NORM-REFERENCED COGNITIVE AND ACHIEVEMENT SCORES AS PREDICTORS OF STATE-WIDE HIGH-STAKES TEST SCORES WITH STUDENTS REFERRED FOR SPECIAL EDUCATION

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE

DOCTOR OF PHILOSOPHY

BY

JAMES M. TRINKLE II

DISERTATION ADVISOR: DR. ANDREW S. DAVIS

BALL STATE UNIVERSITY

MUNCIE, IN

JULY 2013

Acknowledgements

It is a special privilege to earn a Doctoral Degree, and many people do not get the chance to pursue such an achievement. I am honored to be so close to the completion of this journey; however, I realize that the completion of this journey would not have happened without the support of so many family, friends, and colleagues. This dissertation is dedicated to all of those who have believed, encouraged, and supported me in completing my Doctoral Degree.

I owe a great deal of thanks to many people for their support, but my sincerest gratitude goes to the following people for their support and dedication to me. First and foremost, I need to thank my wife, Lora, and our two boys, Zachary and Alex, for their love and patience. I have had to sacrifice much time with them to complete my graduate work, and there have been times that Lora has probably felt like a single parent. Yet she has always been understanding, patient, and supportive. Her commitment and dedication to our family has allowed me to pursue this dream. I love you very much! I also need to thank my parents, Sandy and John Kilmer, and Mike and Jan Trinkle, as well as the rest of my family. Over the past couple of years, I have had to miss out on some family vacations, graduations, and family time because I needed to complete work to meet deadlines. I want to thank them for their understanding and let them know how much I love them.

Next, I am extremely grateful to all of my committee members Dr. Andrew Davis, Dr. Holmes Finch, Dr. George Gaither, and Dr. Raymond Dean for their advice, encouragement, and support. I would like to express my sincerest gratitude to the chair of my committee, Dr. Andrew Davis for his guidance and dedication. Dr. Davis is the reason that I have been able to complete my dissertation. Although I had the capability to complete a dissertation, I was not focused and was floundering. With Dr. Davis' guidance, I not only became focused but energized about my work. I will never be able to give you the thanks you deserve, but know that you are the reason I did not become another ABD statistic. I would also like to give a special thanks to Dr. Holmes Finch for his support throughout graduate school. Dr. Finch went above and beyond to meet with me and help me with my dissertation statistics.

I would also like to thank the Marion Community School system. I have been working as a school psychologist in Marion for several years. Many of my colleagues have been anxiously awaiting me to finally complete my Doctoral Degree. It is not uncommon for me to hear, "Is it Dr. now?" I would like to thank all of the teachers, administrators, and staff who have supported me. I would also like to thank Mr. Stephen Edwards, Superintendent, and Dr. Kenneth Folks, Assistant Superintendent for their support of my work and for allowing me to use archival records to complete my dissertation.

Finally, I would like to thank one more special person, Mrs. Sheila Decaroli. Mrs. Decaroli was my Director of Special Education when I began working in Marion. However, Mrs. Decaroli became much more than my supervisor. She became my mentor and friend. Even after Mrs. Decaroli retired, she continued to ask for updates about my dissertation and genuinely wished for me to complete this degree. Thank you for your continued support.

Table of Contents

Chapter I

Introduction

High-Stakes Testing. Student achievement, particularly from underserved or disadvantaged populations was a topic of great interest during Lyndon B. Johnson's presidency. As part of an educational reform movement, The Elementary and Secondary Education Act (ESEA) of 1965 (PL 89 – 750, 1965) became part of President Johnson's "War on Poverty and Great Society programs." The main purpose of ESEA was improving the educational opportunities afforded to disadvantaged children (Duffy, Giordano, Farrell, Paneque, & Crump, 2008). A provision of ESEA was Title 1, which allowed the United States Department of Education to provide funding to schools and school districts with a high percentage of low-income families. However, according to Stein (2004), there was much debate over who qualified for this funding, and the continued ambiguity of Title 1 led to misuse of the funds. Furthermore, a study conducted between the years of 1976 and 1979, titled the *Sustaining Effects Study* (Carter, 1983), concluded Title 1 was essentially not efficacious with severely disadvantaged children (Fritzberg, 2004). Therefore, ESEA was amended four times between 1965 and 1980 (Thomas & Brady, 2005). Despite criticism of Title 1, Duffy et. al. (2008), noted that education reform continued to move toward more rigorous academic standards with the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983). However, there was still debate regarding who was accountable for the underachievement of disadvantaged students (Fritzberg, 2004). A shift to bring accountability to the forefront was approaching and appeared to become most dramatic after the 1988 Education Summit, "which led to the development of national education goals …" (Braden & Tayrose, 2008, p. 20). Furthermore, it was during the late 1980s to the mid-1990s that the requirement of individual states to document the academic achievement of disadvantaged students was implemented through amendments to Title 1 (Duffy et. al., 2008). In 1994, accountability for the academic achievement of disadvantaged students was continuing to be brought to the forefront. Indeed, President Bill Clinton's administration reauthorized ESEA in 1994 and titled it the *Improving America's Schools Act* (IASA) (Public Law 103 - 382, 1994), which explicitly utilized the term "Adequate Yearly Progress" (AYP) (Fritzberg, 2004). Furthermore, states were now required to assess disadvantaged students on statewide assessments in the same manner as all other children, yet there was still a wide-range of variability in the "rigor" of these assessments (Fritzberg, 2004, p. 10).

The educational reform movement that began with ESEA, and reauthorized as IASA, continues today in its current form known as the No Child Left Behind (NCLB) Act of 2001 (Public Law 107 – 110, 2001), which was passed into law January 28, 2002, and mandates that all public school students, including traditionally underserved populations, such as low income students, special needs students, and minority students, pass a state-administered assessment (Public Law 107 - 110, 2001). NCLB proposed to meet these goals by requiring each state to develop and administer a standardized state-wide assessment in basic skills to all students in certain grades, with the implication that the results of these tests would impact the federal funding these schools received (Hursh, 2005; Bloomfield, 2007; Tichá, Espin, & Wayman, 2009). Thus, the era of statewide high-stakes testing began.

Indiana Statewide Testing for Educational Progress- Plus. Due to the legislation of NCLB, all public schools in Indiana are required to use standardized testing to assess the academic progress of all students, including students in special education. The state administered assessment for Indiana is known as the Indiana Statewide Testing for Educational Progress- Plus (ISTEP+) (Indiana Department of Education, ISTEP+ Program Manual, 2003). In its current form, ISTEP+ measures English/language arts and mathematics to students in grades 3 through 8, science in grades 4 and 6, and social studies in grades 5 and 7 (Indiana Department of Education, ISTEP+ Program Manual, 2003). ISTEP+ is based on Indiana's current academic standards and is comprised of multiple choice questions, as well as short answer and essay questions, which are administered in the spring of each school year. Results from the ISTEP+ are provided in the form of a student score in the areas assessed, which is then compared to a minimum performance standard in each of those areas. Based on the performance standards, known as cut-scores, the student will receive a qualitative statement of Did Not Pass, Pass, or Pass+ (Indiana Department of Education, ISTEP+ Program Manual, 2003). According to IC 20-35-5-1, one of the purposes of ISTEP+ is to "… compare achievement of Indiana students to achievement of students on a national basis." Furthermore, ISTEP is utilized to help make decisions regarding "identifying students who may need remediation…" and "diagnosing individual student needs."

Cognitive Abilities and Achievement. Traditionally, assessment, including the use of norm-referenced tests, for the purposes of identifying students in need of special education has been the primary role for school psychologists (Fagan & Wise, 2000). School psychologists often employ the use of a multifactored assessment comprised of several types of tests including a cognitive ability or intelligence test and achievement tests (Fagan & Wise, 2000). As Burszytn (2007) notes, the primary use of the cognitive ability or intelligence test and the achievement test as part of a psychoeducational evaluation is to aid in educational planning for students who are being assessed for special education programming as prescribed by the Individual with Disabilities Education Improvement Act (IDEIA) of 2004.

Although intelligence testing has a long history in the field of school psychology, the inception of the federal special education law known as Public Law 94-142 (1975) made the use of intelligence tests essential for the identification of cognitive disabilities and specific learning disabilities (SLDs) given that the determination of SLD historically was primarily conducted by identifying a discrepancy between intelligence and achievement (Fletcher, Lyon, Fuchs, & Barnes, 2007). Indeed, in 1977 the U.S. Office of Education defined SLD as "a severe discrepancy between achievement and intellectual ability in one or more of the following areas…" (Fletcher, Lyon, Fuchs, & Barnes, 2007, p. 20).

However, two important events occurred since the original inception of P.L. 94-142 (1975), and the original definition of specific learning disability. First, P.L. 94-142 (1975) went through several reauthorizations culminating in its present form known as the Individuals with Disabilities Education Improvement Act of 2004 (IDEIA, 2004; Bloomfield, 2007). IDEIA generally retained the original definition of SLD but no longer required the use of an ability-achievement discrepancy (Fletcher, Lyon, Fuchs, & Barnes, 2007). However, IDEIA (IDEIA, 2004, section 1401) continued to define SLD as "… a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, which may manifest itself in an imperfect ability to listen, speak, read, write, spell, or to do mathematical calculations."

The second important event leading to a change in the way specific learning disabilities were identified was the culmination of years of research on the nature of intelligence, resulting in the "empirically supported and theoretically sound model" of intelligence known as the Cattell-Horn-Carroll (Horn, 1991; Horn & Noll, 1997) theory of cognitive abilities (Taub & McGrew, 2004; Floyd, Evans, & McGrew, 2003). The contemporary model of CHC theory is a combination of the original Cattell-Horn G*f*-G*c* (Horn, 1991; Horn & Noll, 1997) theory and the Carroll three-stratum theory (Carroll, 1993, 1998), which states that human intelligence is comprised in a hierarchical nature with three layers: stratum III (*g* or general intelligence), stratum II (10 broad cognitive abilities), and stratum I (70 narrow cognitive abilities) (Newton & McGrew, 2010; Floyd, McGrew, & Evans, 2008; Taub, Floyd, Keith, & McGrew, 2008). According to Newton & McGrew (2010), Fluid Reasoning (G*f*), Comprehension-Knowledge (G*c*), Short-term Memory (G*sm*), Visual or Visuospatial Processing (G*v*), Auditory Processing (G*a*), Long-Term Retrieval (G*lr*), Processing Speed (G*s*), and Quantitative Knowledge (G*q*), are the broad cognitive abilities assessed by some modern intelligence tests, with the narrow cognitive abilities subsumed under each. Furthermore, research has supported that CHC abilities have aided in the understanding of reading, math, and written language (Newton & McGrew, 2010; Floyd, McGrew, & Evans, 2008; Taub, Floyd, Keith, & McGrew, 2008; Evans, Floyd, McGrew, & Leforgee, 2001). Even those researchers who have theorized different models of intelligence are espousing that the best way to understand the relationship between cognitive abilities and academic achievement is through measuring the various aspects of intelligence (Naglieri & Otero, 2010).

Since one of the most common referrals for consideration of special education programming is to determine whether a student has a specific learning disability, and learning disabilities account for more than half of all children receiving special education services (Fletcher, Lyon, Fuchs, & Barnes, 2007), the use of cognitive ability tests continues to be a cornerstone of the school psychologist's tool kit. The use of cognitive ability tests to validly predict academic achievement is well documented in the research (e.g., Floyd, McGrew, & Evans, 2008; Rohde & Thompson, 2007; Glutting, Watkins, Konold, & McDermott, 2006; Taub & McGrew, 2004; Evans, Floyd, McGrew, & Leforgee, 2001). Historically, the Wechsler Scales, including its most recent version, The *Wechsler Intelligence Scale for Children- Fourth Edition* (WISC-IV; Wechsler, 2003), and *The Wechsler Individual Achievement Test- Second Edition* (WIAT-II; The Psychological Corporation, 2001) have traditionally been considered the 'gold-standard' or 'go-to' battery in assessing cognitive ability in relation to achievement (Taub & McGrew, 2004). However, over the past decade the *Woodcock-Johnson Third Edition Tests of Cognitive Ability* (WJ-III-COG; Woodcock, McGrew, & Mather, 2001a) has come to prominence as a measure to predict achievement in children because of its empirically-sound underlying theoretical structure, CHC theory (Evans, Floyd, McGrew, & Leforgee, 2001; Floyd, McGrew, & Evans, 2008). Additionally, its academic counterpart, the *Woodcock-Johnson Third Edition Tests of Achievement* (WJ-III-ACH; Woodcock, McGrew, & Mather, 2001b) has also become one of the most frequently used tests of achievement (Taub & McGrew, 2004). Together, the WJ-III-COG and the WJ-

III-ACH are a powerful battery to predict achievement based on cognitive abilities (Gregg, Hoy, Flaherty, Norris, Coleman, Davis, & Jordan, 2005). Not only are they developed on a technically and empirically sound theoretical model, namely CHC theory, but the WJ-III-COG and WJ-III-ACH were co-normed from the same sample of subjects (McGrew & Woodcock, 2001). McGrew & Woodcock (2001) assert that the tests being co-normed together allows for more diagnostic accuracy given that direct comparisons can be made between and within a subject's scores. As noted previously, there is extensive research documenting the capability of cognitive ability testing, such as the WJ-III-COG to predict academic achievement as assessed by the WJ-III-ACH (Taub, Floyd, Keith, & McGrew 2008; Taub & McGrew, 2004; Evans, Floyd, McGrew, & Leforgee, 2001).

The use of cognitive and achievement tests will likely continue to be a foundation of the school psychologists toolkit given the definition of SLD under the federal special education law, IDEIA (2004).Furthermore, Mägi & Kikas (2009) assert that school administrators value the information provided from the assessment and testing role of the school psychologist for whom they supervise, thereby promoting the continued use of this role as one of the school psychologist's primary functions. Some state school psychology associations have also specifically delineated the importance of the use of assessment. Fagan & Wise (2000; p.21) quote the California Association of School Psychologists position on the use of assessment, "Assessment is the cornerstone of educational and psychological services delivery. Assessment is required to define pupil needs, guide children's education, and to provide data, which may be used to evaluate educational outcomes. Educational research shows that assessment is necessary…"

While the research on the relationship between standardized cognitive tests and high-stakes testing is quite sparse, there is a plethora of research regarding the relationship between standardized cognitive and standardized academic achievement tests (McGrew & Wendling, 2010). The consensus of many of these studies is that cognitive abilities are substantially correlated with and good predictors of academic achievement. A study conducted by Glutting, Watkins, Konold, & McDermott (2006, p. 110) suggested that *g* or global intelligence was the greatest predictor of reading achievement; however, beyond *g*, the factor index of verbal comprehension had a "medium effect" on reading achievement. Further, a study by Naglieri, De Lauder, Goldstein, & Schwebech (2006) concluded that cognitive ability was significantly correlated with academic achievement. Indeed, their study indicated that cognitive ability was substantially correlated with reading, math, and written language skills. One of the key points in their study was the use of a cognitive ability measure that did not include "achievement-like" subtests, which they assert "would inflate the correlation between tests of ability and achievement (Naglieri, De Lauder, Goldstein, & Schwebech, 2006, p. 73). Similarly, a study conducted by Evans, Floyd, McGrew, & Leforgee (2001) revealed distinct CHC cognitive abilities were predictive of reading achievement, especially at developmental age ranges. For example, G*c* was significantly correlated with decoding skills, as well as reading comprehension from childhood through adolescence, whereas G*sm* demonstrated a substantial relationship to decoding skills only from childhood through adolescence. However, other factor scores, G*a*, G*s*, and G*lr*, were found to be significant for decoding and comprehension only during the foundational years of reading skill acquisition.

While the area of reading achievement seems to have the largest research base, the relationship between cognitive abilities and mathematics achievement has a much smaller research base (Floyd, Evans, & McGrew, 2003). However, several studies have demonstrated that G*f* and G*c* were consistently the strongest cognitive ability predictors of math achievement during distinct developmental age ranges, namely childhood and adolescence (Floyd, Evans, & McGrew, 2003; Taub, Floyd, Keith, & McGrew, 2008). G*s* also was significantly correlated with mathematics achievement (Taub, Floyd, Keith, & McGrew, 2008). While G*sm* was significantly correlated with math achievement, one of the narrow abilities subsumed under G*sm*, namely working memory was more predictive of math achievement than the broad cluster of G*sm* (Floyd, Evans, & McGrew, 2003). The importance of such studies was that they refuted earlier research that stated *g* was the most significant predictor of mathematics achievement, and these studies also demonstrated the need to assess these cognitive abilities when evaluating a student who has difficulties in mathematics (Floyd, Evans, & McGrew, 2003; Taub, Floyd, Keith, & McGrew, 2008).

Similar to the relationship between standardized cognitive tests and high-stakes, the relationship between standardized academic achievement tests and high-stakes testing is also scarce. However, when considering academic achievement's relationship to high-stakes testing, there is research evaluating the relationship of curriculum-based measurement (CBM) to high-stakes testing. More specifically, CBM reading measures have consistently been shown to be quite predictive of broad reading abilities (Stage & Jacobsen, 2001; McGlinchey & Hixson, 2004; Hintze & Silberglitt, 2005). Indeed, in the Stage & Jacobsen (2001) study, fourth grade students were administered oral reading fluency CBM probes during the fall, winter, and spring, and then administered a state-mandated, high-stakes reading test at the end of the spring term. Results indicated that the oral reading fluency CBM probes, especially those taken in the fall, were highly predictive of passing or failing the high-stakes reading test (Stage & Jacobsen, 2001). One implication from this study was the ability to identify students who are at-risk of failing a high-stakes reading test and provide interventions early in the school year (Stage & Jacobsen, 2001).

