 2006). Digit Span was retained as a core subtest, while an adapted version of the WAIS-III subtest, Letter-Number Sequencing, was added to the WMI in order to enhance the measurement of working memory. Wechsler (2004) indicated that working memory is an important component of higher order cognitive processes (e.g., fluid reasoning) and has been found to be related to achievement and learning. Overall, the WISC-IV WMI places greater emphasis on memory span and working memory, reflecting its stronger focus on Short-Term Memory (*Gsm*) abilities. Previous research has demonstrated a moderate correlation of .60 between the WAIS-III WMI (which contains both the Digit Span and Letter-Number Sequencing subtests) and the WJ III COG *Gsm* factor (McGrew & Woodcock, 2001), suggesting a greater correlation of the WISC-IV WMI with the WJ measure.

The WISC-IV PSI is consistent with the WISC-III PSI as it retains both the Coding and Symbol Search subtests. However, Symbol Search was a supplemental subtest on the WISC-III and is a core subtest on the WISC-IV (Sattler & Dumont, 2004). Wechsler (2004) placed more emphasis on measuring speed of information processing since in children, processing speed demonstrates a relationship to neurological development, other higher-order cognitive abilities, and learning and reasoning. The WISC-IV PSI is expected to be a stronger measure of processing speed (*Gs*) when considering the inclusion of Symbol Search as a core subtest. Moreover, this suggests greater correlations with the WJ III COG *Gs* factor.

The WISC-IV FSIQ is comprised of 10 subtests that contribute to the four factors of the test described above. These 10 subtests include three core subtests that measure verbal comprehension (Similarities, Vocabulary, and Comprehension), three core subtests that assess nonverbal perceptual reasoning (Block Design, Picture Concepts, and Matrix Reasoning), two core subtests that measure working memory (Digit Span and Letter-Number Sequencing), and two core subtests that measure visuomotor processing speed (Coding and Symbol Search) (Sattler & Dumont, 2004). The revisions made to the FSIQ result in an obvious change in the degree to which those constructs measured by the WISC-IV are represented and, therefore, a change in the degree of representation of *g*. While the FSIQ still measures crystallized knowledge, fluid reasoning, working memory, and processing speed, each construct is now more appropriately emphasized (Baron, 2005). While the WISC-IV FSIQ and the WJ III GIA are not equivalent in the extent to which specific lower-order factors contribute to the overall score, both scores are considered to be representations of the single *g* factor (Keith et al., 2006; McGrew & Woodcock, 2001). The WISC-IV's closer alliance with CHC theory provides support for the notion that the WISC-IV FSIQ score will demonstrate significant correlations with the WJ III GIA-Ext score. Research demonstrating high correlations between previous versions of these instruments also supports this hypothesis.

Overall, the numerous changes made to the structure of the four factors comprising the WISC-IV demonstrates the attempt made by The Psychological Corporation to provide more valid operational definitions of the domain specific constructs measured by the battery, as well as to incorporate factors more consistent with the Cattell-Horn-Carroll theory. Although the validity of the new factor structure has been demonstrated, research regarding the convergent and discriminate validity of the WISC-IV and the WJ III COG is lacking, limiting conclusions that can be made. Research is needed to examine these relationships to determine if the WISC-IV does in fact align more closely with CHC theory and to provide further support for the validity of the WISC-IV.

Purpose of Investigation

The purpose of the present research was to examine the relationship between the WISC-IV scores and a theoretically based and standardized measure of cognitive ability, the WJ III COG. The proposed research questions investigated the degree of continuity between these two measures, as well examined those constructs being measured by the WISC-IV. The current research evaluated the strength of the relationship between the WISC-IV Full Scale Intelligence Quotient score and the WJ III COG General Intellectual Ability Standard score, as well examined the comparability of the mean composite scores between the two batteries. Additionally, the current research question served to determine the pattern of convergent and discriminate validity correlations between the WISC-IV and the WJ III COG composite scores. Although previous research has demonstrated significant relationships between the WISC-III and the WJ III COG, significant differences exist between the WISC-III and the WISC-IV, indicating that findings from prior studies cannot extend to the validity research concerning the WISC-IV.

As such, the current research examined those correlational patterns that exist when taking into account the significant changes made to the Wechsler scale. Also, while research of earlier versions of these measures was conducted within clinical samples, more recent research concerning the WISC-III and the WJ III focused on demonstrating the magnitude of the correlations between these measures within a normal child population. As such, the present investigation focused on examining the extent to which these instruments yield similar scores when administered to the same children in a clinical population, which becomes important when making determinations regarding neurological and psychological disorders.

Hypothesis One

It was hypothesized that neuropsychiatric subjects (i.e., children diagnosed with either a neurological disorder or a psychological disorder, or both) would obtain higher mean Full Scale Intelligence Quotient (FSIQ) scores on the WISC-IV than General Intellectual Ability (GIA) scores on the WJ III COG.

This hypothesis was based on the empirical literature that has indicated that both typically developing and referred samples have obtained different full scale scores on previous versions of these tests (e.g., Reeve et al., 1979). In a study conducted by Bracken et al. (1984) in children referred for psychoeducational evaluations, results indicated that children diagnosed as having a learning disability performed 6 to 14 points lower on the WJTCA cluster score than on the Wechsler Intelligence Scales for Children-Revised Full Scale, Verbal, and Performance IQ scores. In contrast, students retained in regular education classrooms scored 4 to 10 points lower on the WJTCA cluster score as compared to the Wechsler Full Scale, Verbal, and Performance IQ scores. Also, their LD sample scored approximately 1 standard deviation below the normative sample on the Woodcock-Johnson. Similar performance discrepancies were found among fourth grade LD students, obtaining scores 7.68 points lower on the WJTCA Broad Cognitive than on the WISC-R Full Scale IQ (Ysseldyke et al., 1981). Although the WJCTA emphasizes academic achievement, making it more sensitive to impairments in cognitive or academic abilities, these findings suggest that differences in performance exist due to the different constructs being measured by each test.

The discrepancy between Full Scale IQ and GIA scores has also been found in adults. Metz (2005) demonstrated that while the WAIS-III FSIQ and WJ III COG scores were substantially related (.82) in a sample of college students being evaluated for a specific learning disability, the two scores were significantly different. Although the average difference between the scores on these measures was only 4.5 points, these findings reflect the need to consider the characteristics of particular tests that appear to provide comparable measurement of a person's general cognitive ability because differences underlying the measures could lead to different scores for a single individual. The hypothesized statistical difference between the mean scores for the WISC-IV and the WJ III COG is further substantiated by research indicating that differences are often found between mean scores of intelligence tests, despite the significant correlations that can be found within referred samples. Such differences can be due to the psychometric properties of the measures, the task demands and underlying structure, and the specific abilities of the students, often leading to different diagnostic impressions (Bracken, 1988; Dumont, Willis, Farr, McCarthy, & Price, 2000; Brown & Morgan, 1991; Prewett & Matavich, 1994).

A study by Thompson and Brassard (1984) examined the relationship between the WJTCA and the WISC-R in normal, mild-to-moderate, and severe LD elementary students. Groups were defined based on discrepancies between ability and achievement performance, with the normal group showing no discrepancy, children in the mild-to-moderate group showing a 30% to 44% discrepancy, and children in the severe group showing between 45% and 74% discrepancy. Results showed a discrepancy of 9.5 points in the mild-to-moderate LD group and a trend of increasing discrepancy between mean scores across the severity of LD. Phelps, Rosso, and Falasco (1984) examined the concurrent validity of the WJTCA Broad Cognitive score with the WISC-R in a sample of adolescents with behavior disorders. Results indicated that discrepancies were more likely to be seen among children with learning difficulties because of the underlying task demands of the tests and the cognitive weaknesses associated with particular diagnostic groups.

While the WJ III is based on a well-validated theory of cognitive abilities, the newly revised WISC-IV is still considered to be largely atheoretical because it is not based upon a structural model of intelligence. However, its substantive validity is strengthened by the incorporation of recent research regarding cognitive functions and learning. With such changes including greater contributions of working memory and processing speed, as well as better measures of fluid reasoning (Wechsler, 2004), it would be expected that there would be less discrepancy across mean scores. This would be due to the WISC-IV's focus on providing stronger measurement of cognitive abilities according to CHC theory, specifically crystallized knowledge (*Gc*), fluid reasoning (*Gf*), processing speed (*Gs*), and short-term memory (*Gsm*), along with measurement of visual-spatial thinking (*Gv*).

Despite the apparent convergence between the measures, the WISC FSIQ score does not take into account the broad cognitive abilities of auditory processing and long-term retrieval (Flanagan, 2000; Flanagan et al., 2007), and demonstrates limited measurement of visual processing abilities (Keith et al., 2006; Phelps et al., 2005). In contrast, all of these domains are well-reflected in the WJ III COG GIA-Ext score. These differences are likely to result in discrepancies in mean scores, particularly when considering that children with reading disorders are likely to perform more poorly on tests of auditory processing and retrieval fluency (see Morris et al., 1998). With this in mind, it was hypothesized that WISC-IV FSIQ scores would be significantly higher than scores on the WJ III COG GIA-Ext. There currently is a paucity of research exploring the statistical difference between the mean scores for the WISC-IV and the WJ III COG. As such, this hypothesis, which was tested using a paired samples *t*-test, was exploratory in nature.

Hypothesis Two

When considering the changes made to the structure of the WISC-IV Full Scale IQ (FSIQ), it was hypothesized that there would be a correlation significantly greater than .76 between the WISC-IV Full Scale IQ (FSIQ) and the WJ COG General Intellectual Ability-Extended (GIA-Ext) score. The WISC-IV FSIQ score and the WJ III COG GIA-Ext score have been considered comparable measures of a person's overall general intellectual or cognitive ability. While previous versions of the WISC and the WJ have demonstrated large correlations (i.e., .65 or higher) in both normal and referred samples, a strong correlation of .76 was found between the WISC-III FSIQ and WJ III COG GIA-Ext (McGrew & Woodcock, 2001). However, it is considered that there will be a stronger relationship between the measured constructs than has been found in the past.

The literature (Wechsler, 2004) has demonstrated that significant relationships exist between the WISC-IV FSIQ score and other Wechsler measures among typically developing children. Significant relationships similarly exist with other measures of intelligence in both referred (Edwards & Paulin, 2007; Fletcher-Janzen, 2003) and normal samples (Kaufman et al., 2005). Research (Bell et al., 2002; Canivez et al., 2005; Law & Faison, 1996; Prewett & Matavich, 1994) has also provided evidence of strong correlations between the WISC-III FSIQ with other tests of intelligence.

Research has demonstrated moderate to strong correlations between previous versions of the Wechsler Scales and the Woodcock-Johnson in normal and neuropsychiatric samples. While one study reported that the WJCTA Broad Cognitive Scale score showed a correlation of .79 with the WISC-R FSIQ in a normal sample, Wechsler (1991) found that the WISC-R had a correlation of .65 with the WJ-R Broad Cognitive Ability score. Among special populations, Reeve, Hall, and Zakreski (1979) compared composite scores in a sample of children with reading and/or math learning disabilities. The correlation between the WJTCA Broad Cognitive Ability score and the WISC-R Full Scale IQ (FSIQ) score was .79. Ysseldyke, Shinn, and Epps (1981) found a correlation of .67 between the two measures in their sample of children diagnosed with a learning disability (LD). Thompson and Brassard (1984) reported correlations of .86, .74, and .93 for samples of normal, mild-to-moderate LD, and severe LD groups, respectively. A more recent study examined the concurrent validity of the WJ III Cognitive Factors with the WISC-III. According to McGrew and Woodcock (2001), research conducted with normal elementary school students in grades 3 through 5 indicated correlations of .71 and .76, respectively, between the WISC-III FSIQ and the WJ III COG GIA-Std. and GIA-Ext scores.

McGrew and Woodcock (2001) suggested that the relationship between the WISC-III and WJ III COG general intellectual ability scores may have been underestimated due to the restriction of ranges for the scores used in the Phelps normal study. However, when considering the changes made to the structure of the WISC-IV Full Scale IQ (FSIQ), it was hypothesized that there would be a correlation significantly greater than .76 between the WISC-IV FSIQ and the WJ COG GIA-Ext score.

Keith, Fine, Taub, Reynolds, and Kranzler (2006) found that, when compared to research conducted by Keith and Witta (1997) concerning the hierarchal structure of the WISC-III, the WISC-IV index scores had similar loadings on *g*. However, the subtests now comprising the WISC-IV FSIQ more accurately represent the constructs that comprise the measure. Subtests with high *g* loadings (i.e., Arithmetic and Information) were removed from the FSIQ score, while subtests with relatively low *g* loadings (i.e., the WMI and PSI subtests) now constitute 40% of the FSIQ (Kaufman et al., 2006). Moreover, it has been demonstrated that the WISC-IV provides better measurement of the five Cattell-Horn-Carroll (CHC) broad abilities measured by the test (see Keith et al, 2006).

More recent research (McGrew, 2010) has also demonstrated that, when compared to the WJ III COG, the WISC-IV shows similar to greater proportions of coverage of the five broad Cattell-Horn-Carroll (CHC) abilities found by Keith et al., (2006). When also considering that the WISC-IV FSIQ now demonstrates a more equal weighting of its indexes compared to its WISC-III predecessor, this information suggests that the WISC-IV and WJ III COG scores will show a higher correlation because they are theorized to show a greater convergence of *g* and CHC ability domains. Therefore, it was hypothesized that the WISC-IV FSIQ would show a correlation significantly greater than .76 with the WJ COG GIA-Ext Index.

Hypothesis Three

It was hypothesized that the correlation between the WISC-IV Verbal Comprehension Index (VCI) and WJ III COG Comprehension-Knowledge (*Gc*) factor would not be significantly different from .78. Research has shown that measures assessing verbal abilities tend to show moderate to high correlations with each other in both normal and referred populations [Bell et al., 2002; Brown & Morgan, 1991; Dumont, et al., 1996; Grados & Russo-Garcia, 1999; McIntosh & Dunham's study (as cited in McGrew & Woodcock, 2001)]. Both the WISC-IV and WJ COG contain subtests that measure similar narrow verbal abilities according to the Cattell-Horn-Carroll (CHC) theory (i.e., language development, lexical knowledge, and general information). Previous work (McGrew & Woodcock) has shown a large correlation of .78 between the WISC-IV VCI and WJ III COG *Gc* factor. While the structure of the WISC-IV VCI has changed with the removal of a subtest measuring the CHC narrow ability of general information, the WISC-IV VCI core subtests have been shown to have loadings of .74 or greater on the *Gc* factor (Keith, Fine, Taub, Reynolds, & Kranzler, 2006), suggesting that its correlation with the WJ III COG *Gc* factor will not be significantly different from that found for the WISC-III.

Research has shown that atheoretical measures of word knowledge and verbal concept formation demonstrate a marked degree of correlation with the WISC-III Vocabulary and Similarities subtests in children with learning disabilities (Dumont et al., 1996) and are good measures of comprehension-knowledge (*Gc*) (Cole & Randall, 2003; Sanders et al., 2007). Literature regarding the Wechsler and Woodcock-Johnson batteries (Carroll, 1993; Flannigan et al., 2000; Flannigan & Ortiz, 2001; Flannigan et al., 2007; McGrew & Flannigan, 1998; Woodcock, 1990) has suggested that these measures contain subtests assessing similar broad and narrow abilities.

The WISC-IV Verbal Comprehension Index (VCI) is considered to be a measure of comprehension-knowledge (*Gc*). The Similarities and Vocabulary subtests have been suggested to measure both language development and lexical knowledge, while the Comprehension subtest has been reported to assess language development and general information. These findings show that the WISC-IV VCI is similar to the WJ III COG *Gc* factor, which includes the Verbal Comprehension subtest as a measure of lexical knowledge and language development, and the General Information subtest, which measures general verbal knowledge (or general information). The Verbal Ability-Std score is a measure of language development, containing only Verbal Comprehension. This is in contrast to the Verbal Ability-Ext score, which includes both Verbal Comprehension and General Information. Confirmatory cross battery investigations have demonstrated that the WISC-III Information, Similarities, Vocabulary, and Comprehension subtests have loadings of greater than .60 on the broad G*c* factor (Phelps et al, 2005; Woodcock, 1990).

McGrew and Woodcock's (2001) report indicated a correlation of .78 between the WISC-III VCI and the WJ III COG *Gc* factor. This finding does not take into account the removal of the Information subtest from the overall VCI score. This subtest measures the narrow ability of general information, which is a narrow ability included on the WJ III Gc factor score.

While the WISC-IV Comprehension subtest has been suggested to be a strong measure of the narrow ability of general information, the subtest is factorially complex because it also provides measurement of language development (Flanagan et al., 2007). Furthermore, Keith, Fine, Taub, Reynolds, and Kranzler's (2006) factor analytic study of the WISC-IV showed that the VCI subtests had loadings of .74 or greater on the *Gc* factor, though the Information subtest was shown to have a stronger loading than the Comprehension subtest. This has also been found with research concerning the WISC-III (see Phelps et al., 2005).

Given the suggested factor loadings of the core WISC-IV VCI subtests and previous correlations between the WISC-III VCI with the WJ III COG *Gc* factor, it is likely that the WISC-IV VCI will continue to show a large correlation with the WJ III COG *Gc* factor, though it is not expected to be greater than that found for the WISC-III. As such, it was hypothesized that the correlation between the WISC-IV VCI and WJ III COG *Gc* factor would not be significantly different from that found between WISC-III and WJ III COG (i.e., .78).

Hypothesis Four

It was hypothesized that there would be a correlation significantly greater than .58 between the WISC-IV Working Memory Index (WMI) and the WJ III COG Short-Term Memory (*Gsm*) factor. The WISC-III and WJ III COG included indices of working memory that demonstrated a medium relationship (i.e., .58) with each other (McGrew & Woodcock, 2001). The WISC-IV now contains a composite measure of working memory that is theorized to be more consistent with that found on the WJ III COG. Specifically, the WISC-IV WMI has been suggested to measure the Cattell-Horn-Carroll (CHC) narrow abilities of working memory and memory span. As such, it appears that it will be more strongly related to the WJ III COG *Gsm* factor than its WISC-III counterpart.

