The Dissertation Committee for Jacqueline Marie Caemmerer Certifies that this is the approved version of the following dissertation:

Beyond Individual Tests: The Effects of Children's and Adolescents' Cognitive Abilities on their Achievement

Committee:

Timothy Z. Keith, Supervisor

Cindy I. Carlson

Stephanie W. Cawthon

Matthew R. Reynolds

Beyond Individual Tests: The Effects of Children's and Adolescents' Cognitive Abilities on their Achievement

by

Jacqueline Marie Caemmerer

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August 2017

Acknowledgements

I am very grateful for the training and mentorship I received throughout my time at UT. I have grown a great deal professionally during the past six years due to the support I have received from my committee members and others within the Educational Psychology Department. Dr. Keith, I am grateful for the countless learning opportunities, excellent research and teaching mentorship, encouragement and support you provided me with these past six years; your mentorship will serve as a guide in my future role as a professor. I am thankful for the mentorship Dr. Cawthon provided me throughout my six years as a member of her collaborative and productive research team. I also appreciate Dr. Carlson's professional guidance and support, and encouragement to explore and develop other areas of interest. Additionally, I sincerely appreciate the support and wise words Dr. Reynolds shared with me throughout my graduate school journey. And I am also grateful to Pearson for providing the data that made this study possible.

I also would not have made it through all these years of schooling without an immense amount of social support. My husband, Josh, provided unwavering support and encouragement throughout all the hurdles, small and big; I could not imagine going through this process without you by my side. I am also grateful for the love and support I received from my mother and brother, and the excellent models of hard work both my parents provided.

Beyond Individual Tests: The Effects of Children's and Adolescents' Cognitive Abilities on their Achievement

Jacqueline Marie Caemmerer, Ph.D.
The University of Texas at Austin, 2017

Supervisor: Timothy Z. Keith

Students' performance across several tests, including both cognitive and achievement tests, is often analyzed together to better understand their learning. This analysis is guided by the assumption that there are specific relations between students' cognitive abilities and their reading, writing, and math skills. The research supporting this assumption is limited because cognitive-achievement research findings are mostly based on a single test, the Woodcock-Johnson tests (McGrew & Wendling, 2010), and previous studies involve analyzing a single intelligence and achievement test in isolation. Thus, findings are limited to the specific tests that are included in those analyses, and are not necessarily generalizable across other tests. Research that incorporates multiple intelligence and achievement tests, cross-battery analyses, can better address questions about the broader influences of children's cognitive abilities on their achievement. Such cross-battery research can extend psychologists' understanding of how intelligence and achievement relate beyond the test-level to the construct level.

Six intelligence tests (KABC-II, WJ III, WISC-III, WISC-IV, WISC-V, and DAS-II) and three achievement tests (KTEA-II, WIAT-II, WIAT-III) were analyzed in a cross-

battery cognitive-achievement analysis in the current study. Data were derived from seven of the tests' standardization or linking samples; participants were 3,930 children and adolescents aged 6 to 16.

In order to simultaneously analyze several tests a planned missingness approach and structural equation modeling were used. Six broad abilities (Gc, Gf, Gv, Gsm, Gs, and Glr) and *g* were modeled as latent variables; each broad ability latent variable was indicated by 7 – 14 subtests. Results suggest Gf and *g* were perfectly correlated and it was impossible to separate the two abilities statistically. The cognitive abilities were predictors of three achievement skills (basic reading, broad writing, and broad math), which were indicated by four to six subtests. Findings indicated Gc influenced all three academic skills; Gsm and Glr influenced basic reading and broad writing; Gs influenced broad writing and broad math; Gf exerted a significant effect on broad math; and Gv was not significantly related to any academic skill. Significant cognitive-achievement relations have implications for diagnostic decision-making regarding specific learning disabilities, assessment planning, and educational recommendations.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

When students struggle to perform adequately in their classes, school psychologists are often called on to assess their functioning. Assessment results are crucial in determining the possible causes of achievement difficulties and in developing appropriate learning environments for students, including whether they meet criteria for special education or section 504 services and related accommodations and interventions. The components of these evaluations vary according to the reason for the assessment referral, but if a specific learning disability is suspected or must be ruled-out, the evaluation typically includes the administration of standardized cognitive and achievement measures along with other measures. Thus, improving school psychologists understanding of the relationships between cognitive and achievement measures is necessary in order to inform evidence-based diagnostic decisions and educational recommendations, with the goal of enhancing students' academic achievement.

Cognitive Assessment Considerations

Interpretation of students' cognitive assessment results is often guided by Cattell-Horn-Carroll (CHC) theory. CHC theory is not only the leading intelligence theory within the field of school psychology (Keith & Reynolds, 2010), but CHC theory is also relevant more broadly to the field of clinical diagnostic assessment because CHC theory provides a common taxonomy for cognitive abilities. The CHC theory taxonomy improves the consistency of the interpretation of assessment results across different types of test batteries, including neuropsychological tests, and across different practitioners (Jewsbury, Bowden, & Duff, 2016). CHC theory posits a three-stratum model of intelligence. A general intelligence factor, *g*, is at the apex of the model, the third stratum; *g* involves reasoning, problem solving, and learning (Colom, Karama, Jung, & Haier, 2010). Moving to the second stratum, general intelligence (*g*) subsumes 8 to 10 broad abilities.

Moving to the first stratum, the broad abilities subsume many narrow abilities—the intelligence subtests themselves. The broad abilities represent abilities such as verbal comprehension/knowledge (Gc), fluid/novel reasoning (Gf), visual-spatial processing (Gv), short-term memory (Gsm), and processing speed (Gs) (Schneider & McGrew, 2012; see Table 1 on page 17 for a definition of each broad ability and the literature review for an in-depth discussion of the development of intelligence theory). Modern intelligence tests measure a variety of these broad abilities, but each ability is not included in every battery. Although different tests purport to measure the same broad ability constructs, the subtests within each test battery vary according to task demands, stimuli, and response format. Due to these subtest specific differences, some school psychologists question whether or not these different tests are actually measuring the same abilities and if results across the tests are comparable (Reynolds, Keith, Flanagan, & Alfonso, 2013). This question raises concerns regarding whether or not estimates of children's abilities vary depending on which test was administered.

In an attempt to answer this question, previous research has tested whether different intelligence tests measure the same constructs via cross-battery confirmatory factor analysis (CB-CFA) (Reynolds et al., 2013). CB-CFA is a useful technique to test theory and establish factorial invariance across tests. The Reynolds study was among the largest CB-CFA analyses to date, simultaneously analyzing four recent and commonly used intelligence tests to determine whether the CHC broad abilities were invariant across different populations and tests (Reynolds et al., 2013). Their sample included children and adolescents ages 6 to 16 from the Kaufman Assessment Battery for Children, Second Edition (KABC-II) concurrent validity studies. Indeed, the CHC broad abilities were found to be invariant, providing evidence that CHC theory is applicable across different tests and the tests measure the same CHC constructs similarly. These findings suggest that school psychologists can administer any of those four tests, or combinations of those tests, and be confident that they are measuring the same abilities in the students they assess regardless of the test(s) selected (Reynolds et al., 2013).

Several other studies provide support for the findings from the Reynolds study and have also shown that the CHC broad abilities represent the same constructs across batteries. These studies, however, were limited to simultaneously analyzing only two intelligence tests, and as a result, the findings were limited to a smaller set of tests (Flanagan & McGrew, 1998; Keith et al., 2001; Keith & Novack, 1987; Phelps, McGrew, Knopik, & Ford, 2005; Roid, 2003; Sanders, McIntosh, Dunham, Rothlisberg, & Finch, 2007; Stone, 1992). One of the earliest CB-CFAs was the largest, and included six intelligence tests, but all of the tests have since been revised (Woodcock, 1990). Thus, CB-CFA analyses that include more than two intelligence tests and the most recent editions of those tests are needed. This will establish that CHC theory explains the relations among cognitive abilities well, using the tests psychologists are likely to administer.

The usefulness of CB-CFA research is not only supported by theoretical rationale, with the purpose of extending CHC theory across batteries, but for practical clinical reasons as well. School psychologists often apply CHC theory to their practice through the lens of the cross-battery assessment approach (Flanagan, Ortiz, & Alfonso, 2013). This approach encourages practitioners to utilize more than one intelligence test when assessing children in order to fully assess the abilities underlying intelligence; a single intelligence test often does not measure all of the possible broad abilities. Practitioners are also encouraged to administer more than one test if the child's scores are discrepant within one or more CHC broad ability composites scores as a means of further investigation. Thus, the theory underlying the cross-battery approach assumes the different intelligence tests are measuring the underlying broad ability constructs equivalently and recommends combining multiple intelligence tests to form one comprehensive evaluation of student's intelligence.

Cognitive-Achievement Assessment and Relations

The cross-battery assessment approach, however, is not limited to cognitive measures. The cross-battery assessment approach also suggests that a more comprehensive picture of students' cognitive abilities will better inform how practitioners relate students' intelligence to their achievement (Flanagan, Ortiz, & Alfonso, 2013). The relations between students' intelligence and achievement scores are often used to inform specific learning disability diagnostic decision making. Interpreting multiple cognitive and achievement tests together simultaneously as part of a cross-battery assessment assumes the relations between different intelligence and achievement tests are equivalent across batteries. School psychologists using this approach assume the CHC broad abilities underlying different intelligence tests are similarly related to reading, mathematics, and writing achievement across batteries; this assumption currently remains untested. Cross-battery research can be used to test these assumptions underlying the cross-battery assessment approach. In order to increase the practical clinical implications of cross-battery research for school psychologists, CB-CFA intelligence test results should also be used to predict standardized academic achievement. Such cross-battery intelligence-achievement research may bolster evidence-based decision making regarding specific learning disabilities diagnoses. Therefore, cross-battery analyses (in this case cross-battery structural equation modeling, CB-SEM) are not only useful for furthering knowledge of the structure of intelligence, but also for clarifying the relations between students' intelligence and their achievement across batteries.

Although explaining students' achievement is the typical use of intelligence tests in schools, there is little cross-battery research examining the relations between CHC abilities (general and specific intelligences) and their effects on achievement. Instead, most of the research using students' performance on intelligence tests to predict performance on achievement tests has involved analyzing a single cognitive test in isolation and then using that test to predict a single standardized achievement test (*e.g.*, the Woodcock Johnson Tests of Cognitive Abilities III (WJ III) may be used to predict students' performance on the Woodcock Johnson Tests of Achievement III). Another limiting factor of previous research in this area is that the majority of studies are based on the Woodcock-Johnson tests (McGrew & Wendling, 2010). Thus, the assumption that the relations between students' intelligence and achievement are stable across different test batteries needs to be tested further. Additionally, analyzing several different tests may broaden school psychologists' understanding of the more *general* effects of intelligence on achievement. Overall, research findings in this area are narrowly focused, and it is questionable whether these relations are generalizable to other tests that were not analyzed. Cross-battery research can address this limitation and more broadly, explain the relations between students' intelligence and achievement performance across multiple tests.

Despite these limitations, previous research provides important insights about the relations between intelligence and achievement. It is well-established that general intelligence, *g*, and standardized general achievement are highly correlated; estimates vary, but tend to be within the .80 range (Deary, Strand, Smith, & Fernandes, 2007; Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012). Research guided by CHC theory suggests that specific broad abilities are important in understanding students' achievement as well. These findings suggest the CHC broad abilities differentially explain variation in students' reading, mathematics, and writing performance; the importance of each ability varies according to the academic area under study (Gustafsson & Balke, 1993; McGrew & Wendling, 2010).

For example, fluid/novel reasoning (Gf) tends to explain more variance in mathematics performance in comparison to reading performance (McGrew & Wendling, 2010; Parkin & Beaujean, 2012). Although fluid reasoning is generally the strongest influence on math performance, other abilities also have significant effects including verbal-comprehension (Gc), short-term memory (Gsm), and processing speed (Gs) (Floyd, Evans, & McGrew, 2003; Fuchs et al., 2010?; Keith, 1999; McGrew, Keith, Flanagan, & Vanderwood, 1997; McGrew & Hessler, 1995; Niileksela, Reynolds, Keith, & McGrew, 2016; Taub, Floyd, Keith, & McGrew, 2008). Two studies have suggested that visual processing (Gv) has significant effects on students' math reasoning abilities using the WJ tests (McGrew & Hessler, 1995; Niileksela et al., 2016). The relations between the broad abilities and achievement vary as a function of age as well. For instance, long-term retrieval (Glr) may be important in explaining young children's math performance, but its effect likely decreases as children develop (Floyd et al., 2003).

In terms of writing performance, there is less research investigating its relations with cognitive abilities. The little available research suggests significant effects for verbal-comprehension (Gc), short-term memory (Gsm), and processing speed (Gs) (Beaujean et al., 2014; Floyd, McGrew, Evans, 2008; McGrew & Knopik, 1991; Niileksela et al., 2016). The effects of fluid reasoning (Gf) were inconsistent (Beaujean et al., 2014; Floyd et al., 2008; McGrew & Knopik, 1991), and dependent on the students' age or the particular writing test and its specific demands. Among younger students, long-term retrieval appeared significant (Floyd et al., 2008) and Gv was significantly related to written expression in a single study (Niileksela et al., 2016).

In contrast, reading is the most studied cognitive-achievement domain. For reading achievement, verbal-comprehension (Gc) has generally shown the largest effect (McGrew & Wendling, 2010). Auditory processing (Ga), processing speed (Gs), and short-term memory

(Gsm) also consistently have significant effects on reading (Beaujean, Parkin, Parker, 2014; Benson, 2008; Elliot, Hale, Fiorello, Dorvil, & Moldovan, 2010; Floyd, Meisinger, Greg, & Keith, 2012; Floyd, Keith, Taub, & McGrew, 2007; Havojsky, Reynolds, Floyd, Turek, & Keith, 2014; McGrew et al., 1997; Niileksela et al., 2016; Vanderwood, McGrew, Flanagan, & Keith, 2002). At a more specific level, in terms of the components of reading, long-term retrieval (Glr) appears important for basic reading (decoding and word recognition skills) (Floyd et al., 2007; Hajovsky et al., 2014), whereas fluid reasoning (Gf) may be important for reading comprehension (Floyd et al., 2012; McGrew, 1993; Niileksela et al., 2016). Most studies suggest that the influence of Gv is negligible, but one study provides contradictory evidence suggesting it may be important for reading comprehension (Hajovksy et al., 2014).

In sum, CHC theory fits well with modern, frequently used intelligence tests, regardless of whether the tests were explicitly developed according to CHC theory (Reynolds et al., 2013). The predictive validity of the CHC broad abilities in explaining students' standardized achievement is well-supported; the broad abilities differentially explain students' reading, mathematics, and writing achievement (McGrew & Wendling, 2010). Some broad abilities, particularly Gc, Gsm, and Gs, exert significant effects across academic domains, while others are particularly salient at certain ages (*e.g.* Glr) or for a narrower range of achievement skills (*e.g.* Gf) (McGrew & Wendling, 2010; Niileksela et al., 2016). It is unclear whether one specific broad ability, Gv, is unimportant in understanding students' achievement or whether Gv is influential for specific skills (Hajovsky et al., 2014; McGrew & Wendling, 2010; Niileksela et al., 2016).

All of the above described studies were limited to the analysis of a single cognitive and achievement test. Reading is the only achievement domain that has been studied using cross-battery research. One study simultaneously analyzed two intelligence tests, the Wechsler

Intelligence Scale for Children, Revised (WISC-R) and Woodcock Johnson-Revised (WJ-R) Cognitive, as predictors of three WJ-R Achievement reading subtests. The sample was limited, however, and included 166 third and fourth grade, mostly Caucasian Texas students (Flanagan, 2000). The results of this CB-SEM study were consistent with previous non-cross-battery studies. Among these elementary school students, Gc was the strongest significant predictor of reading, followed by Ga and Gs (Flanagan, 2000).

Methodological Considerations

One possible reason for the lack of cross-battery cognitive-achievement relations research is the potential time and financial demands of data collection (Enders, 2010). Requiring students to complete multiple intelligence and achievement tests may cause extensive examinee fatigue and be costly for researchers. For this reason, methodology is a critical consideration in CB-CFA research. Fortunately, planned missing data methodology is particularly useful for these purposes (Enders, 2010; Graham, Taylor, Olchowski, & Cumsille, 2006; McArdle, 1994). Planned missing data designs limit examinee fatigue by removing the requirement that all examinees complete each test that will be analyzed. Instead, all examinees complete one test, referred to as the linking test. Then, a subset of tests is given to each examinee. This particular type of missing data design is referred to as a three form design procedure (Enders, 2010). Other designs are possible, however, and may include samples where some participants do not complete a common linking test. These alternative designs allow for the inclusion of broad ability constructs that may not be available in each test, and testing such designs may improve future data collection methods (Reynolds et al., 2013). More research is needed to investigate the feasibility of alternative planned missingness designs.

Purpose of this Study

As previously discussed, the current understanding of cognitive-achievement relations is limited to research analyzing a single intelligence and single achievement test, and the majority of studies are based on the WJ tests. Because much of the research in this area focuses on one specific test or battery, findings are limited to those specific tests and are less generalizable to students' cognitive and achievement abilities more broadly. This is problematic because psychologists assess students using a variety of tests. Psychologists cannot assume the relations between students' intelligence and achievement are reproducible across different tests without empirical evidence validating this assumption. If this assumption is proven false, psychologists need to account for differences across tests when interpreting students' test results.

The current study will address the limitations of previous research and test the assumption that cognitive-achievement relationships are generalizable across tests in two ways. The first purpose is to incorporate additional intelligence tests into a CB-CFA model, which will broaden the scope of cross-battery intelligence research. Secondly, this more comprehensive cross-battery intelligence factor structure will be used to predict students' standardized achievement performance, which will more broadly improve cognitive-achievement relations evidence. Predicting students' achievement by CHC broad abilities that are also representative of several tests will improve school psychologists' understanding of these relations at a construct, as opposed to test-specific, level. Because these results will be generalizable across several test batteries, the cross-battery cognitive-achievement findings may inform school psychologists' recommendations for supporting students' academic achievement and evidence-based diagnostic decision making regarding specific learning disabilities, regardless of the tests they select. The cross-battery cognitive-achievement findings may also influence psychologists' assessment planning and selection, particularly for psychologists who use a cross-battery assessment approach.

Chapter 2: Literature Review

This literature review is organized into two broad sections. The first section is focused on intelligence and achievement theory and research. The second section is focused on planned missing data methodology. Within the first section, the development of intelligence theory, intelligence tests, and their applications are discussed. Then, previous cross-battery confirmatory factor analysis intelligence research is reviewed. The cross-battery assessment approach is described and previous cognitive-intelligence relations research is summarized. The literature review concludes with a discussion of the importance of planned missing data methodology and related issues.

Intelligence and Achievement Theory and Research

Overarching summary. When students struggle to perform adequately in their classes, school psychologists are often called on to assess their functioning. Students' cognitive abilities and achievement are assessed, at a minimum, to better understand their strengths and weaknesses. Such assessment also aids in the determination of what area of their functioning is interfering with their learning. These assessment results often inform school psychologists' decisions about special education eligibility, need for services and accommodations, and whether a student meets criteria for a specific learning disability.

Several tests are available to school psychologists to conduct these evaluations. However, not every intelligence test measures the same CHC broad abilities. The tasks that are part of both intelligence and achievement tests vary across tests and involve different stimuli and response formats. Whether or not tests measure the same abilities and whether the relations between cognitive abilities and achievement are reproducible across batteries are both questions with theoretical and clinical implications for school psychologists. Most of the research that addresses these questions are limited to the specific tests that are included in those analyses. Research that incorporates multiple intelligence and achievement tests, CB-CFA analyses, can better address these questions. Such cross-battery research will extend psychologists' understanding of how intelligence and achievement relate not just at the test-level, but at the broader construct level. The results of cross-battery cognitive-achievement research may have implications for educational recommendations and diagnostic decision-making regarding specific learning disabilities.

Applications of intelligence testing. The purpose of the first cognitive test developed, the Binet-Simon Intelligence Test, was to assist in the identification of students who required special education services in schools (Binet & Simon, 1905). Although the theory, content, and interpretation of intelligence tests has continued to develop over the past century, this original purpose for cognitive testing remains a key reason for assessment in schools. Intelligence tests are often used in schools to diagnose specific learning disabilities or identify students who are gifted or intellectually disabled. Assessment results in general provide information about students' strengths and weakness, and inform academic placements, accommodations, and interventions. The value of assessment, including intelligence tests, is far-reaching. At a broad level, assessment supports the ease of description of individuals and their skills; enhances communication among professionals; aids research and clinical practice by establishing a common terminology; and facilitates program evaluation, development of policy, and advocacy efforts. At a more specific level, assessment results can be used for determining eligibility for special education services, informing diagnostic decision making, identifying a need for services, and informing a treatment plan (Dowdy, Mays, Kamphaus, & Reynolds, 2009). Classifying and diagnosing students represents a core expectation of a school psychologist's role by schools, parents, and the community. Thus, assessment has been and continues to be central to the field of school psychology.

Assessment in general, and the use of intelligence tests in particular, has spread beyond schools and into the military, hospitals, clinics, and other settings. Intelligence testing is an essential piece of clinical diagnostic assessment, and is useful when evaluating individuals with different disorders or brain injuries (Jewsbury, Bowden, & Duff, 2016). Individuals' intelligence have been linked to many important life outcomes beyond academic achievement, including, but not limited to years of education completed, occupational performance, income, and even health behaviors (Gottfredson & Deary, 2004; Neisser, Boodoo, Bouchard, Boykin, Ceci, & Loehlin, 1996). The importance of intelligence is clear and intelligence testing continues to play an important role in the field of psychology.