McGlinchey & Hixson (2004) replicated the Stage & Jacobsen (2001) study due to potential problems with the generalizability of the Stage & Jacobsen (2001) study. Indeed, the current authors assert that the results of the Stage & Jacobsen (2001) study might not be generalizable due to utilizing a single state reading assessment. The current study utilized a different state reading assessment with a more representative student population and a bigger sample of students (McGlinchey & Hixson (2004). Results of the study were similar to the Stage & Jacobsen (2001) study, in which oral reading fluency CBM was highly predictive of performance on a state-mandated, high-stakes reading assessment (McGlinchey & Hixson, 2004).

Similar to the two aforementioned studies, a study by Hintze & Silberglitt (2005) concluded that reading CBM probes were predictive of performance on a state-mandated, high-stakes reading test. During the current study curriculum-based oral reading fluency probes were administered three times per year for students in first through third grade. During third grade, the students were required to participate in a state-mandated, high-stakes reading test to determine the student's level of reading proficiency as set forth by state standards. Of particular interest in the current study was that reading CBM was capable of predicting a student's performance on the third grade high-stakes reading test as early as the first grade (Hintze & Silberglitt, 2005).

Rationale of the Study

Federal legislation such as No Child Left Behind (2001) and IDEIA (2004) contain inconsistent language regarding the assessment of children with disabilities. First, NCLB (2001) requires that children in special education be assessed utilizing the same state standards as their non-special education peers. However, IDEIA (2004) requires special education students to be assessed on goals or standards that are based to the student's individual needs. Second, the purpose of assessment is different under each law. Under NCLB (2001), students in special education are to be assessed to measure adequate yearly progress on the state standards, whereas under IDEIA (2004), students in special education are to be assessed for identification purposes and for monitoring progress on IEP goals (Klotz, 2012).

Next, there has been much criticism related to what statewide high-stakes tests really measure and whether the results are truly indicative of student achievement (Duffy et al, 2008; Rushton & Joula-Rushton, 2008; Zimmerman & Dibenedetto, 2008). According to several studies, there is little to no empirical data to support that statewide high-stakes testing can predict academic achievement, especially in the area of reading, or improve achievement outcomes (Afflerbach, 2005; Schulte, Villwock, Whichard, & Stallings, 2001). Additionally, researchers have asserted that the results of statewide high-stakes tests are either rarely utilized to direct instruction or are not sensitive enough to describe the specifics of the academic deficit (Afflerbach, 2005; Zimmerman & Dibenedetto, 2008).

Most of the criticisms of high-stakes testing have been directed toward the results of statewide high-stakes testing with the general population of students; however, others argue that the implications of statewide high-stakes testing results are worse for students that fall into other categories, such as special education. Johnson (2005) asserts requiring students with disabilities to take high-stakes tests may inadvertently set them up for failure, as the minimum level of standards may be unattainable when the student's disability is not taken into account. Furthermore, others have declared students with disabilities have less exposure to the general education curriculum being tested thereby putting them at a disadvantage when taking high-stakes tests (Katsiyannis, Zhang, Ryan, & Jones, 2007).

The debates regarding the true consequences of high-stakes tests highlight the need for research regarding the relationship between academic achievement and high-stakes testing. A review of the literature suggests there is a dearth of research regarding the relationship between statewide high-stakes testing and norm-referenced standardized achievement testing, such as the *Woodcock-Johnson Third Edition Tests of Achievement* (WJ-III-ACH; Woodcock, McGrew, & Mather, 2001b). Yet, it is these standardized norm-referenced achievement tests which are most readily used by school psychologists and case conference committees when determining whether a student qualifies for special education services. The use of tests, such as the WJ-III-ACH (Woodcock, McGrew, & Mather, 2001b), by school psychologists are due to the rigorous standardization and normative processes these tests undergo. For example, the WJ-III-ACH was normed on a nationally representative sample of 8, 818 subjects, which was stratified for ten specific individual and community variables in order to obtain a sample that was representative of the U.S. population (McGrew & Woodcock, 2001). Furthermore, when a variable, such as race, had a subset with a low percentage of occurrences in the population, that particular subset was oversampled to ensure adequate representation and to control for concerns with reliability (McGrew & Woodcock, 2001). Mather, Wendling, & Woodcock (2001) also point out that at the time of its development, the WJ-III-ACH was the only standardized test of achievement to control for socioeconomic status. To control for errors in testing conditions, the authors chose to utilize a team of trained examiners who were hired by the project and were under the direct supervision of a project staff member (McGrew & Woodcock, 2001).

Significance of the Study

A result of initiatives, such as NCLB (2001) has brought early identification and interventions to the forefront of educational concerns (McGlinchey & Hixson, 2004), but as previously noted there is misalignment in the federal laws when it comes to identifying and assessing students with special needs. Therefore, if students in special education are required to be assessed using statewide high-stakes tests with standards that apply to all children, then it is paramount to study the relationship between these statewide high-stakes tests and empirically sound, theoretically grounded tests of achievement and cognitive abilities. Indeed, such an understanding would be most useful to school psychologists and school personnel when determining identification and interventions for students in special education, especially given that two of the purposes of ISTEP+ are to identify individual student needs and identifying students in need of remediation services (Indiana Department of Education, ISTEP+ Program Manual, 2003). Furthermore, an alignment between common assessment techniques used by school psychologists and high-stakes tests would be important given that school psychologists often provide recommendations for accommodations and modifications to be used on high-stakes tests.

Another important reason for the current study is the lack of research between standardized cognitive and achievement tests and high-stakes testing, namely the ISTEP+. According to Flanagan, Ortiz, Alfonso, & Mascolo (2001) the application of CHC theory to the understanding of academic achievement is limited.

Current Study

The current study used an archival data set of children under the age of 18, who had been referred for special education services and received a psychoeducational evaluation that contained scores on the *Woodcock-Johnson III Tests of Cognitive Abilities* (Woodcock, McGrew, & Mather, 2001a), *Woodcock-Johnson III Tests of Achievement* (Woodcock, McGrew, & Mather, 2001b), and *Indiana Statewide Test for Educational Progress- Plus* (Indiana State Department of Education, I.s., 1988). Current research investigating the ability of the WJ III Battery to predict ISTEP+ scores does not currently exist.

Research Questions

R_1 What is the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ III COG, and English/language arts performance on the ISTEP+?

R_2 What is the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ III COG, and Math performance on the ISTEP+?

R_3 What is the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ III ACH, and reading performance on the ISTEP+?

R_4 What is the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ III ACH, and Math performance on the ISTEP+?

Limitations of the Study

One limitation in the current study was the use of an archival data set. Indeed, the researcher might be constrained by the type of data available for analysis. For example, parental, specifically maternal, education attainment has been shown to be a predictor of a student's academic achievement (Magnuson, 2007); however, information regarding parental education was not able to be gathered. Therefore, extraneous sources of variance and alternative explanations of effects might not be able to be accounted for with the use of the archival data set. Indeed, the data set only included students who were referred for and received special education services after an initial psychoeducational evaluation had been completed. An important comparison group might be a data set of students who had not been referred for special education services.

Another limitation of using an archival data set might be the quality of the archival data that will be made available. The data collected for these analyses might be subject to errors due to changes in record-keeping over time, as well as assessment scoring procedures. Because the archival data will be scribed on separate coding or data collection sheets and subsequently entered into a secure electronic database, there is the possibility the data is subject to error secondary to the initial data collection.

External validity might also be a limitation of the current study. The participants in the archival data set are from a suburban Midwestern area, with a high rate of poverty. Since SES has been shown to have deleterious effects on academic achievement (Cooper & Schlesser, 2006), rates of passing the ISTEP+ might not be generalizable to other regions with higher levels of SES.

Furthermore, the generalizability of the results to samples from other states likely will be unknown given that each state sets its own standards for academic proficiency. Therefore, the use of a different state assessment might affect the hypothesized outcome of the results.

Delimitations of the Study

A strength of the current study is the psychometric validity of the assessments utilized. The WJ-III-ACH is cited as a frequently used achievement battery to assess achievement skills in children and adolescents (Taub & McGrew, 2004). Taub & McGrew (2004) state even though the Wechsler Scales have traditionally been the most frequently used tests to measure intelligence, much interest has surrounded the WJ III Cog due to its "empirically supported and theoretically sound model of the structure of human intelligence" (p.73). Indeed, many recent intelligence test publishers have applied CHC Theory, which the WJ III Battery is drawn from, to their own particular testing batteries (Floyd, McGrew, & Evans, 2008). Furthermore, many studies have cited the ability of the CHC broad (stratum II) and narrow (stratum 1) cognitive abilities to predict academic achievement (Floyd, McGrew, & Evans, 2008; Taub, Floyd, Keith, & McGrew, 2008; Taub & McGrew, 2004; Floyd, Evans, & McGrew, 2003; Evans, Floyd, McGrew, & Leforgee, 2001).

List of Terms

Adequate Yearly Progress (AYP): Adequate Yearly Progress refers to the academic progress students make on a statewide test of academic skill proficiency in a given school district. According to Katsiyannis, Zhang, Ryan, & Jones (2007), AYP is demonstrated by a school when "(a) at least 95% of all enrolled students, including specific subgroups (e.g., African Americans, students with special needs) participate in testing; (b) all students and subgroups score at least proficient on the state's AYP targets; and (c) all students and subgroups meet AYP targets for graduation or attendance."

Bookmark Procedure: The Bookmark procedure is utilized to set cut-scores or a score that is used to separate test takers into different categories, such as passing or failing, for that particular exam. The Bookmark procedure is based on item-response theory, which estimates how well specific items on an exam work (Lin, 2006).

Cattell-Horn-Carroll Theory (CHC theory): The contemporary model of CHC theory is a combination of the original Cattell-Horn G*f*-G*c* theory and the Carroll three-stratum theory, which states that human intelligence is comprised in a hierarchical nature with three layers: stratum III (*g* or general intelligence), stratum II (10 broad cognitive abilities), and stratum I (70 narrow cognitive abilities) (Newton & McGrew, 2010; Floyd, McGrew, & Evans, 2008; Taub, Floyd, Keith, & McGrew, 2008).

Cognitive Abilities: Cognitive ability refers to the brain-based skills needed to carry out a task. In the field of school psychology the use of the term cognitive abilities is often used in place of the term intelligence. Historically, intelligence has had many definitions but was generally thought to be a standard index of general broad cognitive ability. After much research and empirical support, intelligence is now generally thought to encompass approximately eight broad based abilities, which subsume approximately 70 narrow abilities (Newton & McGrew, 2010).

Criterion-referenced Testing: Criterion-referenced testing refers to evaluating an individual student's performance relative to an established criterion. Criterion-referenced testing often focuses on mastery of content, such as whether the student has achieved competence in an academic skill area.

Elementary and Secondary Education Act (ESEA): ESEA was first enacted under President Lyndon B. Johnson's administration to improve the educational opportunities for disadvantaged children (Duffy, Giordano, Farrell, Paneque, & Crump, 2008). ESEA continues through its most reauthorization known as No Child Left Behind (NCLB, 2001).

Free Appropriate Public Education (FAPE): The purpose of FAPE is to ensure that students with disabilities receive regular and special education services designed to meet the student's specific needs in order for the student to receive an educational benefit. Additionally, educational services are to be provided at public expense with the exception of fees that are charged to all students (Turnbull, Stowe, & Huerta, 2007; U.S. Department of Education, 2010).

High-Stakes Tests: The use of the term high-stakes test is somewhat of a misnomer. Some believe that high-stakes testing refers to the test itself; however, high-stakes are not a characteristic of the test itself but rather of the consequences placed on the outcome (Plake, 2011).

Indiana Statewide Testing for Educational Progress- Plus: *Indiana Statewide Testing for Educational Progress- Plus* (ISTEP+) is Indiana's high-stakes test administered yearly to determine whether Indiana students are meeting state academic standards in the areas of English/language arts, math, science, and social studies (Indiana Department of Education, ISTEP+ Program Manual, 2003).

Individuals with Disabilities Education Act of 2004: The Individuals with Disabilities Education Act of 2004 is the United States federal special education law. Its primary focus is the improvement of outcomes, particularly students with disabilities (IDEIA, 2004), and is the reauthorization of PL 94-142.

Least Restrictive Environment (LRE): The purpose of LRE is to ensure that students with disabilities have access to the general education curriculum and other programs afforded to students without disabilities to the greatest extent possible (Turnbull, Stowe, & Huerta, 2007).

No Child Left Behind: The No Child Left Behind (NCLB) Act of 2001 was a legislative act signed into law January 8, 2002, which mandated that all schools be held accountable for student progress, including students in special education. Furthermore, NCLB requires special education students to be assessed using state academic standards that apply to all children (NCLB, 2001).

Norm-referenced Testing: Norm-referenced testing refers to evaluating an individual student's performance relative to a similar group of students' performance on the same measure.

Parallel Forms: Parallel forms refer to having multiple forms of an exam, with the notion that the forms have been developed as similar as possible to one another.

Psychoeducational Evaluation: For purposes of this paper, the term psychoeducational evaluation is used interchangeably with the term educational evaluation as defined by Article 7, Indiana's Special Education Law. Educational evaluation has been defined as: "Procedures used in accordance with 511 IAC 7-40 and 511 IAC 7-41 to provide information about a student's disability or suspected disability for the student's CCC to determine the following:

(1) Whether a student is eligible for special education and related services.

(2) If eligible, the nature and extent of the special education and related services that the student needs.

(b) Based on the suspected disability or disabilities, the educational evaluation may address the following:

(1) Development.

(2) Cognition.

(3) Academic achievement.

(4) Functional performance or adaptive behavior.

(5) Communication skills.

(6) Motor skills and sensory responses.

(7) Available medical and mental health information that is educationally relevant.

(8) Social and developmental history.

(9) Analysis of other factors.

(10) Other assessments or information necessary to determine eligibility and inform the student's CCC.

(Indiana State Board of Education; 511 IAC 7-32-30; filed Jul 14, 2008, 1:24 p.m.: 20080813-IR-511080112FRA).

Public Law 94-142: Public Law 94-142 (PL 94-142) enacted in 1975 was an amendment to the 1970 Education of the Handicapped Act (EHA). The purpose of PL 94-142 was to guarantee the educational rights of children with disabilities, by providing a "*free and appropriate public education* (FAPE) in the *least restrictive environment* (LRE)," under the guidance of the U.S. Department of Education (Bloomfield, 2007).

Chapter II

Review of the Literature

The following review of the literature is organized into several sections that are relevant to the investigation between standardized, norm-referenced cognitive and standardized, norm-referenced achievement measures and their relationship to state-mandated, high-stakes testing. The first section will focus on the history of high-stakes testing, along with an overview of the federal education laws that have mandated the use of high-stakes testing. Additionally, Indiana's high-stakes testing program, the Indiana Statewide Testing for Educational Progress- Plus (ISTEP+) will be reviewed. The second section of this review of the literature will be an overview of the history of standardized, norm-referenced academic achievement testing and its relationship to high-stakes testing. Finally, a synopsis of the history of standardized intelligence testing, including the many theories intended to explain what intelligence/cognitive abilities are and how they have been measured will be addressed. Additionally, the contemporary practice of defining and measuring cognitive abilities and their relationship to academic achievement will be offered.

History of High-Stakes Testing

No Child Left Behind. Psychoeducational and psychological assessment has long been associated with education and has had a multitude of purposes, with the general implication of measuring and guiding student learning (Haladyna, 2002; Vogler & Virtue, 2007). According to Moon (2009), the current iteration of high-stakes testing began in the mid-19th century when such tests were used to compare classrooms. Historically, the use of standardized test scores were relatively benign in that test scores were used to evaluate the school curriculum to change the curriculum when student achievement did not seem to match the curriculum (Haladyna, 2002). However, both internal and external factors are known to play a significant role in how students achieve. Internal factors include, but are not limited to, attitude toward school, motivation, self-concept, and intelligence (Haladyna, 2002). External factors include, but are not limited to, the quantity and quality of teaching instruction, school leadership, school funding, student mobility, parental education, and socioeconomic status (Haladyna, 2002). As reform began to take place in the field of education, many of these external factors came to the forefront. However, separating the internal influences on achievement from external influences can be quite difficult (Haladyna, 2002).

Until the middle of the last century the main use of achievement test scores was to match the curriculum with what students were learning (Haladyna, 2002) and it was the responsibility of state and local policy-makers to determine education curricula and policies (Duffy, Giordano, Farrell, Panque, & Crump, 2008; Fritzberg, 2004; & Hursh, 2005). However, educational reform was taking place, and the federal government was beginning to take a more prominent role in local education policies (Fritzberg, 2004). According to Fritzberg (2004), one of the first major educational reforms in the United States was the Elementary and Secondary Education Act (ESEA) of 1965 (P.L. 89-750), which started to place more control of local education policies at the federal level of government. ESEA (1965) was part of President Johnson's "War on Poverty and Great Society's programs," with the main focus of ESEA (1965) being early access to educational experiences by all students, especially students who came from impoverished or disadvantaged backgrounds (Duffy et. al., 2008, p. 55). One provision of ESEA (1965) provided federal grants to assist states in educating students with disabilities (Turnbull, Stowe, & Huerta, 2007). Another provision of ESEA (1965) was Title 1 (20 U.S.C. 6301 et seq.), whose main purpose was to ensure that all children, especially children living in poverty, have a fair testing opportunity in order to reach at least minimum proficiency on state-mandated academic assessments (Haladyna, 20002). Title 1 (20 U.S.C. 6301 et seq.) allowed the United States Department of Education to provide funding to schools and school districts with a high percentage of low-income families (Duffy et al., 2008; Fritzberg, 2004). Yet, there was much ambiguity over who qualified for this funding, which led to misuse of the funds (Stein, 2004). Furthermore, the efficaciousness of Title 1 programs was being called into question after the *Sustaining Effects Study* (Carter, 1983) was conducted between 1976 and 1979 (Fritzberg, 2004). This study suggested that students in Title 1 programs showed more progress in reading than math and also showed more progress during the elementary years rather than the intermediate years. Furthermore, the study suggested Title 1 programs were generally not efficacious with students considered disadvantaged (Fritzberg, 2004). Moreover, accountability for student achievement was still up for debate, and ESEA (1965) was reauthorized several times between 1965 and 1980 (Thomas & Brady, 2005).