Previous research (Woodcock, 1990) has revealed that the WISC-R Freedom from Distractibility Index (FDI) consisted of subtests that did not load on one common factor, with the Digit Span subtest being the only measure of short-term memory. While revisions were made to the WISC-III FDI, Digit Span continued to be the only valid indicator of short-term memory, having a loading of .70 with Working Memory (Phelps et al., 2005).

On the WISC-IV, the Freedom from Distractibility Index (FDI) was replaced by the Working Memory Index (WMI), which purports to be a more enhanced measure of working memory. Although the Arithmetic subtest was revised to include more age appropriate mathematical knowledge and increased demands for working memory (Wechsler, 2004), it is no longer included as a core subtest. Instead, the WMI retained the Digit Span subtest and incorporates an adapted version of the Letter-Number Sequencing subtest from the WAIS-III, which has been shown to have strong loadings on short-term memory (G*sm*) across distinct age ranges (Wechsler, 1997). Letter-Number Sequencing and Digit Span are hypothesized to be strong measures of the Cattell-Horn-Carroll (CHC) broad ability of short-term memory, with the former measuring the narrow ability of working memory and the latter measuring both working memory and memory span (Flanagan, 2001; Flanagan et al., 2000; Flanagan & Ortiz, 2001; Flanagan et al., 2007; McGrew & Flanagan, 1998). Research by Phelps, McGrew, Knopik, and Ford (2005) showed Digit Span and Letter-Number Sequencing to have moderate to strong loadings on *Gsm* (.65 and .74, respectively).

The WJ III COG Short-Term Memory (*Gsm*) factor contains measures of working memory (Numbers Reversed) and memory span (Memory for Words) (Flanagan et al., 2000; McGrew & Woodcock, 2001; Phelps et al., 2005). Both the WISC-IV WMI and WJ III COG *Gsm* factor incorporate tests that are suggested to assess ability to hold auditory-verbal information in immediate awareness and repeat back the information (memory span) or to recode the information (working memory). Both also exclude a valid measure of the visuospatial sketchpad, as according to Baddeley and Hitch's model of working memory (Leffard et al., 2006). As such, it is speculated that the correlation between the WISC-IV WMI and the WJ III COG *Gsm* factor should be even greater than that found by McGrew and Woodcock (2001). Although this research was exploratory in nature, it was hypothesized that there would be a correlation significantly greater than .58 between the WISC-IV WMI and the WJ III COG *Gsm* factor.

Hypothesis Five

It was hypothesized that the correlation between the WISC-IV Processing Speed Index (PSI) and the WJ III COG Processing Speed (*Gs*) factor would not be significantly greater than .59. Previous research has demonstrated that measures assessing mental and perceptual speed demonstrate high correlations with the WJ III COG *Gs* factor (Keith et al., 2001; Sanders et al., 2007). The WISC-III and WJ III COG included composite measures of processing speed that were shown to demonstrate a medium relationship (i.e., .59) with each other (McGrew & Woodcock, 2001). For the WISC-IV, the Cancellation subtest was added as a supplemental subtest and new ceiling items were added to the Symbol Search subtest. Yet, the core subtests of the WISC-IV PSI remained the same as its WISC-III predecessor. As such, it is expected that similar correlations will be found between the convergent constructs.

The Processing Speed Index (PSI) was not included as a discrete factor in the structure of the Wechsler Intelligence Scales for Children (WISC) until its third revision. Prior to that time, the only available measure of processing speed was the Coding subtest, which was included on the WISC-R Freedom from Distractibility Index and was demonstrated to load strongly on processing speed (*Gs)* (Woodcock, 1990). Beginning with the third revision of the WISC, both the Coding and Symbol Search subtests were included as measures contributing to the PSI. The Coding subtest has been hypothesized to measure the Cattell-Horn-Carroll (CHC) narrow ability of rate-of-test-taking, while the Symbol Search subtest has been hypothesized to measure the narrow abilities of perceptual speed and rate-of-test-taking (Flanagan, 2001; Flanagan et al., 2000; Flanagan & Ortiz, 2001; Flanagan et al., 2007; McGrew & Flanagan, 1998). The WJ III COG *Gs* factor includes measures of perceptual speed and rate-of-test taking (Visual Matching) and speed of reasoning (Decision Speed).

Both the Coding and Symbol Search subtests have been shown to have moderate to strong loadings, respectively, on the *Gs* factor (Phelps et al., 2005). Also, neither has been shown to be better explained by learning or delayed memory, or short-term memory (Keith et al., 2006; Lepach, Petermann, & Schmidt, 2008). As no changes were made to the structure of the PSI on the WISC-IV, it was hypothesized that the relationship between the WISC-IV PSI and WJ III COG *Gs* factor would not be significantly greater than .59.

Hypothesis Six

It was hypothesized that there would be a correlation significantly greater than .46 between the WISC-IV Perceptual Reasoning Index (PRI) and the WJ III COG Fluid Reasoning (*Gf*) factor. The WISC-IV is now considered to include measures of fluid reasoning similar to those observed on the WJ III COG. Moreover, the WISC-III contained a less well-defined measure of fluid reasoning that showed a moderate correlation (i.e. .46) with the WJ III COG *Gf* factor. This research was exploratory in nature because research regarding correlations between both the WISC-IV and the WJ III COG has not yet been conducted with a clinical sample of children.

Research has suggested that the WISC perceptual reasoning subtests are not pure measures of visual processing (*Gv*) (Phelps et al., 2005). The correlation between the WISC-III Perceptual Organization Index (POI) and the WJ III COG *Gv* factor was indicated to be .23, with the comparison complicated by the combination of subtests contained within the POI that measure both fluid reasoning and visual-spatial thinking abilities. However, the POI demonstrated a correlation of .46 with the WJ III COG *Gf* factor (McGrew & Woodcock, 2001).

The WISC-IV places even greater emphasis on fluid reasoning abilities because the subtests that represented these skills on the WISC-III (i.e., Picture Completion, Picture Arrangement, and Object Assembly) were replaced with the Matrix Reasoning and Picture Concepts subtests (Wechsler, 2004). These measures provide better measurement of fluid reasoning by placing more emphasis on nonverbal problem-solving and reasoning and less emphasis on processing speed, visualization, and crystallized abilities (Kaufman et al., 2006). The WISC-IV PRI retained the Block Design subtest, which has consistently been shown to have loadings on the *Gv* factor. The greater loadings on *Gf* can be attributed to the Matrix Reasoning and Picture Concepts subtests (Keith et al., 2006; Phelps et al., 2005; Wechsler, 1997). Block Design is considered to measure the narrow abilities of spatial relations and visualization. In contrast, both Matrix Reasoning and Picture Concepts measure the narrow *Gf* ability of induction, while the Matrix Reasoning subtest may also measure general sequential reasoning. This is similar to the WJ III COG *Gf* factor, which includes the subtests of Concept Formation and Analysis-Synthesis as measures of induction and general sequential reasoning, respectively (Flanagan, 2001; Flanagan et al., 2000; Flanagan & Ortiz, 2001; Flanagan et al., 2007; McGrew & Flanagan, 1998). These findings provide evidence for the hypothesis that there will be a correlation significantly greater than .46 between the WISC-IV PRI and the WJ III COG G*f* factor.

Hypothesis Seven

It was hypothesized that there would be a correlation significantly greater than .10 between the WISC-IV Verbal Comprehension Index (VCI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor. Though research concerning the WISC-III VCI and WJ III COG *Gv* factor demonstrated a negligible relationship between the composite scores, it has been suggested that visual-spatial skills and language ability share some common variance (Bell, Lassiter, Matthews, & Hutchinson, 2001).

The finding regarding the negligible relationship between the WISC-III VCI and WJ III COG *Gv* factor suggests that the WJ II COG subtests are pure measures of visual processing (*Gv*) that do not include components of *Gc*. However, the correlation found may have been truncated when considering that the sample study included restricted ranges of scores (McGrew

& Woodcock, 2001). This provides evidence for the hypothesis that the correlation between the WISC-IV VCI and WJ III COG *Gv* factor will be significantly greater than .10.

Other research concerning the relationship between verbal and visual-spatial thinking abilities has demonstrated a moderate relationship between the divergent constructs (see Ford, Teague, and Tusing Preschool Normal Sample, McGrew & Woodcock, 2001). This may be likely because both language development and visual-spatial processing abilities are related to math achievement (Flanagan et al, 2007). Moreover, research (Anjum, 2004) has shown that a measure of word knowledge demonstrated a moderate correlation with the WJ III COG Spatial Relations subtest. Because the Wechsler Verbal Comprehension Index (VCI) includes a measure of word knowledge (i.e., Vocabulary), this suggests that the divergent Wechsler and Woodcock-Johnson composite scores would be more strongly related than was found by research concerning the WISC-III.

Both the verbal and visual-spatial thinking constructs also appear to exhibit shared content variance. Specifically, while the WJ III COG Picture Recognition subtest involves short-term visual memory, the WISC-IV Vocabulary and Information subtests involve retrieval of stored declarative or semantic knowledge. As such, to some extent, both include a memory component. It is not thought the relationship between the WISC-IV Verbal Comprehension Index (VCI) and WJ III COG Visual-Spatial Thinking (*Gv*) factor will be as significant as would be found for convergent constructs. However, the above findings provide evidence for the hypothesis that there will be a correlation significantly greater than .10 between the WISC-IV VCI and WJ III COG *Gv* factor.

Hypothesis Eight

Based on the previously reviewed literature, it was hypothesized that there would be a correlation significantly greater than .10 between the WISC-IV Processing Speed Index (PSI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor. The WISC-III PSI has been found to be unrelated to visual-spatial abilities (*Gv*) as measured by the WJ III COG (correlation of .10) (McGrew & Woodcock, 2001). However, Keith, Fine, Taub, Reynolds, and Kranzler (2006) found that the WISC-IV Symbol Search subtest demonstrated a moderate loading on the *Gv* factor, suggesting that this subtest measures visual processing abilities.

The Wechsler PSI is comprised of the Symbol Search and Digit-Symbol Coding subtests. Completion of the former subtest involves attention to and processing of visual or figural stimuli in order to make decisions about matching target symbols. The same can also be said of the latter subtest because examinees are required to attend to and process visual stimuli in order to copy simple geometric shapes. When considering this, it would make it likely that the Wechsler subtests measuring processing speed would show some degree of relationship with measures of visual-spatial thinking.

Furthermore, results concerning the WISC-III and WJ III COG may have been truncated when considering that the sample study included restricted ranges of scores (McGrew & Woodcock, 2001). This is further substantiated when considering that other research (see Gregg and Hoy University Normal and Learning Disabled Sample, McIntosh and Dunham Grades 3 through 5 Normal Sample, respectively; McGrew & Woodcock, 2001) has demonstrated small (i.e., .29) to moderate (i.e., .40) correlations between composite measures of visual processing and processing speed. Likewise, Anjum (2004) found large correlations (i.e., greater than .59) between subtests comprising the Differential Ability Scales (DAS; Elliott, 1990) Spatial Ability Cluster (considered a measure of visual-spatial thinking) and the WJ III COG Processing Speed (*Gs*) factor. It is also thought that the divergent processing speed and visual-spatial thinking constructs would show a correlation greater than .10 because both perceptual speed and visual-spatial processing abilities are related to math achievement (Flanagan et al, 2007). As such, in contrast to that found for the WISC-III and WJ III COG, it was hypothesized that the correlation between the WISC-IV PSI and the WJ III COG *Gv* factor would be significantly greater than .10.

Hypothesis Nine

It was hypothesized that the correlation between the WISC-IV Working Memory Index (WMI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor would be significantly greater than .17. Research has shown that the WISC-III Freedom from Distractibility Index (FDI) demonstrated a correlation of .17 with the WJ III COG *Gv* factor (McGrew & Woodcock, 2001). While this provides support for the divergent validity of the WISC-III FDI, the WISC-IV WMI is different in structure from its WISC-III predecessor. Furthermore, the relationship between the WISC-III FDI and WJ III COG *Gv* factor may have been underestimated because the sample study included restricted ranges of scores for both measures.

The finding between the WISC-III and WJ COG III does not take into account that the WISC-IV Working Memory Index (WMI) places even greater emphasis on aspects of working memory with the inclusion of the Letter-Number Sequencing subtest. Research by Leffard, Miller, Bernstein, DeMann, Mangis, and McCoy (2006) has suggested that this subtest does not involve the visual-spatial sketch pad subsystem of Baddeley's model of working memory, with this subsystem being responsible for storing and manipulating visual and spatial information

(Baddeley, 1996). Likewise, other research (Flanagan et al., 2000; McGrew & Flanagan, 1998) has shown Letter-Number Sequencing to load appropriately on short-term memory. However, it has been speculated that performance on Letter-Number Sequencing requires visuospatial imaging (i.e., visuospatial manipulation of the sequence of letters and numbers in immediate awareness), suggesting that the WISC-IV WMI may show greater a correlation with the WJ III COG *Gv* factor than was found for the WISC-III short-term memory composite score. Furthermore, other research (see Phelps and Ford Preschool Normal Sample Study, McGrew & Woodcock, 2001) demonstrated a moderate relationship (i.e., .47) between a measure of short-term memory and the WJ III Visual-Spatial Thinking (*Gv*) factor in preschool children.

Support for the correlation between the WISC-IV Working Memory Index (WMI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor is also provided when considering that tasks of visual-spatial thinking include working memory components. The WJ III COG Picture Recognition subtest appears to demonstrate shared variance because it includes a working memory demand. Specifically, completion of the subtest requires that a presented picture be help in mind in order to identify the visual stimuli when comparing it to its stored representation. Overall, the relationship between the divergent constructs of short-term memory and visual-spatial thinking abilities is not as strong as would be expected between convergent constructs. However, given the above findings, it is likely that the WISC-IV WMI will show a greater correlation with the WJ III COG *Gv* factor than that found for the WISC-III. As such, it was hypothesized that there would be a correlation significantly greater than .17 between the WISC-IV WMI and WJ III COG *Gv* factor.

Hypothesis Ten

It was hypothesized that there would be a correlation significantly greater than .19 between the WISC-IV Perceptual Reasoning Index (PRI) and the WJ III COG Auditory Processing (Ga) factor. The Wechsler structure does not provide a valid measure of auditory processing (*Ga*) (Flanagan, 2000; Flanagan et al., 2007). Previous research (McGrew & Woodcock, 2001) has shown that the WISC-III Perceptual Organization Index (POI) is unrelated to this ability as measured by the WJ III COG *Ga* factor, having a correlation of .19. However, the correlation may have been truncated due to the restricted ranges of scores used. Likewise, the WISC-IV PRI is different in structure from the WISC-III POI.

The WISC-IV Perceptual Reasoning Index (PRI) has been shown to demonstrate loadings on both the WJ III COG Fluid Reasoning (*Gf*) and Visual-Spatial Thinking (*Gv*) factors (Keith et al., 2006), suggesting that the subtests included do not assess auditory processing abilities. However, because the WISC-IV PRI now includes a greater component of fluid reasoning abilities, it may likely show a stronger relationship with the WJ III COG Auditory Processing (*Ga*) factor. This is because both constructs are thought to include subtests reflecting "thinking abilities". On the WJ III COG, the Cattell-Horn-Carroll (CHC) broad abilities of long-term retrieval (*Glr*), visual-spatial thinking (*Gv*), auditory processing (*Ga*), and fluid reasoning (*Gf*) are included on the composite measure of Thinking Ability. This composite measure reflects "process" dominant abilities that constrain new learning because information placed in short-term memory cannot be processed automatically (McGrew, 2009b). Because the WISC-IV PRI has been shown to load on the WJ III COG *Gf* factor, it is speculated that the correlation between the WISC-IV PRI and the WJ III COG *Ga* factor should be even greater than that found by McGrew and Woodcock (2001).

In contrast to findings involving the WISC-III, other research (see McIntosh and Dunham Grades 3 through 5 Normal Sample, McGrew & Woodcock, 2001) has shown that measures of fluid reasoning (*Gf*) and auditory processing (*Ga*) abilities are moderately (i.e., .41) correlated. This may be because fluid reasoning abilities are highly related to general intelligence (*g*). With the WISC-IV Perceptual Reasoning Index (PRI) now measuring a greater proportion of fluid reasoning (*Gf*) abilities as compared to its WISC-III counterpart, it is likely that it will show a stronger degree of relationship with the WJ III COG Auditory Processing (*Ga*) factor than was previously found. This is likely because of the higher *g* loadings of the subtests comprising the index. Also, both fluid reasoning and auditory processing abilities are both significantly related to reading achievement (Flanagan et al., 2007), suggesting some degree of relationship between the constructs. Overall, the above evidence provides support for the hypothesis that the correlation between the WISC-IV PRI and the WJ III COG *Ga* factor will be significantly greater than.19.

Chapter III: Method

Participants

This study was conducted using archival data consisting of patient records at Nova Southeastern University's Neuropsychology Assessment Center. Test scores and demographic data were collected from children and adolescents referred for a comprehensive neuropsychological evaluation of learning, attention, psychological, and/or behavior problems. Each participant's parent provided consent for the results of their evaluation to be utilized anonymously in research. Participants were administered a comprehensive battery of neuropsychological tests that included a measure of general intellectual functioning, cognitive ability, memory, achievement, personality/emotional functioning, and attention. Assessment was conducted over 20 to 25 hours within a two month period by clinical psychology graduate students trained in the standard administration of the measures. Students were expected to have completed coursework in the administration, scoring, and interpretation of the Wechsler Intelligence Scales. Moreover, students were expected to complete further supervised training and pass specific competencies in the administration and scoring of individual intelligence tests prior to participating in practicum assessment experiences. Because of the referral nature of the testing, children were not administered tests in a counterbalanced order. At the conclusion of the comprehensive neuropsychological evaluation, the children in the sample were determined to have met diagnostic criteria for one or more Diagnostic and Statistical Manual of Mental Disorders Text Revision- Fourth Edition (DSM–IV–TR; American Psychiatric Association, 2000) as determined by a doctoral student in clinical psychology under the supervision of a licensed psychologist board certified in clinical neuropsychology. For the purposes of the present research, only variables from the Wechsler Intelligence Scale for Children- Fourth Edition (WISC-IV) and the Woodcock-Johnson III Tests of Cognitive Abilities (WJ III COG) were selected for data analysis.