The development of intelligence theory. "Intelligence can be defined as a general mental ability for reasoning, problem solving, and learning;" overall intelligence is broadly defined and integrates other more specific cognitive functions (Colom et al., 2010, p. 489). The conceptualization and theory of intelligence has progressed greatly over more than a century and refinement of the theory continues to this day. Charles Spearman is credited with being the first to develop a coherent intelligence theory in 1904. He noticed that cognitive tests correlated highly with one another and hypothesized that these strong relations were caused by an underlying common intelligence ability, referred to as *g* (Kamphaus, 2009; Schneider & McGrew, 2012). Additionally, Spearman is credited with developing factor analysis. Factor analysis uses the correlations among items to explain the common underlying constructs or factors (latent, or unobserved, variables). Thus, factor analysis is a useful technique for establishing internal validity of tests, as well as convergent and discriminant validity when applied across tests (Keith, 2015, chapter 15). Spearman's factor analyses led to his two-factor theory of intelligence. According to his theory, the variance of intelligence tests is explained by two parts: variance shared by all tests (*g*) and variance specific to each particular test (Kamphaus, 2009). Spearman's two-factor theory is the foundation for modern, more comprehensive theories of intelligence.

Other theorists emphasized the importance of multiple abilities. L.L. Thurstone proposed multiple cognitive factors that were independent of the *g* factor (1938). Thurstone's multiple-factor method laid the foundation for Cattell's Gf-Gc theory (Schneider & McGrew, 2012). The evolution of this theory began when Raymond Cattell demonstrated that general intelligence, *g*, was better represented by two factors instead of one, referred to as Gf and Gc. Cattell's doctoral student, John Horn, further expanded this theory to incorporate multiple broad abilities. This extended Horn-Cattell Gf-Gc theory, however, excluded a general intelligence, *g*, factor (Schneider & McGrew, 2012). Gf and Gc continue to be important abilities within modern intelligence theory.

These conflicting theories were synthesized as the result of a huge factor analytic investigation. John Carroll's seminal work, *Human Cognitive Abilities: A Survey of Factor Analytic Studies* (1993), presented the results from his reanalysis of more than 460 experimental and clinical datasets. Importantly, his analyses encompassed a reanalysis of many of the key intelligence factor analyses since Spearman's work, with a focus on large batteries, cross-battery data sets, and seminal studies. As a result, Carroll proposed the three-stratum theory of intelligence (Schneider & McGrew, 2012). Carroll unified previous theories by incorporating *g* and specific cognitive abilities into one overall, higher-order, structure of intelligence (Kamphaus, 2009). The third stratum is the most general and represents *g*, the second stratum includes eight specific broad abilities, and the first stratum includes more narrow abilities; each stratum is subsumed by the higher stratum preceding it (Schneider & McGrew, 2012). The stratums start with the most general overall intelligence ability, become increasingly more specific at the broad ability level, and then the most specific at the narrow ability level (each broad ability is measured by more than one narrow ability). The three-stratum theory of intelligence is the best supported structure of intelligence today (Keith & Reynolds, 2010).

Because Carroll's three-stratum theory and Horn-Cattell's Gf-Gc theory share many commonalities, the synthesis of the two theories is frequently referred to as Cattell-Horn-Carroll (CHC) theory (McGrew, 1998). A review of 20 years of recent factor analytic intelligence research demonstrated that CHC theory is currently the most supported intelligence theory, and that tests derived from other theories conform well to a CHC orientation (Keith & Reynolds, 2010). Not only is CHC theory applicable to intelligence tests, but the CHC taxonomy also fits other types of cognitive processes well, often referred to as executive functions within a neuropsychological framework (Floyd, Bergeron, Hamiliton, & Parra, 2010; Jewsbury et al., 2016; Salthouse, 2005). Definitions of seven CHC broad abilities that are relevant to this study are presented in Table 1 and are based on definitions presented in Schneider and McGrew (2012).

As intelligence theory developed over the past century, so too did intelligence tests. Early versions of intelligence batteries were atheoretical or were only loosely based on some sort of theory. The development of current intelligence tests, however, is increasingly guided by theory. The Woodcock-Johnson Revised Test (WJ-R; Woodcock & Johnson, 1989) was among the first to bridge the gap between intelligence theory and practice by applying Horn-Cattell's Gf-Gc theory to test development (Schneider & McGrew, 2012). The WJ-R was unique in this way, and it measured six broad cognitive abilities. Following the lead of the WJ-R, other tests began measuring other broad abilities. Prior to 2000, the majority of intelligence tests, however, measured only two to three broad CHC abilities. As such, several important broad abilities were inadequately measured or neglected altogether, including Gf, Gsm, Glr, Ga, and Gs (Flanagan, Alfonso, & Ortiz, 2013).

Today, this problem is less of a concern because intelligence tests are generally designed to measure multiple broad abilities. Recent revisions of tests generally measure four to five CHC broad abilities (Flanagan et al., 2013). Although CHC theory is regarded as the best supported theory, not all tests were explicitly developed with this theory as the guiding framework. Regardless of whether intelligence tests were explicitly based on CHC theory, however, CHC theory explains the structure of these tests well. Factor analyses of popular intelligence tests, including the Woodcock-Johnson tests, Differential Abilities Scales, Kaufman Scales, and Wechsler scales, indicate that these tests are consistent with CHC theory (Keith & Reynolds, 2010).

Table 1

Descriptions of CHC Broad Abilities

Broad Ability	Definition
Gc	The breadth and depth of acquired cultural knowledge, including language and information learned inside and outside of school. Gc is influenced by "experience, education, and cultural opportunities" and is often referred to as crystallized knowledge (p. 122). Narrow abilities that Gc subsumes include general verbal information, language development, and lexical knowledge.
Gf	Problem solving using unfamiliar information or novel procedures that cannot be performed automatically. Gf involves abstract reasoning, including inferential reasoning and concept formation, which relies less on prior learning. Narrow abilities subsumed by Gf include induction, general sequential reasoning, and quantitative reasoning.
Gv	"The ability to make use of simulated mental imagery (often in conjunction with currently perceived images) to solve problems" (p. 129). Gv involves the mental rotation of images, identification of patterns, or transformation of visual information. Narrow abilities subsumed by Gv include visualization and speeded rotation.
Gsm	The ability to "encode, maintain, and manipulate information in one's immediate awareness" (p. 114). Gsm involves primary memory capacity and efficiency of attentional control in primary memory. Narrow abilities subsumed by Gsm include memory span and working memory capacity.
Gs	The "ability to perform simple, repetitive tasks quickly and fluently" (p. 119). Gs is less important than Gf and Gc in "predicting performance during the learning phase of skill acquisition," but Gs predicts "skilled performance once people know how to do a task." Narrow abilities subsumed by Gs include perceptual speed, reading speed, writing speed, and number facility (also referred to as basic arithmetic speed).
Glr	The "ability to store, consolidate, and retrieve information over periods of time measured in minutes, hours, days, and years" (p. 116). Glr involves the processes of memory. Narrow abilities subsumed by Glr include associate memory, ideational fluency, and naming facility (also referred to as rapid automatic naming in the reading research).
Ga	The ability to detect and process meaningful information in sounds. "Ga is what the brain does with sensory information from the ear" (p. 131). Narrow abilities subsumed by Ga include phonetic coding and speech sound discrimination.

Note. Definitions are adapted from Schneider and McGrew (2012).

Previous cognitive CB-CFA research. The theory, content, and interpretation of intelligence tests varies according to the test. Modern intelligence tests measure a variety of broad abilities and each ability is not included in every battery. Although different tests purport to measure the same broad ability constructs, the subtests within each test battery vary according to task demands, stimuli, and response format. Due to these subtest specific differences, some psychologists question whether or not these different tests are actually measuring the same abilities and if results across the tests are comparable (Reynolds et al., 2013). This question raises concerns regarding whether or not estimates of children's abilities vary depending on which test was administered.

In an attempt to better understand the structure of intelligence tests across batteries, researchers analyze multiple intelligence tests simultaneously. This type of research expands the application of factor analysis from analyzing single intelligence tests to joint analyses of multiple tests, referred to as cross-battery factor analyses (CB-FA). Factor analyzing more than one intelligence test conjointly, particularly when the tests were designed according to different theories, allows researchers to test which theory is best supported. When confirmatory factor analysis (as opposed to exploratory factor analysis) is used in a CB-FA (referred to as CB-CFA), researchers can test and compare models drawn from those different theories (Keith & Reynolds, 2010). In addition to answering questions about the underlying theory of intelligence, CB-CFA analyses can answer questions about the nature of the broad abilities at the construct level. Most tests do not include more than two measures of any one broad ability, but CB-CFA allows for the analysis of several tests for each broad ability. Therefore, simultaneously analyzing multiple tests is advantageous as each broad ability is measured by several indicators (Keith & Reynolds, 2010). Because the content, stimuli, and response format vary across tests, more generalizable conclusions about broad abilities are possible as a result.

In 1990, Richard Woodcock conducted the largest CB-CFA analysis to date that included seven intelligence tests (WJ, WJ-R, WISC-R, WAIS, WAIS-R, K-ABC, and Standford-Binet-IV). Participants were drawn from the WJ and WJ-R concurrent validity studies, and included third graders, fifth graders, and twelfth graders. As previously discussed, many broad abilities were neglected by these pre-2000 era intelligence tests. His synthesis of several tests provided quantitative evidence supporting Cattell-Horn's extended Gf-Gc theory and its application across tests, even though some of the tests were not developed according to this theory. As a result, Woodcock argued for the importance of cross-battery assessment among practitioners; practitioners could supplement one test by administering a second test in order to more completely measure several broad abilities (Woodcock, 1990).

Several other CB-CFA analyses were conducted following Woodcock's analysis based on either Cattell-Horn Gf-Gc theory or CHC theory (Flanagan & McGrew, 1998; Keith et al., 2001; Keith & Novak, 1987; Phelps et al., 2005; Sanders et al., 2007; Stone, 1992). Each of these were limited to only two jointly-analyzed tests due to the time and financial burdens of assessing children with several tests. The sample sizes of five out of six of these CB-CFA analyses were small, ranging from 114 to 155 students (one was an outlier and included 544 students; Keith & Novak, 1987); most tested a somewhat narrow age range, including third through sixth graders and sixth through eighth graders (Flanagan & McGrew, 1998; Keith et al., 2001; Phelps et al., 2005; Sanders et al., 2007). Furthermore, most of the tests have since been revised. Regardless of these limitations, each of these analyses also provided support for the Gf-Gc theory or CHC theory, even though most of the tests, except for the WJ, were not explicitly developed using these theories.

The most recent CB-CFA analysis is the second largest, and doubled the number of tests factor analyzed to four—KABC-II, WJ-IV, WISC-III, WISC-IV (Reynolds et al., 2013). This larger CB-CFA analysis was possible because of its design, using a planned missing data methodology. The planned missing design capitalized on the advantages of CB-CFA analyses, allowing for the analysis of multiple tests. The cost of administering several tests to a large sample of students was low because not every child was required to complete every test (Keith & Reynolds, 2010). This study analyzed a larger sample, a total of 423 students, and included a larger age range, from 6 to 16 years.

The results of the Reynolds analysis are worth discussing in more depth because my analysis will incorporate the same KABC-II convergent validity sample (in addition to six samples taken from additional datasets). Five broad abilities were measured well by the four tests in their analyses. These five broad abilities were Gc, Gf, Gv, Gsm, and Associative Memory, which is a narrow ability of Glr that is specific to remembering unrelated paired information. In order to better understand how well each broad ability is measured by its corresponding subtests, the factor loadings should be examined. The factor loadings indicate the strength of the relation of the subtest to the broad ability. If CHC theory maps well onto the tests, high factor loadings would be expected from the subtests to the broad abilities they are purported to measure (Keith, 2015, chapter 15). In Reynolds and colleagues' analyses the factor loadings of the 12 subtests onto Gc were generally the strongest across the broad abilities (generally ranging from .62 to .87), suggesting Gc was the best measured broad ability among these tests. The magnitude of the factor loadings for the other broad abilities encompassed a similar range. More specifically, the factor loadings of the seven tests that represented Gv ranged from .56 to .74 (one outlier, Gestalt Closure, was .20); seven factor loadings for Gf ranged from .48 to .74; eight Gsm factor loadings ranged from .45 to .76; and six Associative Memory loadings ranged from .53 to .79. The correlations between the broad abilities were strong and ranged from .57 (Gv and Gsm) to .82 (Gf and Gv), meaning they are highly related. In terms of the higher order structure, Gf generally had the strongest relations with *g*, but Gc had the strongest relations when differences due to sex and SES were not controlled. The standardized loadings onto *g* were .98 for Gf, .82 for associative memory, .76 for Gv, .75 for Gc and Gsm. In addition to these strong factor loadings and regression coefficients, the initial model fit the data well, and required minimal modifications, suggesting that the CHC taxonomy explained the broad abilities well. The authors concluded that regardless of what theory was used to design the tests, CHC theory was well supported across these four tests (Reynolds et al., 2013).

Several questions were left unanswered by this study. The analysis was based on the three-form missing data design, meaning that each examinee completed a common linking test (to be described in more detail later), the KABC-II. This design limited the broad ability analyses to only those that were measured by each test. The combination of these four tests prohibited the analysis of Gs, an important broad ability in explaining academic achievement. For this reason, Reynolds and colleagues highlighted the need for further CB-CFA analyses using other designs. Although not part of their final model, preliminary analyses tentatively supported the analysis of broad abilities that were not measured by the reference test (Reynolds et al., 2013). This finding hints at the possibility of CB-CFAs analyses that do not include one common linking test. Such an alternative CB-CFA design will be explored in the current study, and discussed in more detail in a later section.

Cross-battery assessment approach. There are many advantages to both research and clinical findings based on more than one intelligence test. A modern practitioner-oriented approach supports the use of more than one intelligence test when assessing a child. This theory, known as the cross-battery assessment approach, is grounded in CHC theory and allows practitioners to assess a wider range of abilities than is possible when practitioners are constrained to using a single test (Flanagan, Ortiz, & Alfonso, 2013). Based on experts' classifications of tests and CB-CFA analyses (Reynolds et al., 2013), Flanagan and colleagues have classified the subtests of popular intelligence tests according to the CHC broad abilities that they measure (Flanagan et al., 2013). These classifications allow practitioners to comprehensively assess a variety of students' abilities. This means, for example, that a psychologist may primarily evaluate a child using the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V). However, because the WISC-V does not include subtests that measure Ga, the psychologist can supplement the testing results from the WISC-V with the subtests that measure Ga found in the WJ-IV. In addition, if a child's scores are discrepant within a CHC broad ability composite score, the cross-battery assessment approach encourages the practitioner to administer additional measures of that ability, which may be assessed by additional intelligence tests.

Cross-battery assessment extends beyond guiding the assessment of students' intellectual abilities. A key pillar of the cross-battery approach is examining the relations between cognitive abilities and academic skills. This pillar bridges theory and practice (Flanagan et al., 2013). Understanding the relations between intelligence and achievement is important because school psychologists are often trying to understand the reasons for students' learning difficulties. For instance, a common situation is a student referred to a school psychologist due to low math performance. If the student's performance within a particular cognitive ability that is associated with math is low, along with low performance on standardized math tests, the testing results may justify a possible specific learning disability in math. Using the cross-battery assessment approach, a school psychologist may assess a student's cognitive abilities using more than one test and then make inferences between the results from multiple intelligence tests and their relations to standardized achievement test results. This association between intelligence and achievement tests is based on an untested assumption, however. The cross-battery assessment approach assumes the relations between different cognitive and achievement tests are stable across batteries. Research testing this assumption is needed, and extending cross-battery cognitive-achievement research into this area can answer this question.

Cognitive-achievement research. The importance of measuring students' intelligence in order to better support their learning is supported by a wealth of research. It is well-established that there is a strong association between intelligence and standardized academic achievement. At the broadest level, the correlation between general intelligence, *g*, and general academic achievement is high. Some have estimated the correlation to be above .8 (Deary et al., 2007; Kaufman et al., 2012), with variability across studies. This means that approximately 50 to 70% of the variation in standardized general achievement is explained by general intelligence (Deary et al., 2007; Kaufman et al., 2012; McGrew, 1993; McGrew & Hessler, 1995). This percentage is lower for classroom grades, approximately 40% (Gustafsson & Balke, 1993), but intelligence remains a significant predictor regardless of how achievement is measured.

Clearly, *g* is critical in explaining student's achievement. In order to better understand specific academic skills, a more focused examination of cognitive ability, at the broad ability level, has been fruitful. Research guided by CHC theory has demonstrated that the broad abilities likely differentially explain variance in reading, mathematics, and writing achievement and the effects of the broad abilities on achievement are significant above and beyond the effect of *g* (Gustafsson

& Balke, 1993; McGrew & Wendling, 2010). In addition, the strength and significance of the effects of the broad abilities vary according to age (McGrew & Wendling, 2010).

Reading. Early standardized cognitive-achievement relations research focused on the Woodcock-Johnson Revised Tests of Cognitive Abilities, which measured seven broad abilities (Gc, Gf, Gv, Ga, Gs, Gsm, and Glr), and the WJ-R Tests of Achievement. Cognitive-achievement reading relations are the most studied achievement domain. Reading achievement is separated into basic reading skills, including decoding and word recognition skills, and reading comprehension, which is the complex process of making meaning from text (McGrew & Wendling, 2010). Early WJ-R and reading studies analyzed the data from the standardization sample using multiple regression (McGrew, 1993) and then later SEM analyses (Keith, 1999; Vanderwood et al., 2002). Across basic reading skills and reading comprehension, Gc had the strongest influence and Gv was not significantly related to either (Keith, 1999; McGrew, 1993; McGrew et al., 1997; Vanderwood et al., 2002). Results for all CHC cognitive-achievement relations research are summarized in Table 2 on page 34.

The importance of the other variables varied according to the specific reading skills, and some associations were dependent on age. Basic reading skills were most consistently related to Gc, Ga, and Gsm (McGrew et al., 1997; McGrew, 1993), whereas Glr's relations were inconsistent (McGrew, 1993). The effect of Gs on basic reading was more important for young children through late elementary school (Keith, 1993; McGrew, 1993; Vanderwood et al., 2002). Gf, however, had stronger and more consistent relations with reading comprehension, whereas Ga, and Gs were only weakly related (McGrew, 1993). Gsm's effect on reading comprehension varied with age and appeared to increase with age, while Gf was moderately associated with reading comprehension at a young age, but this association declined over time (McGrew, 1993). The possible moderation of these relations by ethnic groups was also studied. Caucasian, African American, and Hispanic groups were compared using the WJ-R (Keith, 1999). Similar relations were found across groups for reading with one exception. Gc and Gs had a stronger relation with the reading comprehension of middle school Hispanic students. Additionally, Caucasian and Hispanic students from low SES backgrounds were also compared using the WJ-R, and Gc and Ga both significantly influenced their basic reading and reading comprehension performance to a similar extent (Garcia & Stafford, 2000).

Despite changes in the revised WJ III test, similar relations have been found to the previous edition (WJ-R). Gc had a strong effect on basic reading and reading comprehension which tended to increase with age (Benson, 2008; Evans et al., 2002; Floyd et al., 2007; Floyd et al., 2012). Gs and Glr were generally important for basic reading and reading comprehension among younger students (Floyd et al., 2007; Floyd et al., 2012); the significant effect of Gsm on basic reading arose by age 7 (Floyd et al., 2007), and appeared consistently stronger for basic reading than reading comprehension (Evans et al., 2002). Ga seemed to have a less prominent effect and was inconsistently significant at different ages (Benson, 2008; Evans et al., 2002; Floyd et al., 2007; Floyd, 2012).

Even in the most recent revision of the WJ, the WJ-IV, Gc, Ga, and Gs significantly influence both basic reading and reading comprehension. Similar to one study using the WJ-R, Gf significantly influenced reading comprehension (Niileksela et al., 2015). In contrast to studies with the WJ-R and WJ-III (McGrew, 1993), Gsm and Glr did not exert a significant effect on either basic reading or reading comprehension. Developmental differences were tested quantitatively and again supported differential effects across ages.

Reading cognitive-achievement relations were also studied using the WISC-IV, WISC-V, KABC-II, and the DAS-II. It is important to note that each of these tests measures fewer than seven CHC broad abilities, and thus, it was not possible to test the effects of some abilities, most often Ga and Glr. More specifically the WISC-IV and –IV included measures of Gc, Gf, Gsm, Gv, and Gs. As measured by the WISC-IV, Gc, Gsm, and Gf significantly influenced students' performance on a composite of reading (basic reading and reading comprehension were not separated; Beaujean et al., 2014); on the WISC-V significant relations between Gc and Gsm and basic reading and reading comprehension were found (Caemmerer, Keith, Maddocks, & Reynolds, 2017). Relations between the KABC-II (including measures of Gc, Glr, Gf, Gsm, and Gv) and the KTEA-II were tested using a CHC model (Hajovksy et al., 2014). In contrast to research with the WJ, Glr had the largest direct effect on reading decoding (a basic reading skill), followed by Gc and Gsm. Surprisingly, Gv, in addition to Gc, was significantly related to reading comprehension. Similar to previous research the effect of Gc on reading decoding and reading comprehension increased with age, and was significantly greater in later grades than earlier grades. Gsm and Glr's effects were less influenced by developmental differences and were small to moderately sized across grades. Relations between the reading skills themselves were tested, and reading decoding was found to have a large direct effect on reading comprehension. (Hajovksy et al., 2014). Lastly, one study analyzed the effects of the seven broad abilities (Gc, Gf, Gv, Gsm, Gs, Glr, Ga) measured by the DAS-II on students' basic reading skills (Elliot et al., 2010). Some of these broad abilities were measured by only one subtest, thus Glr, Ga, and Gs were limited in their representativeness of these broad abilities. Nonetheless, consistent with previous research, Gc, Gsm, and Ga significantly influenced students' basic reading skills. An inconsistent result was the significant effect of Gf on basic reading skills (Elliot et al., 2010).