Due to these concerns with Title 1, the role of the federal government in local education decisions, as well as funding for Title 1 was reduced by President Reagan's administration (Duffy et al., 2008). Despite the reduction in funds and the reduction in the federal government's role in local education policies, the educational reform movement continued to move forward, and appeared to gain momentum after the publication of *A Nation at Risk* (National Commission on Excellence in Education, 1983), which shifted educational decision-making from the local to the state level (Hursh, 2005). Furthermore, educational reform initiatives began to call for more rigorous academic standards and the implementation of standardized tests with the focus of student learning shifting from a focus on the curriculum to a focus on achievement standards, such as student performance (Duffy et al., 2008; Hursh, 2005). Yet, debate continued over who was accountable for the underachievement of disadvantaged students (Fritzberg, 2004).

While assessment of student performance has a long-standing history in the field of education, this shift to bring it to the forefront of educational reform continued from the late 1980s through the mid-1990s, but appeared to become most dramatic after the 1988 Education Summit, "which led to the development of national education goals …" (Braden & Tayrose, 2008, p. 575). Additionally, Title 1 was amended during the late 1980's, and individual states were then responsible for identifying and documenting the achievement standards of disadvantaged students (Duffy et al., 2008; Fritzberg, 2004). Accountability for the academic achievement of disadvantaged students continued as an education reform initiative during President Bill Clinton's administration. Indeed, under President Clinton's administration, ESEA (1965) was reauthorized in 1994 and titled the *Improving America's Schools Act* (IASA) (P.L. 103–382, 1994). Turnbull, Stowe, & Huerta (2007) noted that public education reform under IASA (1994) provided a linkage between general and special education and provided outcomes-based accountability measures and evaluations of schools. IASA (1994) required individual states to not only identify and document the achievement of disadvantaged students, but to also assess disadvantaged students in the same manner as all other children on statewide assessments, and for the first time, the term ‘Adequate Yearly Progress’ (AYP) was explicitly utilized (Fritzberg, 2004). Yet, Hursh (2005) asserted that the implication of these statewide assessments, as well as the implementation of standards and standardized testing varied from state to state. Likewise, Fritzberg (2004) stated there was a wide-range of variability in the “rigor” of these assessments (p. 10).

Finally, the passage of the No Child Left Behind Act (NCLB) of 2001 (P.L. 107–110, 2001) made the consequences of state-based standardized assessments more critical to educational decision-making (Duffy et al., 2008; Fritzberg, 2004; Hursh, 2005). The No Child Left Behind Act of 2001 (NCLB), also known as Public Law 107-110, was proposed by President George W. Bush in 2001 and signed into law on January 8, 2002 (Bloomfield, 2007; Dufy et al., 2008; Turnbull, Stowe, & Huerta, 2007). NCLB (2001) was the most recent reauthorization of ESEA (1965) and proposed to improve the achievement of all students, especially traditionally-underserved populations, such as students from impoverished backgrounds, students with special needs, and minority students, by setting high academic standard and instituting measurable goals (Bloomfield, 2007). NCLB (2001) proposed to implement these improvements to student achievement by requiring each state to establish academic standards, especially in reading and math, for each grade level, as well as to create standardized tests to assess these standards for children in grades three through eight (Bloomfield, 2007; Hursh, 2005). Furthermore, NCLB (2001) required schools to present data aggregated by grade, race, gender, language status, socio-economic status, and students receiving special education

(Bloomfield, 2007; Fritzberg, 2004; Hursh, 2005). Moreover, NCLB (2001) gave the federal government greater power by allowing the federal government to determine the adequacy of each state's standards, accountability systems, and allowing the federal government to determine what interventions should take place at the state level if students consistently failed to meet the standards (Duffy et al., 2008).

Individual with Disabilities Education Act. While one might think that all students traditionally took part in state-wide standardized assessments, it was not until the reauthorization of the Individuals with Disabilities Education Act (IDEA) in 1997 (P.L. 105-17, 1997) that students in special education were mandated to be included in state-wide assessments and that their progress toward reaching the general education academic standards were tracked (Turnbull, Stowe, Huerta, 2007). However, to fully understand the educational rights of students with disabilities, one must review the history of special education in the United States.

According to Turnbull, Stowe, & Huerta (2007), the U.S. federal government did not become involved with the education of students with disabilities until 1966 through a provision of ESEA (1965), which provided federal grants to assist states in educating students with disabilities. Although there was this provision to ESEA (1965), many local education agencies continued to deny or exclude student with disabilities from participating in public schools until 1975 (Scott & Santos de Barona, 2007). Students with disabilities were excluded from public schools in several ways. Students with disabilities were denied access to public schools, were sent to private schools specifically designed for students with disabilities, or parents were billed for the added expense of special education services utilized in the public school (Katzman, 2007; Turnbull, Stowe,

& Huerta, 2007). One of the first federal laws to specifically protect the educational rights of students with disabilities was the Education for All Handicapped Children Act of 1975 (P.L. 94-142) (Scott & De Barona, 2007; Turnbull, Stowe, & Huerta, 2007). P.L. 94-142 (1975) had several key mandates regarding the education of students with disabilities. First, P.L. 94-142 required that students with disabilities receive a "free and appropriate public education" known as FAPE, which is interpreted to mean that students with disabilities cannot be excluded from a public education (Scott & Santos de Barona, 2007; p. 45; Turnbull, Stowe, & Huerta, 2007). Furthermore, school districts may not charge parents for special education services related to a disability; although this was a common practice before the passage of P.L. 94-142 (Turnbull, Stowe, & Huerta, 2007). Additionally, P.L. 94-142 required that students with disabilities be educated in the *least restrictive environment* known as LRE (Turnbull, Stowe, & Huerta, 2007). LRE refers to having students with disabilities educated in the general education classroom with nondisabled peers as much as possible (Scott & Santos de Barona, 2007).

Similar to ESEA (1965), P.L. 94-142 (1975) was also reauthorized several times with amendments to its contents. When P.L. 94-142 was reauthorized in 1990, its name was changed to the Individuals with Disabilities Education Act (1990), also known as IDEA (P.L. 101-476). IDEA (1990) continued to guarantee that students with disabilities were accorded a free appropriate public education emphasizing that their special education services met their individual needs (Bloomfield, 2007). Finally, IDEA (1990) was amended in 1997 and 2004. It has been asserted that one of the key changes to the 1997 amendment of IDEA (P.L. 105-17) was the amount of access to the general education curriculum afforded to students with disabilities (Katzman, 2007). However, it was the 2004 reauthorization of IDEIA (P.L. 108-446), which made significant changes to IDEA, in trying to align itself with NCLB (2001), with one of those changes requiring that students with disabilities were to participate in all general state and district-wide standardized assessments, with appropriate accommodations as specified in the students individualized education plan (Katsiyannis, Zhang, Ryan, & Jones, 2007; Turnbull, Stowe, & Huerta, 2007). Furthermore, IDEIA (2004) required states to track and report the assessment results of students with disabilities for accountability purposes (Katsiyannis, Zhang, Ryan, & Jones, 2007). Similar to the policies mandated in NCLB, IDEIA (2004) required students with disabilities to make progress toward goals that were consistent with the 'AYP' goals of students without disabilities (Katsiyannis, Zhang, Ryan, & Jones, 2007). Turnbull, Stowe, & Huerta (2007) asserted that the 2004 amendment to IDEA was due in part to the *President's Commission on Excellence in Special Education* (President's Commission on Excellence in Special Education, 2002) report, which concluded that special education, along with general education, was ineffective at educating students with disabilities. One of the more salient findings to come out of this presidential commission report (President's Commission on Excellence in Special Education, 2002) was that students with disabilities in special education were "first and foremost general education students" and their education within the general education and special education curriculum was inseparable (Turnbull, Stowe, & Huerta, 2007; p. 35).

Impact of NCLB and IDEA. Due to federal legislation, such as NCLB (2001) and IDEIA (2004) students with disabilities are required to take part in standardized state and district-wide assessments that are administered to all students (Katsiyannis, Zhang, Ryan, & Jones, 2007; Turnbull, Stowe, & Huerta, 2007). Furthermore, the assessment results of students with disabilities are required to be tracked to assure their progress toward 'AYP' goals (Katzman, 2007). The implication of such an accountability system is that the results of these standardized assessments impact the federal funding these schools receive (Hursh, 2005; Bloomfield, 2007; Tichá, Espin, & Wayman, 2009). Thus, the era of high-stakes testing began.

It is the implication of the results of these state-wide assessments that appears to have garnered some of the most interest to researchers and policy-makers. Indeed, Rushton and Joula-Rushton (2008) use the term 'high-stake' to describe the stipulations associated with the No Child Left Behind Act of 2001. Although the term *high-stakes* has been widely used over the past couple of decades, the term itself may be somewhat of a misnomer. Some believe that high-stakes testing refers to the test itself; however, high-stakes are not a characteristic of the test itself, but rather of the consequences placed on the outcome of tests. According to Plake (2011, p. 12), "…the consequences of how test scores are used" is what determines whether the test is high stakes or not. The use of state-wide assessments definitely fall within the realm of high-stakes testing given the results of such tests have high stakes for both students, educators, and school districts alike. Indeed, for students, the stakes might include whether the student is promoted or retained, is placed in remedial classes, or will attain a high school diploma (Haladyna, 2002; Katzman, 2007). The stakes for teachers might include performance pay, personnel evaluations, or even continued employment, whereas the stakes for an entire school district might mean the closure of a school or schools due continuous low scores on state-wide standardized testing or the amount of federal funding the school district receives (Haladyna, 2002; Katzman, 2007).

Due to the implications of the results of high-stakes testing, high-stakes testing has both its proponents and its critics. Proponents of high-stakes testing assert that some research has indicated positive outcomes due to requiring all students to participate in high-stakes testing. For example, in one research study, learning disabled students' reading scores increased and the number of learning disabled students who were proficient with grade level reading standards also increased (Schulte, Villwock, Whichard, & Stallings, 2001). The study by Schulte et. al. (2001) was conducted on 461 students with learning disabilities in grades 3 to 5 to determine the participation and progress of these students across multiple years on a state-wide high-stakes assessment. According to the authors of the study, progress made by the students appeared to be correlated with more participation in the general education curriculum due to the student's inclusion to participate in high-stakes testing (Schulte et. al., 2001). In another study, researchers reviewed multiple studies involving the participation of students with disabilities and noted there had been an increase in participation in high-stakes tests by students with disabilities, as well as better performance on high-stakes tests as the result of higher expectations (Ysseldyke, Nelson, Christenson, Johnson, Dennison, Triezenberg, Sharpe, and Hawes, 2004).

Despite some positive results of including students with disabilities in high-stakes state-mandated assessment, critics point out that the negative outcomes continue to outweigh the positives. Indeed, there has been much criticism related to what high-stakes tests really measure and whether the results are capable of predicting student achievement. (Duffy et al, 2008; Jones, 2001; Rushton & Joula-Rushton, 2008; Zimmerman & Dibenedetto, 2008). According to several researchers, there is little to no empirical data to support that high-stakes testing can predict academic achievement, especially in the area of reading, or improve achievement outcomes (Afflerbach, 2005; Schulte, Villwock, Whichard, & Stallings, 2001). Furthermore, Dutro & Selland (2012) question the validity of high-stakes testing. Although some research studies indicated that, in general, student scores do seem to rise over time, there is not enough evidence to concur that these students have made "true improvements" (Schulte, Villwock, Whichard, & Stallings, 2001). Additionally, researchers have asserted that the results of high-stakes tests are either rarely utilized to direct classroom instruction or are not sensitive enough to describe the specifics of the academic deficit (Afflerbach, 2005; Zimmerman & Dibenedetto, 2008). Others have stated that the focus of instruction may narrow to the point that teachers are preparing students to pass these state-mandated high-stakes tests by focusing on teaching the content that will be on these tests than teaching higher level thinking skills (Fritzberg, 2004; Haladyna, 2002). Moreover, parent groups have criticized the decline in enrichment activities, such as field trips, due to the focus on high-stakes testing results (Haladyna, 2002).

Most of the criticisms of high-stakes testing have been directed toward the results of high-stakes testing with the general population of students; however, others argue that the implications of high-stakes testing results are worse for students that fall into other categories, such as special education (Schulte & Villwock, 2004). While one of the goals of NCLB (2001) and IDEIA (2004) was to promote the education of all students by increasing accountability, Defur (2002) concluded that a negative consequence of high-stakes testing was less collaboration between general education and special education teachers. Johnson (2005) asserted requiring students with disabilities to take high-stakes tests may inadvertently set them up for failure, as the minimum level of standards may be unattainable when the student's disability is not taken into account. Some argue that provisions in IDEA (2004) and NCLB (2001) allow for accommodations on state-mandated high-stakes tests in order to minimize the impact of the student's disability on the test results (Fritzberg, 2004; Haladyna, 2002; Katzman, 2007). However, others argue there is limited research regarding the impact accommodations have on high-stakes testing results despite the requirement that states gather and report data on the effects of accommodations being utilized (Katzman, 2007; Yell, Shriner, & Katsiyannis, 2006; Ysseldyke et al., 2004). Furthermore, others have declared students with disabilities have less exposure to the general education curriculum being tested thereby putting them at a disadvantage when taking high-stakes tests (Katsiyannis, Zhang, Ryan, & Jones, 2007). Additionally, narrowing of the curriculum also affects students with disabilities. Indeed, one researcher concluded that students with disabilities were not able to pursue elective courses due to having to take remedial courses aimed at teaching the students what will be on the test (Ysseldyke et al., 2004).

As reported previously, the results of high-stakes testing usually do not provide enough information or are not sensitive enough to describe specific academic deficits (Afflerbach, 2005; Zimmerman & Dibenedetto, 2008). Although the previous finding was directed toward students in general education, a similar result was found with students with disabilities. Indeed, in a study conducted by Sharpe & Hawes (2003), they noted that the delay of results from high-stakes testing was not useful when developing goals for individualized education plans for students with disabilities.

Not only are students with disabilities at risk of not doing well on high-stakes tests, but some researcher have concluded that students from poverty are one of the most at-risk groups of not doing well on high-stakes tests (Tuerk, 2005; Lopez, 2007). It has been suggested that students from a low-socioeconomic status are not afforded the same opportunities as students from higher socio-economic status (Haladyna, 2002; López, 2007). López (2007) stated that students from poverty typically have lower scores on standardized tests and college aptitude exams. Such claims have been shown in many research studies. Indeed, research by Tuerk (2005) using data, such as achievement scores, educational resources, and SES, from the entire Virginia Public System suggested that socio-economic status was associated with performance on high-stakes tests. Haladyna (2002) reported students from poverty typically scored lower on standardized achievement tests. Similarly, a policy brief by Skiba and Rausch (2004) indicated several factors including poverty were significant predictors of whether students passed the ISTEP. Moreover, studies by Jencks & Phillips (1998) and Rech & Stevens (1996) concluded that socioeconomic status contributed from 17% to 33% of the variance in math performance for children.

The reasons for students in poverty having lower scores on standardized tests are likely many. For example, López (2007) asserted that school districts with a lower socioeconomic base often have inadequate resources, such as underfunding and understaffing. Furthermore, such school districts also have to contend with higher rates of student and teacher mobility (López, 2007). Other researchers have asserted that educational attainment and employment status for parents tends to be lower in children living in poverty (Kiernan & Mensah, 2011). Although the research is clear about the relationship of poverty and test scores, school districts in poverty-stricken areas are still accountable for student achievement under NCLB (2001) and risk having fewer funds afforded to their schools when students do not meet state-mandated standards (López, 2007).

Indiana Statewide Testing for Educational Progress- Plus. Due to federal legislation, such as NCLB (2001), all public schools in Indiana are required to use standardized testing to assess the academic progress of all students, including students in special education. The state administered assessment for Indiana is known as the Indiana Statewide Testing for Educational Progress- Plus (ISTEP+) (Indiana Department of Education, ISTEP+ Program Manual, 2011). According to IC 20-35-5-1, one of the purposes of ISTEP is to "… compare achievement of Indiana students to achievement of students on a national basis." Furthermore, ISTEP+ is utilized to help make decisions regarding "identifying students who may need remediation…" and "diagnosing individual student needs."

The state of Indiana has had some form of basic skills testing since at least 1984; however, it was 1987 when the Indiana Statewide Testing for Educational Progress (ISTEP) was first introduced. Originally, ISTEP was a norm-referenced multiple-choice test administered in the spring with a writing component administered in the winter. Students in specific grade levels participated in ISTEP. Indeed, students in grades 1, 2, 3, 6, 8, 9, and 11 originally participated in ISTEP testing. During this time period, students with disabilities were either exempt from testing or participated in ISTEP for diagnostic purposes. It was 1995 when ISTEP was revised to the Indiana Statewide Testing for Educational Progress- Plus (ISTEP+) (Indiana Department of Education, ISTEP+ Program Manual, 2010). A major revision of ISTEP+ was that it changed from a norm-referenced to a criterion-referenced test. Furthermore, testing was changed from spring to fall to allow for more focused instruction and remediation if warranted. ISTEP+ underwent another change in the 2008-2009 school year, when testing was administered in the fall and the spring with the intention of ISTEP+ being administered only in the spring of the year beginning with the 2009-2010 school year (Indian Department of Education, ISTEP+ Program Manual, 2010).