Selection criteria for the present study included one administration of both the WISC-IV and WJ III COG and the diagnosis of either a neurological or psychological disorder, or both. Exclusion criteria for the present study included incomplete or missing scores on the core WISC-IV and WJ III COG subtests or indexes, hearing or visual impairments, and chronological age less than 6 years, 0 months or greater than 16 years, 11 months. No data were excluded on the basis of race, gender, education level, ethnicity, race, religion or socioeconomic status. Also, common comorbid disorders (e.g. mood disorders, ADHD, learning disabilities, adjustment disorder) and overall IQ scores were allowed to covary naturally and did not serve as exclusionary criteria. Because of the retrospective nature of the study, information about language acquisition was not known for all participants. Thus, this did not serve as exclusionary criteria.

Descriptive Statistics

Two-hundred thirty-nine participants were identified from an archival database as having completed the WISC-IV and WJ III COG during their evaluation. Thirty-five subjects were initially excluded from the data analysis due to failure to meet diagnostic criteria. One subject was excluded due to failure to meet age criteria of 6 years to 16 years, 11 months. An additional 112 subjects were excluded due to incomplete data on one or more core subtests or standard indexes. The final sample ($N = 92$) included clinic-referred children with either a neurological or psychological disorder, or both, who all met study criteria. Demographic data for this final sample were explored. The mean age of children was 9.82 years (*SD* = 2.81) with a mean grade level of 3.95 (*SD* = 2.63). Sixty-one subjects (66.3%) were boys and 31 subjects (33.7%) were girls. A little over 93% of the sample were right-handed. In terms of ethnicity, 60.4% were Caucasian, 6.6% African-American, 17.6% Hispanic, and 15.4% who identified themselves as representing another ethnicity, such as Indian or Asian. Table 2 presents descriptive statistics on demographic variables for the total sample.

Table 2
Descriptive Statistics for Demographic Variables (N = 92)

Variable	Mean or Percent	Standard Deviation
Age (years)	9.82	2.81
Education (grade)	3.95	2.63
<u>Gender</u>		
Male	66.3%	
Female	33.7%	
<u>Race</u>		
Caucasian	60.4%	
African American	6.6%	
Hispanic	17.6%	
Other	15.4%	
<u>Handedness</u>		
Right	93.3%	
Left	6.7%	

In terms of clinical diagnoses, forty children (43.5%) were diagnosed with psychiatric disorders, 24 children (36.1%) were diagnosed with neurological disorders, and 28 children (30.4%) met diagnostic criteria for both a psychiatric and neurological disorder. The most frequently occurring diagnosis in the sample was Adjustment Disorder, with 28.3% of the sample meeting diagnostic criteria. Five of the 26 children diagnosed with an adjustment disorder were also diagnosed with a learning disorder. Furthermore, 18.5% of the sample met diagnostic criteria for Major Depressive Disorder, 14.1% met diagnostic criteria for a Reading Disorder, and 13% met diagnostic criteria for Borderline Intellectual Functioning. Two of the 17 children diagnosed with Major Depressive Disorder were also diagnosed with a learning disorder, while none of the children diagnosed with Borderline Intellectual Functioning were diagnosed with a learning disorder. Additionally, 10.9% of the sample met diagnostic criteria for Disorder of Written Expression, while 10.9% of the sample met diagnostic criteria for Generalized Anxiety Disorder. These diagnoses are reported in Table 3.

Table 3

Diagnoses Represented in the Sample (N = 92)

Diagnosis	Frequency	Percent
Adjustment Disorder	26	28.3
Major Depressive Disorder	17	18.5
Reading Disorder	13	14.1
Borderline Intellectual Functioning	12	13.0
Disorder of Written Expression	10	10.9
ADHD Combined Type	8	8.7
Cognitive Disorder NOS	6	6.5
Oppositional Defiant Disorder	6	6.5
ADHD Inattentive Type	5	5.4
Expressive Language Disorder	5	5.4
Mild Mental Retardation	4	4.3
Dysthymia	4	4.3
Enuresis	2	2.2
ADHD Hyperactive-Impulsive Type	2	2.2
ADHD NOS	2	2.2
Post-Traumatic Stress Disorder	2	2.2
Intermittent Explosive Disorder	1	1.1
Mood Disorder NOS	1	1.1
Disorder of Mathematics	1	1.1
Learning Disorder NOS	1	1.1
Autism	1	1.1
Asperger’s Disorder	1	1.1
Encopresis	1	1.1
Brain Injury	1	1.1

Note. ADHD = Attention-Deficit/Hyperactivity Disorder; LD = Learning Disorder; NOS = Not Otherwise Specified

Measures

Measures selected for the study included the Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) and the Woodcock-Johnson III Tests of Cognitive Abilities (WJ III COG). Both are assessments of intellectual functioning and/or cognitive abilities. The measures were selected based upon their reliability and validity. As reported in the WISC-IV manual, the internal consistency coefficients for the WISC-IV's Full Scale score and composite indexes are generally high (> .88+) and the measure demonstrates average test-retest coefficients ranging from .86 to .93. Likewise, the manual provides evidence of the structural validity of the measure as supported by factor-analytic studies (Wechsler, 2004). The median reliabilities for the WJ III COG clusters are generally .90 or higher and the median retest reliability across all reliability coefficients listed was .94 (McGrew & Woodcock, 2001). Furthermore, the content validity of the battery was addressed by making revisions according to the Cattell-Horn-Carroll (CHC) framework. Developmental growth curves analyses provide further validity evidence regarding the unique abilities measured by the battery, while confirmatory factor-analytic research demonstrates the measure's consistency with CHC theory (Schrank, McGrew, & Woodcock, 2001).

The measures selected yield composite and index/factor scores reported in standard scores, with a mean of 100 and a standard deviation of 15. While the WJ III COG also yields subtest scores reported in standard scores, the WISC-IV yields subtest scores reported in scaled scores, with a mean of 10 and a standard deviation of 3.

Intellectual functioning. The Wechsler Intelligence Scale for Children- Fourth Edition (WISC-IV), a revised version of the WISC-III, is an individually administered, norm-referenced test commonly used to measure general intellectual ability. It was standardized on a sample of 2,200 children ages 6 years to 16 years, 11 months, closely approximating the 2000 U.S. Census on such demographic variables as gender, race, parent education level, and geographic region. The WISC-IV is comprised of 10 core and 5 supplemental subtests that measure different components of intelligence. The core subtests are combined to yield the four factor indices: Verbal Comprehension (VCI), Perceptual Reasoning (PRI), Working Memory (WMI), and Processing Speed (PSI) (Wechsler, 2004).

Extensive reliability and validity evidence were provided by Wechsler (2003b), as well as by Prifitera, Saklofske, and Weiss (2005). The WISC-IV manual provides evidence that the reliability coefficients for the measure's composite scales are .88 or higher and are identical to or slightly better than the corresponding scales in the WISC-III. The corrected correlation coefficients between the WISC-IV and WISC-III ranged from .72 (i.e., WISC-IV WMI -WISC-III FDI) to .89, as shown between the WISC-IV FSIQ and WISC-III FSIQ. Correlations between the WISC-IV composite scales and other Wechsler based measures provides evidence of the convergent and discriminant validity of the measure.

The VCI is derived from subtests that evaluate word knowledge, verbal reasoning, and knowledge of conventional rules and concepts: Vocabulary, Similarities, and Comprehension, respectively. The PRI, a measure of visuoconstruction, nonverbal problem solving, and visual/spatial abilities, is comprised of three subtests: Block Design, Picture Concepts, and Matrix Reasoning. The WMI reflects short-term auditory memory and mental manipulation. The WMI includes two subtests: Digit Span and Letter-Number Sequencing. The PSI is comprised of two subtests: Coding and Symbol Search. The PSI is a measure of intellectual fluency and speed of processing. The WISC-IV also generates a Full Scale Intelligent Quotient (FSIQ) score that reflects overall intellectual functioning. The ten core subtests that comprise the four indices are combined to derive the FSIQ.

The Woodcock-Johnson III Tests of Cognitive Abilities (WJ III COG), an updated and expanded version of the Woodcock-Johnson Psychoeducational Battery-Revised Tests of Cognitive Ability (WJ-R COG), is an individually administered, co-normed battery commonly used as a measure of general intellectual ability and specific cognitive abilities. It was standardized on a sample of 8,818 subjects aged 24 months to age 90 years and older, closely approximating the 2000 U.S. Census on variables including gender, race, parent education level, and geographic region (Schrank, McGrew, & Woodcock, 2001).

The WJ III COG consists of twenty individual subtests, with each measuring a specific primary narrow factor of cognitive ability. These subtests are divided into the Standard Battery (seven standard and three supplemental subtests) and the Extended Battery (10 additional tests). Seven broad CHC cognitive abilities are measured through the combination of two or more individual subtests, including: Comprehension-Knowledge (*Gc),* Long-Term Retrieval (*Glr*), Visual-Spatial Thinking (*Gv*), Auditory Processing (*Ga*), Fluid Reasoning (*Gf*), Processing Speed (*Gs*), and Short-Term Memory (*Gsm*) (Mather & Woodcock, 2001). Technical and independent reviews provide extensive reliability and validity evidence (Cizek, 2003; Mather & Woodcock; 2001; Sandoval, 2003; Schrank et al., 2001). Median reliability coefficients for the factor scores range from .80 [Visual-Spatial Thinking (*Gv*)] to .95 [i.e. Fluid Reasoning (*Gf*)] for ages 5 to 19. For the GIA, the median reliability coefficient alphas for all age groups within the standard battery ranges from .81 to .94. Median coefficients for the Extended battery ranges from .74 to .97 (Mather & Woodcock, 2001).

The Comprehension-Knowledge (*Gc*) domain assesses facets of crystallized intelligence, namely, verbal expression, language development, and general knowledge. It is composed of two subtests: Verbal Comprehension and General Information. The domain of Visual-Spatial Thinking (*Gv*) includes the Spatial Relations and Picture Recognition subtests. This domain assesses the ability to perceive, analyze, synthesize, and think with visual patterns. Auditory Processing (*Ga*) measures aspects of auditory perception, including phonological awareness, acoustic-phonetic processing, and speech-sound discrimination. The subtests within this domain include Sound Blending and Auditory Attention. The Fluid Reasoning (*Gf*) domain includes the Concept Formation and Analysis-Synthesis subtests, and assesses abilities in reasoning, forming concepts, and solving problems using novel information. Processing Speed (*Gs*) examines a subject's efficiency with automatic, cognitive processing under timed conditions. This domain includes Visual Matching, Decision Speed, Rapid Picture Naming, and Pair Cancellation. The Short-Term Memory (*Gsm*) domain, including Numbers Reversed and Memory for Words, assesses the ability to apprehend and hold information in immediate awareness, which is retained to perform a new task. Long-Term Retrieval (*Glr*) measures abilities in memory consolidation, including the ability to acquire, store, and later retrieve information. Visual-Auditory Learning and Retrieval Fluency are included in this domain (Mather & Woodcock, 2001; Shrank, 2006).

The WJ III COG has both a standard and extended version, with each version providing a General Intellectual Ability (GIA) score. The entire WJIII battery of tests is defined by three causally related categories of cognitive performance that are intended to be measures of information processing abilities. These include verbal ability, thinking abilities (or abilities that, although are not processed automatically, depend on short-term memory for processing, such as long-term retrieval, visual-spatial thinking, auditory processing, and fluid reasoning abilities), and cognitive efficiency (i.e., processing speed, short-term memory).

Procedure

Data collection. The data for this study was extracted from a de-identified archival database of children and adolescents clinically referred to the Neuropsychology Assessment Center at Nova Southeastern University. All testing was administered by clinical psychology practicum students enrolled in a doctoral graduate program and who were trained in administration and scoring of standardized psychological test instruments. All students were under the supervision of a licensed, board certified, clinical neuropsychologist. All practicum students completed the Nova Southeastern University CITI Course in the Protection of Human Subjects. Data for all participants were collected following administration of a battery of measures as part of a neuropsychological evaluation, with tests administered in no particular order. Only selected measures as described above were included in the analysis. In addition to relevant test scores, demographic data including age, education, gender, race, diagnosis, and handedness was collected for the entire sample to provide descriptive information.

Institutional Review Board requirements. Before analyses of the data were conducted, approval was obtained to conduct archival research on the clinical sample from the Institutional Review Board (IRB) at Nova Southeastern University. In keeping with the requirements of the IRB, the data was de-identified to maintain strict confidentiality.

Chapter IV: Results

Preliminary Analyses

Data analyses were conducted using the Predictive Analytics Software (PASW) Statistics 18. Prior to data analyses, study variables were scanned for accuracy of data entry and missing values through examination of descriptive statistics, examining the data for outliers, and checking the accuracy of the scores against actual data. Study variables were evaluated to determine if their distributions met assumptions for the proposed statistical procedures to be employed, including tests of departures from normality, presence of significant outliers, linearity, and homoscedasticity. The mean score and standard deviations for all variables (Table 4) were analyzed and appeared consistent with the performance of a clinical population. Mean scores were generally in the average range. Performance on the variables used for the current study were generally significantly below the normative mean of 100, with the exception of the WISC-IV Perceptual Reasoning Index and the WJ III COG clusters of Visual-Spatial Thinking, Auditory Processing, and Fluid Reasoning.

Table 4
Descriptive Statistics (N = 92)

Variable	Mean	Standard Deviation
WISC-IV Full Scale IQ	93.30	16.35
WISC-IV Verbal Comprehension Index	94.79	15.43
WISC-IV Perceptual Reasoning Index	100.12	14.86
WISC-IV Working Memory Index	93.14	15.27
WISC-IV Processing Speed Index	89.79	13.62
WJ III COG GIA-Ext	95.46	15.00
WJ III COG Verbal Ability-Ext	93.58	14.62
WJ III COG Comprehension-Knowledge	93.58	14.62
WJ III COG Visual-Spatial Thinking	103.16	11.93
WJ III COG Auditory Processing	100.97	15.41
WJ III COG Fluid Reasoning	101.29	15.72
WJ III COG Processing Speed	93.10	15.94
WJ III COG Short-Term Memory	95.16	15.75

Note. WISC-IV = Wechsler Intelligence Scale for Children- Fourth Edition; WJ III COG = Woodcock-Johnson III Tests of Cognitive Abilities; GIA = General Intellectual Ability; Ext = Extended

Normality. The independent variables were analyzed for departures from normality using statistical methods (Table 5). The Kolmogorov–Smirnov test and tests of skewness and kurtosis (+/- 1) were reviewed to identify departures from normality. The WISC-IV Perceptual Reasoning Index, Working Memory Index, Processing Speed Index, and the WJ III COG General Intellectual Ability-Extended scores demonstrated a significant departure from normality based on the Kolmogorov–Smirnov test ($p < .05$). Also, inspection of the histogram and normal probability plot for each score suggested that the sample scores were reasonably normally distributed. All scores met the assumption of normality based on skewness and kurtosis.

Table 5
Tests of Normality (N = 92)

Variable	K-S	Skewness	Kurtosis
WISC-IV Full Scale IQ	.20	-.30	.11
WISC-IV Verbal Comprehension Index	.17	-.30	.26
WISC-IV Perceptual Reasoning Index	.04*	.45	.66
WISC-IV Working Memory Index	.02*	-.12	.44
WISC-IV Processing Speed Index	.00*	.04	-.42
WJ III COG GIA-Ext	.04*	-.05	.01
WJ III COG Verbal Ability-Ext	.14	-.58	.81
WJ III COG Comprehension-Knowledge	.14	-.58	.81
WJ III COG Visual-Spatial Thinking	.20	.31	.53
WJ III COG Auditory Processing	.20	.03	-.25
WJ III COG Fluid Reasoning	.20	.24	.36
WJ III COG Processing Speed	.20	-.38	.19
WJ III COG Short-Term Memory	.20	-.11	.01

Note. K-S = Kolmogorov-Smirnov test; WISC-IV = Wechsler Intelligence Scale for Children- Fourth Edition; WJ III COG = Woodcock-Johnson III Tests of Cognitive Abilities; GIA = General Intellectual Ability; Ext = Extended
* Statistically significant at $p < .05$

Outliers. Data points were analyzed for significant departures from the sample means. The cutoff value for extreme outliers was set to ±3 standard deviations. Outliers were also examined by inspecting the histogram, scatterplot, boxplot, and the 5% Trimmed Mean for each score. No extreme outliers exceeding the cutoff were found.

Linearity and homoscedasticity. Linearity was assessed through examination of the normal probability plot and scatterplot. For all independent variables, the normal probability plot evidenced a reasonably straight line from bottom left to top right. In the scatterplot for correlations, scores for independent variables measuring similar constructs showed a roughly eliptical distribution, with data showing an upward trend. Scores for independent variables measuring dissimilar constructs also showed a roughly circular shaped distribution. Regarding homoscedasticity, for all variables, there appeared to be equal variability in y across the different scores of x.

Study Analyses

Hypothesis one. It was hypothesized that neuropsychiatric subjects (i.e., children diagnosed with either a neurological disorder or psychological disorder, or both) would obtain statistically significant higher mean Full Scale IQ scores on the WISC-IV than General Intellectual Ability-Extended (GIA-Ext) scores on the WJ III COG.

A one-tailed paired samples *t*-test was conducted to compare the mean WISC-IV FSIQ score to the mean WJ III COG GIA-Ext score. In contrast to the expected difference, results revealed that the mean WISC-IV FSIQ score was significantly lower than the mean WJ III COG GIA-Ext score, $t(91) = -2.04$, $p = .04$. The standardized mean effect size of the difference between the scores was medium (Cohen's $d = .59$).

Hypothesis two. It was hypothesized that there would be a correlation significantly greater than .76 between the WISC-IV Full Scale IQ (FSIQ) and the WJ III COG General Intellectual Ability-Extended (GIA-Ext) factor. A Pearson product-moment correlation was computed to examine the relationship between the composite scores, with results of a one-tailed test yielding, as per Cohen (1988), a large, significant correlation between the composite pairs, $r(92) = .87$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value under the null hypothesis. Results supported the hypothesis, indicating that the observed correlation differed significantly from the hypothesized value of .76 ($z = 3.18$, $p < .001$).

Hypothesis three. It was hypothesized that the correlation between the WISC-IV Verbal Comprehension Index (VCI) and the WJ III COG Comprehension-Knowledge (*Gc*) factor would not be significantly different from .78. A Pearson product-moment correlation was computed to examine the relationship between the composite scores. Results of a one-tailed test yielded a large, significant correlation, $r(92) = .72$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value under the null hypothesis. Results supported the hypothesis. It was found that the observed correlation was not significantly different from the hypothesized value of .78 ($z = 1.30$, $p = .10$).