While only a few studies analyzed relations between students' cognitive abilities and their reading skills using tests other than the WJ, many of the effects were consistent with research based on the WJ. Consistently, Gc and Gsm seemed to exert important effects on basic reading and reading comprehension. In contrast, Gv (as measured by the KABC-II) significantly influenced reading comprehension and Gf (as measured by the DAS-II) significantly influenced basic reading. These two relations were not consistently supported in WJ research and thus warrant further exploration.

Reading is the only achievement domain that has been studied using CB-SEM. Two intelligence tests, the WISC-R and WJ-R, were used to predict reading tests from the Woodcock Johnson III Tests of Achievement. Once again the strong influence of Gc on reading was supported, followed by Ga and Gs (Flanagan, 2000).

In sum, it is clear that Gc is a strong influence on both basic reading skills and reading comprehension. Ga, Gsm, and Gs have also shown consistent effects across batteries. In contrast, Glr's effects were somewhat inconsistent and appeared more important for basic reading and at younger ages. Gf's influence on reading comprehension was tentatively supported in a few studies. Most studies suggest that Gv's influence is negligible, but one study provides contradictory evidence suggesting it may be important for reading comprehension (Hajovksy et al., 2014).

Mathematics. Another widely researched academic domain is math (see Table 2 for a summary of findings). Math achievement has been conceptualized as two skills, basic math skills, including arithmetic and computation skills, and math reasoning, which involves problem solving with word problems and applying mathematical operations and concepts (McGrew & Wendling, 2010). An early regression study using the WJ-R suggested that Gs, Gc, and Gf were the most consistent predictors of basic mathematics skills and mathematics reasoning (McGrew & Hessler 1995). Gc and Gf's effects on basic math and math reasoning increased with age, and Gsm was also moderately related to Basic Math, but from ages 5 to 10 only. Math reasoning was influenced by Gv for five through eight year olds and by Glr in late adolescence (McGrew & Hessler 1995).

Results based on a SEM analysis of the WJ-R were similar and reinforced the importance of Gs, Gc, and Gf in explaining math achievement (Keith, 1999; McGrew et al., 1997). The SEM results, however, tended to vary more with age. The relative importance of Gs for younger and older students was inconsistent (Keith, 1999; McGrew et al., 1997). Gf appeared more important for math reasoning in elementary and middle school, but was not significant for high schoolers (Keith, 1999; McGrew et al., 1997). Gc, on the other hand, was important for math reasoning across grades (Keith, 1999).

Analyses with the revised WJ III offer contradictory evidence regarding which broad ability exerted the most important influence on math. One study found that Gf was the only broad ability to exert large effects (Taub et al., 2008), whereas another suggested that Gc had the strongest effect on both basic math skills and math reasoning (Floyd et al., 2003). Although Gf was not the strongest effect in this study, Gf moderately influenced basic math skills across all ages, and its effect was stronger for math reasoning (Floyd et al., 2003). Both studies consistently found that Gc's effect increased from moderate to strong in later adolescence (Floyd et al., 2003, Taub et al., 2008). The effects of the other broad abilities were inconsistent across the studies, and tended to vary according to age. Gsm was consistently moderately sized across ages for math reasoning and was significantly related to basic math skills after age seven (Floyd et al., 2003). Gs' effect decreased from strong among five to six year olds to moderate for 9 through 13 year olds (Taub et al., 2008). Glr was important across math skills for ages six through eight only, and

Ga was important for basic math reasoning among young children only (Floyd et al., 2003). Across these studies, Gv consistently had no effect. When analyzing the new WJ-IV, Gc, Gf, and Gs were all significantly related to both basic math skills and math reasoning. Unexpectedly, Gv was significantly related to math reasoning (Niileksela et al., 2015).

Two studies examined cognitive-math achievement relations using a battery other than the WJ, the WISC-IV and -IV (Caemmerer et al., 2017; Parkin & Beaujean, 2012). Both studies support the importance of Gf on math, either an overall math achievement latent variable measured by one basic math and one math reasoning subtest (Parkin & Beaujean, 2012) or math reasoning and basic math separately (Caemmerer et al., 2017). There is also evidence to suggest Gs is important for both math skills, and Gsm may be particularly important for math reasoning (Caemmerer et al., 2017). Overall, much remains to be explored regarding cognitive-math relations using tests other than the WJ.

In sum, Gf and Gc consistently exerted significant effects on students' math achievement, while the effects of Gs and Gsm were consistent, but the strength of their effects varied with age. The influence of Glr and Gv was inconsistently significant and warrants further study.

Writing. While there have been a number of analyses of the relations between intelligence and reading and math achievement, writing is the least understood academic domain in terms of cognitive-achievement relations. Writing achievement is separated into basic writing skills and written expression. Basic writing skills tend to include measures of spelling, knowledge of writing mechanics, and word usage skills. Written expression tends to measure sentence construction, sentence production in response to prompts or pictures, and it may include fluency measures. Conclusions based on the WJ-R suggest that Gc and Gs were consistently related to writing achievement across development (McGrew & Knopik, 1991). The strength of Gc increased with age. The strength of Gs' association was consistent across development for written expression, and it decreased for basic writing around age eighteen. Gf was consistently related to written expression across development, but was mostly related to basic writing from ages six to thirteen. Ga's relation was the most age dependent, as Ga was related to both writing domains before age 11. In contrast, there was little evidence of significant relations for Glr and Gsm, and less for Gv (McGrew & Knopik, 1991).

Writing cognitive-achievement relations were largely similar based on the WJ III (Floyd, McGrew, & Evans, 2008). Again, Gc's effects were moderate to strong, and increased with age, and the effects of Gs were consistently moderate for both writing domains. Similar to the findings with the WJ-R, Ga's influence on written expression was limited to young children. Additionally, Gv's effects were negligible. In contrast to the earlier edition of the WJ, Gsm exerted a moderate effect on both writing domains after age seven, Glr was important for both domains among young children, and Gf's influence did not emerge until age fifteen (Floyd et al., 2008). Research with the new WJ-IV (Niileksela et al., 2015) was more similar to the WJ III than the WJ-R for basic writing skills; Gc and Gsm exerted strong effects on basic writing, and the effect of Gs was moderate. The associations with written expression, however, were more divergent. Gv was strongly related to written expression across all ages, as well as Gs, but Gc did not exert a significant effect (Niileksela et al., 2015).

Two studies analyzed these relations using the WISC-IV and WIAT-II and WISC-V and WIAT-III. Results were mostly in agreement with those suggested by the WJ scales. As measured by the WISC-IV, Gf, Gc, Gsm, and Gs were all important in explaining overall writing achievement (a combined factor of basic writing and written expression; Beaujean, Parkin, Parker, 2014); and according to the WISC-V, Gc and Gsm had a significant influence on basic writing, while Gf had a significant influence on a more complex written expression task (Caemmerer et al., 2017).

In summary, the effects of the broad abilities on writing achievement seem to be more inconsistent than the other achievement domains (see Table 2 for a summary of findings). The influence of Gc and Gs were generally significant, Gf and Gsm exerted significant effects inconsistently, and Gv influenced written expression in one study.

To summarize the research examining cognitive-achievement relations, some broad abilities consistently exert significant effects on multiple academic domains. Gc, Gsm, and Gs significantly influence reading, math, and writing achievement. Glr appears to influence all of these domains as well, but this broad ability seems to be more important for younger ages. Gf, on the other hand, appears to consistently influence math, but the effects of Gf on reading comprehension and written expression are tentative. Across most studies the effects of Gv were negligible; a few significant relations were found, however, on written expression, math reasoning, and reading comprehension. The inconsistent effects of Gf and Gv warrant further study. Also worth further study are the differences in the significance and relative importance of broad abilities across tests. Most of the studies analyzed the WJ tests, and the assumption that cognitive-achievement relations are replicable across different tests requires further study.

Applications to learning disability research. As previously noted, findings from the current study may have implications for students with learning disabilities.[1] Previous studies

---

[1] Federal criteria indicate a student may have an specific learning disability if the "child does not achieve adequately for the child's age or to meet State-approved grade-level standards;" a specific learning disability is defined as "a disorder in one or more of the basic psychological processes involved in understanding or in using language, spoken or written, that may manifest itself in an imperfect ability to listen, think, speak, read, write, spell, or to do mathematical calculations, including conditions such as perceptual disabilities, brain injury, minimal brain dysfunction, dyslexia, and developmental aphasia," and excluding learning problems that are "primarily the result of visual, hearing, or motor disabilities, of mental retardation, emotional disturbance, or of environmental, cultural, or economic disadvantage" (U.S. Department of Education, Office of Special Education Programs, 2006).

suggest different cognitive abilities are more or less salient for students with particular learning disabilities. Overall, working memory, processing speed, and language skills are important predictors for achievement among students with specific learning disabilities. Students at-risk for or classified as meeting criteria for math, reading, or writing disabilities score relatively lower on working memory tests (also referred to as short-term memory in the CHC literature; Fuchs et al., 2010; Geary, Hoard, Byrd, Craven, Nugent, & Numtee, 2007; Geary, Hoard, & Hamson, 1999; Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001; Mayes & Calhoun, 2007; Swanson & Alexander, 1997). Fluency and processing speed abilities also appear to be lower in students with math, reading, and writing disabilities (Burns et al., 2002; Calhoon, Emerson, Flores, & Houchins, 2007; Elliot et al., 2010; Geary et al., 2007; Mayes & Calhoun 2007; Niileksela & Reynolds, 2014). Fluency measures involve a timed component and tend to require students to solve simple mathematical operations (such as addition, subtraction, multiplication) or quickly read passages. These achievement fluency measures are similar to processing speed tasks on intelligence tests given the timed and simple nature of the tasks.

In addition, students with reading disabilities and math problem solving disabilities tend to score lower on language tasks (referred to in the CHC literature as verbal-comprehension (Gc); Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001; Fuchs et al., 2008). Also, students with reading disabilities tend to exhibit deficits in phonological skills (referred to in the CHC literature as audiological processing (Ga); McBride-Chang & Manis, 1996; Vellutino et al., 1996).

Recent cognitive-profile analyses revealed intra-individual differences among students with reading and math disabilities; intra-individual analyses involve comparisons between student's scores on one cognitive ability and another ability (Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2012). Students with basic reading disabilities scored relatively lower on language (which were similar to Gc tasks) and working memory tests, students with reading comprehension disabilities scored relatively lower on language tests, and students with math problem solving disabilities scored relatively lower on a Gf test in comparison to their scores on other cognitive abilities (Compton et al., 2012). Processing speed was a relative strength for students with learning disabilities in comparison to their scores on other cognitive abilities. Similar to the lack of CHC cognitive-writing relations literature, writing disabilities are the least studied and understood specific learning disability category.

Table 2

Summary of Significant Cognitive-Achievement Relations Across Studies

	McGrew, 1993 WJ-R	McGrew et al., 1997 WJ-R	Keith, 1999 WJ-R	Vanderwood et al., 2002 WJ-R	Evans et al., 2002 WJ III	Benson, 2008 WJ III	Floyd et al., 2012 WJ III	Hajovsky et al., 2014 KABC-II	Beaujean et al., 2014 WISC-IV*	Niileksela et al., 2016 WJ-IV	Caemmerer et al., 2017 WISC-V
Reading	**Comp.**										
Gc	Sig. (All: 6 -18)	Sig. (All: Grades 1 – 12)	Sig. (All: grades 1 – 12)	Sig. (All: Grades 1 – 12)	Sig. (All: ages 6 – 18)	Sig. (All: Grades K – 12)	Sig. (All: ages 5 – 18)	Sig. (Grades 4 – 12)	Sig. [a]	Sig. (All: ages 6 – 18)	Sig. (All: ages 6-16)
Gsm	Sig. (Ages 10 – 18)	n.t.	n.t.	n.t.	Sig. (All: ages 6 – 18)	-	Indirect through decoding	Indirect through decoding (Grades 1 – 3, 7 – 12)	Sig. [a]	-	Sig. (interacted with age)
Gs	Sig. (Ages 6 – 12)	Sig. (grades 5 – 6)	Sig. (grades 5 – 8)	Sig. (Grades 5 – 6)	Sig. (ages 6 – 10)	n.t.	Indirect through decoding	N/A	-	Sig. (All: ages 6 – 18)	-
Gf	Sig. (All: ages 6 – 18)	n.t.	n.t.	n.t.	Sig. (ages 11 – 14)	n.t.	-	-	Sig. [a]	Sig. (All: ages 6 – 18)	-
Glr	-	n.t.	n.t.	n.t.	Sig. (ages 6 – 11)	n.t.	Indirect through decoding	Indirect through decoding (Grades 1 – 6)	N/A	-	-
Gv	-	n.t.	n.t.	n.t.	-	-	-	Sig. (Grades 1 – 3)	-	-	-
Ga	Sig. (ages 6 – 10)	n.t.	n.t.	-	Sig. (ages 6 – 9)	n.t.	-	N/A	N/A	Sig. (All: ages 6 – 18)	-

Note. If a broad ability was not measured by a specific test, N/A was entered into the cell. n.t. denotes paths from broad abilities to achievement skills that were not tested. Non-significant effects were indicated by dashes.

[a] Age differences were not tested in these studies.

* Composite scores were used in these studies, therefore it was not possible to separate the effects according to specific achievement skills.

Table 2, cont.

	McGrew, 1993 WJ-R	McGrew et al., 1997 WJ-R	Keith, 1999 WJ-R	Vander-wood et al., 2002 WJ-R	Evans et al., 2002 WJ III	Floyd et al., 2007 WJ III	Benson, 2008 WJ III	Elliot et al., 2010 DAS-II	Floyd et al., 2012 WJ III	Hajovsky et al., 2014 KABC-II	Beaujean et al., 2014 WISC-IV	Niileksela et al., 2016 WJ-IV	Caemmerer et al., 2017 WISC-V
Basic	**Reading**												
Gc	Sig.	n.t.	n.t.	Sig. (All: grades 1 – 12)	Sig. (All: ages 6 – 18)	Sig. (ages 7 – 18)	Sig. (Grades 7 - 12)	Sig. [a]	Sig. (ages 7 – 18)	Sig. (All: grades 1 – 12)	Sig. [a]	Sig. (All: ages 6 – 18)	Sig (All: ages 6 – 16)
Gsm	Sig.	n.t.	n.t.	n.t.	Sig. (All: ages 6 – 18)	Sig. (ages 7 – 18)	Sig. (Grades 7 - 12)	Sig. [a]	Sig. (ages 7 – 18)	Sig. (grades 1 – 3, 7 – 12)	Sig. [a]	-	Sig (All: ages 6 – 16)
Gs	Sig. (All: ages 6 – 18)	n.t.	n.t.	-	Sig. (ages 6 – 10)	Sig. (ages 5 – 8)	n.t.	-	Sig. (ages 5 - 8	N/A	-	Sig. (All: ages 6 – 18)	-
Gf	-	n.t.	n.t.	n.t.		-	n.t.	Sig. [a]	-	-	Sig. [a]	-	-
Glr	Sig. (ages 6 – 8)	n.t.	n.t.	n.t.	Sig. (ages 6 – 9)	Sig. (ages 5 – 6)	n.t.	N/A	Sig. (ages 5 – 6)	Sig. (grades 1 – 6)	N/A	-	-
Gv	-	n.t.	n.t.	n.t.	-	-	n.t.		-	-	-	-	-
Ga	Sig. (All: ages 6 – 18)	Sig (All: Grades 1 – 9)	Sig.	Sig	Sig. (ages 6 – 9)	-	-	Sig. [a]	-	N/A	N/A	Sig. (All: ages 6 – 18)	-

Table 2, cont.

	McGrew & Hessler, 1995 WJ-R	McGrew et al., 1997 WJ-R	Keith, 1999 WJ-R	Floyd et al., 2003 WJ III	Taub et al., 2008 WJ III**	Parkin & Beaujean, 2012* WISC-IV	Niileksela et al., 2016 WJ IV	Caemmerer et al., 2017 WISC-V
Math	**Reasoning**							
Gc	Sig. (All: ages 6 – 18)	Sig. (Grades 3 -12)	Sig. (grades 1 – 12)	Sig. (All: ages 6 – 18)	Sig. (ages 9 – 18)	n.r.	Sig. (All: ages 6 – 18)	-
Gsm	Sig. (ages 6 – 10)	n.t.	n.t.	Sig.(All: ages 6 – 18)	-	n.r.	-	Sig. (interacted with age)
Gs	Sig. (All: ages 6 – 18)	Sig. (Grades 1 – 2, 5 – 6, 10 – 12)	Sig. (grades 1 – 4, 9 - 12)	Sig. (ages 6 – 14)	Sig. (ages 5 – 6, 9 – 13)	n.r.	Sig. (All: ages 6 – 18)	Sig. (interacted with age)
Gf	Sig. (All: ages 6 – 18)	Sig. (Grades 1 – 4, 7 – 9)	Sig. (grades 1 – 8)	Sig. (All: ages 6 – 18)	Sig. (All: ages 6 – 18)	Sig. [a]	Sig. (All: ages 9 – 18)	Sig. (All: ages 6 – 16)
Glr	Sig. (ages 15 – 18)	n.t.	n.t.	Sig. (ages 6 – 8)	-	n.r.	-	-
Gv	Sig. (ages 6 – 8)	n.t.	n.t.	-	-	n.r.	Sig. (All: ages 6 – 18)	-
Ga	-	-	Sig. (ages 6 – 9)	-	-	n.r.	-	-
Basic	**Math**							
Gc	Sig. (All: ages 6 – 18)		n.t.	Sig. (ages 9 – 18)	Sig. (ages 9 – 18)	n.r.	Sig. (All: ages 6 – 18)	-
Gsm	Sig. (age 10, 12)		n.t.	Sig.(ages 7 – 18)	-	n.r.	-	-
Gs	Sig. (All: ages 6 – 18)		Sig. (grades 1 – 12)	Sig. (All: ages 6 – 18)	Sig. (ages 5 – 6, 9 – 13)	n.r.	Sig. (All: ages 6 – 18)	Sig. (All: ages 6 – 16)
Gf	Sig. (All: ages 6 – 18)		n.t.	Sig. (All: ages 6 – 18)	Sig. (All: ages 6 – 18)	Sig. [a]	Sig. (All: ages 6 – 18)	Sig. (All: ages 6 – 16)
Glr	-		n.t.	Sig. (ages 6 – 8)	-	n.r.	-	-
Gv	-		n.t.	-	-	n.r.	-	-
Ga	-		n.t.	Sig. (ages 6 – 7)	-	n.r.	-	-

****** A math reasoning and basic math subtest were loaded onto the same latent variable in Taub et al., 2008. Therefore it was not possible to separate the effects on the two skills. n.r. The effects of these broad abilities were not reported in Parkin & Beaujean, 2012.

Table 2, cont.

	McGrew & Knopik, 1991 WJ-R	Floyd et al., 2008 WJ III	Beaujean et al., 2014 WISC-IV*	Niileksela et al., 2016 WJ IV	Caemmerer et al., 2017 WISC-IV
Written	**Expression**				
Gc	Sig. (All: ages 8 – 18)	Sig. (All: ages 7 – 18)	Sig. [a]	-	Sig. (Sentence Composition, All: ages 6 – 16)
Gsm	Sig. (ages 6 – 10)	Sig. (All: ages 8 – 18)	Sig. [a]	-	Sig. (interacted with age)
Gs	Sig. (All: ages 6 – 18)	Sig. (All: ages 7 – 18)	Sig. [a]	Sig. (All: ages 6 – 18)	-
Gf	Sig. (All: ages 7 – 18)	Sig. (ages 15 – 16)	Sig. [a]	-	Sig. (Essay Composition, All: ages 6 – 16)
Glr	-	Sig. (ages 7 – 8)	N/A	Sig. (ages 6 – 8, 9 – 14)	-
Gv	-	-	-	Sig. (All: ages 6 – 18)	-
Ga	Sig. (ages 6 – 10)	Sig. (age 7, 16 – 17)	N/A	-	-
Basic	**Writing**				
Gc	Sig. (All: ages 8 – 18)	Sig. (All: ages 7 – 18)	Sig. [a]	Sig. (All: ages 6 – 18)	Sig. (All: ages 6 – 16)
Gsm	-	Sig. (ages 8 – 18)	Sig. [a]	Sig.(All: ages 6 – 18)	Sig. (All: ages 6 – 16)
Gs	Sig. (All: ages 6 – 18)	Sig. (ages 7 – 17)	Sig. [a]	Sig. (ages 6 – 18)	-
Gf	Sig. (ages 6 – 13)	Sig. (ages 15 – 18)	Sig. [a]	-	-
Glr	-	Sig. (ages 7 – 10)	N/A	-	-
Gv	-	-	-	-	-
Ga	Sig. (ages 6 – 10)	-	N/A	-	-

Note. If a broad ability was not measured by a specific test, N/A was entered into the cell. n.t. denotes paths from broad abilities to achievement skills that were not tested. Non-significant effects were indicated by dashes.