Initially, ISTEP+ was utilized to measure students' basic academic skills in English/language arts and mathematics but has grown to include science and social studies. In its current form, ISTEP+ measures English/language arts and mathematics to students in grades 3 through 8, science in grades 4 and 6, and social studies in grades 5 and 7 (Indiana Department of Education, ISTEP+ Program Manual, 2011). ISTEP+ is a criterion-referenced test and is based on Indiana's current academic standards, which are established by the Indiana State Board of Education, and define what a student should know and what a student should be able to do in English/language arts, mathematics, science, social studies, and other content areas at specific grade levels (Indiana Department of Education, Guide to Test Interpretation, 2011). ISTEP+ is comprised of multiple choice questions, as well as short answer and essay questions, which are administered in the spring of each school year. Currently, the short answer and essay portion of the test is known as ISTEP+ Applied Skills, which is administered in March. The ISTEP+ multiple-choice portion is administered in April (Indiana Department of

Education, ISTEP+ Program Manual, 2011). Results from the ISTEP+ are provided in the form of a student score in the areas assessed, which is then compared to a minimum performance standard in each of those areas. Based on the performance standards, known as cut-scores, the student will receive one of the three following qualitative statements: 1) Did Not Pass, meaning the student did not meet proficiency standards in that particular subject area; 2) Pass, meaning the student met at least the minimum proficiency in a particular subject area; or 3) Pass+, meaning the student was well-above the proficiency level for that particular subject area and would be considered "high-achieving" in that subject area (Indiana Department of Education, Guide to Test Interpretation, 2011; p. 50).

History of Standardized Cognitive Ability and Standardized Achievement Testing

History of Standardized Testing. Individuals working in the field of education are quite familiar with the term achievement, which is often defined as accomplishment or attainment (Haladyna, 2002). Yet, the purpose for measuring the achievement of students has changed over time based on education laws. According to Haladyna (2002), the main use of standardized achievement testing within school systems should be for the purpose of measuring student learning and asserts that the initial use of achievement tests in schools was to change the curriculum if students did not perform well on achievement testing. Prior to the 20th century, generally only students from financially wealthy families could afford to attend school. However, compulsory education laws were implemented in the late 19th century and all children, regardless of socioeconomic status, were required to attend school (Thorndike, 1997). As more and more disadvantaged students entered the school system, educators recognized that many of these students were not prepared to succeed in the classroom but still did not understand the reasons why these students were not successful (Thorndike, 1990). Hence, researchers began developing hypotheses or models to understand and capture the meaning of ability.

Conceptualizations of Intelligence and Intelligence Testing. The term intelligence has been synonymous with terms such as mental ability, capability, aptitude, and cognitive ability to name a few (Haladyna, 2002; Wasserman & Tulsky, 2005). Furthermore, intelligence has had many different definitions, and the consensus on what the construct of intelligence measures has garnered much disagreement (Dehn, 2006). Historically, interest in human abilities has likely been around for centuries. It has been noted that more than 2,000 years ago the Chinese introduced a system, the civil service exam, to classify individuals based on their abilities in order to place individuals in professions based on those abilities (Kamphaus, Winsor, Rowe, & Kim, 2005). The current use of intelligence or cognitive ability testing is primarily used for educational planning, clinical applications, and forensic applications (Bursztyn, 2007). However, in order to understand how psychologists have arrived at the current theory and understanding of intelligence, one must first consider the history of the conceptualization of intelligence and intellectual assessment. While considering the history of intelligence, one could go back centuries and discuss the philosophical views of intelligence; however, for brevity and to maintain continuity to the topic of this research, this paper will focus on the history of intelligence beginning in the late 19th century to the early 20th century since the modern conception of intelligence testing is often thought to have begun with Alfred Binet (Wasserman & Tulsky, 2005).

Modern intelligence testing appears to be directly related to the work of psychologists during the late 19th century when researchers were attempting to quantify or quantitatively classify individual's intelligence (Sattler, 2008; Wasserman & Tulsky, 2005). Many of these early psychologists believed that intelligence or mental abilities could be understood by measuring physical characteristics, visual judgments, and physiological responses to sensory stimuli, with the belief that those who could respond faster or with more acuity must be better or more intelligent than those who responded slower (Sattler, 2008; Wasserman & Tulsky, 2005).

Sir Francis Galton has been noted to be one of the first researchers to use objective techniques to study individual differences by measuring physical characteristics, reaction time, visual judgment, and other sensory-motor activities, with the main goal of determining who could "produce talented offspring" (Sattler, 2008; Wasserman & Tulsky, 2005, p. 4). Although Galton never had a formal definition of intelligence, he believed that individuals with the best sensory discrimination abilities would produce the most capable offspring since knowledge of the environment comes through the senses (Sattler, 2008; Wasserman & Tulsky, 2005). Beyond the use of objective techniques, Galton also influenced the field of intelligence testing through his development of two statistical concepts, correlation and regression to the mean (Sattler, 2008).

Following in Galton's footsteps was James Cattell, who also believed that intelligence could be measured by evaluating one's sensory and motor abilities. Although Cattell spent some time working with Galton, Cattell eventually moved to the United States and accepted a position at Columbia University after a brief stint at the University of Pennsylvania (Sattler, 2008). Cattell developed a test battery using some of Galton's techniques while adding additional tests. Cattell's testing battery consisted of 10 basic tests, with 50 supplemental tests measuring various sensory and motor abilities (Wasserman & Tulsky, 2005). Cattell utilized his tests to evaluate university students to determine whether there was a relationship between performance on these tests and success in college. However, Clark Wissler, a graduate student of Cattell's, demonstrated that student's test scores from Cattell's test battery had very little correlation to student's grades (Sattler, 2008). Similar research was also being conducted around the United States. Indeed, researchers were conducting studies interested in distinguishing "bright" from "dull" children (Sattler, 2008, p. 217).

Because these early sensory and motor tests were not as successful as previously hoped in measuring intelligence, others were becoming interested in finding other ways to measure intelligence. Indeed, the work of Alfred Binet, Theodore Simon, and Victor Henri in France began to focus on "higher mental" functions rather than sensory functions as a means to measure intelligence (Foschi & Cicciola, 2006; Sattler, 2008). Eventually, through Binet's collaborations with Henri and Simon, Binet concluded that cognitive abilities could not be evaluated as independent functions but when measured together formed the basis of "higher mental processes" (Wasserman & Tulsky, 2005, p. 7). In the early 20th century, Alfred Binet was commissioned by the French Ministry of

Public Instruction to develop techniques for identifying children with mental retardation who might require a special classroom. In 1905 the first intelligence test, the Binet-Simon Scale, was produced leading some to call Alfred Binet the 'father' of intelligence testing (Foschi & Cicciola, 2006; Sattler, 2008; Wasserman & Tulsky, 2005). The Binet-Simon Scale consisted of 30 short tasks that supposedly involved basic processes of reasoning, and unlike, the early sensory and motor tasks, this new scale was age-based with testing items increasing in difficulty to match a specific developmental level (Sattler, 2008; Wasserman & Tulsky, 2005). The essential feature of the Binet-Simon Scale was that all tests at a given level were considered capable of being solved by normal children within that age group (Wasserman & Tulsky, 2005). One of the innovations of the first Binet-Simon Scale was the use of standardized instructions. The Binet-Simon Scale was revised two more times, in 1908 and 1911.

This interest in children's intelligence was not confined to Europe. In the United States, Lewis Terman of Stanford University had developed an interest in the Binet-Simon Scale believing it had theoretical value, as well as pragmatic potential (Sattler, 2008). Terman made several contributions to further the assessment of intelligence by revising the original Binet-Simon Scale in 1912, which culminated in a published copy in 1916. His final revision to this scale was in 1937 and renamed the *New Revised Stanford-Binet Tests of Intelligence* (Terman & Merrill, 1937). Although other English translations of the Binet-Simon Scale were developed, Terman's version was considered advantageous over other versions due to its methodological rigor and large normative sample (Foschi & Cicciola, 2006; Wasserman & Tulsky, 2005). Furthermore, Terman was the first to coin the phrase "Intelligence Quotient" or IQ, which he adopted from

Louis Stern's term "mental quotient" (Sattler, 2008, p. 219). IQ was defined as mental age divided by chronological age and multiplied by 100, with the purpose of being able to quantify intellectual functioning to allow comparison among individuals.

Other intelligence tests, such as the Army Alpha and Army Beta (Yerkes, 1921) tests, were developed during this general time period but usually for specific purposes. However, Charles Spearman began analyzing the relationship of tasks among several tests using an early form of factor analysis and concluded that performance on some intellectual tasks correlated with doing well on other tasks (Brody, 2000; Wasserman & Tulsky, 2005). Based on his work, Spearman hypothesized that intelligence was composed of two factors, *g* or general ability and *s* or specific abilities, hence Spearman's two-factor theory of intelligence (Brody, 2000). Although Spearman proposed that the variance not accounted for by *g* could be accounted for by *s* or the specific abilities, his work tended to focus mainly on *g*. One of the unique features of Spearman's theory was that it was one of the first that offered a theoretical explanation of *g* (Carroll, 1993; Horn & Noll, 1997).

Because Spearman's statistical procedures were commonly used when investigating the constructs of intelligence, *g* or the general factor of intelligence became strongly associated with the IQ score of many intelligence test batteries. However, some researchers disputed the importance of Spearman's *g*. Two of those researchers were Edward L. Thorndike and Louis Thurstone. Thorndike hypothesized that intelligence was composed of distinct but interrelated abilities. Thorndike began to test this two-factor theory using data sets similar to the ones Spearman used, and began to claim that

Spearman's results were inaccurate supporting his own theory of multiple abilities (Cattell, 1987). Interestingly, Thorndike's hypotheses were based more on his theory than on statistical methodology (Sattler, 2008). However, Louis Thurstone was another researcher who disputed *g* but relied on statistical methodology to dispute *g*. Indeed, Thurstone hypothesized that the general factor arose from other primary factors that were related to one another. Through the use of multi-factor analysis, Thurstone obtained 13 factors, seven of which he interpreted as "primary mental abilities" (Gulliksen, 1968; Wasserman & Tulsky, 2005, p. 17). These primary mental abilities were verbal comprehension, word fluency, reasoning, associative memory, spatial visualization, perceptual speed, and number facility. Thrustone eventually accepted the hypothesis of a higher-order general factor or *g*, but disputed its importance (Carroll, 1993; Sattler, 2001). The important feature of Thurstone's contributions was that intelligence was better understood and measured by considering distinct primary abilities rather than a single factor or general ability (*g*) (Gulliksen, 1968; Wasserman & Tulsky, 2005). The idea of a global intelligence being measured by specific abilities was also the focus of the work by David Wechsler. Indeed, Wechsler believed intelligence was a component of one's overall personality (Sattler, 2008; Wechsler, 1950). Wechsler set about designing an intelligence scale taking into account the factors that contribute to the global intelligence (Matarazzo, 1981; Wechsler, 1950; Wechsler, 1975). However, Wechsler appeared more interested in the global intelligence obtained by these factors rather than the factors themselves (Sattler, 2008).

Raymond Cattell, a student of Spearman's, was influenced by the works of Thurstone and Thorndike. He believed that Spearman's one factor of intelligence could not adequately explain intelligence and hypothesized that intelligence was hierarchical in nature (Cattell, 1963, 1987). Indeed, Cattell suggested there were two related but distinct components of *g*, fluid intelligence (G*f*) and crystallized intelligence (G*c*) (Horn, 1968). Furthermore, Cattell believed G*f* and G*c* were related (Cattell, 1987). Fluid intelligence (G*f*) was described as the ability to solve novel problems, which will require different skills depending on the nature of the task, and is influenced by biologically-based factors rather than factual knowledge (Flanagan, Ortiz, & Alfonso, 2007; Wasserman & Tulsky, 2005), whereas crystallized intelligence (G*c*) was described as fact-based knowledge acquired through formal and informal education, as well as acculturation (Wasserman & Tulsky, 2005). As a result of Cattell's work he found that G*f* abilities decline with age, whereas G*c* abilities do not (Brody, 1985; Wasserman & Tulsky, 2005).

Cattell began working with his student John Horn to refine G*f*-G*c* theory using factor analysis and structural equation modeling. Horn had noted that some of the abilities associated with aspects of memory such as visual memory, auditory memory, and basic memory, were different than originally conceived with just G*f*-G*c*. Additionally, through further factor analytic and structural equation modeling studies involving several neurological, intellectual, and academic batteries, Cattell and Horn concluded that several broad and narrow abilities could be identified and these abilities were added to the model (Woodcock, 1990). Hence, the original G*f*-G*c* model included seven more broad abilities short-term memory (G*sm*), long-term storage and retrieval (G*lr*), auditory processing (G*a*), visual processing (G*v*), processing speed (G*s*), correct decision speed (G*t*), and quantitative reasoning (G*q*) (Horn & Noll, 1997). Unlike the earlier researchers that concluded intelligence was comprised of multiple abilities, Cattell and Horn organized the specific abilities in a hierarchical system that was subsumed under the nine broad abilities.

Many researchers were now developing factor analytic theories of intelligence. However, one researcher, J. P. Guilford split from the others by refusing to acknowledge the existence of any general factor. He proposed that intelligence was comprised of 120 elementary abilities or factors that were encompassed by three dimensions he termed, operations, content, and products (Sattler, 2002). He posited that there were six types of operations, four types of content, and six types of products, with a combination of one element from each of the dimensions yielding a factor.

Although researchers such as Guilford rejected the notion of *g*, other researchers included the notion of *g* into their models. One such researcher was John Carroll, who proposed a three stratum model fashioned in a hierarchical design (Carroll, 1993). Carroll (1993) identified approximately 70 narrow or specific abilities, which comprised stratum I. Stratum I was subsumed under stratum II, which was comprised of eight broad based abilities: G*f*, G*c*, general memory and learning (G*y*), broad visual perception (G*v*), broad auditory perception (G*u*), broad retrieval ability (G*r*), broad cognitive speediness (G*s*), and decision speed/reaction time (G*t*) (Flanagan, Ortiz, Alfonso, & Mascolo, 2002). At the apex of the hierarchy is stratum III, a general factor, similar to Spearman's *g*.

While there were differences between the Cattell-Horn G*f*-G*g* model (Horn, 1991; Horn & Noll, 1997) and the Carroll three-stratum theory (Carroll, 1993, 1998) model, there were also similarities. Recognizing the necessity for a single model to classify intelligence from a theoretical basis due to the lack of unity regarding the constructs of intelligence, Kevin McGrew conducted a series of factor analytic studies to resolve the differences between the two models and integrate them into a single model (Flanagan, Ortiz, & Alfonso, 2007; Flanagan, Ortiz, Alfonso, & Mascolo, 2002). This unification of the two models led to the current CHC theory of human intelligence.

Contemporary CHC Theory. The contemporary model of CHC theory is an integration of the original Cattell-Horn G*f*-G*c* (Horn, 1991; Horn & Noll, 1997) theory and the Carroll three-stratum theory (Carroll, 1993, 1998). Current CHC theory holds that human intelligence is comprised in a hierarchical nature with three layers: stratum III (*g* or general intelligence), stratum II (10 broad cognitive abilities), and stratum I (>70 narrow cognitive abilities) (Floyd, McGrew, & Evans, 2008; Newton & McGrew, 2010; Taub, Floyd, Keith, & McGrew, 2008). However, CHC theory tends to focus on stratum I and stratum II, as many of the theorists who prescribe to CHC theory believe that stratum III or *g* does not offer practical relevance in assessment related practices. Indeed, it has been asserted that a significant amount of the variance in specific academic abilities can be accounted for by the broad and narrow cognitive abilities above and beyond *g* (Flanagan, Ortiz, & Alfonso, 2007). According to Flanagan, Ortiz, Alfonso, & Dynda (2008), the following broad cognitive abilities comprise stratum II of the current CHC model: Fluid Intelligence (G*f*), Quantitative Knowledge (G*q*), Crystallized Intelligence (G*c*), Short-Term Memory (G*sm*), Visual Processing (G*v*), Auditory Processing (G*a*), Long-Term Storage & Retrieval (G*lr*), Processing Speed (G*s*), Decision Reaction Time/Speed (G*t*), and Reading and Writing (G*rw*). Stratum I consists of more than 70 narrow cognitive abilities, which are subsumed under stratum II (Flanagan, Ortiz, & Alfonso, 2007; Floyd, McGrew, & Evans, 2008; Newton & McGrew, 2010). See table

2.1 for the list of Stratum I narrow abilities subsumed under their respective Stratum II broad ability.