Hypothesis four. It was hypothesized that there would be a correlation significantly greater than .58 between the WISC-IV Working Memory Index (WMI) and the WJ III COG Short-Term Memory (*Gsm*) factor. A Pearson product-moment correlation was used to examine the relationship between the composite scores. Results of a one-tailed test revealed a large, significant correlation, $r(92) = .72$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value of .58 under the null hypothesis. Results supported the hypothesis, indicating that the observed correlation differed significantly from the hypothesized value ($z = 2.02$, $p = .02$).

Hypothesis five. It was hypothesized that the correlation between the WISC-IV Processing Speed Index (PSI) and the WJ III COG Processing Speed (*Gs*) factor would not be significantly greater than .59. Results of a one-tailed Pearson product-moment correlation yielded a large, significant correlation, $r(92) = .60$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value under the null hypothesis. Results supported the hypothesis. It was found that the observed correlation was not significantly greater than the hypothesized value of .59 (z = .14, p = .44).

Hypothesis six. It was hypothesized that there would be a correlation significantly greater than .46 between the WISC-IV Perceptual Reasoning Index (PRI) and the WJ III COG Fluid Reasoning (*Gf*) factor. A Pearson product-moment correlation was computed to examine the relationship between the composite scores, with results of a one-tailed test yielding a large, significant correlation between the WISC-IV PRI and the WJ III COG *Gf* factor, $r(92) = .68$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value under the null hypothesis. Results indicated that the observed correlation differed significantly from the hypothesized value of .46 ($z = 3.13$, $p < .001$).

Hypothesis seven. It was hypothesized that the correlation between the WISC-IV Verbal Comprehension Index (VCI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor would be significantly greater than .10. Results of a one-tailed Pearson product-moment correlation yielded a large, significant correlation, $r(92) = .54$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value under the null hypothesis. Results supported the hypothesis. It was found that the observed correlation was significantly greater than the hypothesized value of .10 ($z = 4.75$, $p < .001$).

Hypothesis eight. It was hypothesized that there would be a correlation significantly greater than .10 between the WISC-IV Processing Speed Index (PSI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor. A Pearson product-moment correlation was computed to examine the relationship between the composite scores, with results of a one-tailed test yielding a medium, significant correlation between the divergent composite scores, $r(92) = .38$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value under the null hypothesis. Results indicated that the observed correlation differed significantly from the hypothesized value of .10 ($z = 2.83$, $p = .002$).

Hypothesis nine. It was hypothesized that the correlation between the WISC-IV Working Memory Index (WMI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor would be significantly greater than .17. Results of a one-tailed Pearson product-moment correlation yielded a medium, significant correlation, $r(92) = .42$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value under the null hypothesis. Results supported the hypothesis. It was found that the observed correlation was significantly greater than the hypothesized value of .17 ($z = 2.60$, $p = .005$).

Hypothesis ten. It was hypothesized that there would be a correlation significantly greater than .19 between the WISC-IV Perceptual Reasoning Index (PRI) and the WJ III Auditory Processing (*Ga*) factor. A Pearson product-moment correlation was used to examine the relationship between the composite scores. Results of a one-tailed test revealed a medium, significant correlation between the composite pairs, $r(92) = .49$, $p < .001$. Fisher's r to Z transformation was used to determine if there was a statistically significant difference between the observed correlation and the stipulated value of .19 under the null hypothesis. Results supported the hypothesis, indicating that the observed correlation differed significantly from the hypothesized null value ($z = 3.24$, $p < .001$).

Chapter V: Discussion

Limited research is available regarding the WISC-IV's correlational relationship with other measures of intellectual functioning outside of the Wechsler domain (e.g., Edwards & Paulin, 2007). The present investigation sought to explore the construct validity of the WISC-IV within a sample of clinic-referred children by examining its relationship with the Woodcock-Johnson III Tests of Cognitive Abilities (WJ III COG). This study examined the comparability of the mean WISC-IV and WJ III COG general intellectual ability scores, as well as examined the pattern of convergent and discriminate validity correlations between the two measures. Results will be discussed in detail below.

Hypothesis One

It was hypothesized that neuropsychiatric subjects would obtain significantly higher mean WISC-IV Full Scale IQ scores than WJ III COG General Intellectual Ability-Ext scores. Results revealed findings in the opposite direction and did not support the hypothesis.

Past research (e.g., Bracken et al., 1984; Reeve, Hall, & Zakreski, 1979; Thompson & Brassard, 1984; Ysseldyke, Shinn & Epps, 1981) has generally shown that on previous versions of the WISC and Woodcock-Johnson batteries, the Wechsler full scale intelligence score was significantly higher than that derived by the Woodcock-Johnson measure in both referred and normal samples. Differences in performance between the WISC-R and WJCTA were suggested to be a function of the inclusion of skills on the Woodcock-Johnson not assessed by the Wechsler measure. These differences were attributed to the skills measured by the WJCTA and their sensitivity to the deficits among learning disabled children, as well as the significant correlation of the measure with academic achievement.

Revisions made to both the Wechsler and Woodcock-Johnson measures have increased the comparability of cognitive abilities assessed by each battery, suggesting less discrepancy between the general intellectual composite scores. However, more recently, in the Phelps validity research conducted with typically developing elementary school students in grades 3 through 5 (McGrew & Woodcock, 2001), scores on the WISC-III FSIQ were higher than scores obtained on the WJ III COG GIA-Ext. It is important to note that the study utilized special research standard scores and the magnitude of difference between the scores was not reported. Overall though, results of the current study represent a departure from the previous literature and have significant implications for clinicians regarding neuropsychological and psychoeducational assessment.

Current findings demonstrated that the mean WISC-IV FSIQ score was significantly lower than the mean WJ III COG GIA-Ext score. Although test scores can often vary as a function of age groups (Strauss, Spreen, & Hunter, 2000), exploratory *post hoc* analysis results indicated that there was no significant correlation between age and the size of the WISC-IV and WJ III COG difference score, $r(92) = .12$, $p = .27$, failing to account for the difference found between the scores. The hypothesis regarding the discrepancy between the scores was based on previous research findings, with the assumption that while changes have been made to the overall structure of the Wechsler measure, these changes made would not result in a composite measure that was lower than that obtained by the WJ III COG. Instead, it was expected that the mean WISC-IV FSIQ score would continue to be higher, although with less of a difference between the mean scores. Current results suggest that the changes made to the overall structure of the WISC-IV have resulted in a composite index that is remarkably different from its predecessor. As such, this hypothesis was generated without fully examining, a priori, how the specific changes made to the WISC-IV Full Scale IQ would impact findings.

The WISC-III FSIQ was a composite measure of the 10 subtests contributing to the Verbal and Performance IQ scores, creating an unequal weighting of the four constructs measured by the battery. Specifically, the Verbal IQ composite included the four Verbal Comprehension Index (VCI) subtests (Information, Similarities, Vocabulary, and Comprehension) and one subtest from the Freedom from Distractibility Index, Arithmetic, which is a highly *g*-loaded subtest (Kaufman et al, 2006) and is considered to better represent the construct of quantitative knowledge than that of short-term memory. The Performance IQ composite was comprised of the four Perceptual Organization Index (POI) subtests (Picture Completion, Picture Arrangement, Block Design, and Object Assembly) and the Coding subtest from the Processing Speed Index (PSI). Though the POI subtests provided information about processing speed, with a deficit in this area lowering the old Performance IQ, they did not allow for a full appreciation of how slowed or impaired visuomotor performance contributed to a child's overall functioning (Baron, 2005).

Because of the composition of the WISC-III FSIQ score, it received little contribution from the construct of short-term (or working) memory, it provided a mixed understanding regarding psychomotor speed, and it was heavily influenced by crystallized knowledge. This resulted in a full scale composite score that was biased in regards to its representation of a child's cognitive abilities, was sensitive to differences in ethnicity, and perhaps resulted in higher scores than the WJ III COG because of its unequal and limited representation of Cattell-Horn-Carroll theory (CHC) broad abilities and significant influence of crystallized knowledge.

In contrast to the WJ III COG GIA score, which gives differential general intelligence (*g*) weighting to the subtests contributing to the overall score, the Wechsler measure weights all subtests equally (Woodcock & McGrew, 2001). However, as suggested above, the WISC-IV FSIQ gives a more equal weighting to the four indexes that comprise the battery. While the score retains five subtests that were included in the WISC-III FSIQ, specifically, Similarities, Comprehension, Vocabulary, Block Design, and Coding, there are five new subtests. These include two new Perceptual Reasoning Index (PRI) subtests (Matrix Reasoning and Picture Concepts), two Working Memory Index (WMI) subtests (Digit Span and Letter-Number Sequencing), and the Symbol Search subtest from the Processing Speed Index (PSI).

Though still not fully equal in its weighting, the composition of the WISC-IV Full Scale IQ score suggests that it now provides better representation of the constructs of working memory and processing speed, which contribute 20 percent each to the overall score. Also, the composite score has a reduced focus on crystallized knowledge with removal of the Arithmetic and Information subtests and provides improved measurement of the construct of fluid reasoning through the addition of the Matrix Reasoning and Picture Concepts subtests. The addition of these latter subtests also improved interpretations regarding the influence of speeded performance and motor skill to the FSIQ score because it removed the influence of construct-related variance. While the Block Design subtest was retained, which includes motor-dexterity under timed conditions, this subtest is better viewed as a measure of visual-spatial processing. The Matrix Reasoning subtest, in contrast, can be considered a measure of nonverbal fluid reasoning ability, involving perceptual reasoning and matching, attention to detail, concentration, classification, analogic reasoning, and serial reasoning for successful performance. The Picture

Concepts subtest appears to measure abstract, categorical reasoning based on perceptual recognition processes.

Overall, the WISC-IV appears to provide improved measurement of CHC theory broad abilities, which is also likely better reflected in the Full Scale IQ score. Exploratory paired sample *t*-tests were utilized to compare the mean index scores of the WISC-IV with convergent factor scores from the WJ III COG. Because the analyses were exploratory in nature, an alpha level of .05 was used for all *post-hoc* analyses. Results demonstrated a significantly lower mean WISC-IV Processing Speed Index (PSI) score as compared to the WJ III COG Processing Speed (*Gs*) factor, $t(91) = -2.36$, $p = .02$. The standardized mean effect size of the difference between the composite pairs was large (Cohen's $d = .9$). However, no differences were found between the verbal, fluid reasoning, and working memory composite pairs. Given these findings, it appears that the changes made to the overall structure of the Wechsler measure and its resultant Full Scale IQ score have resulted in a measure that is actually more consistent with contemporary theory than was hypothesized. Furthermore, results suggest that clinicians should remain cognizant of how differences in test structure contribute to differences in performance between measures.

Within both the WISC-IV and the WJ III COG, processing speed (*Gs*) is measured by subtests assessing the Cattell-Horn-Carroll (CHC) narrow ability of perceptual speed. The WJ III COG *Gs* factor also assesses the narrow ability of speed of reasoning, resulting in the WJ III COG placing greater emphasis on visual mental speed abilities. In contrast, beyond the narrow ability of perceptual speed, both subtests included on the WISC-IV Processing Speed Index (PSI) measure the narrow ability of rate-of-test-taking, resulting in greater emphasis on writing speed.

So, while both measures include subtests which require examinees to visually scan, locate, and match items presented, the WISC-IV PSI subtests are discriminated by a greater demand upon graphomotor speed. This suggests that the WISC-IV PSI is likely to be more sensitive to neurodevelopmental conditions with slowed psychomotor speed and written performance deficits, allowing for greater determination of a child's cognitive strengths or weaknesses.

Overall, the differences found between the mean general intellectual ability scores on the Wechsler and Woodcock-Johnson measures suggests that clinicians cannot directly equate performance regarding general intelligence on the WISC-IV and WJ III COG within a neuropsychiatric population of children. Not only do findings argue for careful assessment of the domain specific constructs that constitute the overall general ability scores, but differences in IQ points across batteries can have significant implications for qualitative and diagnostic classifications. The mean difference between the WISC-IV FSIQ and WJ III GIA-Ext scores was only 1.64 points, yet the mean standard deviation was 7.72 points. Such differences could result in clinically meaningful differences regarding classifications of a subject's ability level (i.e. Low Average, Average, High Average), has significant implications when making hypotheses regarding a child's cognitive strengths and weaknesses, and may lead to different diagnostic impressions when making determinations regarding intellectual disabilities.

Clinicians needing information regarding general intellectual ability within a neuropsychiatric population of children will need to remain mindful of what the overall derived score may indicate. If tests measure different broad or narrow abilities, then the Full Scale score provides information about divergent sets of abilities, leading to misinterpretations regarding overall intelligence (Baron, 2005).

Hypothesis Two

Hypothesis two proposed that the correlation between the WISC-IV Full Scale IQ (FSIQ) and the WJ III COG General Intellectual Ability-Extended (GIA-Ext) factor would be a significantly greater than .76. Results supported the hypothesis, with the correlation found significantly greater than the hypothesized null value.

Results suggest that the WISC-IV FSIQ and WJ III COG GIA-Ext scores are highly correlated among children with neuropsychiatric disorders, with the overall correlation between the scores reflecting 76 percent shared variance between the two tests. The relationship found was significantly greater than that found for previous research involving the WISC-III FSIQ and WJ III COG GIA-Ext scores. Accordingly, this finding offers evidence that, in comparison to its WISC-III predecessor, the WISC-IV FSIQ score can be interpreted more similarly to that of the WJ III COG GIA-Ext score as a valid screening measure of general intellectual ability within a neuropsychiatric population.

While the relationship between the WISC-III FSIQ score and the WJ III COG GIA-Ext score was hypothesized to be underestimated due to the restriction of ranges for the scores used in the Phelps normal study (McGrew & Woodcock, 2001), the changes made to the structure of the WISC-IV FSIQ offered support for it being a stronger measure of general intelligence similar to that of the WJ III COG composite score. As such, results support the notion that the structure of the WISC-IV FSIQ may be more consistent with the theoretical structure underlying the WJ III COG GIA-Ext score. Specifically, similar to that of the WJ II COG composite score, the WISC-IV FSIQ has been suggested to measure the Cattell-Horn-Carroll (CHC) broad constructs of crystallized knowledge (*Gc*), fluid reasoning (*Gf*), processing speed (*Gs*), and short-term

Memory (*Gsm*), along with measurement of visual-spatial thinking (*Gv*). Overall, findings provide criterion-related validity evidence for the WISC-IV FSIQ score and suggest that the score may be interpretable under the CHC framework.

Importantly, the results suggest that clinicians utilizing the WISC-IV as part of their neuropsychological or psychological assessment battery should not make generalizations with regards to their interpretation of a child's overall general intellectual ability when comparing performance to the WISC-III. In other words, the current study suggests that clinicians cannot apply the WISC-IV FSIQ score similarly to that of its predecessor, the WISC-III. Instead, clinicians should expect to find different results regarding general intellectual ability within a neuropsychiatric population. Furthermore, IQ-based research results for the WISC-III cannot be generalized to the WISC-IV.

Though the WJ III COG GIA-Ext score also includes the broad abilities of auditory processing (*Ga*) and long-term retrieval (*Glr*), it appears that when performance across constructs is generally within consistent ranges (i.e., less than one standard deviation difference between composite scores), clinicians can expect that the scores can be interpreted similarly. Given that the WISC-IV FSIQ demonstrates a stronger relationship with the WJ III COG GIA-Ext score compared to its WISC-III counterpart, clinicians may be benefited in their interpretation of overall general intellectual ability when comparing performance between the WISC-IV and WJ III COG. However, as previously discussed, the current study found differences in mean performance between the two scores. This suggests that performance on one test may not reliability predict scores on the other, despite the convergent validity of the composites.

Hypothesis Three

It was hypothesized that the correlation between the WISC-IV Verbal Comprehension Index (VCI) and the WJ III COG Comprehension-Knowledge (*Gc*) factor would not be significantly different from .78. The relationship between the composite pairs was examined and results were as expected, not shown to be significantly different from the hypothesized null value.

The relationship found provides support for correlations between measures assessing verbal abilities and is relatively consistent with previous research demonstrating a correlation of .78 between the WISC-III VCI and the WJ III COG *Gc* factor. As such, the present investigation demonstrated that the WISC-IV VCI correlates at a similar level with the WJ III COG *Gc* factor, offering evidence that it can be applied similarly to its WISC-III counterpart as a valid measure of crystallized knowledge (*Gc*) abilities among children with neuropsychiatric disorders.

The pattern of convergent validity found suggests that even with removal of the Information subtest from the WISC-IV core battery, which provides measurement of the Cattell-Horn-Carroll narrow ability of general information, the WISC-IV and WJ III COG verbal composite scores continue to demonstrate a consistent relationship with each other among clinic-referred children. As such, clinicians may expect to find similar results on the WISC-IV Verbal Comprehension Index (VCI) as was found for the WISC-III. Furthermore, research findings from the WISC-III can likely be generalized to the newest WISC within this area. Likewise, when comparing performance between the WISC-IV and WJ III COG, interpretations will be similar to that of the WISC-III because both batteries provide measurement of the Cattell-Horn-Carroll (CHC) narrow abilities of lexical knowledge, language development, and general information.

Although a consistent relationship was found, clinicians should nevertheless remain cognizant of the changes made to the WISC-IV Verbal Comprehension Index (VCI). The Cattell-Horn-Carroll narrow ability of general information is represented across the WJ III COG Comprehension Knowledge (*Gc*) factor through inclusion of the General Information subtest. The Wechsler Comprehension subtest has been suggested to be a strong measure of the narrow ability of general information, yet the subtest is factorially complex because it also provides measurement of language development (Flanagan et al., 2007). As was previously discussed, the current study found no significant differences between mean scores for the WISC-IV VCI and WJ III COG *Gc* factor. However, further research is needed to determine if retention of the Information subtest from the WISC-IV VCI changes interpretations made when comparing performance to the WJ III COG *Gc* factor among other diagnostic populations.

Overall, the correlation between the WISC-IV VCI and the WJ III COG *Gc* factor remains relatively high, providing evidence of convergent validity. Similar to that of the WJ III COG composite score and its WISC-III predecessor, findings provide strong support for the interpretation of the WISC-IV VCI as a measure of verbal knowledge and comprehension abilities (*Gc*) among clinic-referred children. In other words, results argue that the structure of the WISC-IV VCI continues to be consistent with its predecessor, as well as with the theoretical structure underlying the WJ III COG *Gc* factor score, suggesting that the score may be interpretable under the CHC framework.