[a] Age differences were not tested in these studies.

* Composite scores were used in these studies, therefore it was not possible to separate the effects according to specific achievement skills.

Missing Data by Design Methodology

Overarching summary. Planned missing data designs are beneficial when researchers want to analyze many items, but the burden of collecting the data from participants is high. My study involved analyzing several intelligence and achievement tests simultaneously. A planned missing data design with multiple linking tests was utilized in order to maximize the number of CHC broad abilities and achievement skills that can be studied. Concerns about bias and power due to missing data were addressed by the use of FIML.

Planned missing data methodology. A major advantage of CB-CFA intelligence-achievement analyses is that these relations can be analyzed at a construct, rather than test, level. A challenge with CB-CFA research, however, is the potential time and financial demands (Enders, 2010). If every participant were required to complete each test included in the analyses, the number of measures would be small due to the time it takes to complete multiple intelligence tests, examinee fatigue, and the financial burdens of administering several tests to a large sample of students. One possible solution is to focus on a smaller number of tests. The problem with this approach, however, is that CB-CFA analyses that focus on two tests are less generalizable to the construct level of intelligence or achievement because the findings are limited to those specific tests. Thus, CB-CFA analyses incorporating several tests are preferred.

In order to capitalize on the benefits of CB-CFA analyses and overcome the inherent time and resource difficulties of this approach, a different type of data collection method is necessary. One such method is referred to as planned missing data methodology (Enders, 2010; McArdle, 1994), which is well suited for large CB-CFA analyses. Using this methodology, every participant is not required to complete each test, thus reducing examinee fatigue and the financial demands that are associated with administering large numbers of tests. Participants will be missing data for some tests, but the missing data is spread out across participants and is under the control of the researcher (Enders, 2010; McArdle, 1994). For this reason, planned missing data designs are considered “efficiency-of-measurement designs” (Graham, Taylor, Olchowski, & Cumsille, 2006, p. 323).

One popular type of planned missing data designs is the three-form design procedure (Enders, 2010; Graham et al., 2006). In this design, all examinees complete one test, referred to as the linking test. Then, a subset of tests is given to each examinee. The three-form design allows a researcher to collect data on, for example, four tests, while each participant may only complete two tests. Furthermore, an equal number of items is not required in each set (Grahman et al., 2006). The three form design was used by Reynolds and colleagues (2013) in their large intelligence CB-CFA analyses. As discussed earlier, however, this type of design is not without its limitations. For example, if participants are not required to complete the same common linking test, it is possible to incorporate more tests and samples. Also, the three form design limited Reynolds and colleagues (2013) analysis to the broad abilities that each of the four intelligence tests shared in common. Reynolds and colleagues (2013) noted that testing other types of planned missing designs may improve data collection methods and allow for the examination of an increased number of broad abilities.

Methodologists also acknowledge that one common linking test may not be necessary. The rationale for incorporating a common linking test is that the questions in the linking test are vital to the research questions. Failure to have these vital questions answered could result in less power when answering the research questions involving those variables. Despite the advantage of avoiding such a concern, there may be scenarios when a common linking test is not needed (Graham et al., 2006). Cognitive-achievement relations research may be such a scenario. The constructs measured by intelligence (Reynolds et al., 2013) and achievement tests are similar across batteries. Thus, the items within any particular test are not the focus of the analysis. Instead, conclusions are aimed at the construct level and reach across batteries. Therefore, it is reasonable that CB-CFA analyses without a common linking test are worth exploring for cognitive-achievement relations research. The current analysis will contribute to the literature by testing an alternative planned missing data design that is not based on one common linking test.

Missing data mechanisms. Beyond the type of planned missing design, two issues may concern researchers who are considering using a planned missing data design: bias and power. In order to discuss bias, a discussion of how missing data are conceptualized is necessary. In 1976, Rubin and colleagues proposed a classification system for missing data that is currently still in use (Enders, 2010). They proposed three missing data mechanisms that explain how missing data relate to variables (Enders, 2010). One mechanism, data that is classified as missing completely at random (MCAR), is considered the ideal scenario. When data are MCAR, the probability of "missing data on variable Y is unrelated to other measured variables and the values of Y itself" (Enders, 2010, p. 7). Said differently, there is no association between the variable that caused the missingness and the variable containing the missingness (Graham et al., 2006). Although ideal, MCAR is the most restrictive missing data condition. In this scenario observed data points are considered a random sample of the scores that would have been analyzed if the data were complete. Thus, data that are MCAR are considered unbiased. There are several scenarios in which MCAR might arise. For example, in the process of collecting test data from students, students may become ill or move to another school and miss the test day (Enders, 2010). Data are missing for those students who were randomly missing during the data collection process.

Two other missing data mechanisms are less desirable, but more common in research. Missing at random (MAR) is the more desirable of the two. MAR means that missing data on variable Y are related to other measured variables in the model, but are not related to the values of Y itself, once the other variables are controlled (Enders, 2010). In other words, data are considered MAR when there is no relationship between the missing data and the incomplete outcome variable (Enders, 2010). An example involves a scenario in which students are administered an intelligence test as part of the screening process for entry to a charter school and then students' physical fitness is measured 9 months later. Suppose the charter school does not admit students who scored below 80 on the intelligence test; all students who performed in the low cognitive range are missing physical fitness scores. Thus, the probability of missing achievement data is a function of intelligence scores, but is unrelated to students' physical fitness performance.

The final missing data mechanism is data that are missing not at random (MNAR). Data that are MNAR occur when the likelihood of missing data on variable Y is related to the values of Y itself, even after controlling for other relevant variables. An example of MNAR is a case where students who are missing achievement data are also those who are below average on reading comprehension (Enders, 2010). The missing achievement data is related to the students' performance on reading comprehension, and this missing data pattern is not ameliorated by the other variables in the model.

Importantly, data that are missing in planned missing designs are under the researcher's control; the data are intentionally missing. There is no correlation between the cause of the missingness and the variable under study. As a result it is reasonable to assume that the MCAR assumptions are met in planned missing designs (Enders, 2010; Graham et al., 2006). As a result of this missingness, some effects will be tested without the full sample, meaning there is less power to detect these effects. Power concerns are minimized, however, because researchers can restrict the missing data to certain variables. Essentially, deciding to use planned missing data involves a cost-benefit analysis. Researchers consider whether collecting additional variables compensates for the resulting loss of power (Enders, 2010).

Another benefit of planned missing data designs is that modern techniques for dealing with missing data are applicable, such as maximum likelihood estimation (MLE). The advantage of MLE is that it allows researchers to analyze the data without discarding incomplete cases. More specifically, MLE maximizes power by borrowing information from the observed data (Enders, 2010; Schafer & Graham, 2002; described in more detail below). When MLE is applied to data that are MCAR or MAR, Rubin demonstrated that all parameters estimates are consistent and unbiased (McArdle, 1994; Rubin, 1987). The effect on power depends on the magnitude of the correlations, however. Weaker correlations between variables limit the effectiveness of MLE by decreasing the amount of information MLE can borrow from the observed data (Enders, 2010). Simulation studies using MLE assuage some of the concerns regarding loss of power though (Enders, 2010; Graham et al., 2006; McArdle, 1994). The results of these studies demonstrate that the decrease in power is not proportional to the decrease in sample size. For example, in terms of data with a medium effect size, power decreased by just ten percent between a pair of variables with only one-third complete data (Enders, 2010).

Although researchers were cautious, and unsure of incorporating planned missing designs into their research two decades ago, missing data analysis is now well established and researchers are more confident in the merits of such an approach (Graham et al., 2006). Power and bias issues are relatively minor and do not limit the usefulness of planned missing data designs in any substantial way. Therefore, such designs are well suited for studying cognitive-achievement relations at the construct level.

MLE for Missing Data. Methods for handling missing data are important in maximizing the effectiveness of planned missing data designs. Thus, understanding how missing data is handled is critical when selecting such designs. MLE for missing data handling is often referred to as full information maximum likelihood (FIML) estimation. Like most missing data analyses, FIML requires an iterative process. FIML operates by trying different combinations of population parameters until it identifies the combination of values that produces the best fit to the data (Enders, 2010; Schafer & Graham, 2002).

FIML begins by specifying a distribution for the population data. The squared standardized distance between an individual's data points and the center of the normal distribution (known as Mahalanobis distance) determines the magnitude of the log-likelihood value (Enders, 2010). Small distances produce less negative, large log-likelihood values, while large distances produce small log-likelihoods. Thus, larger log-likelihood values are preferred and suggest a better fit to the data. When this process is applied to missing data specifically, it is slightly different for each missing data pattern and the log-likelihood values depend only on the variables for which an individual has complete data (Enders, 2010). The variables that are missing are ignored during the iterative estimation process, referred to as the EM algorithm, which continues until the highest log-likelihood is produced (Enders, 2010; Schafer & Graham, 2002). The earlier steps in the iterative process produce the largest changes in the log-likelihood, whereas changes in later steps are much smaller. The iterative process continues until the difference between the steps falls below a very small threshold (referred to as the convergence criterion). At this point, the iterative process stops as the estimates have converged on the maximum likelihood estimates (Enders, 2010).

In this way, the log-likelihood remains a summary measure of the probability of drawing the observed data from a normally distributed population with a particular mean and covariance matrix. It is important to emphasize that FIML does not impute or replace the missing values. Instead, the log-likelihood values for the incomplete cases "serve as a correction factor that steers the estimator to a more accurate set of parameter estimates" (Enders, 2010, p. 94). Throughout the iterative process, regression equations are built that predict the incomplete variables from the observed variables. By using the information from the observed data, the standard errors account for the missing data patterns. FIML, in comparison to listwise deletion for handling missing data (or discarding all participants who have missing data), produces smaller standard errors, which results in higher power for FIML. Even under ideal MCAR conditions, the standard errors for listwise deletion are approximately seven to forty percent larger than FIML, according to simulation research (Enders, 2010). FIML produces unbiased estimates under MAR conditions as well, and thus, FIML is effective under conditions that cause traditional approaches to fail (Enders, 2010). In contrast, FIML will produce biased estimates under MNAR conditions. This bias is more likely to be limited to a subset of variables using FIML, whereas the bias is more likely to be dispersed throughout the model when listwise deletion is used (Enders, 2010). In sum, FIML allows researchers to maximize the data that are available and not discard important information provided by variables with missing data, which thereby increases the accuracy the estimation process.

Concluding Summary

In sum, CHC theory fits well with modern intelligence tests, regardless of whether the tests were developed according to CHC theory (Reynolds et al., 2013). The predictive validity of the CHC broad abilities in explaining students' standardized achievement is well-supported; the broad abilities differentially explain students' reading, mathematics, and writing achievement (McGrew & Wendling, 2010). The current understanding of cognitive-achievement relations, however, is limited to research simultaneously analyzing a single intelligence and single achievement test, and the majority of studies are based on the Woodcock-Johnson tests. Therefore, findings are limited to those specific tests and are less generalizable to students' cognitive and achievement abilities more broadly. The current study addressed these limitations by utilizing a planned missingness design and incorporating additional tests into an intelligence CB-CFA. Then, the cross-battery intelligence factor structure was used to explain students' reading, mathematics, and writing achievement. Using CHC broad abilities to explain students' achievement skills, both of which are representative of several tests, may improve school psychologists' understanding of these relations at a construct, as opposed to test-specific, level.

Chapter 3: Method

Participants

Participants were 3,930 children and adolescents aged 6 to 16 drawn from seven standardization linking samples. Sample sizes within each sample ranged from 88 to 2,223. These samples included the Kaufman Assessment Battery for Children, Second Edition (KABC-II) concurrent validity studies (referred to as the KABC-II XBA sample, *n* = 350), the KABC-II and Kaufman Test of Educational Achievement, Second Edition (KTEA-II) linking sample (*n* = 2,223), the Wechsler Intelligence Scale for Children, Fifth Edition (WISC-V) and KABC-II linking sample (*n* = 88), the WISC-V and Wechsler Individual Achievement Test, Third Edition (WIAT-III) linking sample (*n* = 181), the Wechsler Intelligence Scale for Children Fourth Edition (WISC-IV) and Wechsler Individual Achievement Test, Second Edition (WIAT-II) linking sample (*n* = 532), the WISC-IV and Differential Abilities Scale, Second Edition (DAS-II) linking sample (*n* = 202), the DAS-II and WIAT-II linking sample (*n* = 370). Sample sizes within the KABC-II concurrent validity studies also varied because data collection resembled a planned missingness design (Reynolds et al., 2013).

Participant identification numbers were checked across samples to determine whether the same child participated in multiple standardization samples; 16 duplicates were identified. One duplicate had two entries for the KABC-II and 15 duplicates had two entries for the DAS-II. Duplicate intelligence test entries were deleted, resulting in a total sample size of 3,930 children.

These samples were created to be representative of United States children according to sex, racial group, parental educational level, and geographic region. Demographic information for the sample is shown in Table 3.

Table 3

Percentages for Each Demographic Variable Across the Four Samples

Test	KABC XBA	WISC4/ DAS	DAS/ WIAT2	KABC/ KTEA	KABC/ WISC5	WISC5/ WIAT3	WISC4/ WIAT2
Gender							
Male	49.4	50.0	48.5	50.1	48.9	55.2	50.8
Female	50.6	50.0	51.5	49.9	51.1	44.8	49.2
Ethnic Background							
White, non-Hispanic	63.1	35.5	58.8	62.2	46.6	50.3	61.8
Hispanic	19.6	27.2	20.0	17.7	35.2	21.0	17.5
African American	9.2	24.8	16.2	14.9	10.2	19.9	15.4
Asian	4.7	6.4	3.8	-	2.3	1.7	4.3
Native American	0.7	n.r.	n.r	-	n.r.	n.r.	n.r.
Other	1.4	6.4	1.4	5.2	5.7	7.2	0.6
Parents' Highest Level of Education							
8th grade or below	-	4.5	5.1	-	1.1	2.2	5.6
9th - 11th grade	9.0*	12.4	11.3	14.4*	12.5	8.3	11.7
High school diploma	19.1	25.7	25.6	32.5	18.2	24.9	26.7
Some college	35.7	28.7	30.2	30.1	36.4	35.4	31.8
Bachelor's or higher	33.6	21.2	27.8	23.0	31.8	29.3	24.2

Note. SES was not categorized the same across samples. Asterisks denote samples in which a percentage was reported for 11th grade and below only. Values that were not reported are denoted by n.r.

Across the different samples, children and adolescents were administered specific sets of tests for validity purposes. Convergent validity was evaluated by examining the correlations between one intelligence test and other intelligence tests, or predictive validity was evaluated between an intelligence test and a standardized achievement test. Because data collection for the KABC-II concurrent validity sample resembled a planned missingness design, all the participants in this sample completed the KABC-II (n = 350), which served as the linking test, and then select intelligence and achievement tests in a counterbalanced order. The other intelligence tests included from the KABC-II concurrent validity sample are the WISC-III, WISC-IV, and the WJ III.

The other six samples did not share the same linking test. In this study those six samples were linked to each other through tests the samples shared in common. Specifically, the KABC-II-KTEA and KABC-II-WISC-V samples were linked to the KABC-II concurrent validity sample through the KABC-II; the WISC-IV-WIAT-II and WISC-IV-DAS-II samples were linked to the KABC-II concurrent validity sample using the WISC-IV; the DAS-II-WIAT-II sample was linked to the WISC-IV-DAS-II sample through the DAS-II and to the WISC-IV-WIAT-II sample through the WIAT-II; and the WISC-V-WIAT-III sample was linked to the KABC-II-WISC-V through the WISC-V (see Table 4 for more details about how the tests were linked).

Table 4

Samples and How They are Linked

	KABC-II	WJ III	WISC-III	WISC-IV	WISC-V	WISC-V	DAS-II	KTEA-II	WIAT-II	WIAT-III
SAMPLES:										
KABC-II XBA	350*	89	123	58*	-	-	-	-	-	-
KABC-II/ KTEA-II	2,223*	-	-	-	-	-	-	2,223	-	
WISC-IV/ DAS-II	-	-	-	202*	-	-	202*	-	-	
DAS-II/ WIAT-II	-	-	-	-	-	-	370*	-	370*	
WISC-IV/ WIAT-II	-	-	-	532*	-	-	-	-	532*	
WISC-V/ WIAT-III	-	-	-	-	181*	-	-	-	-	181
WISC-V/ KABC-II	88*	-	-	-	88*	-	-	-	-	-

Note. Values represent the sample sizes for each test. Asterisks indicate the linking tests across samples.

Measures

Six intelligence tests and three achievement tests were included in the cross-battery analyses. The intelligence tests included the KABC-II (Kaufman & Kaufman, 2004), WISC-V (Wechsler, 2014), WISC-IV (Wechsler, 2003), WISC-III (Wechsler, 1991), WJ III (Woodcock et al., 2001), and DAS-II (Elliot, 2007). The achievement tests included the KTEA-II (Kaufman & Kaufman, 2004), WIAT-II (Wechsler, 2001), and WIAT-III (Wechsler, 2009). Refer to Table 5 for the number of tests per broad ability that each test measures and to Table 6 for a description of all the subtests.

KABC-II. The KABC-II was developed using the CHC taxonomy and Lurian theory (Kaufman & Kaufman, 2004). The KABC-II measures five CHC broad abilities: Gf, Gc, Gv, Gsm, and Glr. All 16 KABC-II subtests were analyzed in this study. Age-referenced standardized subtest scores range from 1 to 19, with a mean of 10 and a standard deviation of 3. Average internal consistency estimates ranged from .74 to .93 in the norming sample.

Participants who completed the KABC-II within the current study were drawn from three samples (KABC-II/KTEA-II, KABC XBA, and KABC-II/WISC-V). The KABC-II and KTEA-II were co-normed, meaning participants completed both tests. Participants within the standardization sample included a national representation of children aged 3 to 18 who spoke English, were not institutionalized, and did not "have physical or perceptual impairments that would prevent them from being able to perform the tasks" (Kaufman & Kaufman, 2004, p. 78). Participants were representative according to sex, ethnicity, parental education, geographic location, special education, and gifted placement (Kaufman & Kaufman, 2004). The KABC-II

XBA sample included students with special education and gifted placements as well (other demographic details are reported in Table 3).

WISC. The Wechsler Intelligence Scales for Children, normed for six to 16 year olds, were not initially developed using CHC taxonomy. Instead the factor structure of the WISC has evolved over time. Research with the WISC-IV and WISC-III editions suggests that they do, however, adhere to the constructs in CHC theory (Keith, Fine, Taub, Reynolds, & Kranzler, 2006; Keith & Witta, 1997). The most recent revision, WISC-V, is more consistent with CHC theory than previous editions. Regardless, a CHC five factor structure was analyzed for all three Wechsler tests. The five broad abilities included Gc, Gf, Gv, Gsm, and Gs. Twenty WISC subtests were analyzed in this study. Age-referenced standardized subtest scores range from 1 to 19, with a mean of 10 and a standard deviation of 3.

The 16 WISC-V subtest scores evidenced high reliability. Average internal consistency estimates ranged from .80 to .96 in the norming sample (Wechsler, 2014). Ten highly reliable subtests from the WISC-IV were analyzed. Average internal consistency estimates ranged from .81 to .91 in the norming sample (Wechsler, 2003). In addition, twelve subtests from the WISC-III were analyzed. Average internal consistency estimates ranged from .69 to .87 in the norming sample (Wechsler, 1991).

Participants who completed a WISC measure within the current study were age six through 16 years 11 months and were drawn from five samples, including the KABC-II XBA sample (which included the WISC-III and –IV), WISC-IV/WIAT-II, WISC-IV/DAS-II, WISC-V/KABC-II, and WISC-V/KABC-II. The 123 WISC-III participants included one child with a special education classification and the 58 WISC-IV participants included 16 children with a special education placement. Information regarding the WISC-IV/WIAT-II sample was unavailable and the WISC-IV/DAS-II sample is described below.

Participants within the WISC-V standardization sample excluded those whose primary language was not English, those who were "primarily nonverbal or uncommunicative," had an "uncorrected visual impairment" or "uncorrected hearing loss," an "upper extremity disability that would affect motor performance," disruptive behavior that would prevent a valid assessment, and "previously or currently diagnosed with any physical condition, neurological condition, psychological condition, or illness that might depress test performance, such as epilepsy, traumatic brain injury, or mood disorder" (Wechsler, 2014, p. 42). "A representative proportion of children from various special education classifications was added" to the standardization sample, which included children with developmental delays, intellectual disabilities, specific learning disabilities, speech/language impairment, attention-deficit/hyperactivity disorder, and gifted and talented (Wechsler, 2014, p. 48); however, the WISC-V/KABC-II and WISC-V/WIAT-III samples were "nonclinical samples," meaning children with special education classifications were not included. Other demographic information about the WISC samples is presented in Table 3.

WJ III. The WJ III is appropriate for a wide age range, from ages two to 90 or above. The WJ III was developed using CHC theory and is the most complete measure of the range of CHC broad abilities in this study. The WJ III is the only test in these analyses to assess auditory attention (Ga). Therefore, Ga will not be included in these analyses because it was not measured by multiple batteries. Instead, eleven subtests representing Gc, Gf, Gv, Gsm, Gs, and Glr were analyzed. Age-referenced standardized subtest scores are on a standard intelligence scale with a mean of 100 and standard deviation of 15. Median internal reliability estimates for these subtests ranged from .74 to .94 in the norming sample.