Table 2.1

Narrow Cognitive Abilities

G*f*	G*c*	G*q*	G*sm*	G*lr*	G*v*	G*s*	G*a*	G*t*	G*rw*
General Sequential Reasoning	Language Development	Math Know.	Memory Span	Assoc. Memory	Visualization	Perceptual Speed	Phon. Coding: Analysis	Simple Reaction Time	Reading Decoding
Induction	Lexical Knowledge	Math Ach.	Working Memory	Mngful. Memory	Spatial Relations	Rate of Test Taking	Phon. Coding: Synthesis	Choice Reaction Time	Reading Comp.
Quantitative Reasoning	Listening Ability		Learning Abilities	Free Recall Memory	Visual Memory	Number Facility	Speech Sound Discrim.	Mental Comparison Speed	Verbal Language Comp.
Piagetian Reasoning	General Information			Ideational Fluency	Closure Speed	Semantic Processing	Res. to Auditory Stimulus Distortion	Correct Decision Speed	Cloze Ability
Speed of Reasoning	Information about Culture			Assoc. Fluency	Flexibility of Closure	Speed	Memory for Sound Patterns		Spelling Ability
	General Science Information			Expressional Fluency	Spatial Scanning		General Sound Discrim.		Writing Ability
	Geography Achievement			Naming Facility	Serial Perceptual Integration		Temporal Tracking		English Usage Know.
	Communication Ability			Word Fluency	Length Estimation		Musical Discrim. & Judgment		Reading Speed
	Oral Production & Fluency			Figural Fluency	Perceptual Illusions		Maintaining & Judging Rhythm		
	Grammatical Sensitivity			Figural Flexibility	Perceptual Alternations		Snd-Freq. Discrim.		
	Foreign Language Proficiency			Sensitivity to Problems	Imagery		Hearing & Speech Threshold		
	Foreign Language Aptitude			Originality/ Creativity			Absolute Pitch		
				Learning Abilities			Sound Localizatio		

Source: (Table adapted from Flanagan, Ortiz, Alfonzo, & Mascolo, 2002)

The following section describes the stratum II or broad cognitive abilities of the current CHC theory. As noted previously, stratum III or *g* will not be defined as it tends to be de-emphasized and its definition as a general ability factor has not changed with the modifications to CHC theory (Flanagan, Ortiz, & Alfonso, 2007). Stratum II consists of ten broad cognitive abilities. Fluid Intelligence (G*f*) has been described as the ability to solve novel problems, which might include drawing inferences, generalizations, identifying relationships, classification, hypothesis generation and testing, concept formation, transforming information, and problem solving, whereas Crystallized Intelligence (G*c*) is the depth and breadth of acquired knowledge. G*c* consists of declarative and procedural knowledge, which is acquired through formal and informal educational experiences, as well as through acculturation (Flanagan, Ortiz, Alfonso, & Mascolo, 2002; Newton & McGrew, 2010). An ability similar to G*c* is Quantitative Knowledge (G*q*), which is the depth and breadth of acquired numerical knowledge (Newton & McGrew, 2010). Short-Term Memory (G*sm*) is the ability to capture, store, preserve, and use information immediately before it disappears from the memory trace (Flanagan, Ortiz, Alfonso, & Mascolo, 2002; Newton & McGrew, 2010). Visual Processing (G*v*) is the ability to store and recall visual stimuli, as well as the ability to perceive, synthesize, analyze, and manipulate visual patterns (Newton & McGrew, 2010). Auditory Processing (G*a*) is the ability to discriminate, analyze, and synthesize auditory stimuli (Newton & McGrew, 2010). Long-Term Storage & Retrieval (G*lr*) is the ability to encode information by storing and consolidating this information into long-term memory and then retrieving this information from memory fluently. This ability to store and fluently retrieve information from memory occurs through association (Flanagan, Ortiz, Alfonso, & Mascolo, 2002; Newton & McGrew, 2010). Processing Speed (G*s*) is the ability to perform simple cognitive tasks automatically and fluently under conditions requiring sustained attention and concentration (Newton & McGrew, 2010). Decision Reaction Time/Speed (G*t*) is the ability to make simple decisions fluently or to react quickly to simple stimuli (Newton & McGrew, 2010). Reading and Writing (G*rw*) is similar to G*q* and G*c*, but is the depth and breadth of acquired reading and writing knowledge (Newton & McGrew, 2010).

CHC theory has had several important implications for the understanding of intelligence, as well as the measurement of intelligence. Indeed, CHC theory has extensive empirical support and has bridged the gap between theory and practice by giving both practitioners and researchers a common language for understanding cognitive abilities (Flanagan, Ortiz, Alfonso, & Dynda, 2007). Moreover, one of the chief practical consequences of CHC theory has been its influence on cognitive assessment. Indeed, several of the CHC broad abilities, namely G*f*, G*a*, G*s*, G*sm*, and G*lr*, have been shown to provide predictive information regarding academic achievement skills (Evans, Floyd, McGrew, & Leforgee, 2001; Flanagan, Ortiz, Alfonso, & Dynda, 2008; Floyd, Evans, & McGrew, 2003; Floyd, McGrew, & Evans, 2008; & Taub, Floyd, Keith, & McGrew, 2008).

Although CHC theory appears to have influenced many of the intelligence test batteries that have been developed since 2000, prior to that time very few intelligence test batteries measured any of the broad cognitive abilities, with generally only two or three broad cognitive abilities being measured (Flanagan, Ortiz, Alfonso, & Dynda, 2008).

However, prior to 2000 the *WJ-R Tests of Cognitive Ability* (WJ-R COG; Woodcock & Johnson, 1989) was a primary cognitive battery that measured seven broad abilities (Schrank, Flanagan, Woodcock, & Mascolo, 2002). The subsequent revision of the WJ-R COG, the *Woodcock-Johnson III Tests of Cognitive Abilities* (WJ III COG; Woodcock, McGrew, & Mather, 2001a) was based entirely on CHC theory providing the WJ III COG with "greater generalizability (validity)…" (Schrank, Flanagan, Woodcock, & Mascolo, 2002, p. 6). Giving the WJ III COG greater diagnostic power is the fact that it was co-normed with its academic achievement counterpart, the *Woodcock-Johnson Third Edition Tests of Achievement* (WJ III ACH; Woodcock, McGrew, & Mather, 2001b). According to Gregg et al. (2005) the WJ III COG and WJ III ACH, when used together, form a powerful battery in which to predict achievement based on cognitive abilities. Having such a powerful battery is important given the school psychologists vital role in providing assessment within the school system. The use of comprehensive, standardized cognitive and standardized achievement tests is paramount to school psychologists, given that assessment of learning disabilities is one of the most common psycho-educational evaluation referrals (Fletcher, Lyon, Fuchs, & Barnes, 2007). Given federal legislation, such as NCLB (2001) and IDEA (2004), the use of cognitive and achievement tests will likely continue to be a cornerstone of the school psychologists toolkit.

While there are many different achievement tests to choose from, the *Woodcock-Johnson Third Edition Tests of Achievement* (WJ III ACH; Woodcock, McGrew, & Mather, 2001b) is one of the most utilized standardized achievement tests (Taub & McGrew, 2004). Furthermore, the WJ III COG has come to prominence as a measure to predict achievement in children because of its empirically-sound underlying theoretical structure, namely CHC theory (Evans, Floyd, McGrew, & Leforgee, 2001; Floyd, McGrew, & Evans, 2008). Moreover, research has supported that CHC abilities have aided in the understanding of reading, math, and written language (Newton & McGrew, 2010; Floyd, McGrew, & Evans, 2008; Taub, Floyd, Keith, & McGrew, 2008; Evans, Floyd, McGrew, & Leforgee, 2001). Therefore, the WJ III COG and WJ III ACH will likely continue to be integral in research on the relationship between cognitive abilities and achievement.

Chapter III

Methodology

Chapter 3 is organized into four sections: (1) Participant Selection; (2) Instrumentation, Validity, and Reliability; (3) Procedures; and (4) Statistical Procedures and Data Analysis. The purpose of Chapter 3 is to provide a detailed explanation of participant selection, as well as the proposed procedures that will be used to collect and analyze the data.

Participant Selection. Participants in the current study were students in grades 3 through 8 in an urban school district in a Midwestern city who had been referred for a psychoeducational evaluation. All information from potential participants in this study was gathered from an archival database in which the psychoeducational evaluations had been completed by masters and specialist level school psychologists who had been licensed by the Indiana Department of Education. The school psychologists were employed by a school district in a Midwestern city and worked as part of a multi-disciplinary team including special education teachers, speech/language pathologists, occupational therapists, and physical therapists in determining eligibility for special education services. Referral sources for the students included the following: school personnel, caregivers, pediatricians, foster care facilities and homes, and the juvenile justice system.

A total of 45 participants were selected from archived data, which originally included a sample of 98 participants. Only 45 of the original 98 participants were selected due to missing data. The original sample was collected from archived data in which students had been referred for a psychoeducational evaluation due to academic difficulties. Participants for the current study were selected if they had been administered a *Woodcock-Johnson III Tests of Cognitive Abilities* (WJ-III-COG; Woodcock, McGrew, & Mather, 2001a), a *Woodcock-Johnson III Tests of Achievement* (WJ-III-ACH; Woodcock, McGrew, & Mather, 2001b), and had taken the *Indiana Statewide Test for Educational Progress- Plus* (ISTEP+; Indiana State Department of Education, I.s., 2011). The original sample of 98 participants was narrowed down to 45 participants due to missing data on one or more of these instruments for 44 of the participants. The final sample included 24 males and 21 females. Forty of the participants were identified with a special education label and five were identified as not eligible for a special education label. The participants ranged in age from 7 years to 16.3 years (mean = 11.1 years, SD = 2.3 years). The sample varied with regard to ethnicity, with Caucasians representing a majority of the sample at 51% (N = 23). African-Americans accounted for 29% (N = 13) of the sample, while Multi-Racial and Hispanic participants accounted for 16% (N = 7) and 4% (N = 2) of the sample, respectively. The study sample closely matched the state of Indiana 2011 census for ethnicity with regard to the Hispanic population, with Indiana's Hispanic population reported as 6.2% of the population of the state (U.S. Census Bureau, 2012). The sample of Caucasian participants was smaller than the Indiana population of Caucasians at 86.8%, whereas the sample of African-American participants was larger than the state of Indiana population of African-Americans at 9.4% (U.S. Census Bureau, 2012). With regard to socio-economic status, 16 (36%) of the participants received free/reduced lunch.

Demographic information for the sample is summarized in Table 3.1.

Table 3.1.

Descriptive Statistics of the Sample

	Total (Percents)
Gender	
Males	24 (53%)
Females	21 (47%)
Race	
Caucasian	23 (51%)
African-American	13 (29%)
Multi-Racial	7 (16%)
Hispanic	2 (4%)
Socio-Economic Status	
Paid Lunch	29 (64%)
Free/Reduced	16 (36%)
Special Education Label	
Specific Learning	24 (53%)
Disability	10 (22%)
Other Health Impairment	3 (7 %)
Emotional Disability	1 (2%)
Cognitive Disability	1 (2%)
Autism Spectrum Disorder	1 (2%)
Deaf/Hard of Hearing	5 (11%)

Did Not Qualify	

Instrumentation.

Woodcock-Johnson III Tests of Cognitive Abilities. The *Woodcock-Johnson III Tests of Cognitive Abilities* (WJ-III-COG; Woodcock, McGrew, & Mather, 2001a) is an individually administered clinical instrument used to assess the cognitive ability of children and adults aged 2 years, 0 months through 95+ years of age. The WJ-III-COG is a valid and reliable measure of cognitive functioning (Schrank, Flanagan, Woodcock, McGrew, & Mather, 2001). The mean index score for the WJ-III-Cog is 100 with a standard deviation of 15, with the mean subtest score being 10 with a standard deviation of 3.

The WJ-III-COG, which was developed from CHC (Horn, 1991; Horn & Noll, 1997) theory, is comprised of two parts containing 20 subtests, with each subtest measuring a different aspect of cognitive ability. The first part of the WJ-III-COG is the Standard Battery consisting of tests 1 through 10, whereas the second part of the WJ-III-COG is the Extended Battery consisting of tests 11 through 20. Each of the 20 subtests on the WJ-III-COG represents a specific narrow cognitive ability. These narrow abilities represent Stratum I of CHC theory. Combinations of these narrow abilities are grouped together to form the broad abilities represented on the WJ-III-COG (Schrank, Flanagan, Woodcock, & Mascolo, 2002; Woodcock, McGrew, & Mather, 2001a). According to McGrew & Woodcock (2001), the broad abilities are the component of the WJ-III-COG that provides the most practical utility of the instrument. These broad abilities represent the Stratum II CHC factors. Stratum II CHC factors measured by the WJ-III-COG include Comprehension-Knowledge (G*c*), Long-Term Retrieval (G*lr*), Visual-Spatial Thinking (G*v*), Auditory Processing (G*a*), Fluid Reasoning (G*f*), Processing Speed (G*s*), and Short-Term Memory (G*sm*). Additionally, combinations of these broad factors form three clusters on the WJ-III-COG, namely Verbal Ability, Thinking Ability, and Cognitive Efficiency. Finally, the WJ-III-COG provides an overall intelligence score, the General Intellectual Ability (GIA), which can be obtained by administering seven specific subtests from the standard battery, or by administering the seven standard battery subtests along with seven specific subtests from the extended battery (Schrank, Flanagan, Woodcock, & Mascolo, 2002). The GIA represents Stratum III of CHC theory.

Validity. According to McGrew & Woodcock (2001), strong evidence of construct validity is evidenced with the WJ-III-COG. Further, the *Technical Manual* (McGrew & Woodcock, 2001) stated that the validity of the WJ-III-COG is supported by the 8,818 norming sample subjects, as well as 775 validity study subjects. Each of the seven CHC cognitive factors demonstrated divergent growth and decline patterns, which is reported to be one type of evidence that distinct cognitive abilities exist (McGrew & Woodcock, 2001; Schrank, Flanagan, Woodcock, & Mascolo, 2002). Moreover, according to Schranck, Flanagan, Woodcock, & Mascolo (2002), the intercorrelations of the seven broad CHC factors ranged from 0.20 to 0.60, suggesting that the even though the WJ-III-COG factor clusters are slightly related to each other, they are diverse enough to reveal distinct cognitive abilities. Furthermore, a *g* + seven broad-factor model provided the best fit to the standardization data, which would suggest that the WJ-III-COG follows CHC theory’s broad and narrow abilities and general intellectual ability.

Additionally, the WJ III *Technical Manual* (2001) reports several concurrent validity studies conducted using scores obtained on the WJ-III-COG as compared to the full scale and composite scores of other intelligence tests, namely the *Wechsler Preschool and Primary Scale of Intelligence – Revised* (WPPSI-R; Wechsler, 1989), the *Wechsler Intelligence Scale for Children – Third Edition* (WISC-III; Wechsler, 1991), the *Differential Ability Scales* (DAS; Elliott, 1990), the *Stanford-Binet Intelligence Scale – Fourth Edition* (SB-IV; Thorndike, Hagen, & Sattler, 1986), the *Kaufman Adolescent and Adult Intelligence Test* (KAIT; Kaufman & Kaufman, 1993), and the *DAS·Naglieri Cognitive Assessment System* (CAS; Naglieri & DAS, 1997). Statistical analyses indicated the WJ-III-COG strongly correlated with these intelligence tests as the correlations consistently fell within the 0.70s when comparing the General Intellectual Ability index to the full scale and composite scores of these tests (McGrew & Woodcock, 2001).

A study was conducted by Taub & McGrew (2004) investigating the use of 14 of the cognitive subtests that provide the seven broad CHC cognitive ability clusters, as well as the General Intellectual Ability – Extended, across five age groups. The authors examined a portion of the standardization sample from the WJ-III-COG, specifically ages 6 through 90+, to determine if the instrument measures the same constructs across age ranges. Results of Taub & McGrew's (2004) study indicated that indeed, the WJ-III-COG appears to measure the same CHC cognitive constructs from age 6 through 90+. This was an important finding in that the ability of the WJ-III-COG to measure the same trait across different age groups supports its ability to be used to examine the pattern of correlations between variables in different age groups.

Reliability. McGrew & Woodcock (2001) indicated that "reliability characteristics of the WJ III meet or exceed basic standards for both individual placement and programming decisions," (p. 48) and reported median reliability coefficients of .97 on the General Intellectual Ability – Standard and .98 on the General Intellectual Ability – Extended. Moreover, median reliability coefficients of the WJ-III broad cognitive abilities included .95 on Comprehension-Knowledge (G*c*), .88 on Long-Term Retrieval (G*lr*), .81 on Visual-Spatial Thinking (G*v*), .91 on Auditory Processing (G*a*), .95 on Fluid Reasoning (G*f*), .92 on Processing Speed (G*s*), and .88 on Short-Term Memory (G*sm*) (McGrew & Woodcock, 2001).

Woodcock-Johnson III Tests of Achievement. The *Woodcock-Johnson III Tests of Achievement* (WJ-III-ACH; Woodcock, McGrew, & Mather, 2001b) is an individually administered instrument used to assess the academic achievement of children and adults aged 2 years, 0 months through 95+ years of age. The WJ-III-ACH is a valid and reliable measure of academic achievement. The mean index score for the WJ-III-ACH is 100 with a standard deviation of 15, with the mean subtest score being 10 with a standard deviation of 3.

The WJ-III-ACH, which was developed from CHC theory, contains 22 subtests divided into two batteries, the Standard Battery and the Extended Battery. The Standard Battery consists of subtests 1 through 12, whereas the Extended Battery consists of subtests 13 through 22. Each subtest measures a specific aspect of academic achievement. Additionally, the battery is available in two forms, which are parallel in content, Form A and Form B.

As noted previously, each of the subtests on the WJ-III-ACH measures a specific or primary narrow ability. Combinations of these primary narrow abilities are grouped together to form the broad CHC abilities represented on the WJ-III-ACH. The broad abilities measured by the WJ-III-ACH include Reading-Writing (G*rw*), Mathematics (G*q*), as well as Comprehension-Knowledge (G*c*), Long-Term Retrieval (G*lr*), and Auditory Processing (G*a*). Additionally, combinations of these narrow abilities from the standard battery are grouped together to form academic clusters, namely Broad Reading, Oral Language – Standard, Broad Math, and Broad Written Language (Mather & Woodcock, 2001).

Validity. Strong evidence of construct validity is evidenced with the WJ-III-ACH. The WJ-III-ACH was co-normed with the WJ-III-COG on 8,818 participants, which McGrew & Woodcock (2001) stated supports the validity of the WJ-III-ACH. Co-norming of the WJ-III-ACH with the WJ-III-COG supports construct validity in that discrepancy scores are not influenced by unknown factors.

A concurrent validity study between the WJ-III-ACH and the *Kaufman Test of Educational Achievement* (KTEA; Kaufman & Kaufman, 1985) and the *Wechsler Individual Achievement Test* (WIAT; Wechsler, 1992) are provided within the WJ-III *Technical Manual* (2001). This study used a sample of 52 students in grades 1 through 8 with a mean age of approximately 10 years of age. This particular study noted the Broad Reading Cluster of the WJ-III-ACH had correlations of 0.76 with the Reading Composite of the KTEA and 0.67 with the Reading Composite of the WIAT. The Broad Reading Cluster of the WJ-III-ACH had correlations of 0.67 on Reading Decoding on the KTEA and 0.63 with Basic Reading on the WIAT. The Basic Reading Skills Cluster of the WJ-

III-ACH had a correlation of 0.82 with Basic Reading on the WIAT. Reading Decoding on the KTEA had a correlation of 0.67 with the Broad Reading Cluster, 0.66 with Basic Reading Skills, and 0.74 with Reading Comprehension on the WJ-III-ACH. Reading Comprehension on the WJ-III-ACH had a correlation of 0.62 with Reading Comprehension on the KTEA and a correlation of 0.79 with Reading Comprehension on the WIAT. The Broad Math Cluster on the WJ-III-ACH had a correlation of 0.66 with the Math Composite of the KTEA and a correlation of 0.70 with the Math Composite of the WIAT. Math Computation on the KTEA had a correlation of 0.65 with Broad Math and 0.67 with Math Calculation Skills of the WJ-III-ACH. Numerical Operations on the WIAT had correlations of 0.57 with Broad Math and 0.59 with Math Calculation Skills of the WJ-III-ACH. Math Applications on the KTEA had a correlation of 0.52 with Broad Math. Mathematics Reasoning on the WIAT had correlations of 0.66 with Broad Math, 0.60 with Math Calculation Skills, and 0.60 with Math Reasoning of the WJ-III-ACH. The Broad Written Language Cluster of the WJ-III-ACH had correlations of 0.47 with the Writing Composite and 0.47 with Written Expression on the WIAT. Finally, the Basic Writing Skills Cluster had correlations of 0.69 with the Writing Composite and 0.57 with Written Expression on the WIAT (McGrew & Woodcock, 2001).