Hypothesis Four

Hypothesis four proposed that the correlation between the WISC-IV Working Memory Index (WMI) and the WJ III COG Short-Term Memory (*Gsm*) factor would be a significantly greater than .58. Results supported the hypothesis, with the correlation found significantly greater than the hypothesized null value.

Results suggest that the correlation between the WISC-IV WMI and the WJ III COG *Gsm* factor was significantly greater than that found between the WISC-III Freedom from Distractibility Index (FDI) and WJ-III COG *Gsm* factor, reflecting 52 percent shared variance between the two tests. Compared to its WISC-III predecessor, the WISC-IV WMI may function as a more valid screening measure of the Cattell-Horn-Carroll broad ability of *Gsm* (short-term memory) within a neuropsychiatric population.

Findings suggest that the changes made to the structure of the WISC-IV Working Memory Index (WMI) have resulted in a composite measure that appears to be more consistent with the underlying structure of the WJ III COG Short-Term Memory (*Gsm*) factor. Both the Wechsler and Woodcock-Johnson measures now include a greater convergence of Cattell-Horn-Carroll (CHC) narrow short-term memory abilities of memory span and working memory. Specifically, both measures now include subtests requiring examinees to hold auditory-verbal information in their immediate awareness and then ask them to repeat back the information (memory span) or to recode the information (working memory). Accordingly, results provide convergent validity evidence for the WISC-IV WMI according to CHC theory. In other words, clinicians can expect that the WISC-IV WMI can be interpreted more similar to that of the WJ III COG *Gsm* factor score within a neuropsychiatric population.

As discussed above with regards to the WISC-III and WISC-IV FSIQ scores, the current study proposes that clinicians should not make generalizations about a child's short-term memory abilities on the WISC-IV when comparing performance to the WISC-III. As such, though the Working Memory Index (WMI) and Freedom from Distractibility Index (FDI) show a strong correlation with each other, it may be unwise for clinicians to make assumptions about performance between the two measures within a neuropsychiatric population. Rather, clinicians should remain cognizant of the subtest composition of the overall working memory composite scores to better interpret differences between the measures. Furthermore, findings provide evidence that research concerning the WISC-III FDI cannot be generalized to the WISC-IV.

In contrast, results suggest that the WISC-IV WMI demonstrates a stronger relationship with the WJ III COG *Gsm* factor compared to its WISC-III counterpart. As such, clinicians can expect that the WISC-IV WMI can be interpreted as a measure of *Gsm* abilities similar to that of the WJ III COG. Furthermore, when considering that the present study demonstrated no differences in the mean scores for each measure, clinicians may be better able to make predictions about performance between the WISC-IV and WJ III COG.

Hypothesis Five

It was hypothesized that the correlation between the WISC-IV Processing Speed Index (PSI) and the WJ III COG Processing Speed (*Gs*) factor would not be significantly greater than .59. When the relationship between the composite scores was examined, results showed that the correlation was not significantly greater than the hypothesized null value, providing support for the hypothesis.

Previous research evidenced a moderate correlation (i.e., .59) between the WISC-III PSI and the WJ III COG (*Gs*) factor, providing support for the interpretation of the WISC-III PSI as a measure of *Gs* abilities. In contrast to the WISC-IV Working Memory and Perceptual Reasoning Indexes, the WISC-IV PSI was not substantially changed in the most recent revision. As such, it was expected that the correlation between the WISC-IV PSI and WJ III COG *Gs* factor would not be significantly greater than prior findings. Accordingly, results argue that the WISC-IV PSI can be interpreted similarly to that of the WISC-III PSI in a neuropsychiatric population.

Results suggest that the correlation between the WISC-IV Processing Speed Index (PSI) and the WJ III COG Processing Speed (*Gs*) factor reflects 36 percent shared variance between the two tests. The WISC-IV PSI contains the same core subtests as its WISC-III predecessor. As such, it appears to function equally as a valid screening measure of the Cattell-Horn-Carroll broad ability of processing speed (*Gs*).

Because of the amount of variance unaccounted for between the Wechsler and Woodcock-Johnson measures, clinicians would be wise to consider the differences in narrow abilities and task demands between measures that may contribute to differences in performance. As previously discussed, both the Wechsler and the WJ III COG processing speed composite scores provide measurement of the Cattell-Horn-Carroll (CHC) narrow ability of perceptual speed. The WJ III COG Processing Speed (*Gs*) factor places greater emphasis on visual mental speed abilities with the inclusion of another subtest that measures the narrow ability of speed of reasoning (Decision Speed). On this subtest, the examinee is asked to locate and circle conceptually similar pictures. In contrast, the WISC-IV Processing Speed Index (PSI) places greater emphasis on graphomotor speed because both subtests included are considered to measure the narrow ability of rate-of-test-taking. While the Symbol Search subtest does not have as great a demand for graphomotor performance, the Digit-Symbol Coding subtest requires examinees to draw symbols paired with numbers. Such differences may contribute to inconsistencies in the diagnostic specificity and sensitivity of the measures.

Despite the differences between the Wechsler and Woodcock-Johnson measures, clinicians can expect that the WISC-IV PSI can be interpreted as a valid measure of processing speed (*Gs*) abilities similar to that of the WJ III COG *Gs* factor score within a neuropsychiatric population. Likewise, practitioners can apply the WISC-IV PSI similarly to its WISC-III predecessor. In other words, clinicians can expect to find similar results when comparing performance between the two measures and research concerning the WISC-III PSI can be applied similarly to the newest WISC.

Hypothesis Six

It was hypothesized that there would be a correlation significantly greater than .46 between the WISC-IV Perceptional Reasoning Index (PRI) and the WJ III COG Fluid Reasoning (*Gf*) factor. Results of the analysis supported the hypothesis, yielding a correlation significantly greater than the hypothesized value of .46.

Previous research demonstrated a moderate correlation of .46 between the WISC-III Perceptual Organization Index (POI) and the WJ III COG Fluid Reasoning (*Gf*) factor. The result of the present analysis provides stronger convergent validity evidence for the WISC-IV Perceptional Reasoning Index (PRI) in comparison to its WISC-III predecessor. Consequently, this argues for the interpretation of the index as a more valid measure of the Cattell-Horn-Carroll (CHC) broad construct of fluid reasoning (*Gf*).

The convergent validity of the WISC-III POI was hampered by the fact that it was not a well-defined measure of fluid-reasoning. Whereas the WJ III COG *Gf* factor includes subtests that measure the narrow fluid reasoning abilities of induction (Concept Formation) and general sequential reasoning (Analysis-Synthesis), the WISC-III POI was a factorially complex combination of fluid reasoning and visual-spatial thinking abilities, as well as a measure of processing speed and verbal knowledge and comprehension abilities. In contrast, the changes made to the WISC-IV Perceptional Reasoning Index (PRI) were cited to allow for improved measurement of fluid reasoning by placing more emphasis on nonverbal problem-solving and reasoning and less emphasis on processing speed, visualization, and crystallized abilities. While the Block Design subtest was retained, which has been shown to have loadings on the visual-spatial thinking (*Gv*) factor, previous work has shown that that the two new subtests added to the index, specifically Matrix Reasoning and Picture Concepts, have greater loadings on the Fluid Reasoning (*Gf*) factor.

Similar to the WJ III COG Fluid Reasoning (*Gf)* factor, both of the new WISC-IV Perceptional Reasoning Index (PRI) subtests measure the Cattell-Horn-Carroll (CHC) narrow *Gf* abilities of induction, while the Matrix Reasoning subtest also allows for measurement of general sequential reasoning. In other words, though the WISC-IV PRI provides measurement of the visual-spatial thinking (*Gv*) abilities of spatial relations and visualization due to the inclusion of the Block Design subtest, both measures now include subtests that require examinees to analyze visual stimuli to assess their ability to start with stated rules or conditions and to engage in steps in order to reach a solution to a novel problem (general sequential reasoning). Likewise, both measures include subtests that ask examinees to categorize visual stimuli to determine how well they are able to discover the underlying characteristics that govern a problem (induction) (Flanagan & Kaufman, 2004). As such, the relationship found offers evidence that the changes made better reflects measurement of fluid reasoning (*Gf*) abilities.

Similar to results found for the WISC-IV Full Scale IQ and Working Memory Index, the current analysis speaks to the inability to interpret scores on the WISC-IV Perceptual Reasoning Index (PRI) similarly to that of its WISC-III predecessor. In other words, clinicians comparing performance between these two measures should expect to find different results, limiting their ability to make interpretations of a child's fluid reasoning abilities. Instead, clinicians need to remain aware of the changes made to the structure of the index that will account for differences found within a neuropsychiatric population. Likewise, research results concerning the WISC-III Perceptual Organization Index (POI) cannot be generalized to the new WISC composite.

Overall, results suggest that the correlation between the WISC-IV Perceptual Reasoning Index (PRI) and the WJ III COG Fluid Reasoning (*Gf*) factor accounts for 36 percent of the shared variance between the two measures. When compared to the WISC-III Perceptual Organization Index (POI), results suggest that structure of the WISC-IV PRI may be more consistent with the theoretical structure of the WJ III COG *Gf* factor. Accordingly, clinicians can be expect that the WISC-IV PRI can be interpreted as a valid measure of *Gf* abilities more similar to that of the WJ III COG *Gf* factor within a neuropsychiatric population. However, further research is needed to separate the continued factor complexity inherent in the WISC-IV PRI given the amount of variance unaccounted for between the two measures. As the subtests included on the WISC-IV PRI involve other Cattell-Horn-Carroll (CHC) abilities, clinicians should remain cognizant of how these abilities may impact differences in performance across the composite pairs.

Hypothesis Seven

Hypothesis seven proposed that the correlation between the WISC-IV Verbal Comprehension Index (VCI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor would be a significantly greater than .10. Results supported the hypothesis, with the correlation found significantly greater than the hypothesized null value.

The relationship found was significantly greater than that found by previous research involving the WISC-III VCI and the WJ III COG *Gv* factor. It was previously shown that these constructs demonstrated a negligible relationship (i.e., .10). While it was suggested that this relationship may have been underestimated, the result of the current analysis offers evidence that verbal abilities are more strongly related to visual-spatial thinking abilities than was suggested by the previous research. Accordingly, it can be inferred that the research findings concerning the WISC-III and WJ III COG should not be applied to interpret findings between the WISC-IV and WJ III COG among clinic-referred children.

The current study suggests a stronger relationship between the WISC-IV VCI and the WJ III COG *Gv* factor than was found for the WISC-III. Although interpretation of the divergent validity patterns for the WISC-IV VCI is complicated by this finding, the relationship found does not fully argue against the divergent validity for the index. Results of the current study suggest a significantly greater correlation with the WJ III COG Comprehension-Knowledge (*Gc*) factor ($z = 2.86$, $p = .002$). As such, in contrast to previous findings, the current study suggests that clinicians should remain mindful of the relationships that exist between divergent constructs that may account for findings among clinic-referred children.

Results of the current research argue for a relationship between verbal and visual-spatial thinking abilities, consistent with previous literature (see Ford, Teague, and Tusing Preschool Normal Sample, McGrew & Woodcock, 2001). As such, it can be inferred that because verbal subtests are highly associated with general intelligence (*g*), these constructs likely demonstrate a relationship due to the extent to which they both measure *g*. Moreover, results suggest that clinicians may be able to evaluate a child's pattern of performance across these constructs to guide intervention efforts. Specifically, the Cattell-Horn-Carroll (CHC) crystallized knowledge (*Gc*) narrow abilities of language development and lexical knowledge, as well as visual-spatial thinking abilities, have been reported to be significantly related to math achievement (Flanagan et al, 2007). As such, results support the notion that clinicians evaluating cognitive processes germane to mathematics abilities can make interpretations about a child's academic skills based on performance across these constructs.

The relationship found also suggests that clinicians may need to consider the cognitive strategies used by examinees to complete specific tasks. As reported above, previous research (Anjum, 2004) has demonstrated a moderate correlation between subtests involving lexical knowledge and both spatial relations and visualization. The WJ III COG Spatial Relations subtest requires examinees to detect visual features and to manipulate visual images in space to identify the pieces needed to form a complete shape. However, it may need to be considered that some examinees may use verbal abilities to assist in the completion of this spatial visualization task. Furthermore, divergent constructs may exhibit shared content variance. In other words, as both the WJ III COG Picture Recognition subtest and WISC-IV Vocabulary and Information subtests both involve a memory component to some extent, clinicians may be able to make interpretations regarding a child's specific memory abilities by comparing performance across tests.

Hypothesis Eight

It was hypothesized that there would be a correlation significantly greater than .10 between both the WISC-IV Processing Speed Index (PSI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor. When the relationship between the divergent constructs was examined, the correlation found was significantly greater than .10, providing support for the hypothesis.

Similar to the previous analysis, the relationship found between the divergent constructs argues that the research findings concerning the WISC-III and WJ III COG should not be applied similarly to the WISC-IV among clinic-referred children. Previous research (McGrew & Woodcock, 2001) has evidenced a negligible relationship between the WISC-III Processing Speed Index (PSI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor. However, this is in contrast to other research (McGrew & Woodcock; Anjum, 2004) that has demonstrated small (i.e., .29) to large (i.e., greater than .59) correlations between composite measures of processing speed and visual-spatial thinking. The results of the present investigation offer evidence that the divergent composite pairs are more strongly related than was previously suggested.

The WISC-IV Symbol Search subtest has been reported to include visual-spatial thinking (*Gv*) processes (i.e., Keith et al., 2006). This subtest involves the ability to quickly discern similarities and differences among visual stimuli, suggesting that it has shared variance with *Gv* processes. The result of the present analysis may suggest a previously unidentified relationship between the constructs, and may therefore point to declines on tasks of perceptual mental speed because of weaknesses in processing visual details. Accordingly, in contrast to that suggested by previous findings, the current study proposes that clinicians should remain mindful of the relationships that exist between divergent constructs that may account for findings among clinic-referred children. In other words, clinicians should consider the cross-loadings of underlying subtests from divergent constructs when interpreting performance findings among children with neuropsychiatric disorders.

Research has indicated that both perceptual speed and visual-spatial processing abilities are related to math achievement (Flanagan et al, 2007). However, previous findings for the WISC-III suggested that these constructs were relatively unrelated among typically developing children (McGrew & Woodcock, 2001). As such, results of this analysis argue that clinicians may be able to examine findings across the WISC-IV Processing Speed Index and WJ III COG Visual-Spatial Thinking factor when attempting to identify specific or narrow cognitive processes associated with specific academic difficulties.

As previously discussed, the relationship found does not fully argue against the divergent validity patterns for the WISC-IV Processing Speed Index. The current study found that the index demonstrated a significantly greater correlation with the WJ III COG Processing Speed factor ($z = 2.62$, $p = .004$). However, when considering the correlation found between the WISC-IV PSI and WJ III COG *Gv* factor during the current study, clinicians should expect that the narrow abilities or method/performance demands of underlying subtests may account for the shared variance between divergent constructs.

Hypothesis Nine

Hypothesis nine proposed that the correlation between the WISC-IV Working Memory Index (WMI) and the WJ III COG Visual-Spatial Thinking (*Gv*) factor would be significantly greater than .17. The result of the analysis provided support for this hypothesis. The relationship between the divergent constructs was indicated to be significantly greater than the hypothesized null value.

The result found argues for an association between short-term memory and visual-spatial thinking abilities. Previous research (McGrew & Woodcock, 2001) indicated a low correlation of .17 between the WISC-III Freedom from Distractibility Index and the WJ III COG Visual-Spatial Thinking (*Gv*) factor. However, other research (see Phelps and Ford Preschool Normal Sample Study, McGrew & Woodcock) demonstrated a moderate relationship (i.e., .47) between short-term memory and visual-spatial thinking abilities in preschool children. The current results offer evidence that the divergent WISC-IV and WJ III COG constructs are more strongly related among children with neuropsychiatric disorders than was found for research involving the WISC-III.

Although the correlation found between the WISC-III and WJ III COG scores may have been truncated due to the restriction or range of scores on both measures found in the Phelps normal study, results suggest that clinicians should consider the differences between the WISC-III and WISC-IV short-term memory composites that may account for differences in divergent validity patterns. The WISC-IV Working Memory Index (WMI) includes a subtest that requires listening to and quickly storing information in short-term memory and then moving to working memory where mental manipulation and visualization of information is needed (i.e., Letter-Number Sequencing). Accordingly, the relationship between the short-term memory and visual-spatial thinking composites may be due to the extent of this shared process variance. As such, the correlation found may suggest a previously unavailable relationship between the divergent constructs. In other words, findings indicate that research between the WISC-III and WJ III COG should not be applied similarly to the WISC-IV.

The current finding also suggests a probable role of working memory needed to complete tasks of visual-spatial thinking. Specifically, the WJ III COG Picture Recognition subtest is a task of visual memory. Because subjects are asked to recognize a subset of previously presented pictures among a subset of distracters, this requires that the original stimuli be held in mind in order to make such comparisons. Therefore, the relationship found argues for shared process variance between the WISC-IV Working Memory Index (WMI) and WJ III COG Visual-Spatial Thinking (*Gv*) factor.

The relationship between the divergent constructs was significantly greater than that found for the WISC-III Freedom from Distractibility Index, complicating divergent validity interpretations for the WISC-IV. However, it can be inferred that the extent of the relationship found still argues for the divergent validity of the WISC-IV WMI. The present study found that the relationship between the WISC-IV and WJ III COG short-term memory composite scores was significantly greater than that found between the divergent constructs ($z = 4.34$, $p < .001$). As such, the results suggest that clinicians should consider how the changes made to the structure of the WISC-IV WMI changes its relationship with the WJ III COG *Gv* factor. Furthermore, it can be inferred that clinicians should consider how similar processing demands among divergent subtests may contribute to weaknesses in performance.

Hypothesis Ten

It was hypothesized that there would be a correlation significantly greater than .19 between the WISC-IV Perceptual Reasoning Index (PRI) and the WJ III COG Auditory

Processing (*Ga*) factor. Similar to the previous findings, the result of this study supported the hypothesis. The correlation found was significantly greater than the hypothesized value of .19.