Participants who completed the WJ III were drawn from one sample, the KABC-II XBA sample. WJ III participants were 89 children aged 7 to 16, none of which had a special education classification; additional demographic information is reported in Table 3.

DAS-II. The development of the DAS-II was guided by multiple theoretical orientations, including CHC theory. The DAS-II is appropriate for ages two to 17. Fourteen subtests from the DAS-II measure six broad abilities: Gc, Gf, Gv, Gsm, Gs, and Glr. Age-referenced standardized subtest scores are *t*-scores with a mean of 50 and standard deviation of 10. Average internal consistency estimates ranged from .68 to .97 in the norming sample.

Participants who completed the DAS-II in the current study were drawn from two samples, the DAS-II/WISC-IV and DAS-II/WIAT-II samples. Children in these samples included those ages six to seventeen whose primary language was English and children who were not part of the clinical samples. Other demographic information about these samples is presented in Table 3.

Table 5

Number of Subtests per CHC Broad Ability

	Gc	Gf	Gsm	Gv	Glr	Gs
KABC II	3	2	3	4	4	0
WISC-III, IV, V	4	4	4	4	0	3
WJ III	2	2	2	2	1	2
DAS-II	3	4	3	5	2	2

WIAT. The WIAT-II and WIAT-III measure reading, writing, and mathematics achievement via nine and ten subtests, respectively. An additional writing subtest was present on the WIAT-III because written expression was measured differently across the two WIAT editions.

The WIAT-II included one subtest, named Written Expression, but the WIAT-III included two written expression subtests, Sentence Composition and Essay Composition; the content and organization of the WIAT-III subtests was revised to provide more in-depth written expression skill coverage (Breaux, 2010). The oral language subtests were excluded from these analyses because they are less relevant to the achievement analyses, an oral language specific learning disability does not exist, and the oral language subtests tend to overlap with Gc tasks.

Age-referenced standardized subtest scores have a mean of 100 and standard deviation of 15. Reliability estimates were generally above .90 in the norming samples of both tests (Breaux, 2010; Wechsler, 2005). Participants who completed the WIAT in the current study were drawn from three samples, and their demographic information was presented above and in Table 3.

KTEA-II. The KTEA-II assesses students' writing, mathematics, and reading achievement via six subtests. Age-referenced standardized subtest scores have a mean of 100 and standard deviation of 15. Split-half reliability estimates ranged from .89 to .97 in the norming sample. Participants who completed the KTEA-II in the current study were drawn from one sample, and their demographic information was presented above and in Table 3.

Total sample. A total of 66 subtests were included in the intelligence CB-CFA model, and 16 in the achievement model. See Table 6 for the names of each subtest and its corresponding instrument and descriptions of every subtest analyzed in this study. Age-referenced standardized scores were in this study. Mean subtest scores vary according to the test. The mean subtest score for the WJ-III and all of the achievement tests (KTEA, WIAT-II, and WIAT-III) is 100 with a standard deviation of 15, the mean subtest score for the DAS-II is 50 with a standard deviation of

10 (*T*-scores), and the mean subtest score is 10 with a standard deviation of three for the KABC-II and the Wechsler scales.

Table 6

Descriptions of the Tasks Involved in Each Subtest

Cognitive Tests	
DAS-II	
Subtest	Task Description
Copying (Gv)	The child draws a reproduction of abstract, geometric line designs.
Digits Backward (Gsm)	The child repeats, in reverse order, increasingly long series of digits.
Digits Forward (Gsm)	The child repeats increasingly long series of digits.
Early Number Concepts (Gf)	The child answers basic quantitative questions including counting, number concepts, and arithmetic.
Matching Letter Like Forms (Gv)	The child is shown a figure and then must select the identical shape for a several options.
Matrices (Gf)	The child solves visual puzzles by selecting a missing image from a picture matrix.
Pattern Construction (Gv)	The child is presented with a pattern by the examiner and then must use blocks or tiles to reproduce the pattern.
Rapid Naming (Gs)	The child, working as fast as possible while avoiding mistakes, must name colors and images that are presented to the child by the examiner.
Recall of Designs (Gv)	The child is shown an abstract geographic pattern for five seconds and then must recreate the pattern from memory by drawing it.
Recognition of Pictures (Gv)	The child is shown multiple images for a specified period of time and then must choose the images viewed from a larger group of pictures that includes pictures not viewed by the child.
Recall - Digits Forward (Gsm)	The child repeats a series of numbers in the order the child heard them from the examiner.
Recall – Digits Backwards (Gsm)	The child repeats a series of numbers in the inverse order the child heard them from the examiner.
Recall of Objects-Immediate (Glr)	The child is exposed to an array of objects, and then is asked to recall as many as possible.
Recall of Objects-Delayed (Glr)	A delayed version of Recall of Objects.
Recall of Sequential Order (Gsm)	The child is required to recall a series of verbal information and pictures in the order that the child saw them.
Sequential and Quantitative Reasoning (Gf)	The child completes a sequential pattern involving figures or numbers.

Table 6, cont.

Speed of Information Processing (Gs)	The child, under timed conditions, views a series of figures that have parts in rows and must choose the figure with the most parts in each row.
Verbal Comprehension (Gc)	The child follows verbal instructions to point to, manipulate, or select objects.
Verbal Similarities (Gc)	The child must describe the similarities between three words that describe common objects or concepts.
Word Definitions (Gc)	The child must define given words.
KABC-II	
Atlantis (Glr)	The child is taught nonsense names for pictures of fish, shells, and plants and recalls that information and points to the corresponding picture.
Atlantis Delayed (Glr)	A delayed recall version of Atlantis.
Block Counting (Gv)	The child views pictures of stacks of blocks, some hidden, and counts the exact number of blocks.
Expressive Vocabulary (Gc)	The child must name pictures of objects.
Gestalt Closure (Gv)	The child is shown a partially completed drawing and provides the name of the drawing as if it were complete.
Hand Movements (Gsm)	The child copies a series of taps demonstrated by the examiner, involving the fist, palm, or side of the hand.
Number Recall (Gsm)	The child listens to strings of numbers, increasing in length, and repeats the numbers back verbatim.
Pattern Reasoning (Gf)	The child is shown a matrix of pictures and points to one stimulus out of several options that completes the logical, linear pattern.
Rebus (Glr)	The child is taught a word associated with a symbol and then reads phrases using these symbols.
Rebus Delayed (Glr)	A delayed recall version of Rebus.
Riddles (Gc)	The child listens to characteristics of concepts and either points to a picture of the concept (early items) or verbally names the concept (later items).
Rover (Gv)	The child manipulates a toy dog on a grid with obstacles and attempts to move the dog to a given spot in the fewest moves.
Story Completion (Gf)	The child is shown a row of pictures that tell a story with missing parts and is required to select other pictures to complete the story in the correct order.
Triangles (Gv)	The child is shown an abstract design and recreates the design with several plastic or foam shapes.
Verbal Knowledge (Gc)	The child selects a picture that illustrates a given word or answers a general information prompt.
Word Order (Gsm)	The child touches a series of pictures in order after listening to the examiner read the names of the pictures. More difficult items include an interference task before the child can respond.
WISC (Version)	
Arithmetic (Gsm) (III - V)	The child solves orally presented arithmetic problems, without the use of paper and pencil, under timed conditions.
Block Design (Gv) (III - V)	The child must reproduce two-dimensional geographic patterns using blocks in a specified amount of time.

Table 6, cont.

Cancellation (Gs; IV & V)	The child is shown arrays of pictures and must select target symbols under timed conditions.
Coding (Gs) (III - V)	The child must use a key to copy symbols that correspond to shapes or numbers within a specific amount of time.
Comprehension (Gc) (III - V)	The child answers questions based on general knowledge and social conventions.
Digit Span (Gsm) (III - V)	The child must perform two tasks. Digit span forward requires the child to repeat a series of numbers in the order the child heard them from the examiner. Digit span backward requires the child to repeat a series of numbers in the inverse order the child heard them from the examiner.
Figure Weights (Gf) (V)	The child is presented with a key and selects a response option that balances a scale.
Letter-Number Sequencing (Gsm) (IV & V)	The child must listen to a set of numbers and letters. The child must then repeat the numbers back from smallest to largest and the letters back in alphabetical order.
Matrix Reasoning (Gf) (IV & V)	The child is provided with five response options and must select one to complete a picture with a missing portion.
Information (Gc) (III - V)	The child is required to answer general knowledge questions.
Object Assembly (Gv) (III)	The child must complete puzzles using pieces without outlines.
Picture Arrangement (Gf) (III)	The child must sequence picture cards to complete a story.
Picture Concepts (Gf) (IV & V)	The child must choose a series of pictures from separate rows to create a group that shares common characteristics.
Picture Completion (Gc) (III & IV)	The child looks at a picture and identifies the essential missing piece of the picture under timed conditions.
Picture Span (Gsm) (V)	The child is shown pictures and then must recall those pictures in sequential order from a response page.
Similarities (Gc) (III - V)	The child must describe the similarities between two words that describe common objects or concepts.
Symbol Search (Gs) (III - V)	The child must determine whether or not a specified symbol is present or absent in a group of other symbols within a specified amount of time.
Visual Puzzles (Gv) (V)	The child is presented with images and must mentally manipulate them to form a complete picture.
Vocabulary (Gc) (III - V)	The child must define given words or provide a name for a picture.
WJ III	
Analysis-Synthesis (Gf)	The child must analyze parts of an incomplete logic puzzle and identify missing parts.
Auditory Working Memory (Gsm)	The child is required to listen to a series of numbers and words and then reorder the string of information.
Concept Formation (Gf)	The child is required to derive a set of rules pertaining to a set of pictures.
Decision Speed (Gs)	The child is required to quickly process concepts by circling two images that are related.

Table 6, cont.

General Information (Gc)	The child is required to respond to a series of questions identifying where she would find and how she would use specified objects.
Numbers Reversed (Gsm)	The child is required to repeat a series of numbers in the inverse order the child heard them from the examiner.
Picture Recognition (Glr)	The child must recognize previously presented target pictures within a field of distracting pictures.
Spatial Relations (Gv)	The child identifies puzzle pieces that form a complete shape.
Verbal Comprehension (Gc)	The child is required to identify pictures of familiar and unfamiliar objects, listen to words presented by the examiner and provide an appropriate synonym or antonym, and complete four-part verbal analogies based on three parts already given.
Visual-Auditory Learning (Glr)	The child is required to learn unfamiliar symbols that represent familiar words, and then translate sequences of symbols into sentences that she read aloud.
Visual Matching (Gs)	The child must quickly circle matching numbers in an array of numbers.
Achievement Tests	
WIAT (Version)	
Essay Composition (III)	The child writes words, sentences, or a paragraph/short essay in response to prompts.
Math Reasoning/Problem Solving (II &III)	The child solves orally presented math word problems that may require multiple steps and may be related to time, money, measurement, geometry, probability, or reading graphs.
Numerical Operations (II &III)	The child is required to solve written math problems involving addition, subtraction, multiplication, and division.
Pseudoword Decoding (II &III)	The child is required to sound out nonsense words.
Reading Comprehension (II &III)	The child reads sentences or short passages and then answers questions about the main idea, details, or is asked to make inferences.
Spelling (II & III)	The child is required to spell a word based on definitions and its use in a sentence that are presented orally.
Sentence Composition (III)	The child must build sentences using target words and combine multiple sentences into one sentence while maintaining the meaning.
Word Reading (II &III)	The child identifies letters, sounds, or reads words from a list.
Written Expression (II)	The child writes words, sentences, or a paragraph/short essay in response to a prompt.
KTEA-II	
Mathematics Applications/ Math Concepts & Applications	The child answers math problems that are read to them that involve both math concepts and math applications used to solve real-world problems. The child can rely on visual aids to assist in solving the problem.
Mathematics Computation	The child must solve written math problems that involve basic math concepts.
Nonsense Word Decoding	The child must pronounce non-sense words aloud.
Reading Comprehension	The child is initially given commands in written sentences that the child must respond to either orally or by gesturing. The child then is required to read

Table 6, cont.	provided material and must answer literal and inferential questions about the reading.
Spelling	The child must spell a word after hearing the word from the examiner and having the word used in a sentence by the examiner.
Word Recognition	The child begins the subtest by identifying specific letters and then reads words (both phonetic and nonphonetic) that get more difficult as the child progresses.
Written Expression	The child completes a story booklet with age-dependent content. Earlier grades write letters and fill in writing mechanics, while older grades write sentences, complete dialogue, etc.

Note. The latent ability that each subtest measures is in parentheses following the name of the subtest.

Data Analyses and Research Questions

Three statistical programs were used to conduct the SEM analyses. The Statistical Package for the Social Sciences (SPSS, version 21, 2012) was used to select variables and participants and check the data. Following data preparation, invariance was tested via SPSS Amos, Version 23.0 (Arbuckle, 2015). Then, Mplus (Muthén & Muthén, 2012), version 7 was used to analyze the CB-CFA and SEM models. Amos and Mplus handle missing data through the Full Information Maximum Likelihood (FIML) procedure. Currently, FIML is a strongly recommended procedure for handling missing data (Enders, 2010; Enders & Bandalos, 2001; Schafer & Graham, 2002). For a detailed description of missing data considerations, including how FIML operates, refer to the literature review. This study addressed three broad questions.

Question one: Are the different samples of participants who completed the same test invariant? Measurement invariance was tested across different samples of youth who completed the same test (sample invariance) to determine if the intelligence constructs were measured in the same way across samples. For example, two samples of youth completed the WISC-V: 88 youth were included in the WISC-V/KABC-II sample and 181 youth were included in the WISC-V/WIAT-III sample. In order to establish the WISC-V broad abilities were measured similarly across the two separate groups of youth, measurement invariance was tested. Four other sample invariance tests included: (1) three WISC-IV samples (WISC-IV/DAS-II, WISC-IV/WIAT-II, WISC-IV/KABC-II XBA), (2) three KABC-II samples (KABC-II XBA, KABC-II/WISC-V, KABC-II/KTEA-II), (3) two DAS-II samples (DAS-II/WIAT-II, DAS-II/WISC-IV), and (4) two WIAT-II samples (WIAT-II/DAS-II and WIAT-II/ WISC-IV). If measurement invariance was established across samples, equivalent subtests across the samples were merged in later steps of the analyses, which allowed for larger combined sample sizes.

The process of testing for invariance involves a series of steps using a multi-group confirmatory factor analysis. Each step introduces more constraints into the model, and therefore, each step becomes progressively more stringent (Keith, 2014, chapter 19). The first step involved testing configural invariance at the first order (broad abilities and subtests) level because the measurement model is of interest, not the structural model (which includes the relations between *g* and the broad abilities). If the model fit well for both samples, configural invariance was accepted. Next, weak factorial invariance (also known as metric invariance) was tested. In this step, the subtest factor loadings were constrained to be equal across groups. If weak invariance was supported, this means that the scaling of the latent variables was the same across groups (Keith, 2015, chapter 19). In the third step, the factor loading constraints remained plus the intercepts of the subtests were constrained to be equal across groups (strong factorial invariance or intercept invariance). Finally, strict invariance were tested by retaining all previous constraints plus constraining the residual variances of the subtests to be equal across groups (also known as residual invariance). Establishing residual invariance means that differences in the means and variances of the observed scores are fully explained by differences in the latent variables means and variances (Keith, 2015). Together, these four invariance steps constitute measurement invariance.

Beginning with the second step, each model was compared with the previous model using two criteria: the change in chi-square test and the change in the comparative fit index (CFI) (Keith, 2015). When comparing models, invariance is accepted if the change in chi-square is not significant and if the change in CFI is equal to or less than -.01 (Cheung & Rensvold, 2002). The analysis progresses to the next step only if invariance is supported at the previous step. It is possible to establish partial invariance, which allows a limited number of differences across groups at one step. If partial invariance is accepted, then invariance testing can proceed to stricter constraints (Keith, 2015). If measurement invariance was supported across samples, the subtests were merged.

Question two: Are the different editions of the same test invariant? Measurement invariance is also of concern across editions of the WISC (WISC-III, WISC-IV, and WISC-V) and WIAT (WIAT-II and WIAT-III) because many of the subtests are identical across the editions of the tests. The merged invariant samples of the WISC-IV, WISC-V, and WIAT-II were used to test for edition invariance in this step of the analysis.

WISC edition invariance was first tested between the merged KABC-II WISC-III and data and data from the three WISC–IV samples (KABC-II XBA WISC-IV, WISC-IV/DAS-II, and WISC-IV/WIAT-II). If those data were invariant, they were merged and invariance was tested between the combined WISC-III/-IV combined data and the merged WISC-V data. If invariance was established, data for the 13 subtests that the WISC-III,-IV, and –V shared in common were merged into one combined dataset.

Finally, invariance was tested for the merged WIAT-II sample data and the single sample of the WIAT-III. If the six subtests that the two WIAT tests shared in common were invariant, those six subtests were merged. Edition invariance was evaluated according to the same criteria used for sample invariance, the likelihood ratio test and change in CFI were compared between more constrained invariance models.

Question three: How well will a CB-CFA model represent data from six IQ tests? After establishing invariance across the samples and editions of the same tests, the next analysis step tested a first order model across all six intelligence tests, including the six CHC broad ability latent variables and excluding *g*. In this step, correlations were included between each of the six broad abilities. The six broad ability latent variables were Gc, Gf, Gv, Gsm, Gs, and Glr. Each broad ability was estimated by 7 to 15 subtests. This resulted in 15 measured variables representing Gv, 12 subtests each represented Gc, Gf, and Gsm, and Glr and Gs were each estimated by 7 subtests. Three correlations were drawn between pairs of Glr subtests that included a delayed counterpart (KABC-II Atlantis, KABC-II Rebus, and DAS-II Recall of Objects). In order to assess delayed recall, the same test was administered twice, but after a delay of a specified time. These correlations are referred to as correlated errors and indicate that the subtests share something in common beyond the Glr latent variable.

Cognitive models were evaluated according to multiple measures of fit, as suggested by methodologists (Hu & Bentler, 1998, 1999). Chi-square, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), the comparative fit index (CFI), and the Tucker-Lewis index (TLI) were used to assess the fit of single models (Keith, 2015). Cut-off values that suggest good fit are RMSEA below .05, SRMR below .08, and CFI and TLI values above .95 (Hu & Bentler, 1999).

After an acceptable first order model was established, *g* was introduced and a second-order model (with *g* subsuming the six broad ability latent variables) was tested. Model fit was evaluated according to the fit indices identified above. Parameters of interest included the factor loadings of the subtests on their respective latent broad abilities and the factor loadings of the broad abilities on *g*.

Question four: How well will a CB-CFA model represent data from three achievement tests? Participants' reading, mathematics, and writing achievement was initially analyzed separately in a CB-CFA. Previous research suggests that the specific skills within these achievement domains are differentially influenced by the CHC broad abilities (McGrew & Wendling, 2010). Therefore, six separate latent variables were created for the specific skills within reading, mathematics, and writing. In terms of reading, basic reading and reading comprehension were tested. Math was represented by basic math and math reasoning, while writing was represented by written expression and spelling. Each of these skills were defined in the literature review. The basic reading and written expression latent variables were estimated by four subtests and the other achievement skill latent variables were estimated by two subtests each (see Table 5 for descriptions of each of these subtests). These six specific achievement skills were correlated with each other and included in one large achievement model.

The CB-CFA achievement involved testing one comprehensive cross-battery achievement model including all six specific achievement skills. Model modifications based on modification indices, if theoretically justifiable, were investigated during this step. The fit of single models will be evaluated as described earlier under question three. Parameters of interest included the factor loadings of the subtests onto their respective latent specific achievement variables.

Question five: Do the CHC broad abilities differentially affect reading, writing, and mathematics achievement? The final CB-CFA intelligence was used to predict the final achievement models. These models, where intelligence explains achievement, were referred to as cross-battery SEM models, CB-SEM. Across these models, the parameters of interest were the paths from each broad ability to the achievement variables.

Analyses were completed in a series of steps. First, paths from all six broad abilities were tested in these initial broad ability-achievement models. These paths were then examined for statistical significance (alpha level = .05). Second, all non-significant paths were deleted from the cognitive-achievement in one step. Third, *g* was introduced into the model with only significant cognitive-achievement paths. A higher order model was tested in which *g* subsumed the six broad abilities, plus a path was added from *g* to the achievement latent variable, in addition to all of the significant broad ability-achievement paths. If the path from *g* to the achievement skill was significant, the path was retained in the final model.

When only significant and positive paths remained, a path from *g* to the achievement latent variables was tested. The paths from *g* cannot be included earlier in the analysis because this will result in model underidentification, meaning that the model is unsolvable unless additional constraints are added (Keith, 2015). Previous research supports the order of this analysis approach (Niileksela et al., 2016). If the path from *g* to the achievement variable was non-significant it was removed in a subsequent model.

Some researchers in the field argue that *g* should be the central ability in explaining achievement, however (Parkin & Beaujean, 2012). In order to address this current debate, an additional model was tested which included a second-order model of intelligence and included only one path, from *g* to each achievement skill (referred to as *g* only-achievement models). Therefore, a total of six CB-SEM cognitive-achievement models were tested.

Model evaluation. As described earlier, several fit measures were used to evaluate the fit of single models. Cut-off values that suggest good fit are RMSEA below .05, SRMR below .08, and CFI and TLI values above .95 (Hu & Bentler, 1999). Alternative, nested models were compared using the likelihood ratio test; change in CFI was used to compare invariance models (Cheung & Rensvold, 2002). For non-nested competing models, the Akaike's information criterion (AIC) was examined; smaller AIC values indicate better fitting models (Keith, 2015).