An independent study of concurrent validity was conducted using the WJ-III-ACH and the *Test of Silent Word Reading Fluency* (TOSWRF; Mather, Hammill, Allen, & Roberts, 2004). This study used a sample of 98 students ranging in age from 7 years to 13 years, who were either diagnosed with a reading disorder or were identified as being at-risk for reading failure as evidenced by below grade-level reading scores “based on group achievement data” (pg. 3). Results of the study provided evidence of concurrent validity between the TOSWRF and specific subtests from the WJ-III-ACH; namely, Letter-Word Identification (0.58), Reading Fluency (0.59), Passage Comprehension (0.55), Broad Reading Cluster (0.66), and Spelling (0.58).

Reliability. As previously reported for the WJ-III-COG, the following statement also holds true for the WJ-III-ACH, the "reliability characteristics of the WJ III meet or exceed basic standards for both individual placement and programming decisions" (McGrew & Woodcock, 2001; p. 48). The WJ III *Technical Manual* (McGrew & Woodcock, 2001) reported median reliability coefficients of .94 for Broad Reading, .95 for Broad Math, .94 for Broad Written Language, and .87 for Oral Language – Standard.

Indiana Statewide Test of Educational Progress-Plus. The *Indiana Statewide Test of Educational Progress-Plus* (ISTEP+; Indiana Department of Education, ISTEP+ Program Manual, 2011) is a criterion-referenced test based on Indiana's Academic Standards. ISTEP+ purports to measure which Indiana Academic Standards that an individual student in the state of Indiana has mastered, particularly in the areas of English/language arts, mathematics, science, and social studies. Furthermore, ISTEP+ states that information obtained from the testing is to identify student strengths and weaknesses in order to plan for instruction (Indiana Department of Education, ISTEP+ Guide to Test Interpretation, 2011).

Validity. The ISTEP+ purports that staff from the Department of Education in conjunction with "educators from around the state, continually strive for the best match between that which should be taught and that which is tested" (Indiana Department of Education, ISTEP+ Program Manual, 2011; p. 78). The ISTEP+ *Program Manual* (2011) reports that Exploratory and Confirmatory Factor Analyses have been conducted on the ISTEP+. As part of the Exploratory Factor Analysis, the Kaiser-Meyer Olkin (KMO; Kaiser, 1970, 1974) measure of sampling adequacy was used. KMO values reportedly ranged from 0.90 to 0.98. Next, the manual reports the Kaiser criterion was used to compute eigenvalues in order to determine how much variability is accounted for by each factor. The reader is referred to specific tables reporting eigenvalues; however, such tables were not present in the manual. The ISTEP+ indicates that a single model factor is present for all grades and in each of the content domains. Finally, the ISTEP+ *Program Manual* (2011) states, "Summary inspection across all the criteria – variance, ratio of eigenvalues, scree plots, and goodness-of-fit indices – seems to indicate that the tests for each grade and content area, and for each subgroup, are essentially unidimensional" (p. 108).

Reliability. The ISTEP+ *Program Manual* (2011) reports that one measure of its reliability is that the test is administered in a standard manner. Furthermore, the *Program Manual* (2011) reports it obtained several measures of reliability including item-level reliability, test-level reliability, and proficiency-level reliability. Although the ISTEP+ *Program Manual* (2011) refers the reader to specific tables with data, such tables were not included within the manual. However, some reliability data was provided in the narrative of the manual. It is noted that the correlation data provided in the ISTEP+ *Program Manual* (2011) was condensed to include all items within a subject area, as well as condensed to include all grade areas. For example, the mean intra-class correlation provided for mathematics would have included the means for all items at each specific grade level within the mathematics subject domain.

The ISTEP+ *Program Manual* (2011) reports item-level reliability data. One form of item-level reliability data provided includes differential-item functioning, which purports to measure bias in the items through errors made by specific subgroups. The ISTEP+ purports to be a relatively fair test with minimal item bias. According to the ISTEP+ *Program Manual* (2011) approximately "7% of the operational items exhibited gender or ethnic differential-item functioning" (p. 103).

Another form of item-level reliability provided by the ISTEP+ is inter-rater reliability. Indeed, the ISTEP+ reports extensive use of hand scorers, and therefore, is concerned with the consistency between scorers. Table 3.2 references inter-rater reliability by providing intra-class correlations for each of the four subject areas.

Table 3.2.

ISTEP+ Intra-Class Correlations

Subject	*M*
English/language arts	0.89
Mathematics	0.94
Social Studies	0.87
Science	0.94

Source: (Indiana Department of Education, ISTEP+ Program Manual, 2011)

Finally, the ISTEP+ reports measuring the internal consistency of the ISTEP+ through the use of point-biserial correlations as a final way of measuring item-level reliability. Table 3.3 references point-biserial correlations for each of the four subject areas.

Table 3.3.

ISTEP+ Point-Biserial Correlations

Subject	*M*
English/language arts	0.14 to 0.58
Mathematics	0.13 to 0.70
Social Studies	0.15 to 0.53
Science	0.11 to 0.53

Source: (Indiana Department of Education, ISTEP+ Program Manual, 2011)

In addition to item-level reliability, the ISTEP+ reports test-level reliability data as a means of measuring how well the overall ISTEP+ questions consistently measure the domain constructs. As with the item-level statistics the ISTEP+ *Program Manual* (2011) refers readers to tables which are not present in the manual. Table 3.4 provides total test reliability coefficients, specifically Cronbach's α (Cronbach, 1951) for each of the content domains.

Table 3.4.

ISTEP+ Reliability

Subject	*α*
English/language arts	0.91
Mathematics	0.91
Social Studies	0.86
Science	0.87

Source: (Indiana Department of Education, ISTEP+ Program Manual, 2011)

The ISTEP+ provides proficiency-level reliability data as an indication that students are correctly classified as proficient or not proficient in content areas by their scores on the ISTEP+. The ISTEP+ reports the proficiency-level reliability as ".088 across all grades and content areas" (Indiana Department of Education, ISTEP+ Program Manual, 2011; p. 106). NCLB (2001) requires that student performance on state standardized testing be reported in terms of at least three performance levels, determined by the State Board of Education, that describe what a student can do with regard to the content and skills tested by a state's particular test. The ISTEP+ has three performance levels labeled as Pass+, Pass, and Did Not Pass, with the Pass level labeled as the level at which a student demonstrated proficiency within a particular subject area (Indiana Department of Education, Guide to Test Interpretation, 2011). Additionally, the ISTEP+ uses cut scores, determined by using the Bookmark procedure, to separate the three performance levels, meaning students must achieve a particular score in order to attain the Pass or Pass+ level (Indiana Department of Education, ISTEP+ Program Manual, 2011). The ISTEP+ Program Manual (2011) refers the reader to a technical guide for specifics on how cut scores were determined using the Bookmark procedure; however, an extensive literature review search by this writer for the technical guide produced no results for the guide. The ISTEP+ Program Manual states, "…the Bookmark procedure (Mitzel, Lewis, Patz, & Green, 2001) is a well-documented and highly regarded procedure that has been demonstrated by independent research to produce reasonable cut scores on tests across the country" (p. 104).

Since the ISTEP+ *Program Manual* (2011) states that reliability data is comparable to data obtained during past reliability studies, Table 3.5 references reliability data for the 2003 ISTEP+.

Table 3.5.

ISTEP+ Reliability

Grade	*English/language arts*		*Mathematics*	
	N	Reliability	N	Reliability
3	6007	0.90	6025	0.90
6	7159	0.92	7171	0.92
8	6978	0.92	7064	0.94

N refers to the number of students on which the statistic was based
Source: (Jacobs, 2005)

Procedures. Participants in this study had been administered a psychoeducational evaluation consisting of the *Woodcock-Johnson III Tests of Cognitive Abilities* (WJ III COG; Woodcock, McGrew, & Mather, 2001a) and the *Woodcock-Johnson III Tests of Achievement* (WJ III ACH; Woodcock, McGrew, & Mather, 2001b) among other measures using a flexible battery approach. Additionally, participants had taken the *Indiana Statewide Testing for Educational Progress-Plus* (ISTEP+; Indiana Department of Education, ISTEP+ Program Manual, 2011). The psychoeducational evaluations were completed by licensed school psychologists in an elementary or middle school and testing time for each student varied depending on the amount of assessment that was deemed necessary by the school psychologist administering the psychoeducational evaluation in order to fulfill requirements set forth under Article 7. The psychoeducational evaluations were conducted with individual students in a one-on-one setting using standardized procedures set forth in the manuals of the respective assessment measures. However, administration of the ISTEP+ (Indiana Department of Education, ISTEP+ Program Manual, 2011) was conducted in a group format utilizing standardized procedures set forth in the ISTEP+ Program Manual.

Statistical Procedures and Data Analysis.

The current study plans to answer questions about whether several variables collectively predict some outcome and what the relative contribution of each is. More specifically, the current study will investigate: (1) the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ III COG, and English/language arts performance on the ISTEP+; (2) the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ III COG, and Math performance on the ISTEP+; (3) the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ III ACH, and reading performance on the ISTEP+; and (4) the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ III ACH, and Math performance on the ISTEP+.

Multiple regression (Meltzoff, 1998) will be used to address the research hypotheses, as the use of a multiple regression technique would be the appropriate statistical analysis to answer these questions. Additionally, regression techniques can be used when the Independent Variables (IV) are correlated with each other, as well as with the Dependent Variable (DV) (Tabachnick & Fidell, 2001). Therefore, to answer the questions of the current study, the use of multivariate regression will be utilized since the purpose of the study will be to determine what each IV, specific CHC broad (stratum II) cognitive abilities, adds to the prediction of the DV, performance on the ISTEP+.

It should be noted that the ISTEP+ was changed from the dependent variable to become the independent variable in this part of the analysis. The theoretical decision to make this change was due to the interest in examining what the ISTEP+ measures with regard to academic achievement.

Chapter IV

Results

The purpose of this study was to examine the relationship between a standardized norm-referenced cognitive and achievement test and a statewide high-stakes test. This is an important consideration given the lack of research in this area and the need to provide validation of high-stakes tests. Following a presentation of the descriptive data for the study participants, the results of the inferential analyses are discussed in terms of the research questions.

Descriptive Statistics

Means and standard deviations for the WJ-III-COG subtests and the WJ-III-ACH subtests for the sample are presented in Table 4.1. The means of the WJ-III-COG subtests were lower than the mean scores of the standardization sample means of 100 and standard deviations of 15. Similarly, the means of the WJ-III-ACH subtests were lower than the mean scores of the standardization sample means of 100 and standard deviations of 15.

Table 4.1

Means and Standard Deviations for the WJ-III-COG and WJ-III-ACH

Test	*M*	*SD*
WJ-III-COG		
General Intellectual Ability	80.73	12.12
Verbal Comprehension	82.58	13.36
Visual-Auditory Learning	83.29	15.23
Spatial Relations	93.64	9.57
Sound Blending	97.98	12.90
Concept Formation	81.95	12.95
Visual Matching	83.45	19.92
Numbers Reversed	84.50	17.31
WJ-III-ACH		
Letter-Word Identification	85.85	13.39
Reading Fluency	83.44	11.90
Passage Comprehension	81.27	13.59
Calculation	84.88	16.25
Math Fluency	81.14	15.41
Applied Problems	85.13	12.82

Note: $N = 45$

Means and standard deviations for the ISTEP+ English/Language Arts and Math subtests for the sample are presented in Table 4.2.

Table 4.2

Means and Standard Deviations for the ISTEP+

Test	*M*	*SD*
English/Language Arts	401.05	38.724
Math	375.91	62.778

Note: *N* = 45

Inferential Analyses

The purpose of the present study was to answer four primary questions: (1) What is the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ-III-COG, and English/Language Arts performance on the ISTEP+; (2) What is the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ-III-COG, and Math performance on the ISTEP+; (3) What is the relationship between specific CHC broad (stratum II) achievement abilities, as measured by the WJ-III-ACH, and English/Language Arts performance on the ISTEP+; and (4) What is the relationship between specific CHC broad (stratum II) achievement abilities, as measured by the WJ-III-ACH, and Math performance on the ISTEP+? To address these questions, a multivariate regression analysis was performed for all four of the questions using IBM SPSS 20 (2011).

To answer the first research question: What is the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ-III-COG, and English/Language Arts performance on the ISTEP+, a multivariate regression analysis was conducted using the English/Language Arts test of the ISTEP+ (Indiana Department of Education, ISTEP+ Program Manual, 2011) as the dependent variable, and the first seven subtests of the WJ-III-COG as the independent variables (i.e. Verbal Comprehension, Visual-Auditory Learning, Spatial Relations, Sound Blending, Concept Formation, Visual Matching, and Numbers Reversed). The analyses indicated that the dependent variable, English/Language Arts, was not statistically significantly related to the independent variables (IVs), the seven subtests of the WJ-III-COG ([R^2 = .199, F(7, 33) = 1.174, p = .344)].

To answer the second research question: What is the relationship between specific CHC broad (stratum II) cognitive abilities, as measured by the WJ-III-COG, and Math performance on the ISTEP+, a multivariate regression analysis was conducted using the math test of the ISTEP+ as the dependent variable, and the first seven subtests of the WJ-III-COG as the independent variables (i.e. Verbal Comprehension, Visual-Auditory Learning, Spatial Relations, Sound Blending, Concept Formation, Visual Matching, and Numbers Reversed). The analyses indicated that the dependent variable, math, was not statistically significantly related to the independent variables (IVs), the seven subtests of the WJ-III-COG ([R^2 = .168, F(7, 33) = .953, p = .480)].

To answer the third research question: What is the relationship between specific CHC broad (stratum II) achievement abilities, as measured by the WJ-III-ACH, and English/Language Arts performance on the ISTEP+, a multivariate regression analysis was conducted using the Letter-Word Identification, Reading Fluency, and Passage Comprehension subtests of the WJ-III-ACH as the dependent variables, and the

English/Language Arts test of the ISTEP+ as the independent variable. The Wilks' Lambda test statistic demonstrated the combined dependent variables (DVs), Letter-Word Identification, Reading Fluency, and Passage Comprehension were significantly related to the independent variable (IV), English/Language Arts of the ISTEP+ (Indiana Department of Education, ISTEP+ Program Manual, 2011) [$F(3, 53) = 2.959$, $p = .040$] at the $\alpha = .05$.

As a post hoc follow up to identify which of the independent variables are associated with the significant English/Language Arts multivariate result, individual univariate regression analyses were conducted in which the English/Language Arts test was related to each of the cognitive measures individually. The univariate results are displayed in Table 4.3 and indicated that English/Language Arts was significantly related to Passage Comprehension ($p = .004$) at the $p <.0167$ (Bonferroni corrected alpha), with a parameter estimate of $B = .059$, whereas English/Language Arts was not significantly related to Letter-Word Identification ($p =.036$), with a parameter estimate of $B = .047$ and Reading Fluency ($p = .096$) at the $p < .0167$ (Bonferroni corrected alpha), with a parameter estimate of $B = .035$. This positive coefficient value for Passage Comprehension indicates that examinees that had higher Passage Comprehension scores also had higher scores on the English/Language Arts test.

Table 4.3 includes the univariate regression analyses showing the relationship between English/Language Arts and each of the individual cognitive measures.

Table 4.3

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Eng/Lang	Lttr-Wrd ID	713.667	1	713.667	4.628	.036
	Read Fl	392.565	1	392.565	2.866	.096
	Pass Comp	1129.475	1	1129.475	9.151	.004*

* denotes significance at the $p \leq .0167$ (Bonferroni corrected alpha)

To answer the fourth research question: What is the relationship between specific CHC broad (stratum II) achievement abilities, as measured by the WJ-III-ACH, and Math performance on the ISTEP+, a multivariate regression analysis was conducted using the Calculation, Math Fluency, and Applied Problems subtests of the WJ-III-ACH as the criterion, and the Math test of the ISTEP+ as the predictor. Results indicated that Wilks' Lambda criterion demonstrated the combined dependent variables (DVs), Calculation, Math Fluency, and Applied Problems were not significantly related at the $\alpha = .05$ to the independent variable (IV), Math of the ISTEP+ ([$F(3, 53) = 2.673$, $p = .057$)].

Chapter V

Discussion

This chapter is divided into five sections: (1) summary of the present study, (2) discussion and implications of the findings, (3) limitations and delimitations of the study, (4) directions for future research, and (5) conclusions.