Previous research (McGrew & Woodcock, 2001) demonstrated a low correlation of .19 between the WISC-IV Perceptual Organization Index (POI) and the WJ III COG Auditory Processing (*Ga*) factor. While it was suggested that this relationship may have been underestimated, this finding does not take into account the changes made to the structure of the WISC-IV Perceptual Reasoning Index (PRI), with a demonstrated increase in its measurement of fluid reasoning abilities. Therefore, it can be inferred that the research findings for the WISC-III and WJ III COG should not be applied to findings obtained on the WISC-IV among clinic-referred children. In other words, the results offer evidence that the divergent WISC-IV and WJ III COG constructs are more strongly related among clinic-referred children than was true for the WISC-III POI.

While interpretation of the divergent validity patterns for the WISC-IV PRI is complicated by the current findings, results are not fully against the divergent validity for the index. The current study found that the index demonstrated a significantly greater correlation with the WJ III COG *Gf* factor ($z = 2.76$, $p = .003$). Overall, the current study argues for an association between fluid reasoning (*Gf*) and auditory processing (*Ga*) abilities, consistent with other research (see McIntosh and Dunham Grades 3 through 5 Normal Sample, McGrew & Woodcock, 2001). In contrast to that suggested by previous findings, the current study suggests that clinicians should remain mindful of the relationships that exist between divergent constructs that may account for results among clinic-referred children. Because fluid reasoning abilities are highly representative of general intelligence (*g*), these constructs likely demonstrate a relationship due to the extent to which they both measure *g*.

Furthermore, the results suggest that clinicians may be able to evaluate a child's pattern of performance across these constructs to guide intervention efforts. Specifically, because the WISC-IV PRI now provides greater measurement of fluid reasoning abilities, it may be that the constructs are related due to the extent to which they both reflect process-dominant thinking abilities. As such, clinicians evaluating a child's ability to access stored acquired knowledge may be able to evaluate performance across these constructs to determine if there are any learning difficulties related to weaknesses in the different thinking abilities used when information placed in short-term memory cannot be automatically processed (Woodcock, McGrew, & Mather, 2001). Also, the Cattell-Horn-Carroll (CHC) narrow fluid reasoning (*Gf*) abilities of inductive reasoning (Picture Concepts) and general sequential reasoning (Matrix Reasoning) have been suggested to play a moderate role in reading comprehension, whereas the narrow auditory processing (*Ga*) ability of phonetic coding (Sound Blending) is significantly related to reading achievement during elementary school years (Flanagan et al, 2007). As such, clinician's may be able to evaluate a child's processing abilities in deduction, induction, and phonological awareness to determine potential causes of observed reading difficulties.

Implications of Findings

As revisions to the WISC-IV have been hypothesized to align the measure more closely with the Cattell-Horn-Carroll (CHC) theory of cognitive abilities, and as cross-battery assessment and the application of CHC theory to intelligence test interpretation increases, so too does the importance of discovering the nature of the relationship between the WISC-IV and a well-validated measure based upon the CHC theory. The WJ III COG provides the avenue for such explorations. Inherent to the study of construct validity is the examination of the external relations of a measure's focal construct and observed measures (McGrew, 2009b) to help understand the similarities and differences between the measures and how best to make interpretations of test performance. As such, making comparisons between the WISC-IV and WJ III COG not only adds to the validity evidence for the WISC-IV, but also provides clarification for interpretations of test performance, particularly with regards to the CHC framework.

The current study focused on examining the comparability of performance of the global general intellectual ability scores for the WISC-IV and the WJ III COG among a clinic-referred population. Another primary goal was an attempt to address the convergent and discriminant relationships of the WISC-IV index scores with a scale specifically designed to measure cognitive abilities according to the Cattell-Horn-Carroll (CHC) theory (i.e., the WJ III COG). To that end, this study also attempted to determine the extent to which previous research findings surrounding the WISC-III could be generalized to the current version of the Wechsler scale.

Findings from the present study provided support for the convergent validity of the WISC-IV index scores. Importantly, when considering the relationships found between the composite measures of general intelligence, fluid reasoning, and working memory, results reinforced the hypothesis that the underlying structure of the WISC-IV is more consistent with the Cattell-Horn-Carroll (CHC) framework. This suggests that clinicians can be more confident in making interpretations of these scales according to CHC theory. Yet, convergent validity findings resulted in mixed interpretations regarding the ability to generalize research findings for the WISC-III to the WISC-IV.

Results offer evidence that, in contrast to previous versions of the Wechsler scale, the WISC-IV Perceptual Reasoning Index (PRI) and Working Memory Index (WMI) can be interpreted more similarly to that of the WJ III COG as indicators of the Cattell-Horn-Carroll (CHC) broad abilities of fluid reasoning (*Gf*) and short-term memory (*Gsm*), respectively. This underscores the importance of considering the substantive changes made to the structure of the indexes when making interpretations of performance. It appears that clinicians would be at a disadvantage in applying research findings for the WISC-III when making interpretations of these WISC-IV indexes.

In contrast, the relationship between the WISC-IV and WJ III COG composite verbal and processing speed measures was consistent with previous research findings for the WISC-III. As no significant differences were found, this suggests that research concerning the WISC-III Verbal Comprehension Index and Processing Speed Index can be applied similarly to the WISC-IV. However, the current study also suggests that professionals who utilize the both the WISC-IV and the WJ III COG as assessment instruments need to be aware of the underlying psychometric characteristics of each test. This is not just in regards to differences that may arise due to using measures with different production dates and normative groups, or even when comparing performance across tests with different item gradients. While large, significant correlations were found between the WISC-IV and WJ III COG convergent constructs, the current study demonstrated a significant difference in mean performance between the global ability and processing speed composite scores. This suggests that interpretations of test performance should be made with caution.

Though no changes were made to the WISC-IV Processing Speed Index (PSI) construct, clinicians may expect to find differences in the mean scores when comparing performance to the WJ III Processing Speed (*Gs*) factor among clinic-referred children. As is suggested by the significant difference found, clinicians should remain aware that minor differences between tests might result in performance differences across constructs intended to measure the same cognitive ability. As such, minor variations in task demands may lead to differences in test performance for children with specific learning, cognitive processing, or motor difficulties, thereby altering interpretations made.

Research regarding previous versions of the Wechsler and Woodcock-Johnson measures (e.g., Bracken et al., 1984; McGrew & Woodcock, 2001; Ysseldyke et al., 1981) has typically resulted in higher general intellectual functioning scores on the Wechsler Scales. In contrast, results from the current study indicated scores on the WISC-IV FSIQ to be significantly below that of the WJ III COG GIA-Ext score. However, the global composite scores evidenced a significantly greater correlation than has been shown for previous versions of the measures, specifically when comparing the findings to research concerning the WISC-III.

When considered together, these results have several important implications for neuropsychological and psychoeducational assessment. It appears that clinicians cannot expect to apply their understanding and interpretation of the WISC-III FSIQ to results obtained on the WISC-IV for children with neuropsychiatric disorders. Furthermore, because the WISC-IV FSIQ correlated more highly with the WJ III COG GIA-Ext than has been found in previous iterations of the measure, this suggests that the WISC-IV overall ability score can be interpreted more similarly to the WJ III COG GIA-Ext score as an indicator of general intelligence than was true for previous versions of the Wechsler scale.

It appears that the structure of the WISC-IV FSIQ can be viewed as being more theoretically similar to the underlying structure of the WJ III COG GIA-Ext score. Because the WISC-IV has a number of significant departures from the WISC-III, as well as older versions of the measure, this has resulted in an alteration in the degree of representation of general intelligence (*g*) on the WISC-IV. The current study provides support for the notion that the battery reflects a more equal weighting of crystallized knowledge, fluid reasoning, working memory, and processing speed abilities, with these abilities also more equally represented in the Full Scale IQ score.

Furthermore, the current study demonstrated that the average difference between the WISC-IV FSIQ and WJ III COG GIA-Ext scores was only 1.64 points (which appears to be less than has been found in the past). However, the significant difference between the mean scores suggests that clinicians should not assume that referred students similar to those in this study will exhibit identical performance across these measures. In other words, despite the support for the convergent validity of the WISC-IV FSIQ and the considerable shared variance between the scores demonstrated by this study, clinicians may not be able to make reliable predictions about global intelligence from one test to the other.

Because it might be expected that lower overall IQ scores will be seen when administering the WISC-IV when compared with the WJ III COG, clinicians should consider how performance differences on the broad ability factors of the measures contribute to differences in global intelligence. Likewise, a lack of equivalence between the two measures will affect interpretations of cognitive performance, as well as the nature of decisions made for determination of disabilities and appropriate intervention or treatment recommendations. As such, future research is needed to understand and determine differences in global ability scores that may arise between more specific diagnostic populations.

Examination of the divergent validity patterns for the WISC-IV and WJ III COG revealed a reliable pattern of significantly greater correlations between divergent constructs than was found for research concerning the WISC-III and WJ III COG. Such findings underscore the difficulty that will arise in applying research findings for the WISC-III and WJ III COG to the WISC-IV. The current study contributed to the empirical literature supporting a relationship between visual-spatial thinking abilities and verbal comprehension facility (see Ford, Teague, and Tusing Preschool Normal Sample, McGrew & Woodcock, 2001), processing speed [Anjum, 2004; McGrew & Woodcock (see Gregg and Hoy University Normal and Learning Disabled Sample study, McIntosh and Dunham Grades 3 through 5 Normal Sample Study), and working memory abilities (see Phelps and Ford Preschool Normal Sample, McGrew & Woodcock). Likewise, results of the current research support a relationship between fluid reasoning and auditory processing abilities (see McIntosh and Dunham Grades 3 through 5 Normal Sample, McGrew & Woodcock). This is in contrast to the research concerning the WISC-III and WJ III COG, where negligible to low correlations were shown between the above-mentioned divergent constructs.

Although the current findings appear to complicate divergent validity interpretations for the WISC-IV, the current study found that the correlations between the WISC-IV and WJ III COG convergent constructs were significantly higher than correlations found between the examined divergent constructs. This provides support for the convergent validity of the WISC-IV index scores with the WJ III COG. However, findings can help to inform decisions made regarding measurements included when evaluating specific learning difficulties, as well as contributes to the understanding of test interpretations.

When considering results regarding divergent validity patterns between the WISC-IV and WJ III COG, findings suggest that clinicians should consider the shared variance between composite or cluster scores that appear to measure dissimilar broad abilities when making interpretations of test performance. In other words, the pattern of divergent correlations provides evidence that clinicians should consider that constructs or tests may reflect cognitive processes other than those it purports to measure. For example, the WISC-IV Processing Speed Index (PSI) Symbol Search subtest is regarded as a measure of the Cattell-Horn-Carroll (CHC) narrow ability of perceptual speed. Yet, the subtest has been cited to include visual-spatial thinking (*Gv*) processes. As such, the apparent cross-loading of the subtest may need to be considered when making interpretations of performance for the WISC-IV PSI among children with neuropsychiatric disorders. Further research is needed to examine the multifactorial nature of the subtests included on the batteries because such complexity is likely to result in considerable differences in performance intra-individually.

Results for the divergent validity correlations found between the WISC-IV and WJ III COG further points to the importance of considering how changes made to the substantive structure of the WISC-IV changes the nature of the relationship for divergent constructs. This is particularly important when considering the relationships found for the WISC-IV Perceptual Reasoning Index (PRI) and Working Memory Index (WMI). As such, clinicians should be aware of the process-related variance that may contribute to test findings. For instance, underperformance on measures of visual-spatial thinking may be due to weaknesses in working memory, so comparison of performance across these divergent constructs may contribute to interpretations of test performance. Moreover, the divergent validity findings support the use of the WISC-IV and WJ III COG for evaluation of a child's cognitive processing deficits germane to specific learning difficulties. However, further research is needed to examine the construct-irrelevant variance at the construct and subtest level to help clinicians better organize their test batteries.

The results from this study provide insight into the similarities between the WISC-IV and WJ III COG. Furthermore, results help to clarify the differences between the WISC-III and WISC-IV. In general, findings suggest that clinicians should not attempt to generalize performance on previous versions of the WISC to the most recent version. Findings support the hypothesis that a relationship exists between the WISC-IV index scores and WJ III COG cluster scores purporting to measure conceptually similar constructs, specifically, comprehension-knowledge (*Gc*), fluid reasoning (*Gf*), short-term memory (*Gsm*), processing speed (*Gs*), and general intelligence (*g*). Practically speaking, given the current as well as the previous research, the data appear to support the WISC-IV as more interpretable under the CHC framework in comparison to previous versions of the measure. Accordingly, it appears that the WISC-IV evidences clinical validity for use as a core battery when conducting a cross-battery assessment.

Importantly, clinicians can expect that the changes made to the substantive structure of the WISC-IV FSIQ will result in differences in the mean scores when comparing performance to the WJ III COG GIA-Ext score. Specifically, the current study found no significant differences in mean performance between the WISC-IV and WJ III COG composite measures of verbal, fluid reasoning, and working memory abilities. This provides evidence that the global full scale scores for each measure demonstrate a considerable proportion of convergence. However, the significant differences found between the processing speed composite measures points to the importance of considering that when the Full Scale score provides information about divergent sets of abilities, with tests including minor variations in task demands or narrow abilities, this may lead to misinterpretations regarding overall intelligence.

Overall, the use of the WISC-IV appears to change the nature of neuropsychological assessment and interpretations of general intelligence and cognitive abilities. The current results suggest that the WISC-IV is an improved measure of global ability, with improved measurement of the second-order factors comprising the battery. Likewise, the results suggest that the WISC-IV provides closer examination of particular functions germane to the cognitive assessment of children. Results similarly highlight the importance of engaging in careful assessment of the domain specific constructs that constitute the overall general ability scores to increase the validity of clinical practice.

The merit of the current study may be to inform clinicians regarding cautions warranted when making interpretations of performance across separate measures of intelligence among children with neuropsychiatric disorders, and to provide a necessary link to interpretations of test performance under the Cattell-Horn-Carroll framework. As such, the empirical data presented here helps to inform clinical practice by providing understanding regarding composite measures than can be administered to allow for the best possible measurement and interpretation of specific cognitive abilities. Also, the findings help to elucidate the usefulness and appropriateness of the WISC-IV as a measure of cognitive abilities, particularly as it relates to the CHC framework.

Limitations

Internal validity. One of the primary limitations of this study was that it was archival in nature. The participants were part of a clinically referred sample and were not matched by age, gender, or ethnicity to represent the composition of the general population. Also, due to selection criteria, the sample is neither representative of other client populations nor from the population from which it was drawn. In addition, all participants were clinically referred for a neuropsychological evaluation at a community mental health center, which may have biased the clinically referred sample.

At the same time, because of the archival nature of the study, several important pieces of demographic data were not available, thus were not used as exclusionary criteria. For example, the native/primary language of examinees was not accounted for. Intelligence tests presume that a level of language proficiency is present, and both the WISC-IV and WJ III COG are reliant on verbal and receptive language abilities (Flanagan et al., 2007). As such, it is unclear as to what extent differences in language competency may have affected results.

Other extraneous factors that may have influenced the results were also considered. Because of the archival nature of the study, attempts could not be made to ensure that the tests were administered in a counterbalanced or random order to minimize the effect of order of test administration. Also, because the study utilized a repeated measures design, it is possible that the examinees' exposure to tasks measuring similar cognitive abilities are including similar tasks demands may have affected their performance across tests.

Furthermore, the participants in this study were administered the tests by graduate students who had taken coursework in the in the administration, scoring, and interpretation of the Wechsler Intelligence Scales. While the students completed further supervised training and evaluation of competencies in the administration and scoring of individual intelligence tests, there is no guarantee that every graduate student trained on the measures administered all subtests properly. Because of the archival nature of the study, inter-rater reliability could not be evaluated to determine the effects of testing across different administrators, and to determine if the administrators had scored the performance of a participant similarly. Although there are guidelines for administration and scoring, with instructions to be read verbatim, due to human error, there is no guarantee that these instructions and guidelines were strictly followed.

Another important limitation is the heterogeneous nature of the sample, which consisted of children with a variety of psychological and neurological diagnoses. In contrast to previous research that has examined differences in groups of children with learning disorders, the mixed nature of the sample may have impacted the relationships found between the measures.

Sample size. Another limitation included the relatively small sample size of the study. Although the sample size used for this study was adequate for the preliminary investigation, the sample was drawn from archival data of clinic-referred children assessed at one community clinic, with the research being limited by the number of participants in the database who had completed both measures under scrutiny. Studies that compare the correlations of major intelligence tests are benefited by using larger, nationally represented samples. As such, a larger sample size from a larger geographic region would increase the assurance that the findings are stable and the sample size is adequately representative of the population.

External validity. The generalizability of the study findings to other populations is questionable. The findings are more likely to generalize to other clinical populations than to the general population because a clinical sample was utilized. Additionally, the number of participants in the sample who had been diagnosed with a DSM-IV-TR disorder was substantially larger than in the general population. In fact, over 25% of the sample met diagnostic criteria for an adjustment disorder and almost 20% of the sample met diagnostic criteria for Major Depressive Disorder. The sample utilized in the study had a higher prevalence of Major Depressive Disorder than would be found in the general population. The study may also be limited due to the lack of distinction among the various disorders and when considering that several of the participants were diagnosed with more than one disorder. Additionally, the participants in the study all sought out an evaluation at a university-based Mental Health Center, which may limit the generalization of the sample to other clinically referred populations.

When considering the selection criteria, it should be noted that it is unclear how results will generalize to minority populations. The study sample was comprised of a larger percentage of Caucasian children (60.4%) as compared to African American (6.6%) and Hispanic children (17.6%). In comparison to the WISC-III, where there was an 11 point difference between Caucasians and African Americans, ethnic differences on the WISC-IV were smaller (9 points) (Kaufman et al., 2006). Despite the smaller ethnic differences found on the WISC-IV, this study may present results more typical for Caucasian children.

Regarding theoretical limitations, the current study involved examination of the WISC-IV from the theoretical standpoint of the Cattell-Horn-Carroll (CHC) framework. While most other intelligence tests can also be interpreted using a CHC interpretative model, not all measures are explicitly (or implicitly) based on this framework, including the WISC-IV. The results of this study may be used to understand the convergent and divergent validity of the WISC-IV from an interpretive CHC model standpoint. However, results of the study provide limited understanding of the WISC-IV's relationship with measures based on different substantive theories [i.e., Cognitive Assessment System (CAS; Naglieri & Das, 1997).