Chapter 4: Results

This results section is divided into several sections. First, descriptive statistics for all six datasets are presented. Second, model results for sample and test edition invariance are described. Third, CB-CFA results for the intelligence and achievement models are presented. Finally, CB-SEM cognitive-achievement models are described.

Descriptive Statistics

Subtest sample sizes, means, standard deviations, skewness, and kurtosis estimates are presented in Table 7. The means and standard deviations of the subtests were mostly similar to those of their respective norming samples. As evidenced in Table 7, the subtests were normally distributed; skewness and kurtosis values were within the acceptable ranges and were well below suggested cut-off points for univariate normality (below 2 and 7, respectively; Curran, West, & Finch, 1996).

Table 7

Descriptive Statistics for Cognitive and Achievement Tests

Tests and Subtests	N	*M*	*SD*	Skew	Ku
Cognitive Tests					
DAS-II					
Copying (Gv)	178	51.23	8.412	.432	1.141
Digits Backward (Gsm)	557	49.822	8.556	-.321	.515
Digits Forward (Gsm)	557	49.810	9.777	.131	.934
Early Number Concepts (Gf)	178	51.101	8.636	.398	-.189
Speed of Information Processing (Gs)	557	51.068	9.205	-.008	.374
Matching Letter Like Forms (Gv)	178	51.708	9.081	-.289	.404
Matrices (Gf)	557	50.206	9.193	.090	-.038
Pattern Construction (Gv)	557	49.969	8.647	.686	1.605
Rapid Naming (Gs)	557	50.470	8.969	.836	1.974
Recall of Designs (Gv)	556	50.245	8.778	.014	.039

Table 7, cont.

Recognition of Pictures (Gv)	557	50.260	9.294	.654	1.114
Recall - Digits Forward (Gsm)	557	49.810	9.777	.131	.934
Recall – Digits Backwards (Gsm)	557	49.822	8.556	-.321	.515
Recall of Objects-Immediate (Glr)	557	49.057	10.324	-0.075	0.924
Recall of Objects-Delayed (Glr)	557	50.165	9.458	-0.066	0.33
Recall of Sequential Order (Gsm)	557	50.092	9.482	0.03	0.743
Sequential and Quantitative Reasoning (Gf)	556	50.545	9.086	0.598	1.201
Speed of Information Processing (Gs)	557	51.068	9.205	-0.008	0.374
Verbal Comprehension (Gc)	178	50.152	8.956	1.094	2.15
Verbal Similarities (Gc)	557	50.659	8.476	-0.286	1.244
Word Definitions (Gc)	556	50.192	8.839	0.159	1.14
KABC					
Atlantis (Glr)	2654	10.019	3.085	-0.181	-0.061
Atlantis Delayed (Glr)	2435	9.933	2.804	-0.339	-0.131
Block Counting (Gv)	2655	9.972	2.998	-0.019	-0.113
Expressive Vocabulary (Gc)	2656	9.848	2.950	-0.032	0.045
Gestalt Closure (Gv)	619	9.997	2.893	0.051	0.232
Hand Movements (Gsm)	2656	10.075	2.878	0.038	0.11
Number Recall (Gsm)	2657	10.235	2.860	-0.063	-0.073
Pattern Reasoning (Gf)	2656	10.187	2.957	-0.087	0.039
Rebus (Glr)	2657	10.144	3.042	-0.155	0.01
Rebus Delayed (Glr)	2407	10.026	2.962	-0.275	-0.266
Riddles (Gc)	2657	10.144	3.042	-0.155	0.01
Rover (Gv)	2652	10.149	3.017	-0.053	-0.018
Story Completion (Gf)	2653	10.098	2.980	0.019	-0.021
Triangles (Gv)	2656	10.003	2.913	-0.083	-0.136
Verbal Knowledge (Gc)	2657	10.012	2.944	-0.003	-0.069
Word Order (Gsm)	2657	9.925	2.834	0.109	0.128
WISC (Version)					
Arithmetic (Gsm) (III – V)	880	10.318	2.773	0.161	-0.334
Block Design (Gv) (III – V)	1178	10.175	2.844	0.069	0.052
Cancellation (Gs; IV – V)	998	10.032	3.019	0.048	0.062
Coding (Gs) (III – V)	1178	10.055	2.902	0.204	-0.007
Comprehension (Gc) (III – V)	1174	10.248	2.894	-0.062	0.188
Digit Span (Gsm) (III – V)	1167	10.041	2.863	0.156	-0.01
Figure Weights (V)	269	9.918	2.691	-0.109	0.078

Table 7, cont.

Information (Gc) (III – V)	1124	10.254	2.836	0.061	-0.199
Letter-Number Sequencing (Gsm) (III – V)	1050	9.996	2.816	-0.464	0.762
Object Assembly (Gv) (III)	123	10.341	2.946	-0.221	0.523
Picture Arrangement (Gf) (III)	123	10.65	3.445	0.203	-0.236
Matrix Reasoning (Gf) (IV – V)	1060	10.205	2.862	0.132	-0.231
Picture Completion (Gc) (III & IV)	324	10.33	3.003	-0.082	0.477
Picture Concepts (Gf) (IV – V)	1060	10.282	2.916	-0.239	0.188
Picture Span (Gsm; V)	269	9.747	2.663	0.063	-0.584
Similarities (Gc) (III – V)	1179	10.179	2.873	-0.088	-0.069
Symbol Search (Gs) (III – V)	1143	10.206	2.934	-0.145	0.736
Visual Puzzles (Gv; V)	268	10.063	2.623	0.002	-0.53
Vocabulary (Gc) (III – V)	1178	10.14	2.911	-0.164	0.079
WJ III					
Analysis-Synthesis (Gf)	87	102.908	17.188	-0.293	0.404
Auditory Working Memory (Gsm)	88	105.398	13.909	0.341	-0.109
Concept Formation (Gf)	89	105.36	13.904	-0.083	0.535
Decision Speed (Gs)	88	100.557	16.183	-0.706	3.748
General Information (Gc)	89	98.371	16.156	-0.313	0.315
Numbers Reversed (Gsm)	89	100.618	14.308	-0.04	0.518
Picture Recognition (Glr)	89	100.787	12.576	0.037	2.841
Spatial Relations (Gv)	89	100.618	11.329	-0.697	1.326
Verbal Comprehension (Gc)	89	102.551	14.237	-0.685	0.556
Visual-Auditory Learning (Glr)	89	94.652	19.760	-0.468	1.861
Visual Matching (Gs)	89	95.843	13.372	0.342	0.082
Achievement Tests					
WIAT (Version)					
Essay Composition (III)	151	100.755	15.719	-0.157	-0.487
Math Reasoning/Problem Solving (II & III)	1083	101.727	15.327	-0.348	0.34
Numerical Operations (II & III)	1081	102.087	15.313	-0.218	0.293
Pseudoword Decoding (II & III)	1054	102.201	13.67319	-0.38	0.078
Spelling (II & III)	1083	101.587	14.263	-0.204	0.44
Sentence Composition (III)	179	99.587	12.828	-0.117	-0.12
Word Reading (II & III)	1078	101.915	14.27557	-0.412	0.407
Written Expression (II)	871	101.901	15.487	0.032	-0.193

Table 7, cont.

KTEA-II					
Nonsense Word Decoding	2021	99.883	15.05088	-0.014	-0.084
Math Concepts & Applications	2223	100.102	14.999	0.072	0.126
Mathematics Computation	2222	99.958	14.071	-0.063	0.198
Spelling	2021	99.739	14.905	-0.004	0.045
Word Recognition	2223	99.829	14.66346	0.025	0.239
Written Expression	2223	100.029	15.158	-0.049	0.055

Invariance Testing

Measurement invariance was tested to determine whether the intelligence and achievement constructs were measured in the same way across different samples and different editions of the tests. This analysis was a necessary precursor to combining and simultaneously analyzing the data for the cognitive, achievement, and cognitive-achievement models. Eight separate invariance models were tested.

First, invariance was tested across different samples of youth who completed the same test (sample invariance). Five sample invariance tests were conducted: (1) two samples WISC-V samples (WISC-IV/KABC-II and WISC-IV/WIAT-III), (2) three WISC-IV samples (WISC-IV/DAS-II, WISC-IV/WIAT-II, WISC-IV/KABC-II XBA), (3) three KABC-II samples (KABC-II XBA, KABC-II/WISC-V, KABC-II/KTEA-II), (4) two DAS-II samples (DAS-II/WIAT-II, DAS-II/WISC-IV), and (5) two WIAT-II samples (WIAT-II/DAS-II and WIAT-II/WISC-IV). After invariance was established across all samples of youth who completed the same test, scores from those samples were merged into one large total sample of all students who completed that specific test.

The merged data were then used to test invariance across test editions (edition invariance). Two tests in the current analysis included multiple editions of the same test, the WIAT (WIAT-II and WIAT-III) and WISC (WISC-III, WISC-IV, and WISC-V). Invariance was tested across different editions of these tests to determine whether the data could be merged across the subtests each edition shared in common. Merging data across test editions and participant samples allowed for larger sample sizes, which increased the power of the subsequent cross-battery analyses.

As described previously, invariance testing was completed in a series of steps, with more stringent constraints added at each step if the constraints in the previous step were supported. If invariance was not established across samples or editions, partial invariance was tested (described further below). The sequence of these steps began with configural invariance (a test of whether the same factor structure fits the data across groups, meaning subtests are associated with the same CHC broad abilities across groups), followed by metric invariance (factor loadings are constrained to be equal across groups), scalar invariance (intercepts plus factor loadings are constrained to be equal across groups), and finally strict invariance (residuals plus intercepts and factor loadings are constrained to be equal across groups). The invariance testing sequence was slightly different for the two WIAT-II samples and two WIAT editions (described further below). Across all invariance models, change in CFI was used to test whether the additional constraints at each step were supported (Cheung & Rensvold, 2002). Absolute values equal to or less than .01 suggest the constraints did not degrade the model, and provide support for invariance.

Sample invariance. Configural invariance was supported across all four models; each of the configural models fit well with the two WISC-V samples, three WISC-IV samples, three KABC-II samples,[2] and two DAS-II samples (refer to Table 8 for model fit indices). In the WISC-

[2] It was not possible to test one KABC-II subtest, Gestalt Closure, from one sample, KABC-II/KTEA-II, for invariance due to a small sample size (n = 193 out of 2,223 total participants) related to a significant amount of missing data for Gestalt Closure in that specific sample. Invariance was supported for Gestalt Closure between the two other KABC-II datasets, KABC-II XBA and KABC-II/WISC-V.

IV invariance model Gv and Gf were combined into one factor because there was only one measure of Gv, Block Design, and the model would have been under-identified otherwise.[3] Accordingly, the structure of what the WISC-V, WISC-IV, KABC-II, and DAS-II measures is the same across the different samples of youth.

Given that configural invariance was supported, metric invariance was tested next. Metric invariance was supported across all four models as well; the change in CFI was well below the cut-off value for each of these models (refer to Table 8 for fit indices and change in CFI values). Thus, the scaling of the latent variables (the broad abilities) is the same across the different samples of youth, meaning that each unit change in the latent broad ability results in the same change in the subtests that estimate that latent variable across the different samples.

Because metric invariance was established, scalar invariance was then tested. Scalar invariance was supported across all four models—the starting point for each of the subtests was the same across the different samples of youth. Finally, the most stringent level of invariance, strict invariance, was supported across three of the models: the WISC-V, WISC-IV, and KABC-II models (see Table 8). Therefore, the subtests in these tests measured the broad abilities with the same degree of measurement error across the multiple samples of the WISC-IV and –V and KABC-II. Strict invariance was not supported for the DAS-II, however, because the change in CFI was slightly above the cut-off value (.011). Modification indices and a comparison of the residuals across the two samples, derived from the scalar invariance output, suggested the Early Number Concepts subtest was the largest contributor to the lack of invariance. Partial strict

[3] Picture Completion is also a measure of Gv, but only one of the three WISC-IV samples included this subtest; therefore it was not possible to include the subtest here. Picture Completion was also completed by participants in the single WISC-III sample (taken from the KABC-II XBA sample), and invariance was tested across the two editions of the tests and is described below.

invariance was tested by allowing the residual of the Early Number Concepts to freely vary across the two samples. This modification resulted in a reduction of the change in CFI (.009), thus supporting partial strict invariance for the DAS-II samples. This finding means differences across the two samples in DAS-II subtest means and variances, excluding Early Number Concepts, are due to differences in latent broad ability means and variances and covariances (Keith, 2015). Of note, the sample size for Early Number Concepts was considerably lower than many of the other DAS-II subtests because Early Number Concepts is only administered to the youngest participants, those aged 2:6 – 6:11. The smaller sample size of the Early Number Concepts subtest may have resulted in more variability in scores. Some methodologists do not consider strict invariance a necessary component of measurement invariance (Vandenberg & Lance, 2000; Widaman & Reise, 1997). For this reason, all DAS-II data was merged, including scores for the Early Number Concepts subtest, because the partial residual invariance issue was not thought to influence later analyses.

Initially, the WIAT-II model included three latent variables (one reading, one math, and one writing latent variable), which were estimated by two to three subtests each. Configural invariance was not supported for this WIAT-II model, however—the initial proposed factor structure was inadequate. Although CFI suggested good fit (.965) and TLI suggested adequate fit (.910), the adjusted RMSEA indicated poor fit (.130) ($\chi^2(df) = 185.198(22)$, $p < .001$). Additional analyses were conducted to explore the inadequate model fit. Neither analyzing the two WIAT-II samples separately nor examining the modification indices resulted in improvements in the model; the WIAT-II factor structure remained inadequate. Due to poor model fit (which is not problematic for later analyses because reading, math, and writing performance will be modeled separately) a different approach was used to test invariance. Invariance was instead tested using the covariance matrices, rather than the raw data. Only subtests were included in these invariance models, no achievement latent variables were modeled. Each of the seven WIAT-II subtests were correlated with each other. First, each of these covariances was constrained to be equal across the two samples; this model fit the data well (see Table 8 for model fit indices and change in CFI values). Next, the subtest variances were also constrained to be equal, which was supported by no change in CFI. Lastly, the subtest means were also constrained to be equal. These constraints were also supported. These steps allowed a more stringent test of invariance than those imposed during measurement invariance testing, but without specifying a known factor structure; these strict invariance steps were supported for the WIAT-II.

Invariance was supported across the multiple samples of the five tests. As a result, these multiple samples were merged into one total sample for each of the five tests. Merging data involved combining subtest data from each sample into a single data column. For example, DAS-II Matrices subtest scores from the DAS-II/WIAT-II and DAS-II/WISC-IV samples were combined into one data column within a combined dataset.

Table 8

Invariance Testing Across Samples

Model Name	**$\chi^2(df)$**	***p***	**$\Delta\chi^2$ (Δdf)**	**Δp**	**CFI**	**Δ CFI**	**Adj. RMSEA**
<u>WISC-V</u> (2 samples)							
Configural Invariance	227.845(188)	.025	-	-	.968	-	.040
Metric Invariance	241.667(199)	.021	13.822(11)	.243	.965	.003	.040
Intercept Invariance	248.995(210)	.034	7.328(11)	.772	.968	-.003	.037
Residual Invariance	263.053(226)	.046	14.058(16)	.594	.970	.002	.035
<u>DAS-II</u> (2 samples)							
Configural Invariance	389.887(310)	.001	-	-	.970	-	.031
Metric Invariance	405.244(324)	.001	15.357(14)	.354	.970	.000	.030
Intercept Invariance	415.799(338)	.002	10.555(14)	.721	.971	.001	.028
Residual Invariance	467.622(358)	.000	51.823(20)	<.001	.959	.011	.033
Partial Invariance	459.893(357)	.000	44.094(19)	.001	.962	.009	.033
<u>WIAT-II</u> (2 samples)							
Covariance Invariance	33.794(21)	.038	-	-	.997	-	.037
Variance Invariance	44.672(28)	.000	10.878(7)	.144	.997	.000	.037
Mean Invariance	68.715(35)	.000	24.043(7)	.001	.993	.004	.047
<u>WISC-IV</u> (3 samples)							
Configural Invariance	282.713(157)	.000	-	-	.968	-	.047
Metric Invariance	307.314(172)	.000	24.557(15)	.056	.965	.003	.047
Intercept Invariance	354.512(187)	.000	47.198(15)	<.001	.957	.008	.049

Table 8, cont.

Residual Invariance	384.303(210)	.000	29.791(23)	.155	.955	.002	.047
KABC-II (3 samples)							
Configural Invariance	841.003(280)	.000	-	-	.972	-	.045
Metric Invariance	865.834(302)	.000	24.831(22)	.305	.972	.000	.043
Intercept Invariance	931.721(324)	.000	65.887(22)	<.001	.970	.002	.043
Residual Invariance	1011.046(356)	.000	79.325(32)	<.001	.968	.002	.043

Test edition invariance.

WISC edition invariance. First, WISC invariance was tested between the WISC-III and WISC-IV data. This invariance model included data from the merged WISC-III and –IV KABC-II XBA concurrent validity study data and merged data from the other two WISC–IV samples (WISC-IV/DAS-II, WISC-IV/WIAT-II; see Figure 1). The subtests that were administered in only one version of the test were included in the model, but were not tested for invariance. Two such subtests, Object Assembly and Picture Arrangement, were part of the WISC-III but not the WISC-IV, and thus were not included in the WISC-IV merged data group (see Figure 1). For the purposes of invariance testing only, Gv and Gf were combined into one factor to avoid model underidentification due to too few Gv subtests. Configural, metric, scalar, and residual invariance were each supported; change in CFI was below the .01 cut-off value (see Table 9). Thus, data for the 14 subtests that were administered in both the WISC-III and –IV were merged resulting in the merging of data from three samples.

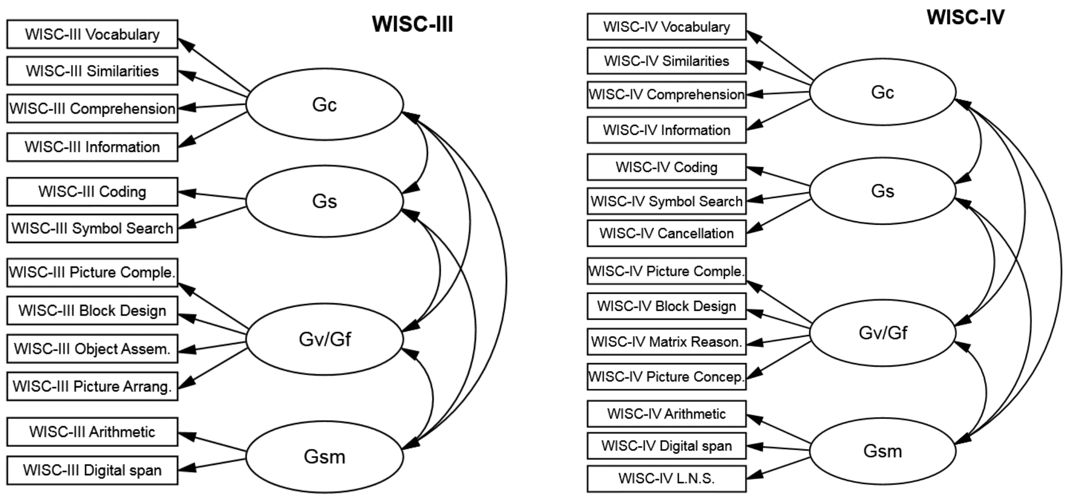

Figure 1. WISC-III and WISC-IV edition invariance models

Next, invariance was tested for the merged WISC-III/-IV data and the merged WISC-V data. Here, three subtests were unique to the WISC-V (Figure Weights, Picture Span, and Visual Puzzles) and thus were not included in the WISC-III/-IV group; one subtest was not included in the recently revised WISC-V test (Picture Completion) and thus was only included in the WISC-III/-IV group in this model; and the two WISC-III only subtests, Object Assembly and Picture Arrangement, were not included in this model because the amount of missing data impeded the analysis (see Figure 2). These unique subtests are not tested for invariance because the subtests are not compared across groups. Again, configural, metric, scalar, and residual invariance were each supported with the change in CFI well below the .01 cut-off value (see Table 9). Data for the 13 subtests that were administered in both the WISC-III/–IV and WISC-V were merged resulting in the merging of data from five samples. These 13 merged subtests are referred to as “WISC” subtests in later analyses.