Summary of the Study

The purpose of the present study was twofold. First, the study sought to investigate the relationship between a norm-referenced cognitive ability measure, the *Woodcock-Johnson III Tests of Cognitive Abilities* (WJ-III-COG; Woodcock, McGrew, & Mather, 2001a), and a mandated statewide high-stakes test, the *Indiana Statewide Test for Educational Progress- Plus* (ISTEP+; Indiana State Department of Education, I.s., 2011). This is an important area of investigation as the question of whether intelligence is a determining factor in mandated high-stakes testing may lead to an increased understanding of the role that cognitive factors play in high-stakes tests. This is particularly important for children being considered for special education as many childhood conditions which would lead to a special education diagnosis have associated cognitive processing concerns, and school psychologists may be able to use this information to provide recommendations of accommodations and modifications for high-stakes tests. The Cattell-Horn-Carroll theory of cognitive processing has been well supported in the research for predicting academic achievement (Evans, Floyd, McGrew,

& Leforgee, 2002; Floyd, Evans, & McGrew, 2003; Floyd, McGrew, & Evans, 2008). However, there is a dearth of research investigating the relationship between CHC theory and high-stakes testing.

The second purpose of the study was to examine the relationship between a norm-referenced achievement measure, the *Woodcock-Johnson III Tests of Achievement* (WJ-III-ACH; Woodcock, McGrew, & Mather, 2001b), and a state mandated high-stakes measure, the *Indiana Statewide Test for Educational Progress- Plus* (ISTEP+; Indiana State Department of Education, I.s., 2011). The increased mandate for the use of high-stakes testing by states has raised questions about what these tests are measuring. The WJ-III-ACH is a well-validated measure of academic achievement and performance on this measure has been linked to academic success (Evans, Floyd, McGrew, & Leforgee, 2002; Floyd, Evans, & McGrew, 2003; Floyd, McGrew, & Evans, 2008). As such, establishing a relationship between high-stakes testing and the WJ-III-ACH would help determine if two commonly used achievement measures, one used for diagnostic purposes, and the other used to benchmark progress, have commonalities in the way they measure academic achievement. If they are found to be related, as was found in the current study with reading comprehension, perhaps school psychologists could use the results from the WJ-III-ACH to better inform accommodations, interventions, modifications, and recommendations made in a student's IEP in regards to high-stakes testing. Yet, there currently do not appear to be any earlier studies investigating the relationship of the WJ-III-ACH and any high-stakes tests including the ISTEP+.

All of the participants in this study were administered the WJ-III-COG and WJ-III-ACH as part of a psychoeducational evaluation, and had also participated in the ISTEP+. Data for this study were collected from an archival data set retrospectively after the psychoeducational evaluations and the ISTEP+ had been completed following approval by the Ball State University Institutional Review Board (IRB). The data from this study were collected from 45 participants, 24 males and 21 females (mean age = 11.1 years; SD = 2.3 years).

Discussion and Implications

Four main questions were proposed for this study. First, it was the aim of this study to investigate the relationship between CHC broad abilities and performance on the ISTEP+. More specifically, the study investigated the relationship between specific CHC broad cognitive abilities, as measured by the WJ-III-COG, and English/Language Arts performance on the ISTEP+. Second, the study examined the relationship between specific CHC broad cognitive abilities, as measured by the WJ-III-COG, and Math performance on the ISTEP+. Third, the relationship between specific CHC broad cognitive abilities, as measured by the WJ-III-ACH and English/Language Arts performance on the ISTEP+ was investigated. Finally, the current study investigated the relationship between specific broad cognitive abilities, as measured by the WJ-III-ACH, and Math performance on the ISTEP+.

To answer these four questions, a multivariate regression analysis was conducted. Multivariate regression is a statistical technique used when determining whether several variables collectively predict some outcome and what the relative contribution of each variable is to the outcome. Additionally, given that the independent variables are theoretically related to one another, as well as to the dependent variables, the use of regression techniques is appropriate (Tabachnick & Fidell, 2001). Results of the multivariate regression analysis conducted using the English/Language Arts and Math tests of the ISTEP+ as the criterion, and the first seven subtests of the WJ-III-COG as the predictors revealed the criterion variables were not significantly related to the predictor variables. This suggests that the findings indicated that achievement, as measured by the ISTEP+, is not statistically significantly related to cognitive abilities, as measured by the first seven subtests of the WJ-III-COG. Although the finding that English/Language Arts and Math from the ISTEP+ was not statistically significantly related to cognitive abilities was somewhat surprising given the well-established relationship between intelligence and academic achievement, there are several possible reasons that a significant relationship did not exist between English/Language Arts and Math and cognitive abilities from the WJ-III-COG. First, and perhaps most obvious, it could be that no significant relationship exists. This was somewhat unexpected as intelligence has shown to be a strong predictor of academic success (Evans, Floyd, McGrew, & Leforgee, 2001; Floyd, Evans, & McGrew, 2003; Floyd, Keith, Taub, & McGrew, 2007; Floyd, McGrew, & Evans, 2008), and the ISTEP+ should theoretically be measuring some aspect of what a student has learned in school. Another reason no statistically significant relationship was found between cognitive abilities and English/Language Arts and Math from the ISTEP+ is that ISTEP+ may not measure achievement skills in the same way that other achievement tests measure those abilities. ISTEP+ is a criterion-referenced test and it is possible that the achievement skills being evaluated are different from the achievement skills measured on norm-referenced tests. Additionally, another reason that a statistically significant finding did not exist between English/Language Arts and Math and the cognitive abilities might have been due to the sample being one of convenience. Only cases containing the first seven subtests of the WJ-III-COG were chosen for the current study due to the small number of cases that had been administered the *Woodcock Johnson Tests of Cognitive Abilities- Third Edition, Extended Battery*; therefore, utilizing only one subtest each as measures of the seven broad CHC cognitive abilities: Crystallized Knowledge (G*c*), Fluid Intelligence (G*f*), Auditory Processing (G*a*), Processing Speed (G*s*), Long-Term Memory (G*lr*), Short-Term Memory (G*sm*) and Visual-Spatial Thinking (G*v*). It is not uncommon, when measuring each of the broad CHC cognitive abilities, to administer the first seven subtests of the WJ-III-COG, as well as seven more subtests from the WJ-III-COG Extended to have two qualitatively different tests comprise each of the seven CHC cognitive abilities. However, according to Schrank (2006), not only is each test a measure of a specific narrow ability, but it can also be considered a single measure of a broad cognitive ability. Indeed, since a plethora of research has indicated that broad CHC cognitive abilities are related to specific areas of achievement, one would expect that the specific abilities that comprise the broad abilities would also be related to achievement. A vast amount of research has indicated that each of the broad CHC cognitive abilities is related to academic achievement, particularly reading and math (e.g., Evans, Floyd, McGrew, & Leforgee, 2002; Floyd, Keith, Taub, & McGrew, 2007; Garcia & Stafford, 2000; McGrew, Flanagan, Keith, & Vanderwood, 1997). A third reason why a significant finding may not have been found between English/Language Arts and Math and the cognitive abilities was due to the sample being only students referred for special education. It was noted that the mean scores and standard deviations for the study participants were lower than the mean scores and standard deviations of the standardization sample. Indeed, intelligence tests are a common component of evaluations that attempt to determine the presence of a learning disability or learning disorder because they represent an excellent metric of a student's potential which can then be compared to a student's actual current academic performance. Given that the majority of the sample qualified for special education, they may have had different cognitive processing abilities than would have been found in a healthy sample; this is supported by the pattern of below average scores on the WJ-III-COG. It is possible that a significant result would have occurred if the sample included children who had not been referred for special education and who were not expected to have learning difficulties or other conditions which qualify for placement in special education.

To answer the question regarding the relationship between the WJ-III-ACH and the English/Language Arts portion of the ISTEP+, a multivariate regression analysis was used with Letter-Word Identification, Reading Fluency, and Passage Comprehension subtests of the WJ-III-ACH as the dependent variables, and the English/Language Arts tests of the ISTEP+ as the independent variables. A significant relationship was found between the Letter-Word Identification, Reading Fluency, and Passage Comprehension subtests and English/Language Arts at the $\alpha = .05$. Individual univariate analyses were conducted as part of a post hoc follow up in which the English/Language Arts test was related to each of the WJ-III-ACH CHC cognitive abilities individually. To control for committing a Type 1 error (rejecting the null hypothesis when it's true), Bonferroni correction was used (0.05/3) to counteract the problem of multiple comparisons, meaning as the number of hypotheses increases, the likelihood of finding a rare event increases. The univariate results indicated that only Passage Comprehension was significantly related to English/Language Arts $\alpha = 0.0167$. The result of Passage Comprehension being the only variable statistically significantly related to English/Language Arts was surprising in that the ISTEP+ English/Language Arts portion of the test measures not only reading comprehension, but purports to measure word recognition, fluency, and vocabulary development (Indiana Department of Education, ISTEP+ Program Manual, 2011). Interestingly, previous research has indicated oral reading fluency Curriculum Based Measurement is highly predictive of reading performance on state-mandated, high-stakes tests (McGlinchey & Hixson, 2004; Stage & Jacobson, 2001). Therefore, one would have expected Reading Fluency from the WJ-III-ACH to be statistically significantly related to English/Language Arts. However, it is possible that the reading skills tested in the state-mandated reading tests in those studies vary significantly from the ISTEP+. This has implications for Indiana school districts that use oral reading fluency CBM as a predictor of overall reading ability. It is not uncommon for some school districts to use reading fluency CBM scores as benchmarks for student achievement and to also use these scores for progress monitoring a student's progress to reading interventions. The possibility that reading skills tested in state-mandated reading tests might vary significantly from the ISTEP+ is important in that school districts will likely utilize reading interventions that focus more on reading fluency rather than on interventions for reading comprehension. Yet, according to the current study, ISTEP+ has a stronger relationship with reading comprehension than reading decoding or reading fluency as measured by the WJ-III-ACH. School districts might want to incorporate reading comprehension interventions when monitoring the progress of at-risk students who will be taking ISTEP+.

To answer the question regarding the relationship between the WJ-III-ACH and the Math portion of the ISTEP+, a multivariate regression analysis was used using the Calculation, Math Fluency, and Applied Problems subtests of the WJ-III-ACH as the dependent variables, and the Math test of the ISTEP+ as the independent variable. Results of the analysis indicated no significant relationship exists between these subtests of the WJ-III-ACH and the Math portion of the ISTEP+. This result was unexpected in that the Math portion of the ISTEP+ purportedly measures number sense, computation, and problem solving, which are math skills assessed by the WJ-III-ACH. However, the ISTEP+ is a criterion-referenced test and is based on the math standards deemed important by the Indiana Board of Education. It is possible that the since the WJ-III-ACH is a norm-referenced test, it does not measure the same criteria as on the ISTEP+. Indeed, the WJ-III-ACH provides a measure of math for a number of different age levels and grade levels, and to be time-efficient, only a limited number of math questions can be evaluated at each level. Another possibility for the non-significant result between the WJ-III-ACH and the Math portion of the ISTEP+ is that a significant result does not exist. It's possible that the two measures are not related. For example, the WJ-III-ACH might be measuring qualitatively different math skills than the ISTEP+ or one of the tests might be measuring a student's aptitude for math, whereas the other test is measuring how much math a student has mastered. Furthermore, the ISTEP+ was designed to measure math skills that match the curriculum taught in Indiana schools, whereas the WJ-III-ACH was note designed to measure any particular curriculum.

Limitations and Delimitations of the Study

There were several limitations in the current study. The first limitation is that the data from this study were gathered from an archival data set. Gathering data in such a way poses several limitations in and of itself. Data on assessments on children in this study were subject to selective data gathering procedures, and as such, data such as parental level of education was not gathered on every participant. According to Magnuson (2007), maternal educational attainment has been shown to be a predictor of a student’s academic achievement but by using data from an archival data set, maternal level of education was not gathered and could not be controlled for to determine its impact on achievement in the current study.

Another limitation of using an archival data set is that although the administrators of the psychoeducational evaluation were school psychologists licensed by the Indiana Department of Education, multiple licensed practitioners had completed the evaluations. Therefore, some of the data that were collected for this study may have been at a higher level of possible data entry error, such as record-keeping and assessment scoring procedural errors.

An additional limitation of the current study was the use of a sample of convenience. Indeed, the current study was conducted with students who had been referred for a psychoeducational evaluation and subsequently received a special education label. Students in special education receive an Individualized Education Plan

(IEP) and often receive accommodations on high-stakes testing, such as extended time or use of a calculator, whereas the same types of accommodations would not be used during the psychoeducational evaluation. Additionally, the same types of accommodations are not afforded students not in special education. It is possible that the study sample participant's scores on ISTEP+ might have been influenced by the use of accommodations although this information was not collected for this study. For example, an appropriate accommodation for a student in special education is time and a half on high-stakes testing. Even if that student finishes the test questions in the state-allotted time for that test, the student might use that time to review questions of difficulty, which might enhance that student's chance of catching potential errors. A data set with students who had not been referred for a psychoeducational evaluation and did not receive accommodations would serve as an important comparison group.

Another limitation of the study was the data from the participants in this study were gathered from an urban Midwestern area; therefore, the generalizability of the study may be limited. Furthermore, the area from which the data were collected has a high rate of poverty, which might impact the generalizability to other regions with higher level of socioeconomic status. Indeed, Cooper & Schlesser (2006) noted socioeconomic status, particularly low socioeconomic status, has negative effects on academic achievement. It was noted that the participants' cognitive and achievement standard scores were lower than those of the standardization sample, which might have been impacted by socioeconomic status. However, socioeconomic status was not controlled in the current study.

It is important to consider that limitations in the design of the current study possibly prevented a significant relationship from being found between the WJ-III-COG and the English/Language Arts and Math portions of the ISTEP+, as well as between the WJ-III-ACH and the Math portion of the ISTEP+. For example, the sample size was a limitation to the current study. Given the sample size, the use of Multivariate Analysis of Variance (MANOVA) to determine if special education placement, socio-economic status, and gender independently affect scores on the WJ-III-COG, WJ-III-ACH, and ISTEP+ were not conducted due to the possibility that the statistical power was likely to be very small and the chance of finding a significant effect, even when there was one, was more unlikely.

The greatest delimitation to the current study is that there currently are no published studies investigating the relationship between the WJ-III Battery and the ISTEP+. The current study will lend itself to the literature with regard to the CHC Theory's relationship to academic achievement. Furthermore, the WJ-III Battery has been shown to predict academic achievement (Floyd, McGrew, & Evans, 2008; Floyd, Keith, & McGrew, 2009; Taub & McGrew, 2004; Floyd, Evans, & McGrew, 2003; Evans, Floyd, McGrew, & Leforgee, 2001). Although only one achievement subtest, Passage Comprehension, was found to be statistically significantly related to English/Language Arts of the ISTEP+, the current study suggests that Passage Comprehension appears to be useful in predicting ISTEP+ English/Language Arts achievement and would warrant further investigation. The finding that Passage Comprehension was significantly related to English/Language Arts of the ISTEP+ is important because despite the limitations of the current study, a significant relationship was found to exist, and it might signal which areas of reading are most important to investigate with relation to the ISTEP+. It is possible that significant results may have been found if the limitations to the current study could have been controlled.

Directions for Future Research

The directions for future research are designed to improve on any limitations in the current study and draw upon areas of the greatest need in the study of the relationship between CHC theory and high-stakes testing. First and foremost, a larger sample size where other important variables are gathered would be useful. For example, it would be useful to know how socio-economic status, parental level of education, and special education status affect scores on the WJ-III Battery and ISTEP+. These variables would be able to be explored with the current study design with the addition of a larger sample. Additionally, rather than using a sample of convenience, it would be useful to conduct the research with students who had been referred for special education, as well as students who had not been referred for special education. By conducting research with students who had and had not been referred for special education might help determine if special education status affected scores on the measures. It would also be useful to examine the effect on the ISTEP+ of the 14 subtests of the WJ-III-COG that comprise the CHC model rather than just the first seven subtests, given that they are only single measures of the broad CHC abilities. Because CHC theory has relatively recently been applied to many intelligence tests, it might be useful to use other intelligence batteries to see if different results are found. Moreover, the use of other achievement tests might be beneficial to see if similar results to the WJ-III-ACH and English/Language Arts are found. Similarly, most of the research in the literature regarding high-stakes testing and academic achievement focuses on CBM and high-stakes tests. Therefore, it might be useful to examine the relationship between CHC theory and CBM.

Implications for School Psychologists

The current study has implications for practicing school psychologists. School psychologists are being asked more frequently to predict which students might be at-risk for not passing state-wide high-stakes tests. By determining which CHC abilities are most closely linked to different academic areas, school psychologists can make more accurate predictions about possible achievement performance. For example, a school psychologist might make predictions during a case conference regarding a student's ability to pass the reading portion of a high-stakes test. But, more importantly, school psychologists might identify strengths and weaknesses with a student's cognitive abilities and help design appropriate interventions. Additionally, since the WJ-III-ACH Passage Comprehension is a norm-referenced, well-supported, empirically sound test of achievement, it measures multiple dimensions of reading comprehension, whereas the English/Language Arts portion of the ISTEP+ is a criterion-referenced test of reading, it is supposedly measuring specific reading criterion or reading skills. Therefore, interventions tied to results of Passage Comprehension might cover more comprehensive or general reading comprehension skills, whereas interventions based on reading fluency CBM will likely be based on a single criterion and may miss a student's true difficulty in reading. Therefore, school psychologists need to elucidate the type of reading problem a student has and recommend targeted interventions that are correlated with that reading deficit.

Conclusions

In conclusion, due to federal education mandates, such as No Child Left Behind, there has been increased accountability for student performance on state-wide high-stakes tests. Therefore, the current study sought to investigate the relationship between CHC cognitive abilities and performance on a state-mandated high-stakes achievement test, ISTEP+. Research has shown that CHC cognitive abilities, as measured by the WJ-III-COG and WJ-III-ACH, have been highly predictive of academic achievement. However, with no current studies examining the relationship between the WJ-III Battery and ISTEP+, the current study was important at studying this relationship. Because the current study was interested in measuring multiple independent and dependent variables, multivariate regression was used to conduct the statistical analyses. Analyses suggested that there was no statistically significant relationship between the seven subtests of the WJ-III-COG and ISTEP+ with a small, non-normative sample. Additionally, analyses suggested there was no statistically significant relationship between the WJ-III-ACH and the math portion of the ISTEP+. The only statistically significant finding was between Passage Comprehension of the WJ-III-ACH and the English/Language Arts portion of the ISTEP+.