Statistical limitations. A statistical limitation to the study is the alpha level used in determining the statistical significance for the differences between the mean scores of the two measures. The current preliminary study set the alpha level for significance at the .05 level, allowing for statistically significant findings. Future studies that are able to counterbalance the order of tests may choose to employ different statistical methods that can account for test order (i.e., Multivariate Analysis of Variance), resulting in a more stringent alpha level for post-hoc analyses. Adjustments to the alpha level could result in different findings regarding differences between the mean scores, thereby limiting the generalizability of the current investigation.

It is also worth noting the structure of hypothesis three is slightly different from the other hypotheses examining the expected correlations between the convergent and divergent constructs in comparison to stipulated values under the null hypothesis. Rather than testing the statistical significance of the difference between the observed correlation and the stipulated null value, this hypothesis essentially makes use of the stipulated null value for the hypothesis test.

Future Research

Future research in this area should seek to address the limitations of this study. As such, researchers might repeat this study using a larger, representative sample of the general population, rather than children referred for a neuropsychological evaluation. Likewise, future research studies should go beyond the use of archival data to allow for counterbalanced assessment procedures in order to ensure that correlations accurately reflect the relationship between the tests.

Since this investigation included subjects with multiple diagnoses, with the various types of diagnostic disorders collapsed into one group for comparison, future research is needed to determine the pattern of convergent and divergent validity of the WISC-IV with the WJ III COG in samples with clearly defined diagnostic populations, such a specific learning disabilities, separate psychiatric diagnoses (i.e., depression, anxiety disorder), ADHD, and intellectual impairment. With the prevalence of the use of a cross-battery approach to test interpretation, it will be important for future research to focus on how other clinical samples perform on the tests utilized for the current study, as well as the intercorrelations between the tests. Additionally, it may be of importance to determine how differences in English language competency contribute to differences in performance and patterns of correlations between the two measures.

The results of this study suggest that, in comparison to its WISC-III predecessor, the WISC-IV Perceptual Reasoning Index (PRI) and the WJ III COG Fluid Reasoning (*Gf*) factor are measuring similar broad constructs from a theoretical standpoint. However, future research is needed to understand the factorial nature of the WISC-IV PRI, including examining its relationship with the WJ III COG Visual-Spatial Thinking (*Gv*) factor.

Given that the current study provided evidence of increased correlations between composite measures of fluid reasoning and working memory, more research is needed to understand the pattern of correlations that exist at the subtest level in order to contribute to interpretations made. Likewise, future research is needed to increase understanding of the relationship between the WISC-IV and WJ III COG when more divergent performance is shown across constructs comprising the overall ability scores. In other words, it would be important to understand how differences in constructs important to the understanding and prediction of academic achievement [i.e., long-term retrieval (*Glr*), auditory processing (*Ga*)] affect interpretation of the overall ability scores.

In addition, given the mean differences found between the global ability scores, further study is needed to determine if there are mean differences in the global ability and composite constructs among more specific diagnostic populations. Similarly, future research should determine the degree to which new and revised intelligence tests influence educational decision making as a function of their mean IQ differences and/or construct divergence in comparison to other intelligence measures. It would also be important to examine the extent to which the WISC-IV Full Scale and Index scores predict important clinical outcomes.

It is also recommended that this study be replicated with refinements made to the assessment instruments used. For example, future researchers might consider including the administration of an individually-administered brief or short-form screening measure in order to broaden the exploration of construct validity. Likewise, it may be of value to understand how an individually-administered brief screening measure relates to making predictions about performance on both the WISC-IV and the WJ III COG.

Bearing in mind that measures such as intelligence tests are used to help understand and to provide validity regarding decisions made about individuals' mental abilities, test revisions carry with them the consequence of changing what information is collected and assessed (Strauss, Spreen, & Hunter, 2000). Considering that the WISC-IV and the WJ III COG are used for similar purposes and appear to demonstrate a greater convergence of constructs, it is be important to determine how well the WISC-IV predicts performance on the WJ III COG, and vice versa. Though best practice stipulates using the same version of a test under instances of serial testing, it is of clinical value to understand score differences between these widely used instruments.

References

Alfonson, V. C., Flanagan, D. P., & Radwan, S. (2005). The impact of the Cattell-Horn-Carroll theory on test development and interpretation of cognitive abilities and academic abilities. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 185–202). New York: Guilford.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

Anjum, A. (2004). The relationship between the Differential Ability Scale and the Woodcock Johnson III Tests of Cognitive Abilities for children diagnosed with attention deficit hyperactivity disorder. *Dissertations Abstracts International, 65(09),* 4859. (UMI No. 795970521).

Baddeley, A.D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 49*, 5-28. doi: 10.1080/713755608

Baron, I. S. (2005). Test review: Wechsler Intelligence Scale for Children – Fourth Edition (WISC-IV). *Child Neuropsychology, 11,* 471-475. doi: 10.1080/09297040590951587

Bell, N. L., Lassiter, K. S., Matthews, T. D., & Hutchinson, M. B. (2001). Comparison of the Peabody Picture Vocabulary Test-Third Edition and the Wechsler Adult Intelligence Scale-Third Edition with university students. *Journal of Clinical Psychology, 57,* 417-422. doi: 10.1002/jclp.1024

Bell, N. L., Rucker, M., Finch, Jr., A. J., & Alexander, J. (2002). Concurrent validity of the Slosson Full-Range Intelligence Test: Comparison with the Wechsler Intelligence Scale for Children-Third Edition and the Woodcock Johnson Tests of Achievement Revised. *Psychology in the School, 39*, 31-38. doi: 10.1002/pits.10002

Benson, J. (1998). Developing a strong program of construct validation: A test anxiety example. *Educational Measurement: Issues and Practice, 17,* 10-2. doi: 10.1111/j.1745-3992.2009.00169.x

Beres, K. A., Kaufman, A. S., & Perlman, M. D. (2000). Assessment of child intelligence. In G. Goldstein & M. Hersen (Eds.) *Handbook of psychological assessment.* (pp. 65-96). Kidlington, Oxford, United Kingdom: Elsevier Science Ltd.

Bodin, D., Pardini, D., Burns, T. G., & Stevens, A. (2009). Higher order factor structure of the WISC-IV in a clinical neuropsychological sample. *Child Neuropsychology, 15(5),* 417-424. doi: 0.1080/09297040802603661

Bracken, B. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology, 26,* 155-166. doi: 10.1016/0022-4405(88)90017-9

Bracken, B. A., Prasse, D. P., & Breen, M. J. (1984). Concurrent validity of the Woodcock-Johnson Psycho-Educational Battery with regular and learning disabled students. *Journal of School Psychology, 22,* 185-192. doi: 10.1016/0022-4405(84)90038-4

Braden, J. P. (1995). Review of Wechsler Intelligence Scale for Children—Third Edition. In J. V. Mitchell (Ed.), *The tenth mental measurement yearbook* (vol. 1, pp. 1098-1103). Lincoln, NE: Buros Institute of Mental Measurement.

Brody, N. (1999). What is intelligence? *International Review of Psychiatry, 11 (1),* 19-25. doi: 10.1080/09540269974483

Brown, T. L., & Morgan, S. B. (1991). Concurrent validity of the Stanford-Binet, 4th Edition: Agreement with the WISC-R in classifying learned disabled children. *A Journal of Consulting and Clinical Psychology, 3*, 247-253. doi: 10.1037/1040-3590.3.2.247

Camara, W. J., Nathan, J. S., & Peunte, A. E. (2000). Psychological test usage: Implications in professional psychology. *Professional Psychology: Research & Practice, 31,* 141-154. doi: 10.1037/0735-7028.31.2.141

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81-105. Retrieved from http://www.mendeley.com/research/convergent-and-discriminant-validation-by-the-multitraitmultimethod-matrix/

Canivez, G. L., Neitzel, R., & Martin, B. E. (2005). Construct Validity of the Kaufman Brief Intelligence Tests, Wechsler Intelligence Test for Children-Third Edition, and the Adjustment Scales for Children and Adolescents. *Journal of Psychoeducational Assessment, 23(15),* 15-34. doi: 10.1177/073428290502300102

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

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford.

Carroll, J. B. (1998). Human cognitive abilities: A critique. In J.J. McArdle, & R.W. Woodcock (Eds.), *Human cognitive abilities in theory and practice* (pp. 5-24). Mahwah, NJ: Lawrence Erlbaum.

Cattell, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin, 38(7),* 592. doi: 10.1037/h0050099

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40(30)*, 153-193. doi: 10.1037/h0059973

Cattell, R. B. (1957). *Personality and motivation structure and measurement.* New York: World Book.

Cizek, G. J. (2003). [Review of the Woodcock-Johnson III.]. In B. S. Plake & J. C. Impara (Eds.), The fifteenth mental measurements yearbook (pp. 1020-1024). Lincoln, NE: Buros Institute of Mental Measurements.

Coalson, D. & Weiss, L. (2002). The evolution of Wechsler Intelligence Scales in historical perspective. *Focus, 11,* 1-6. Retrieved from http://www.pearsonassessments.com/pai/ai/research/publications/asmntfocus/reslst

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2^{nd} ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Cohen, A., Fiorello, C. A., & Farley, F. H. (2006). The cylindrical structure of the Wechsler Intelligence Scale for Children — IV: A retest of the Guttman model of intelligence. *Intelligence, 34,* 587-591. doi: 10.1016/j.intell.2006.05.003

Cole, J. C., & Randall, M. K. (2003). Comparing the cognitive ability models of Spearman, Horn and Cattell, and Carroll. *Journal of Psychoeducational Assessment, 21(2),* 160-179. doi: 10.1177/073428290302100204

Cronbach, L. (1971). Construct validation after thirty years. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy. Proceedings of a Symposium in Honor of Lloyd Humphreys* (pp. 147-167). Urban, IL: University of Chicago Press.

Cronbach L., & Meehl, P. (1955). Construct validity of psychological tests. *Psychological Bulletin, 52,* 281-302. Retrieved from http://psychclassics.yorku.ca/Cronbach/construct.htm

Das, J. P., Naglieri, J. A., & Kirby, J. R. (1994). *Assessment of cognitive processes: The PASS theory of intelligence.* Boston: Allyn & Bacon.

DiCerbo, K. E., & Barona, A. (2000). A convergent validity study of the Differential Ability Scales and the Wechsler Intelligence Scale for Children-Third Edition with Hispanic children. *Journal of Psychoeducational Assessment, 18(4)*, 344-352. doi: 10.1177/073428290001800404

Donders, J., & Warschausky, S. (1996). A structural equation analysis of the WISC-III in children with traumatic head injury. *Child Neuropsychology, 2(3),* 185–192. doi: 10.1080/09297049608402251

Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature, 427,* 311-312. doi: 10.1038/427311a

Dumont, R., Cruse, C. L., Price, L., & Whelley, P. (1996). The relationship between the Differential Ability Scales (DAS) and the Wechsler Intelligence Scale for Children Third Edition (WISC-III) for students with learning disabilities. *Psychology in the Schools, 33(3),* 203-209. doi: 10.1002/(SICI)1520-6807(199607)33:3<203::AID-PITS3>3.0.CO;2-Q

Dumont, R., Willis, J. O., Farr, L. P., McCarthy, T., & Price, L. (2000). The relationship between the Differential Ability Scales (DAS) and the Woodcock-Johnson Tests of Cognitive Abilities-Revised (WJ-R COG) for students referred for special education evaluations. *Journal of Psychoeducational Assessment, 18(1),* 27-38. doi: 10.1177/073428290001800103

Edwards, O. W., & Paulin, R. V. (2007). Referred students' performance on the Reynolds Intellectual Assessment Scales and the Wechsler Intelligence Scale for Children Fourth Edition. *Journal of Psychoeducational Assessment, 25(4),* 334-340. doi: 10.1177/0734282907300453

Elliott, C.D. (1997). The Differential Ability Scales. In D.P. Flanagan, J.L. Genshaft, & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 183–208). New York: Guilford Press.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell-Horn-Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31(2),* 246-262. Retrieved from http://proquest.umi.com.ezproxylocal.library.nova.edu/pqdweb?did=140548071 sid=2&Fmt=1&clientId=17038&RQT=309&VName=PQD

Fiorello, C. A., Hale, J. B., McGrath, M., Ryan, K., & Quinn, S. (2001). IQ interpretation for children with flat and variable test profiles. *Learning and Individual Differences, 13(2),* 115–125. doi: 10.1016/S1041-6080(02)00075-4

Flanagan, D. P. (2000). Wechsler-based CHC cross-battery assessment and reading achievement: Strengthening the validity of interpretations drawn from Wechsler test scores. *School Psychology Quarterly, 15(3),* 295-329. Retrieved from http://proquest.umi.com.ezproxylocal.library.nova.edu/pqdweb?did=62519982& id=4&Fmt=1&clientId=17038&RQT=309&VName=PQD

Flanagan, D. P. (2001). *Comparative features of the WJ III Tests of Cognitive Abilities (Woodcock-Johnson III assessment service bulletin No. 1).* Itasca, IL: Riverside Publishing.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment.* New York:John Wiley and Sons.

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

Flanagan, D. P., McGrew, K. S., & Ortiz, S. O. (2000). *The Wechsler Intelligence scales and Gf-Gc theory: A contemporary approach to interpretation.* Boston: Allyn & Bacon.

Flanagan, D.P., & Ortiz, S.O. (2001). *Essentials of cross-battery assessment*. New York: John Wiley & Sons.

Flanagan, D. P. & Ortiz, S. O. & Alfonso, V.C. (2007). *Essentials of Cross-Battery Assessment, Second Edition.* New York: Wiley Press.

Fletcher, J. M., Lyon, G. R., Barnes, M., Stuebing, K. K., Francis, D. J., Olson, R. K., et al. (2001, August). *Classification of learning disabilities: An evidence-based evaluation.* Paper presented at the U.S. Department of Education LD Summit, Washington, DC. Retrieved from http://www.ldaofky.org/LD/Classification%20of%20LD.pdf

Fletcher-Janzen, E. (2003). *A validity study of the Kaufman Assessment Battery for Children, Second Edition (KABC-II) and the Taos Pueblo Indian Children of New Mexico.* Circle Pines, MN: AGS Publishing. Retrieved from http://www.pearsonassessments.com/pai/ca/RelatedInfo/KABCIIValidityStudy/htm

Floyd, R. G., Evans, J. J., & McGrew, K. S. (2003). Relations between measures of Cattell Horn-Carroll (CHC) cognitive abilities and mathematics achievement across the school age years. *Psychology in the School, 40 (2),* 155-171. doi: 10.1002/pits.10083

Floyd, R. G., Keith, T. Z., Taub, G. E., & McGrew, K. S. (2007). Cattell–Horn–Carroll cognitive abilities and their effects on reading decoding skills: *g* has indirect effects, more specific abilities have direct effects. *School Psychology Quarterly, 22(2),* 200-233. doi: 10.1037/1045-3830.22.2.200

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Garrett, H. E. (1937). *Statistics in psychology and education.* New York: Longmans, Green.

Gläscher, J., Tranel, D., Paul, L. K., Rudrauf, D., Rorden, C., Hornaday, H., et al. (2009). Lesion mapping of cognitive abilities linked to intelligence. *Neuron, 61(5),* 681-91. doi: 10.1016/j.neuron.2009.01.026

Grados, J. J., & Russo-Garcia, K. A. (1999). Comparison of the Kaufman Brief Intelligence Test and the Wechsler Intelligence Scale for Children-Third Edition in economically disadvantaged African American youth. *Journal of Clinical Psychology, 55(9)*, 1063-1071. doi: 10.1002/(SICI)1097-4679(199909)55:9<1063::AID-JCLP4>3.0.CO;2-U

Grice, J. W., Krohn, E. J., Logerquist, S., (1999). Cross-validation of the WISC-III factor structure in two samples of learning disabilities. *Journal of Psychoeducational Assessment, 17,* 236-248. doi: 10.1177/073428299901700304

Hale, J. B., Fiorello, C. A., Kavanaugh, J. A., Hoeppner, J. B., & Gaither, R. A. (2001). WISC III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly, 16(1),* 31-35. doi: 10.1521/scpq.16.1.31.19158

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Holdnack, J. A., & Aloe, A. M. (2007). Is the demise of IQ interpretation justified? A response to special issue authors. *Applied Neuropsychology, 14(1),* 37-51. doi: 10.1080/09084280701280445

Hebben, N. (2004). Review of special group studies and utility of the process approach with the WISC-IV. In D. P. Flanagan & A. S. Kaufman (Eds.), *Essentials of WISC-IV assessment* (pp. 183-199). New York: John Wiley and Sons.

Holdnack, J. A., & Weiss, L. G. (2006). IDEA 2004: Anticipated implications for clinical practice-Integrating assessment and intervention. *Psychology in the Schools, 43(8),* 871-882. doi: 10.1002/pits.20194

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade (Ed.), *Handbook of multivariate psychology* (pp. 645-685). New York: Academic Press.

Horn, J. L. (1989). Models for intelligence. In R. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29-73). Urbana, IL: University of Illinois Press.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, & R. W. Woodcock (Eds.), *WJ-R technical manual*. Chicago: Riverside.

Horn, J. L. (1994). Theory of fluid and crystallized intelligence. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (pp. 443-451). New York: Macmillan.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 53-91). New York: Guilford.

Kamphaus, R. W. (1993). *Clinical assessment of children's intelligence.* Boston: Allyn & Bacon.

Kamphaus, R. W. (2005). *Clinical assessment of child and adolescent intelligence.* New York: Springer.

Kaufman, A. S., Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2006). Test review: Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV). *Journal of Psychoeducational Assessment, 24(3),* 278-295. doi: 10.1177/0734282906288389

Kaufman, A. S., & Kaufman, N. L. (1993). *Manual for the Kaufman Adolescent and Adult Intelligence Test (KAIT).* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004). *Kaufman Assessment Battery for Children: Technical manual* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., Kaufman, N. L. (2005). *Essentials of KABC-II assessment*. Hoboken, NJ: John Wiley & Sons, Inc.

Kavale, K. A., & Forness, S. R. (1995). *The nature of learning disabilities: Critical elements of diagnosis and classification.* Hillsdale, NJ: Erlbaum.