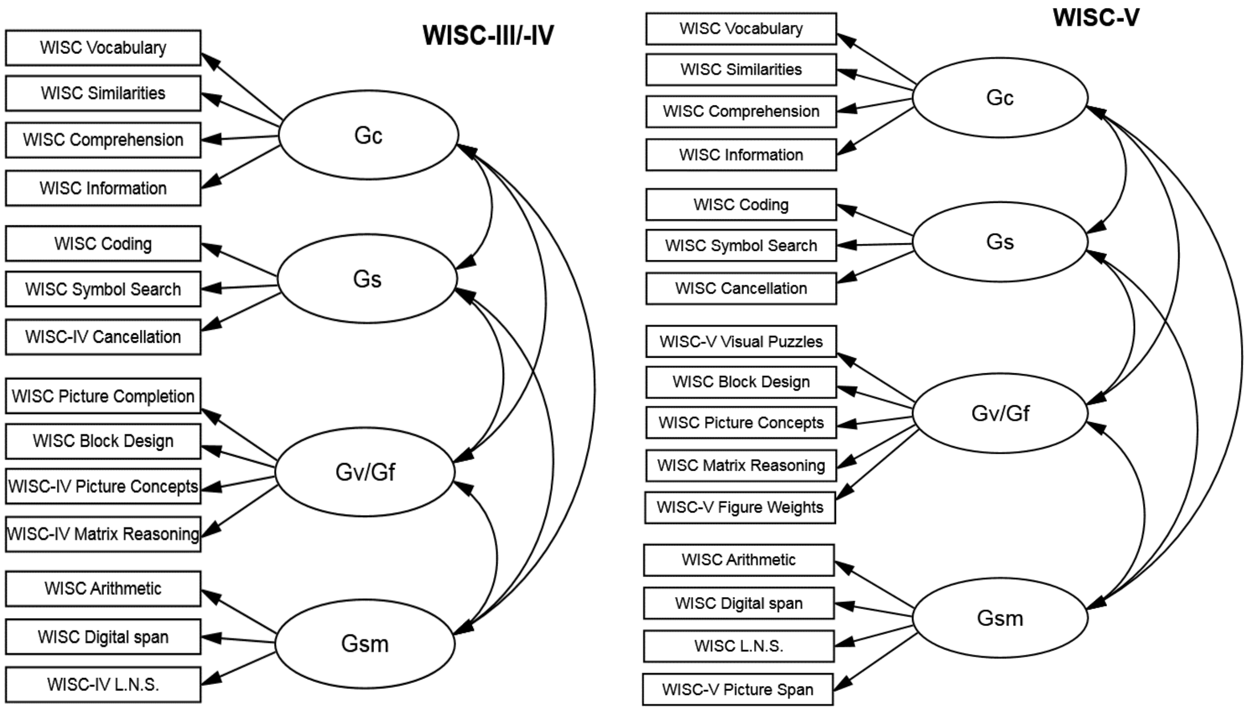

Figure 2. WISC-III/-IV and WISC-V edition invariance models

WIAT edition invariance. Lastly, invariance was tested for the merged WIAT-II data and the single sample of the WIAT-III. The same covariance invariance approach used across WIAT samples was used across editions. The two tests shared six subtests (see Figure 3); again no latent variables were included in the analyses. Subtests measuring written expression were not tested for invariance or merged because these three subtests, WIAT-II Written Expression, WIAT-III Sentence Composition, and WIAT-III Essay Composition, differed according to task content and organization. Covariance, variance, and mean invariance were all supported; the change in CFI was minimal (see Table 9). Accordingly, the six WIAT-II and WIAT-III subtests were merged; thus, data from three samples were merged. These six merged subtests are referred to as "WIAT" subtests in later analyses.

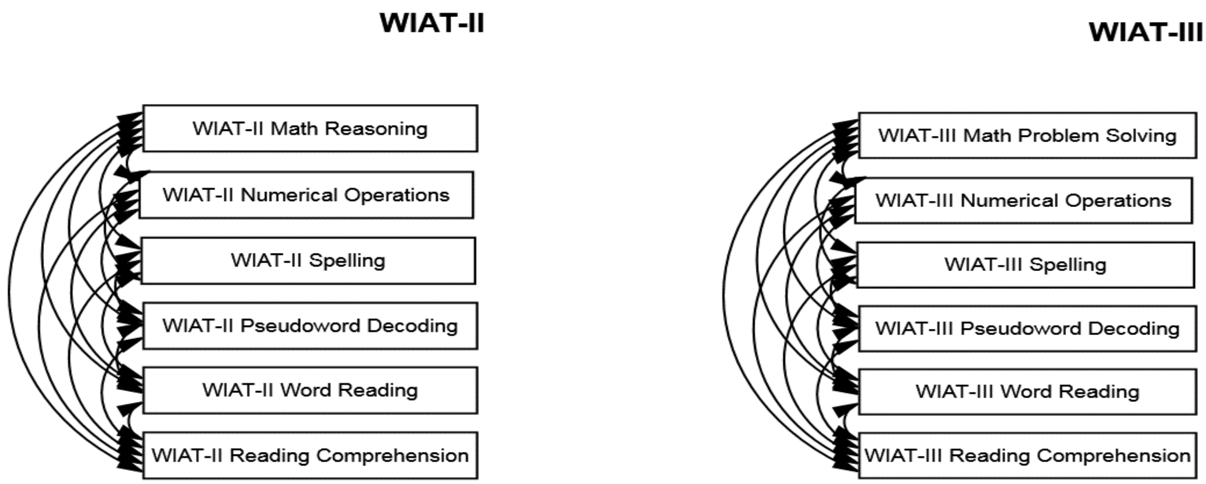

Figure 3. WIAT-II and WIAT-III edition invariance models

Table 9

Invariance Testing Across Editions

Model Name	**χ^2(df)**	***p***	**$\Delta\chi^2$ (Δdf)**	**Δp**	**CFI**	**Δ CFI**	**Adj. RMSEA**
WISC-III & -IV							
Configural Invariance	249.021(119)	.000	-	-	.971	-	.050
Metric Invariance	254.026(125)	.000	5.005(6)	.543	.972	.001	.048
Intercept Invariance	277.805(131)	.000	23.779(6)	<.001	.968	-.004	.050
Residual Invariance	319.563(141)	.000	41.758(10)	<.001	.961	-.007	.052
WISC-III/IV & -V							
Configural Invariance	323.902(169)	.000	-	-	.973	-	.037
Metric Invariance	334.838(178)	.000	10.936(9)	.280	.972	.001	.037
Intercept Invariance	344.747(187)	.000	9.909(9)	.358	.972	.000	.035
Residual Invariance	391.221(200)	.000	46.474(13)	<.001	.966	.006	.038
WIAT-II & -III							
Covariance Invariance	53.406(15)	-	-	-	.992	-	.069
Variance Invariance	56.195(21)	.000	2.789(6)	.835	.992	.000	.055
Mean Invariance	60.587(27)	.000	4.392(6)	.624	.993	.001	.048

Cognitive CB-CFA

Cognitive CB-CFA first-order model. A first-order CB-CFA with six broad ability latent variables was tested. In this first-order model each broad ability variable was correlated with all other broad ability variables. A first-order model was specified prior to testing a second-order model (one that includes *g*) in order to first establish the best fitting model that explained the relations between the subtests and the broad abilities. A priori classification of the subtests by factor is show in Figure 4: Gc was measured by 14 subtests, Gf was measured by 12 subtests, Gv was measured by 13 subtests, Gs was measured by 7 subtests, Gsm was measured by 12 subtests, and Glr was measured by 8 subtests. Six subtests (KABC-II Gestalt Closure, KABC-II Hand Movements, WISC Picture Completion, WISC Arithmetic, DAS-II Verbal Comprehension, and WJ III Picture Recognition) were cross-loaded onto two broad ability factors based on results from previous studies (Keith, Low, Reynolds, Patel, & Ridley 2010; Reynolds et al., 2013). Three correlated residual variances were included for the four KABC-II and two DAS-II subtests that included a delayed recall version of the initial measurement of the subtests (KABC-II Atlantis and Atlantis Delayed, KABC-II Rebus and Rebus Delayed, and DAS-II Recall of Objects Immediate and Recall of Objects Delayed).

Results of the initial cognitive CB-CFA model are presented in Table 10. The fit of the initial cognitive CB-CFA model was acceptable to well-fitting. The RMSEA, CFI, and TLI values were considered excellent, and the SRMR value was adequate, but slightly exceeded the good fit threshold. Almost all of the factor loadings and the three correlated variances were statistically significant. One cross-loading was not statistically significant, the WJ III Picture Recognition subtest onto Gv; this cross-loading was subsequently deleted. A reduced model without that cross-loading was tested, and the fit indices were also acceptable to well-fitting. The reduced model was compared to the initial model using the likelihood ratio test. The change in chi-square was not statistically significant, thus supporting the reduced model; the reduced model is the final first-order model (Table 10 for fit indices).

All of the factor loadings of the subtests onto their respective broad ability latent variable factors were statistically significant. The standardized factor loadings of the subtests on Gc ranged from .716 (WISC Comprehension) to .872 (WISC Vocabulary), Gf factor loadings ranged from .353 (DAS-II Picture Similarities) to .738 (DAS-II Sequential and Quantitative Reasoning), Gv factor loadings ranged from .441 (DAS-II Recognition of Pictures) to .787 (DAS-II Pattern Construction), Gs factor loadings ranged from .391 (WISC Cancellation) to .745 (WISC Symbol Search), Gsm factor loadings ranged from .600 (WJ III Auditory Working Memory) to .779 (KABC-II Word Order), and Glr factor loadings ranged from .462 (DAS-II Recall of Objects Delayed) to .790 (KABC-II Rebus Immediate; excluding subtests that were cross-loaded, as cross-loadings are expected to be smaller given loadings onto two factors rather than one factor). Overall, these results suggest the subtests are generally good measures of the six broad abilities. Thus, the cognitive CB-CFA first-order model results suggest that the six CHC broad ability factors were invariant across the six intelligence tests analyzed in this study.

In addition, all of the six broad abilities significantly correlated with each other. The strongest relation was between Gf and Gv (r = .902), followed by Gf and Gc (r = .782), Gf and Glr (r = .749), Gc and Gv (r = .669), Gc and Glr (r = .676), Gf and Gsm (r = .647), Glr and Gv (r = .643), Gc and Gsm (r = .601), Glr and Gs (r = .592), Gf and Gs (r = .577), Gs and Gv (r = .565), Gv and Gsm (r = .551), Gsm and Glr (r = .540), Gsm and Gs (r = .507), and Gs and Gc (r = .422).

Cognitive CB-CFA second-order model. The second-order model, which included *g*, expanded on the final cognitive CB-CFA first-order model. The addition of *g* resulted in model fit which was also acceptable to well-fitting (see Table 10). The inclusion of *g* caused one cross-loading to become non-significant, the DAS-II Verbal Comprehension subtest onto Gf. As a result, that cross-loading was subsequently deleted from the second-order model and all cognitive-achievement SEM models. The likelihood ratio test was not statistically significant, which supported the reduced second-order model without one cross-loading. This final second-order model was compared to the final first-order model; the change in chi-square and aBIC both supported the final first-order model, however, this finding is not unusual.

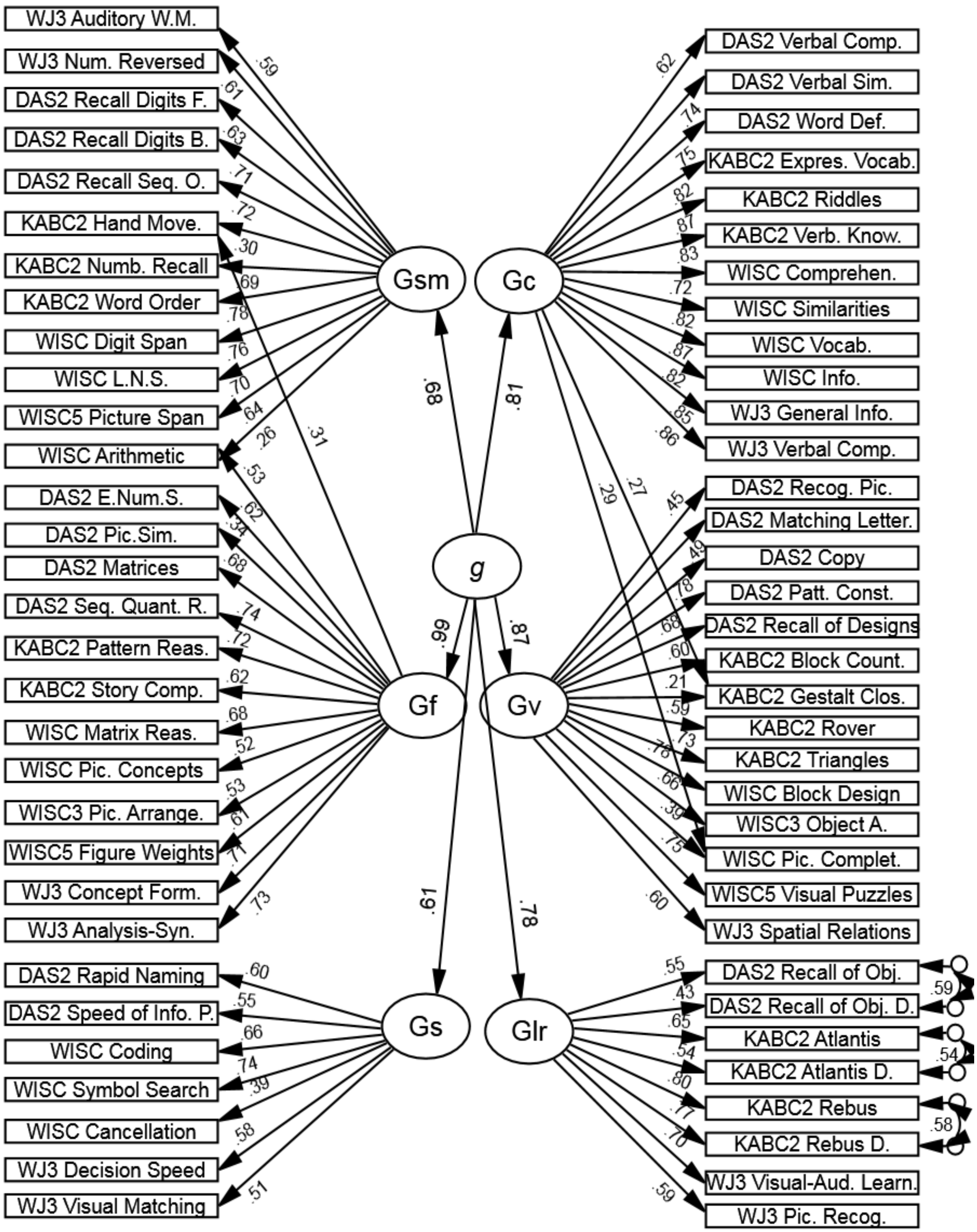

Figure 4. Final cognitive CB-CFA second-order model

Second-order loadings of the broad abilities on *g* were large and statistically significant: Gf = .993, Gv = .870, Gc = .806, Glr = .783, Gsm = .682, and Gs = .612 (see Figure 4 for the complete model). Unlike the other broad abilities, Gf's unique variance was not statistically significant from zero, which, along with Gf's very strong factor loading on *g* (β = .993), suggests that Gf and *g* were perfectly correlated and statistically indistinguishable.

Table 10

Fit Indices of CB-CFA Cognitive Models

Model Name	χ^2	***df***	$\Delta\chi^2$	Δdf	Δp	**CFI**	**TLI**	**RMSEA**	**SRMR**	**aBIC**
Initial First-Order	2496.097	1321	-	-	-	.959	.956	.015	.088	323685.134
Final First-Order	2497.078	1322	.981	1	.322	.959	.956	.015	.086	323681.017
Initial Second-Order	2631.565	1331	134.487	9	<.001	.955	.952	.016	.087	323769.624
Final Second-Order	2634.793	1332	3.228[a]	1[a]	.072[a]	.955	.952	.016	.087	323767.755
	-	-	137.715[b]	10[b]	<.001[b]	-	-	-	-	-

[a]Compared to initial second-order model.
[b]Compared to final measurement model.

Achievement CB-CFA measurement model. Auxiliary variables, which are variables that are not included in the analysis model, were included in all of the achievement measurement models. The auxiliary variables in the achievement measurement models were the cognitive subtests; auxiliary variables were used as missing data correlates in the models, along with the achievement subtests (Muthén & Muthén, 2012). Auxiliary variables were required due to the significant amount of missing data in the achievement models because participants only completed one of the achievement tests, either the WIAT or KTEA-II; no participant completed both tests.

First, a measurement model that included six specific achievement skills was tested (see Figure 5). These specific achievement skills included Basic Reading (estimated by four subtests), Reading Comprehension (estimated by two subtests), Basic Math (estimated by two subtests), Math Reasoning (estimated by two subtests), Basic Writing (estimated by two subtests), and Written Expression (estimated by four subtests). The six specific achievement latent variables were each correlated with each other.

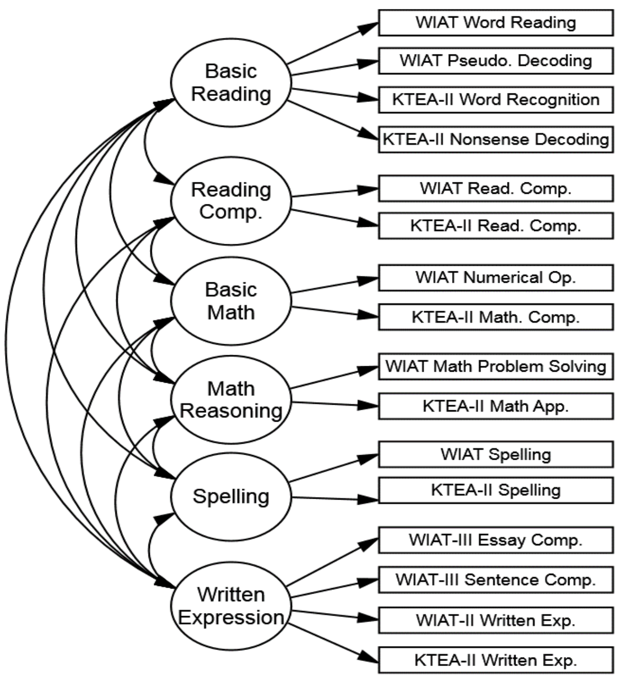

Figure 5. Initial achievement CB-CFA model

The specific achievement skills measurement model did not converge. Variations of the specific achievement skills measurement model were tested in order to specify a model that might converge. One such model included correlations between all of the subtests of each of the tests. All WIAT subtests were correlated with each other and all KTEA subtests were correlated with each other in order to account for shared variance between subtests within the same test; this model also did not converge. Another model tested broad achievement domains, with all the reading, writing, and math subtests loading onto three respective factors; this model also did not converge.

The convergence difficulties are likely due to the preponderance of missing data among the achievement variables. Unlike the cognitive variables, no individuals completed more than one achievement test; individuals either completed the WIAT or the KTEA. Also, individuals who completed the WIAT or KTEA did not complete any of the same intelligence tests. Thus, the linking between the achievement tests was far removed; achievement subtests were linked through intelligence tests that were linked to other intelligence tests. The convergence errors suggest there is likely a threshold amount of missing data that is permissible in planned missingness designs, and these achievement measurement models exceeded this threshold.

Cognitive-Achievement CB models

Because it was impossible to test a combined CB-CFA achievement measurement model across all of the achievement skills, each achievement skill was individually tested in a separate CB cognitive-achievement model. Three cognitive-achievement models were tested: (a) a cognitive-broad reading model (a combination of basic reading and reading comprehension; estimated by six subtests), (b) a broad writing model (a combination of basic writing and written expression; estimated by six subtests), and (c) a broad math model (a combination of basic math and math problem solving; estimated by four subtests).

The cognitive-broad reading model, however, did not converge. A cognitive-basic reading model (estimated by the four basic reading subtests), however, did converge; thus, the two reading comprehension subtests were excluded from the reading model. Therefore, the results for a cognitive-basic reading, cognitive-broad writing, and cognitive-broad math model are interpreted below.

Analyses were completed in a series of steps, which were described earlier, and two types of models were examined. One type of model was a broad ability only (first-order)-achievement model. The cognitive piece of the broad ability-achievement models included the final cognitive first-order model (without *g*). Broad ability only models were first tested (before including *g*) because findings from most studies suggest that *g* affects achievement through the broad abilities; this suggests *g* is not a common cause of the broad abilities and achievement, and because only common causes are needed to create a valid model, *g* was excluded (Keith, 2015). Also, as demonstrated earlier, *g* and Gf were statistically indistinguishable in the second-order model, and it would therefore be difficult to separate their influences on the achievement skills.

Another set of cognitive-achievement models were also tested. These models are referred to as *g* only-achievement models. A higher-order model was tested in these models, and the achievement skill was regressed on only one cognitive ability, *g*. The *g* only-achievement and broad ability-achievement models were then compared.

Cognitive-Basic Reading Model. The factor loadings of the four basic reading subtests (WIAT Word Reading, WIAT Pseudoword Decoding, KTEA-II Nonsense Word Decoding, KTEA-II Word Recognition) on the basic reading latent variable were statistically significant. The standardized factor loadings ranged from .803 (WIAT Pseudoword Decoding) to .987 (WIAT Word Reading), which suggests the subtests are generally good measures of basic reading.

First, a first-order broad ability model was tested which included paths from all of the broad abilities to basic reading (referred to as Reading All Broad model in Table 11). Paths from Gf, Gv, and Gs were non-significant, thus those paths were simultaneously deleted from the model. The fit of the reduced model, which included paths from Gc, Gsm, and Glr, was good to acceptable (see Table 11, referred to as Reading Final model). The reduced model did not result in a significant change in chi-square, which supports the deletion of the non-significant paths (see Table 11 for model comparisons). Next, *g* was incorporated into the reduced model, creating a second-order model, and a path from *g* to basic reading was added. The path from *g* to basic reading was not significant (b = -.089; β = -.062, SE = .056, p = .269), and according to the aBIC the broad ability only-basic reading (first-order model, without *g*) fit better than the second-order model with *g* (see Table 11 for model comparisons).