References

Afflerbach, P. (2005). National reading conference policy brief: High stakes testing and Reading assessment. *Journal of Literacy Research*, 37(2), 151-162.

Bloomfield, D. C. (2007). No child left behind act (NCLB). In Bursztyn, A. M. (Ed.), *The Praeger Handbook of Special Education,* (pp. 43-45).

Braden, J. P., & Tayrose, M. P. (2008). Best practices in educational accountability: High-stakes testing and educational reform. In Thomas, A., & Grimes, J. (Eds.), *Best Practices in School Psychology*, (pp. 575-588).

Brody, N. (1985). The validity of tests of intelligence. In B.B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 353-390). New York: John Wiley & Sons.

Brody, N. (2000). History of theories and measurements in intelligence. In R. Sternberg's (Ed.), *Handbook of Intelligence* (pp. 16-33). New York: Cambridge University Press.

Bursztyn, A. M. (2007). A brief history of intelligence testing. In A. M. Bursztyn (Ed.), *The Praeger Handbook of Special Education* (pp. 3-5). Westport, CT: Praeger Publishers.

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

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: Erlbaum.

Carter, L. F. (1984). The sustaining effects of compensatory and elementary education. *Educational Researcher*, 13(7), 4-13.

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

Cattell, R. B. (1987). *Intelligence: Its structure, growth, and action.* New York: Elsevier Science Publishing Co.

Cooper, J. A., & Schleser, R. (2006). Closing the achievement gap: Examining the role of cognitive developmental level in academic achievement. *Early Childhood Education Journal*, 33(5), 301-306.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334.

Dehn, M. J. (2006). *Essentials of processing assessment.* Hoboken, NJ: John Wiley & Sons, Inc.

Defur, S. H. (2002). Education reform, high-stakes assessment, and students with disabilities. *Remedial & Special Education*, 23(4), 203.

Duffy, M., Giordano, V. A., Farrell, J. B., Paneque, O. M., & Crump, G. B. (2008). No child left behind: Values and research issues in high-stakes assessments. *Counseling and Values*, 53(1), 53.

Dutro, E., & Selland, M. (2012). "I like to read, but I know I'm not good at it": Children's perspectives on high-stakes testing in a high-poverty school. *Curriculum Inquiry*, 42(3), 340-367.

Education of All Handicapped Children Act of 1975 (P.L. 94-142), 20 U.S.C. Sec. 1400(d).

Elementary and Secondary Education Act of 1965, Pub. L. No. 89-10, 79 Stat. 27 (1965).

Elementary and Secondary School Act 0f 1994, Pub. L. No. 103-382, 108 Stat.3519 (1995).

Elliott, C. D. (1990). Differential Ability Scales. San Antonio, TX: The Psychological Corporation.

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2001). 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.

Fagan, T. K., & Wise, P. S. (2000). *School psychology: Past, present, and future* (2nd ed.). Bethesda, MD: National Association of School Psychologists.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). *Essentials of cross-battery assessment* (2nd ed.). Hoboken, NJ: John Wiley & Sons, Inc.

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C., & Dynda, A. (2008). Best practices in cognitive assessment. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed.). Washington DC: The National Association of School Psychologists.

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C., Mascolo, J. T. (2002). *The achievement test desk reference (ATDR): Comprehensive assessment and learning disabilities.* Boston: Allyn & Bacon.

Fletcher, J. M., Lyon, G. R., Fuchs, L. S., & Barnes, M. A. (2007). *Learning disabilities: From identification to intervention.* New York: The Guilford Press.

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 Schools*, 40(2), 155-171.

Floyd, R. G., McGrew, K. S., & Evans, J. J. (2008). The relative contributions of the Cattell-Horn-Carroll cognitive abilities in explaining writing achievement during childhood and adolescence. *Psychology in the Schools,* 45(2), 132-144.

Foschi, R., & Cicciola, E. (2006). Politics and naturalism in the 20th century psychology of Alfred Binet. *History of Psychology*, 9(4), 267-289.

Fritzberg, G. J. (2004). No child left behind? Assessing President Bush's assessment law. *Educational Foundations,* 18(3), 7-24.

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *The Journal of Special Education*, 4(2), 103-114.

Gregg, N., Hoy, C., Flaherty, D. A., Norris, P., Coleman, C., Davis, M., & Jordan, M. (2005). Documenting decoding and spelling accommodations for postsecondary students with dyslexia—It's more than processing speed. *Learning Disabilities: A Contemporary Journal*, 3, 1-17.

Gulliksen, H. (1968). Louis Leon Thurstone, experimental and mathematical psychologist. *American Psychologist*, 23(11), 786-802.

Haladyna, T. M. (2002). *Essentials of standardized achievement testing: Validity and accountability*. Boston: Allyn & Bacon.

Hintze, J. M., & Silberglitt, B. (2005). A longitudinal examination of the diagnostic accuracy and
predictive validity of R-CBM and high-stakes testing. *School Psychology Review*, 34(3), 372-386.

Horn, J. L.(1968). Organization of abilities and the development of intelligence. *PsychologicalReview*, 75, 242-259.

Horn, J. L. (1991). Measurement of intellectual capabilities: A review of theory. In K. S. McGrew, J. K. Werder, and R. W. Woodcock, *WJ-R Technical Manual* (pp. 197-232). Itasca, IL: Riverside.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: G*f*-G*c* 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.

Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline in educational equality. *British Educational Research Journal*, 31(5), 605-622.

Indiana Administrative Code 511 IAC 7-17 et.seq. 2002.

Indiana Department of Education Center for Exceptional Learners Office of Special Education. (2010). *Special Education Rules, Title 511 Article 7, Rules 32-47.* Indianapolis, IN: Author.

Indiana State Department of Education, I. s. (1988). *Indiana Statewide Testing for Educational Progress—ISTEP—Program Manual.* Indianapolis, IN: Author.

Indiana State Department of Education. (2003). *Indiana Statewide Testing for Educational Progress Plus-ISTEP+-Program Manual.* Indianapolis, IN: Author.

Indiana State Department of Education. (2010). *Indiana Statewide Testing for Educational Progress Plus-ISTEP+-Program Manual.* Indianapolis, IN: Author.

Indiana State Department of Education. (2011). *Indiana Statewide Testing for Educational Progress Plus-ISTEP+-Program Manual.* Indianapolis, IN: Author.

Indiana State Department of Education. (2011). *ISTEP+ Guide to Test Interpretation.* Indianapolis, IN: Author.

Indiana State Department of Public Instruction, I. s., Indiana State Commission for Higher Education, I. s., CTB/McGraw-Hill, M. A., & Indiana University, B. n., (1998). *ADescriptive guide to the 10th Grade ISTEP+ and High School Graduation Qualifying Exam for Indiana Parents and Students.* Indianapolis, IN: Author.

Individuals with Disabilities Act Amendments of 1997, Pub. L. No.105-17, 20 U.S.C. §1400 et.seq.

Individuals with Disabilities Education Improvement Act of 2004, Pub. L. No.108-446, §614 (2004).

Jencks, C., & Phillips, M. (Eds.) (1998). *The black-white test score gap.* Washington, DC: Brooking Institution Press.

Johnson, L. (2005). First, do no harm--An argument against mandatory high-stakes testing for students with intellectual disabilities. *Mental Retardation*, 43(4), 292-298.

Jones, L. V. (2001). Assessing achievement versus high-stakes testing: A crucial contrast. *Educational Assessment*, 7(1), 21-28.

Kaiser, H. F. (1970). A second-generation Little Jiffy. *Psychometrika*, 35, 401-415.

Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika,* 39, 31-36.

Kamphaus, R. W., Winsor, A. P., Rowe, E. W., & Kim, S. (2005). A history of intelligence test interpretation. In D. P. Flanagan and P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues (2nd ed.)* (pp. 23-38). New York: Guilford Press.

Kaufman, A. S., & Kaufman, N. L. (1985). The Kaufman Test of Educational Achievement. Circle Pines, MN: American Guidance Service.

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

Katsiyannis, A., Zhang, D., Ryan, J. B., & Jones, J. (2007). High-stakes testing and students with disabilities: Challenges and promises. *Journal of Disability Policy Studies,* 18(3), 160-167.

Katzman, L. (2007). High-stakes testing. In A. M. Bursztyn (Ed.) *The Praeger Handbook of Special Education* (pp. 127-129). Westport, CT: Praeger Publishers.

Kiernan, K. E., & Mensah, F. K. (2011). Poverty, family resources and children's early educational attainment: The mediating role of parenting. *British Educational Research Journal,* 37(2), 317-336.

Lin, J. (2006). The bookmark procedure for setting cut-scores and finalizing performance standards: Strengths and weaknesses. *Alberta Journal of Educational Research,* 52(1), 36-52.

Lopez, S. L. (2007). Social class and special education. In A. M. Bursztyn (Ed.) *The Praeger Handbook of Special Education* (pp. 26-28). Westport, CT: Praeger Publishers.

Mägi, K., & Kikas, E. (2009). School psychologists' role in school: Expectations of school principals on the work of school psychologists. *School Psychology International*, 30(4), 331-346.

Magnuson, K. (2007). Maternal education and children's academic achievement during middle childhood. *Developmental Psychology,* 43(6), 1497-1512.

Mather, N., Hammill, D. D., Allen, E. A., & Roberts, R. (2004). TOSWRF: *Test of Silent Word Reading Fluency:* Examiner's Manual. Austin TX: Pro-Ed.

Mather, N., Wendling, B. J., & Woodcock, R. W. (2001). *Essentials of WJ III tests of achievement assessment.* New York, NY: John Wiley & Sons, Inc.

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

Matarazzo, J. D. (1981). David Wechsler (1896-1981). *American Psychologist,* 36(12), 1542-1543.

McGlinchey, M. T., & Hixson, M. D. (2004). Using curriculum-based measurement to predict performance on state assessments in reading. *School Psychology Review*, 33(3), 193-203.

McGrew, K. S., & Wendling, B. J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools,* 47(7), 651-675.

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

Meltzoff, J. (1998). *Critical thinking about research: Psychology and related fields.* Washington DC: American Psychological Association.

Mitzel, H. C., Lewis, D. M., Patz, R. J., & Green, D. R. (2001). The bookmark procedure: Psychological perspectives. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 249-281). Mahwah, NJ: Erlbaum.

Moon, T. R. (2009). Myth 16: High-stakes are synonyms with rigor and difficulty. *Gifted Child Quarterly*, 53(4), 277-279.

Naglieri, J. A., & Das, J. P. (1997). *Cognitive Assessment System.* Itasca, IL: Riverside Publishing.

Naglieri, J. A., De Lauder, B. Y., Goldstein, S., & Schwebech, A. (2006). WISC-III and CAS: Which correlates higher with achievement for a clinical sample? *School Psychology Quarterly*, 21(1), 62-76.

Naglieri, J. A. & Otero, T. (2010). Cognitive assessment system: Redefining intelligence from a neuropsychological perspective. In A. Davis (Ed.). *Handbook of pediatric neuropsychology.* New York: Springer.

National Center for Education Evaluation and Regional Assistance. (2011). National assessment of IDEA overview. NCEE 2011-4026. *National Center for Education Evaluation and Regional Assistance.*

National Commission on Excellence in Education. (1983). *A nation at risk: A report to the nation and Secretary of Education.* Washington, DC: U.S. Department of Education.

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

No Child Left Behind Act of 2001. Retrieved October 22, 2011, from http://www2.ed.gov/policy/elsec/leg/esea02/index.html

Plake, B. S. (2011). Current state of high-stakes testing in education. In J. A. Bovaird, K. F. Geisinger, C. W. Buckendahl, J. A. Bovaird, K. F. Geisinger, C. W. Buckendahl (Eds.), *High-stakes testing in education: Science and practice in K–12 settings* (pp. 11-26).

President's Commission on Excellence in Special Education. (2002, July). A new era: Revitalizing special education for children and their families. In H. R. Turnbull, M. J. Stowe, & N. E. Huerta (Eds.), *Free appropriate public education: The law and children with disabilities* (4th ed.).

Psychological Corporation, The (2001). *Wechsler Individual Achievement Test-Second Edition.* San Antonio, TX: Author.

Resch, J. F., & Stevens, D. J. (1996). Variables related to mathematics achievement among black students. *The Journal of Educational Research,* 89(6), 346-350.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence,* 35(1), 83-92.

Rushton, S., & Juola-Rushton, A. (2008). Classroom learning environment, brain research and the "No Child Left Behind" initiative: 6 years later. *Early Childhood Education Journal*, 36(1), 87-92.

Sattler, J. M. (2002). *Assessment of children: Behavioral and clinical applications.* San Diego, CA: Author.

Sattler, J. M. (2008). *Assessment of children: Cognitive foundations (5th ed.).* San Diego, CA: Author.

Schrank, F. A., Flanagan, D. P., Woodcock, R. W., & Mascolo, J. T. (2002). *Essentials of WJ III cognitive abilities assessment.* New York, NY: John Wiley & Sons.

Schulte, A. C., & Villwock, D. N. (2004). Using high-stakes tests to derive school-level measures of special education efficacy. *Exceptionality*, 12(2), 107-126.

Schulte, A. C., Villwock, D. N., Whichard, S., & Stallings, C. F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district. *School Psychology Review*, 30(4), 487-506.

Scott, A. N., & Santos de Barona, M. (2007). Disability and civil rights. In A. M. Bursztyn (Ed.) *The Praeger Handbook of Special Education* (pp. 5-6). Westport, CT: Praeger Publishers.

Sharpe, M. N., & Hawes, M. E. (2003). *The intended and unintended consequences of large-scale assessments on the development of individualized education plans* (Tech. Report. No. 1). Minneapolis: University of Minnesota, Institute on Community Integration.

Skiba, R., & Rausch, M. K. (2004). The relationship between achievement, discipline, and race: An analysis of factors predicting ISTEP scores. *Center for Evaluation and Education Policy*, 1-5.

Stage, S. A., & Jacobsen, M. D. (2001). Predicting student success on a state-mandated performance-based assessment using oral reading fluency. *School Psychology Review*, 30(3), 407-419.

Stein, S. J. (2004). *The culture of education policy*. New York: Teachers College Press.

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Boston: Allyn & Bacon.

Taub, G. E., Floyd, R. G., Keith, T. Z., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly*, 23(2), 187-198.

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.

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence.* Boston: Houghton Mifflin.

Thomas, J. Y., & Brady, K. P. (2005). The Elementary and Secondary School Act at 40: Equity, accountability, and the evolving federal role in public education. *Review of Research in Education*, 29, 51-67.

Thorndike, R. L., Hagen, E. P., & Sattler, J. S. (1986). Stanford-Binet Intelligence Scale, Fourth Edition. Itasca, IL: Riverside.

Thorndike, R. M. (1990). Origins of intelligence and its measurement. *Journal of Psychoeducational Assessment*, 8(3), 223-230.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education (6th ed.).*Columbus, OH: Merrill Publishing Co/Prentice-Hall.

Tichá, R., Espin, C. A., & Wayman, M. (2009). Reading progress monitoring for secondary-school students: Reliability, validity, and sensitivity to growth of reading-aloud and maze-selection measures. *Learning Disabilities Research & Practice*, 24(3), 132-142.

Title I, Improving the Academic Achievement of the Disadvantaged, 34 C.F.R.§ 200.1et.seq. 2002.

Tuerk, P. W. (2005). Research in the high-stakes era: Achievement, resources, and No Child Left Behind. *Psychological Science*, 16(6), 419-425.

Turnbull, H. R., Stowe, M. J., & Huerta, N. E. (2007). *Free appropriate public education: The law and children with disabilities* (4th ed.). Denver, CO: Love Publishing Company.

U.S. Department of Education, Office for Civil Rights, *Free Appropriate Public Education for Students With Disabilities: Requirements Under Section 504 of the Rehabilitation Act of 1973,* Washington, D.C., 2010.

Vogler, K. E., & Virtue, D. (2007). "Just the facts, ma'am": Teaching social studies in the era of standards and high-stakes testing. *Social Studies,* 98(2), 54-58.

Wasserman, J. D., & Tulsky, D. S. (2005). A history of intelligence assessment. In D.P. Flanagan & P.L. Harrison's (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (2nd ed.) (pp. 3-22). New York: The Guilford Press.

Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist,* 5(3), 78-83.

Wechsler, D. (1975). Intelligence defined and undefined: A relativistic appraisal. *American Psychologist,* 30(2), 135-139.

Wechsler, D. (1989). *Manual for the Wechsler Preschool and Primary Scales of Intelligence-Revised.* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler Intelligence Scale for Children-Third Edition (WISC-III).* San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1992). *Manual for the Wechsler Individual Achievement Test.* San Antonia, TX: The Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children-Fourth Edition (WISC-IV) administration and scoring manual.* San Antonia, TX: The Psychological Corporation.

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

Woodcock, R. W., & Johnson, M. B. (1989). Woodcock-Johnson Psycho-Educational Battery-Revised. Allen, TX: DLM Teaching Resources.

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

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001b). *Woodcock-Johnson III Tests of Achievement.* Itasca, IL: Riverside Publishing.

Yell, M. L., Shriner, J. G., & Katsiyannis, A. (2006). Individuals with disabilities education improvement act of 2004 and IDEA regulations of 2006: Implications for educators, administrators, and teacher trainers. *Focus on Exceptional Children,* 39(1), 1-24.

Yerkes, R. M. (Ed.) (1921) Psychological examining in the United States Army. *Memoirs of the National Academy of Sciences, 15,* 1-890.

Ysseldyke, J., Nelson, J., Christenson, S., Johnson, D. R., Dennison, A., Triezenberg, H., & Hawes, M. (2004). What we know and need to know about the consequences of high-stakes testing for students with disabilities. *Exceptional Children*, 71(1), 75.

Zimmerman, B. J., & Dibenedetto, M. K. (2008). Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing. *Psychology in the Schools*, 45(3), 206-216.