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,* 108-127. Retrieved from http://www.nasponline.org/publications/spr/pdf/spr351keith.pdf

Keith, T. Z., Kranzler, J. H., Flanagan, D. P. (2001). What does the Cognitive Assessment System measure? Joint Confirmatory Factor Analysis of the CAS and the Woodcock Johnson Tests of Cognitive Ability (3rd Edition). *School Psychology Review, 30(1),* 89-118. Retrieved from http://proquest.umi.com.ezproxylocal.library.nova.edu/pqdweb?did=76964556&sid=6&Fmt=1&clientId=17038&RQT=309&VName=PQD

Keith, T. Z., & Witta, E. L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Quarterly, 12,* 89-107. doi: 10.1037/h0088950

Konold, T. R., Kush, J. C., & Canivez, G. L. (1997). Factor replication of the WISC-III in three independent samples of children receiving special education. *Journal of Psychoeducational Assessment, 15(2),* 123–137. doi: 10.1177/073428299701500203

Kline, Paul. (1991). *Intelligence: The psychometric view.* London, England: Routledge.

Law, Jr., J. G. & Faison, L. (1996). WISC-III and KAIT results in adolescent delinquent males. *Journal of Clinical Psychology, 52(6),* 699-703. doi: 10.1002/(SICI)1097-4679(199611)52:6<699::AID-JCLP12>3.0.CO;2-H

Leffard, S. A., Miller, J. A., Bernstein, J., DeMann, J. J., Mangis, H. A., & McCoy, E. L. B. (2006). Substantive validity of working memory measures in major cognitive functioning test batteries for children. *Applied Neuropsychology, 13(4),* 230-241. doi: 10.1207/s15324826an1304_4

Lepach, A. C., Petermann, F., & Schmidt, S. (2008). Comparisons of the BASIC-Memory and Learning Test and the WISC-IV under developmental aspects. *Zeitschrift für Psychologi/Journal of Psychology, 216(3)*, 180-186. doi: 10.1027/0044-3409.216.3.180

Little, S. G. (1992). The WISC-III: Everything old is new again. *School Psychology Quarterly,* 7 (2), 148–154. Retrieved from http://www.psycnet.apa.org/journals/spq/7/2/148.pdf

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports, 3,* 635-694. Retrieved from http://www.mres.gmu.edu/readings/PSYC557/Loevinger_1957.pdf

Macmann, G. M., & Barnett, D. W. (1994). Structural analysis of correlated factors: Lessons from the Verbal-Performance dichotomy of the Wechsler Scales. *School Psychology Quarterly, 9(3),* 161-197. doi: 10.1037/h0088287

Mather, N., & Gregg, N. (2006). Specific learning disabilities: Clarifying, not eliminating, a construct. *Professional Psychology: Research and Practice, 37 (1),* 99-106. Retrieved from http://www.iapsych.com/articles/mather2006.pdf

Mather, N. & Wendling, B. J. (2005). Linking cognitive assessment results to academic interventions for students with learning disabilities. In D. P. Flanagan & P. L Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.) (pp. 269-294) New York: Guilford Press.

Mather, N., & Woodcock, R.W. (2001). Examiner's Manual. *Woodcock-Johnson Tests of Cognitive Abilities.* Itasca, IL: Riverside Publishing.

Matson, J. L., Andrasik, F., & Matson, M. L. (Eds.). (2008). *Assessing childhood psychopathology and developmental disabilities.* New York: Springer.

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, and P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 151–179). New York: The Guilford Press.

McGrew, K. S. (2004, October 11). Intelligence and Achievement: A lesson from Forrest Gump regarding appropriate expectations for students with cognitive disabilities. Retrieved from http://www.iapsych.com/PPT.htm

McGrew, K. S. (2005). The Cattell-Horn-Carroll (CHC) theory of cognitive abilities: Past, present, and future. In D. Flanagan, & Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues-Second Edition (p.136-202)*. New York: Guilford Press.

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

McGrew, K.S. (2009, November 8b). What does the WAIS-IV measure? CHC analysis and beyond. *Applied Psychometric 101: IQ Test Score Differences Series,* #2. Retrieved from http://www.iqscorner.com/

McGrew, K. S. (2010, February 2). IQ test DNA fingerprints: Comparison of WJ III, WISC-IV and WAIS-IV. Retrieved from http://www.iqscorner.com/2010/03/iq-test-dna-fingerprints-comparison-of.html

McGrew, K. S., & Flanagan, D. P. (1998). *The intelligence test desk reference:* Gf-Gc *cross battery assessment.* Boston: Allyn & Bacon.

McGrew, K. S., Werder, J. K., & Woodcock, R. W. (1991). *Woodcock-Johnson Psychoeducational Battery-Revised technical manual.* Chicago: Riverside.

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

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13- 103). Washington, DC: American Council on Education.

Metz, B. L., (2006). *A comparison of the WJ-III test of cognitive abilities and the WAIS-III.* (Master's thesis, Marshall University Graduate College, 2005). *Masters Abstract International, 44*, 1-11. Retrieved from http://proquest.umi.com/pqdweb?did=1136090701&sid=1&Fmt=2&clientId=2079RQT=309&VName=PQD

Morris, R. D., Stuebing, K. K., Fletcher, J. M., Shaywitz, S. E., Lyon, G. R., Shankweiler, D. P., et al. (1998). Subtypes of reading disability: variability around a phonological core. *Journal of Educational Psychology, 90 (3),* 347-373. doi: 10.1037/0022-0663.90.3.347

Naglieri, J., Salter, C., & Rojahn, J. (2005). Cognitive disorders of childhood and adolescence: Specific learning disabilities and mental retardation. In J. E. Maddux & B. A. Winstead (Eds.), *Psychopathology: Foundations for a contemporary understanding* (pp. 377-392). Mahwah, NJ: Lawrence Erlbaum.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51(2)*, 77-101. Retrieved from http://www.gifted.uconn.edu/siegle/research/Correlation/Intelligence.pdf

Oakland, T., & Hu, S. (1992). The top 10 tests used with children and youth worldwide. *Bulletin of the International Test Commission, 19,* 99-120.

Pfeiffer, S. I., Reddy, L. A., Kletzel, J. E., Schmelzer, E. S., & Boyer, L. (2000). The practitioner's view of IQ testing and profile analysis. *School Psychology Quarterly, 15 (4),* 376-385. doi: 10.1037/h0088795

Piaget, J. (1972). *The psychology of intelligence.* Totowa, N J: Littlefield Adams.

Phelps, L., McGrew, K. S., Knopik, S. N., & Ford, L. (2005). The general (g), broad, and narrow CHC stratum characteristics of the WJ III and the WISC III tests: A confirmatory cross battery investigation. *School Psychology Quarterly, 20(1),* 66-88. doi: 10.1521/scpq.20.1.66.64191

Phelps, L., Rosso, M., & Falasco, S. L. (1984). Correlations between the Woodcock Johnson and the WISC-R for a behavior disordered population. *Psychology in the Schools, 21(4),* 442-446. doi: 10.1002/1520-6807(198410)21:4<442::AID-PITS2310210407>3.0.CO;2-6

Prewett, P. N., & Matavich, M. A. (1994). A comparison of referred students' performance on the WISC-III and the Stanford-Binet Intelligence Scale: Fourth Edition. *Journal of Psychoeducational Assessment, 12(1),* 42-48. doi: 10.1177/073428299401200104

Prifitera, A., Saklofaske, D. H., & Weiss, L. (2005). *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives*. San Diego, CA: Elsevier Science.

Prifitera, A., Weiss, L. G., Saklofaske, D. H., & Rolfhus, E. (2005). The WISC-IV in the clinical assessment context. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.) *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives.* (pp. 3-32). New York: Elsevier Academic Press.

Proctor, B. E., Floyd, R. G., & Shaver, R. B. (2005). Cattell-Horn-Carroll broad cognitiveability profiles of low math achievers. *Psychology in the Schools, 42(1),* 1-12. doi: 10.1002/pits.20030

Raskin, L. M., Bloom, A. S., Klee, S. H., & Reese, A. (1978). The assessment of developmentally disabled children with the WISC-R, Binet, and other tests. *Journal of Clinical Psychology, 34 (1),* 111-114. doi: 10.1002/1097-4679(197801)34:1<111::AID-JCLP2270340125>3.0.CO;2-Y

Reitan, R. M., & Wolfson, D. (1992). *Neuropsychological evaluation of older children.* South Tucson, AZ: Neuropsychology Press.

Reeve, R. E., Hall, R. J., & Zakreski, R. S. (1979). The Woodcock-Johnson Tests of Cognitive Ability: Concurrent validity with the WISC-R. *Learning Disability Quarterly, 2,* 63-69. Retrieved from http://www.jstor.org/stable/1510646

Reynolds, C. R., & Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales (RIAS) and the Reynolds Intellectual Screen Test (RIST) professional manual.* Lutz, FL: Psychological Assessment Resources.

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. A. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children-Second Edition: Consistency with Cattell-Horn-Carroll Theory. *School Psychology Quarterly, 22(4),* 511-539. doi: 10.1037/1045-3830.22.4.511

Riccio, C.A., & Hynd, G.W. (2000). Measurable biological substrates to verbal-performance differences in Wechsler Scores. *School Psychology Quarterly, 15(4)*, 386-399. doi: 10.1037/h0088796

Saklofske, D. H., Prifitera, A., Weiss, L. G., Rolfhus, E., & Zhu, J. (2005). Clinical interpretation of the WISC-IV FSIQ and GAI. In A. Prifitera, D. H. Saklofske, & L. G.

Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives.* New York: Academic Press.

Sanders, S., McIntosh, D. E., Dunham, M., Rothlisberg, B. A., & Finch, H. (2007). Joint confirmatory factor analysis of the Differential Ability Scales and the Woodcock Johnson Tests of Cognitive Abilities-Third Edition. *Psychology in the Schools, 44(2),* 119-138. doi: 10.1002/pits.20211

Sandoval, J. (2003). [Review of the Woodcock-Johnson III.]. In B. S. Plake & J. C. Impara (Eds.), The fifteenth mental measurements yearbook (pp. 1024-1027). Lincoln, NE: Buros Institute of Mental Measurements.

Sanville, D. & Cummings, J (1983). Concurrent validity of the Woodcock-Johnson tests of Cognitive Ability with the WISC-R: EMR children. *Psychology in the Schools, 20(3),* 298-303. doi: 10.1002/1520-6807(198307)20:3<298::AID-PITS2310200308>3.0.CO;2-W

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Sattler, J. M. (2008). *Assessment of children: Cognitive applications* (5th ed.). La Mesa, CA: Jerome M. Sattler, Publisher, Inc.

Sattler, J. M., & Dumont, R. (2004). *Assessment of children: WISC-IV and WPPSI-III supplement.* San Diego, CA: Jerome M. Sattler.

Schrank, F. A., McGrew, K. S., & Woodcock, R. W. (2001). Technical Abstract (Assessment Service Bulletin No. 2). Itasca, IL: Riverside Publishing. Retrieved from http://www.assess.nelson.com/pdf/asb-2.pdf

Shaughnessy, M. F. (2006). An interview with Amy Gabel about the WISC-IV. *North American Journal of Psychology, 8(1),* 135-143. Retrieved from http://proquest.umi.com.ezproxylocal.library.nova.edu/pqdweb?did=121335344 1&sd=3&Fmt=1&clientId=17038&RQT=309&VName=PQD

Shaw, S. R., Swerdlik, M. E., & Laurent, J. (1993). Review of the WISC-III. In B. A. Bracken, & R. S. McCallum (Eds.), *Journal of Psychoeducational Assessment, WISC-III Monograph*. (pp. 151-160). Brandon, VT: Clinical Psychology Publishing Co.

Silver, C. H. Ruff, R. M., Iverson, G. L., Barth, J. T., Broshek, D. K., Bush, S. S., et al. (2008). Learning disabilities: The need for neuropsychological evaluation. *Archives of Clinical Neuropsychology, 23(2),* 217-219. doi:10.1016/j.acn.2007.09.006

Spearman, C. E. (1904). “General intelligence,” objectively determined and measured. *American Journal of Psychiatry, 15(2),* 201-293. Retrieved from http://www.jstor.org/stable/1412107

Spearman, C. E. (1932). *The abilities of man*. New York: AMS Press.

Strauss, E., Sherman, E. M. S., & Spreen, O. (2006). *A Compendium of Neuropsychological Tests.* Oxford: University Press.

Strauss, E., Spreen, O., & Hunter, M. (2000). Implications of Test Revisions for Research. *Psychological Assessment, 12(3),* 237-244. doi: 10.1037//1040-3590.12.3.237

Sternberg, R. J., & Berg, C. A. (1986). Quantitative integration: Definitions of intelligence: A comparison of the 1921 and 1986 symposia. In R. J. Sternberg & D. K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition* (pp. 155-162). Norwood, NJ: Ablex.

Tabachnick, B. G., & Fidell, L. S. (1996). Using multivariate statistics (3rd ed.). Mahwah, NJ: Lawrence Erlbaum.

Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll Theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly, 19(1),* 72-87. Retrieved from http://proquest.umi.com.ezproxylocal.library.nova.edu/pqdweb?did=639207631&sid=3&Fmt=2&clientId=17038&RQT=309&VName=PQD

Thompson, P. L., & Brassard, M. R. (1984). Validity of the Woodcock-Johnson Tests of Cognitive Ability: A comparison with the WISC-R in ld and normal elementary students. *Journal of School Psychology, 22,* 201-208. doi: 10.1016/0022-4405(84)90040-2

Thompson, P. M., Cannon, T. D., Narr, K. L., van Erp, T., Poutanen, V. P., Huttunen, M.,et al. (2001). Genetic influences on brain structure. *Nature Neuroscience, 4(12),* 1253−1258. doi: 10.1038/nn758

Thorndike, R. M. (1997). The early history of intelligence testing. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 3-16). New York: Guilford.

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

Tupa, D. J., Wright, M. O., & Fristad, M. A. (1997). Confirmatory factor analysis of the WISC III with child psychiatric inpatients. *Psychological Assessment, 9(3),* 302-306. doi: 10.1037/1040-3590.9.3.302

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P. & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences*, *13(2)*, 159-188. doi: 10.1016/S1041-6080(02)00077-8

Vellutino, F. R. (2001). Further analysis of the relationship between reading achievement and intelligence: A response to Naglieri. *Journal of Learning Disabilities, 34(4),* 306-310. doi: 10.1177/002221940103400404

Vo, D. H., Weisenberger, J. L., Becker, R., & Jacob-Timm, S. (1994). Concurrent validity of the KAIT for students in grade six and eight. *Journal of Psychoeducational Assessment, 17(2)*, 152-162. doi: 10.1177/073428299901700205

Warner, T. D., Dede, D. E., Garvan, C. W., & Conway, T. W. (2002). One size does not fit all in specific learning disability assessment across ethnic groups. *Journal of Learning Disabilities, 35(6),* 500-508. Retrieved from http://proquest.umi.com.ezproxylocal.library.nova.edu/pqdweb?did=235887801&sid=2&Fmt=2&clientId=17038&RQT=309&VName=PQD

Wasserman, J. D., & Bracken, B. A. (2003). Psychometric characteristics of assessment procedures. In J. R. Graham & J. A. Naglieri (Eds.), *Handbook of psychology: Assessment psychology* (Vol. 10, pp. 43–66). Hoboken, NJ: Wiley.

Wasserman, J. D., & Tulsky, D. S. (2005). A history of intelligences assessment. In D. P. Flanagan & P. L. Harrison (Eds.). *Contemporary intellectual assessment (2nd ed.).* (pp.3-22). NY: Guildford Press.

Watkins, M. W. (2006). Orthogonal higher order structure of the Wechsler Intelligence Scale for Children—Fourth Edition. *Psychological Assessment, 18(1),* 123-125. doi: 10.1037/1040-3590.18.1.123

Watkins, M. W., Wilson, S. M., Kotz, K. M., Carbone, M. C., & Babula, T. (2006). Factor structure of the Wechsler Intelligence Scale for Children–Fourth Edition among referred students. *Educational and Psychological Measurement, 66(6),* 975-983. doi: 10.1177/0013164406288168

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

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. *American Psychologist, 30*, 135-139. doi: 10.1037/h0076868

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition.* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children–Fourth Edition*: *Technicaland interpretative manual.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (2004). *Wechsler Intelligence Scale for Children Integrated-Fourth Edition: Technical and interpretative manual.* San Antonio: TX: Psychological Corporation.

Witt, J. C., & Gresham, F. M. (1985). Review of Wechsler Intelligence Scale for Children-Revised. In J. V. Mitchell (Ed.), *The ninth mental measurement yearbook* (vol. 2, pp. 1716–1719). Lincoln, NE: Buros Institute of Mental Measurement.

Woodcock, R. W. (1990). Theoretical foundations of WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment, 8(3),* 231-258. doi: 10.1177/073428299000800303

Woodcock, R.W., & Johnson, M. G. (1977). *Woodcock-Johnson Psycho-Educational Battery*. Allen, TX: DLM Teaching Resources.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psychoeducational Battery Revised.* Itasca, IL: Riverside Publishing.

Woodcock, R. W., McGrew, K. S. & Mather, N. (2001). Examiner's Manual. *Woodcock Johnson III Tests of Cognitive Abilities.* Itasca, IL: Riverside Publishing.

Yeates, K. O. & Donders, J. (2005). The WISC-IV and neuropsychological assessment. In A. Prifitera, D. H. Saklofske, & L. G. Weiss (Eds.), *WISC-IV clinical use and interpretation: Scientist-practitioner perspectives* (pp. 415-434). *New* York: Elsevier Academic Press.

Ysseldyke, J., Shinn, M., & Epps, S., (1981). A comparison of the WISC-R and the Woodcock-Johnson tests of cognitive ability. Psychology in the Schools, 18(1), 15-19. doi: 10.1002/1520-6807(198101)18:1<15::AID-PITS2310180104>3.0.CO;2-W

Zhu, J., & Weiss, L. (2005). The Wechsler scales. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed., pp. 297-324). New York: Guilford Press.

Zimmerman, I. L., & Woo-Sam, J. M. (1997). Review of the criterion-related validity of the WISC-III: The first five years. *Perceptual and Motor Skills, 85(2),* 531-546. Retrieved from http://ammons.ammonsscientific.com/php/toc.php