Accordingly, the broad ability only-basic reading model was interpreted (see Figure 6). The significant effects of the broad abilities on basic reading were moderate to large in size (using the criteria in Keith, 2015, chap 4). The largest standardized effect was from Gc to basic reading (b = .463; β = .400, SE = .023, p < .001), which means that each standard deviation increase in Gc resulted in a .40 standard deviation increase in Basic Reading, controlling for the other broad abilities in the model. In addition, Gsm (b = .396; β = .249, SE = .022, p < .001) and Glr had statistically significant influences on basic reading (b = .218; β = .215, SE = .027, p < .001).

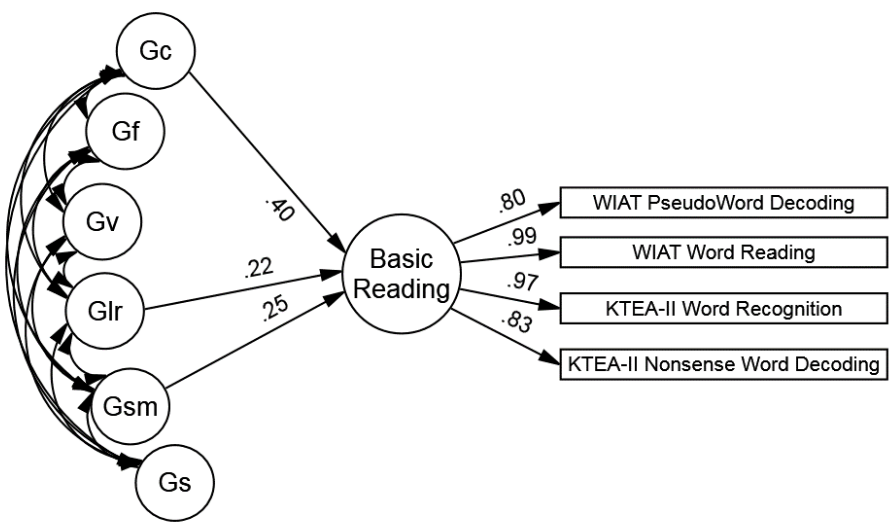

Figure 6. Cognitive-basic reading SEM model

Finally, a second-order model was tested, which only included a path from *g* to basic reading. The fit of the *g* only-basic reading model ranged from poor (RMSEA = .090) to good (CFI = .952; see Table 11 for all fit indices). The path from *g* to basic reading, however, was significant and large (b = 1.022; β = .746, SE = .011, p < .001). According to the aBIC, in comparison to the first-order broad ability-basic reading model, the *g* only-basic reading model fit worse. The *g* only model was also compared to the higher-order model with paths from the significant broad abilities and without a path from *g* (referred to as the Reading SO Final model in Table 11). The change in chi-square was statistically significant, which supports the relations between the broad abilities and basic reading in the higher-order model.

Cognitive-Broad Writing Model. The factor loadings of the six broad writing subtests (WIAT-III Essay Composition, WIAT-III Sentence Composition, WIAT Spelling, WIAT-II Written Expression, KTEA-II Spelling, KTEA-II Written Expression) onto the broad writing latent variable factor were statistically significant. The standardized factor loadings ranged from .515 (WIAT-III Essay Composition) to .842 (KTEA-II Written Expression), which suggests the subtests are generally good measures of broad writing.

Initially, a first-order broad ability model that included paths from all of the broad abilities to broad writing was tested (referred to as Writing All Broad model in Table 11). The only non-significant path was from Gv; this path was subsequently deleted. The reduced model resulted in a significant change in chi-square (see Table 11 for model comparisons), but the path from Gv to broad writing was negative in the initial model, which is uninterpretable. Therefore, the reduced model without Gv was accepted. In the reduced model, the path from Gf to broad writing became non-significant so this path was also subsequently deleted. The two reduced models were compared using the likelihood ratio test; the reduced model without paths from Gf and Gv did not result in a significant change in chi-square, which supports the removal of the Gf path (see Table 11 for model comparisons). The fit of this reduced model, which included paths from Gc, Glr, Gs, and Gsm, was good to acceptable (see Table 11). Next, a higher-order model was tested that incorporated *g* and added a path from *g* to broad writing. The path from *g* to broad writing was not significant (b = -.015; β = -.013, SE = .069, p = .851), and according to the aBIC, the broad ability only-broad writing (first-order model, without *g*) fit better than the second-order model.

Thus, the broad ability only-broad writing model was interpreted (see Figure 7). The significant effects of the broad abilities on broad writing were moderate to large in size (using the criteria in Keith, 2015, chap 4). The largest standardized effect was from Gc to broad writing (b = .276; β = .287, SE = .026, p < .001), which means that each standard deviation increase in Gc resulted in a .29 standard deviation increase in broad writing, controlling for the other broad abilities in the model. In addition, Glr (b = .240; β = .285, SE = .032, p < .001), Gs (b = .375; β = .228, SE = .040, p < .001), and Gsm (b = .293; β = .222, SE = .025, p < .001) significantly predicted broad writing.

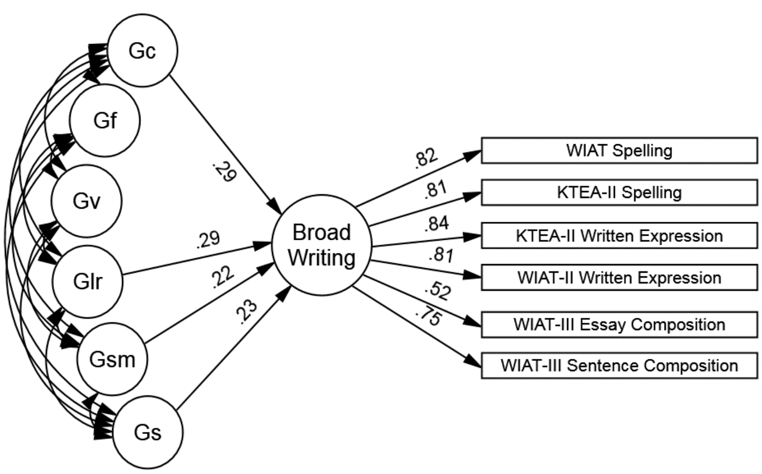

Figure 7. Cognitive-broad writing SEM model

Finally, a second-order model was tested, which only included a path from *g* to broad writing. The fit of the *g* only-broad writing model ranged from adequate to good (see Table 11 for all fit indices). The path from *g* to broad writing was significant and large ($b = .973$; $\beta = .815$, SE $= .011$, $p < .001$). According to the aBIC, in comparison to the final broad ability only-broad writing model, the *g* only-broad writing model fit worse. The *g* only model was also compared to the higher-order model with paths from the significant broad abilities and without a path from *g* (referred to as the Writing SO Final model in Table 11). The change in chi-square was statistically significant, which supports the relations between the broad abilities and broad writing in the higher-order model.

Cognitive-Broad Math Model. The factor loadings of the four broad math subtests (WIAT Math Problem Solving/Reasoning, WIAT Numerical Operations, KTEA-II Math Applications, KTEA-II Math Computations) onto the broad math latent variable factors were statistically significant. The standardized factor loadings ranged from .774 (KTEA-II Math Computation) to .937 (WIAT Math Problem Solving), which suggests the subtests are generally good measures of broad math.

First, a first-order broad ability model that included paths from all of the broad abilities to broad math was tested (referred to as Math All Broad model in Table 11). Paths from Gv, Gs, and Glr were non-significant; thus, those paths were simultaneously deleted from the model. The fit of the reduced model, which included paths from Gf, Gc, and Gsm, was good to acceptable (see Table 11); the reduced model did not result in a significant change in chi-square, which supports the removal of the non-significant paths (see Table 11 for model comparisons). Next a higher-order model was tested that incorporated *g* and added a path from *g* to broad math. The path from *g* to broad math was not significant ($b = 1.876$; $\beta = 1.504$, $SE = 2.282$, $p = .510$), and according to the aBIC, the broad ability only-broad math (first-order model, without *g*) fit better than the second-order model (see Table 11 for model comparisons). Of note, the effects of Gf and *g* on broad math in the second-order model were impossible to disentangle as a result of the perfect correlation between Gf and g. In other models, Gf had no effect on achievement areas, but for broad math the effect of Gf was significant. Although *g* had no effect on broad math beyond that of Gf, because these two constructs are virtually inseparable, these findings can be interpreted as the effects of *g* on broad math, or as the effects of Gf on broad math.

Nevertheless, because the effect of *g* on broad math was not significant, the broad ability only-broad math model was interpreted (see Figure 8). The significant effects of the broad abilities on broad math ranged from small to large in size (using the criteria in Keith, 2015, chap 4). The largest standardized effect was from Gf to Broad Math ($b = .879$; $\beta = .705$, $SE = .038$, $p < .001$), which means that each standard deviation increase in Gf resulted in a .71 standard deviation increase in broad math, controlling for the other variables in the model. In addition, Gc ($b = .136$; $\beta = .133$, SE = .032, $p < .001$) and Gsm ($b = .075$; $\beta = .053$, $SE = .025$, $p < .001$) significantly predicted broad math.

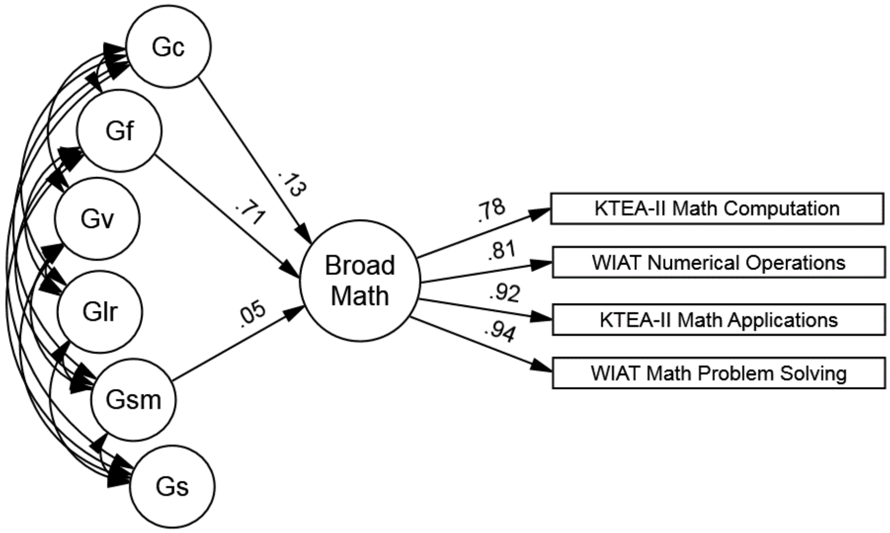

Figure 8. Cognitive-broad math SEM model

Finally, a second-order model was tested, which only included a path from *g* to broad math. The fit of the *g* only-broad math model ranged from adequate to good (see Table 11 for all fit indices). The path from *g* to broad math was significant and large (b = 1.094; β = .869, SE = .008, p < .001). According to the aBIC, in comparison to the broad ability only-broad math model, the *g* only-broad math model fit worse. The *g* only model was also compared to the higher-order model with paths from the significant broad abilities and without a path from *g* (referred to as the Math SO Final model in Table 11). The change in chi-square was statistically significant, which supports the relations between the broad abilities and broad math in the higher-order model.

Table 11

Fit Indices of Cognitive-Achievement Models

Model Name	**$\chi^2(df)$**	**$\Delta\chi^2$ (Δdf)**	**Δp**	**CFI**	**TLI**	**RMSEA**	**SRMR**	**aBIC**
Reading All Broad	2781.158(1467)	-	-	.962	.959	.015	.088	370728.813
Reading Broad Final	2781.589(1470)	.431(3)	.934	.962	.959	.015	.088	370713.949
Reading SO+*g* path	2916.540(1478)	-	-	.958	.955	.016	.089	370808.115
Reading SO Final	2953.995(1479)	37.455(1)	<.001	.957	.954	.016	.087	372907.797
Reading *g* only	3135.560(1481)	181.565(2)	<.001	.952	.949	.017	.090	371011.841
Writing All Broad	2815.943(1497)	-	-	.960	.957	.015	.087	373581.318
Writing (no Gv)	2824.551(1498)	8.608(1)	.003	.960	.957	.015	.087	373584.828
Writing Final	2827.396(1499)	2.845(1)	.092	.960	.957	.015	.087	373582.574
		11.453(2)[a]	.003 [a]					
Writing SO+*g* path	2954.271(1507)	-	-	.956	.953	.016	.088	373668.665
Writing SO Final	2954.307(1508)	.036(1)	.850	.956	.954	.016	.088	373663.603
Writing *g* only	3091.326(1511)	137.019(3)	<.001	.952	.949	.016	.088	373785.327
Math All Broad	2808.625(1467)	-	-	.961	.958	.015	.087	372823.605
Math Broad Final	2810.780(1470)	2.155(3)	.541	.961	.958	.015	.087	372810.465
Math SO+*g* path	2953.120(1478)	-	-	.957	.954	.016	.087	372912.020
Math SO Final	2953.995(1479)	.875(1)	.350	.957	.954	.016	.087	372907.797
Math *g* only	2971.708(1481)	17.713(2)	<.001	.957	.954	.016	.087	372915.314

Note. [a] The writing final (broad ability only-broad writing) model was first compared to the model without Gv, and then compared to the model with paths from all broad abilities.

Chapter 5: Discussion

The primary purpose of this study was to examine cognitive-achievement relations across several tests using cross-battery SEM analyses and planned missing data methodology. A more comprehensive, cross-battery understanding of cognitive-achievement relations can address inconsistencies found in previous studies that were caused by examining these relations separately for individual tests. Generalized cognitive-achievement relations, across tests, may guide practitioners' specific learning disability diagnostic decisions, eligibility determinations, educational recommendations, and assessment planning, as well as provide empirical support to the current practice of cross-battery assessment, in which practitioners administer multiple intelligence and achievement tests to students and interpret the results conjointly.

Several additional purposes were subsumed within this overarching purpose, and were necessary precursors before addressing the overarching purpose. Invariance across different samples of participants who completed the same test was evaluated. After sample invariance was established, invariance across different editions of the same test was established in order to ensure equivalent constructs were measured across editions and allow for the merging of data across the different samples and editions. Next, CB-CFA intelligence models were tested to determine if six intelligence tests measured the same cognitive abilities similarly, which provided further support for CHC theory independent of the test under study. Finally, after the precursory steps of invariance across samples and editions and a comprehensive CHC model of intelligence were supported, CB-SEM basic reading, broad math, and broad writing models were tested.

The organization of this section follows a similar sequence as the analysis sequence described above. First, theoretical and methodological implications related to cross-battery intelligence models are presented. Second, cognitive-achievement relations are compared and contrasted to previous research. Third, implications for practice are discussed; fourth, limitations and future research directions are discussed; and finally, the section concludes with a brief summary.

Cross-Battery Models

Theoretical implications. A cross-battery CHC model consisting of six broad abilities, including Gc, Gf, Gv, Gsm, Glr, and Gs, fit data from six different intelligence tests well. The factor loadings of all six broad abilities on *g* were large. Results suggest Gf had the strongest loading on g, specifically there was a perfect correlation between the two abilities which will be discussed further below. Gv had the second strongest loading on *g*, followed by Gc, Glr, Gsm, and finally Gs. At the subtest level, the majority of the subtests loaded on the broad abilities in accordance with prior CHC classifications (Flanagan et al., 2013). One exception was the WJ III Visual Recognition subtest, which was classified as a Gv subtest, but was found only to significantly load on Glr; these results replicate those based on the other CB-CFA study (Reynolds et al., 2013). Taken together, the loadings of the broad abilities on *g* and the consistent loadings of the subtests on the broad abilities in accordance with a priori CHC classifications supports the applicability of CHC theory across intelligence tests and suggests practitioners can be confident in CHC classifications of these different tests.

An advantage of the current CB-CFA CHC model is the inclusion of two currently used tests, the DAS-II and WISC-V, and one additional broad ability, Gs, which extends the CB-CFA model presented by Reynolds and colleagues (2013). This larger CB-CFA CHC model provides further evidence for the applicability of CHC theory to the development of modern intelligence tests, CHC-based interpretation of test results from these six tests, and cognitive research guided by CHC theory (Reynolds et al., 2013). Despite differences across tests in regard to subtest task demands, stimuli, and response formats, these six intelligence tests are measuring the CHC broad abilities similarly, and thus, practitioners and researchers can assume broad ability scores from any of these six intelligence tests are measuring the same underlying cognitive ability. On a practice focused note, the invariance of these six CHC broad abilities across the six intelligence tests provides empirical support for the cross-battery assessment approach, which encourages practitioners to supplement subtest scores from one intelligence test with scores from the same or different broad ability on another intelligence test.

Another important theoretical implication taken from the CB-CFA intelligence model is the perfect correlation between Gf and *g*; the loading of Gf on *g* was .99 and Gf's residual was non-significant. This perfect correlation between Gf and *g* is supported by previous research (Reynolds et al., 2013; Gustafasson & Balke, 1993), and suggests Gf and *g* constructs are redundant and may be used interchangeably. This interchangeable relationship suggests subtests designed to measure Gf may also be considered primarily as measures of g (Reynolds et al., 2013). Alternatively, these findings suggest that a hierarchical *g* factor may be unnecessary in intelligence models. The perfect relation between Gf and *g* raises questions about the structure of intelligence and which abilities are redundant (Reynolds et al., 2013, Gustafasson & Balke, 1993). Further research is needed to better understand the overlap between Gf and *g*.

Methodology implications. The current study is the first known study to examine a cross-battery intelligence model and cross-battery cognitive-achievement relations using several datasets, which did not include one shared linking test as is common in planned missingness designs; a shared linking test is a test that all participants complete and is used in the well-established three-form planned missingness design. An alternative planned missingness design was explored in the current study, and rather than one shared linking test, the seven datasets were linked to each other through various configurations of tests they shared in common; there was no single direct link between all of the datasets. Such alternative planned missingness designs have been proposed as a possibility by previous researchers (Graham et al., 2006; Reynolds et al., 2013), but have been rarely studied. Results from the current study, however, provide support for the use of an alternative planned missingness design, which does not require one shared linking test that all participants completed. These results are encouraging for researchers because they suggest researchers can merge several datasets without the inclusion of a single test that all participants completed. Eliminating the need for a shared linking test in planned missingness designs may reduce examinee test fatigue, data collection demands, and encourage the merging of several different invariant datasets, which may result in larger sample sizes of merged data and allow for analyses that were not previously possible.

Cognitive-Achievement Relations

Results from the current study provide evidence of cognitive-achievement relations across tests, which is lacking in the field, and indicates which relations are generalizable across different tests. Broadly, the current results suggest verbal-comprehension (Gc) was the only broad ability to influence significantly all three academic skills in this study; both short-term memory (Gsm) and long-term retrieval (Glr) influenced basic reading and broad writing; processing speed (Gs) influenced broad writing and broad math; fluid reasoning (Gf) only exerted a significant effect on broad math; and visual-spatial processing (Gv) did not influence any of the three academic skills examined in the current study, controlling for the other variables in the model. Below, these cross-battery cognitive-achievement results are compared and contrasted with previous research which has examined cognitive-achievement relations based on single cognitive and achievement test pairings.

Basic Reading. In the current study, verbal-comprehension (Gc) was most strongly related to children and adolescent's basic reading performance. Children and adolescents with stronger acquired knowledge of vocabulary, including language and cultural knowledge, scored higher on a latent factor of word reading and pseudoword decoding tasks. The strength of the relation between Gc and basic reading is a consistent finding across different tests (Benson, 2007; Caemmerer et al., 2017; Evans et al., 2002; Flanagan, 2000; Floyd et al., 2012; Keith, 1999; McGrew, 1993; McGrew et al., 1997; Niileksela et al., 2016; Oh et al., 2004; Vanderwood et al., 2002). Similarly, the effect of short-term memory (Gsm) on basic reading, a relation observed across several different tests, was also supported in the current study (Beaujean et al., 2014; Benson, 2007; Caemmerer et al., 2017; Cormier, McGrew et al., 2016; Elliot, Hale, Fiorello, Dorvil, & Moldovan, 2010; Evans et al., 2002; Floyd et al., 2007; Hajovsky et al., 2014; McGrew, 1993). Better developed Gsm abilities may allow children and adolescents to hold phonological information in their minds and manipulate that information to more accurately identify words and non-words (even after controlling for other relevant cognitive abilities). In addition, the relation between long-term retrieval (Glr) and basic reading observed in the current study is supported by some evidence (Evans et al., 2002; Floyd et al., 2007; Floyd et al., 2012; Hajovsky et al., 2014; McGrew, 1993), although this relation was often only observed among younger children in previous studies. The current findings suggest children and adolescents with stronger abilities to store, consolidate, and retrieve information efficiently are better able to identify words and recognize non-word patterns. Unlike past research with the WJ batteries, Gs did not exert a significant effect on basic reading skills (Evans et al., 2002; Floyd et al., 2007; McGrew, 1993; Niileksela et al., 2016); however, the lack of a significant relation between Gs and basic reading is supported by previous research using the WISC-IV and WISC-V (Beaujean et al., 2014; Caemmerer et al., 2017). These discrepant Gs findings suggest there may be something specific about the Gs and basic reading measures of the WJ, which results in a significant relation, and practitioners using the WJ tests may anticipate such a relation, but practitioners using other tests may be less concerned with a Gs-basic reading relation.

Broad Writing. Gc had the strongest influence on children and adolescents' broad writing performance. Children and adolescents with stronger acquired knowledge of vocabulary, including language and cultural knowledge, scored higher on a writing composite that included spelling (basic writing skill) and more complex sentence and essay writing (written expression skills) tasks. Although writing is the least studied achievement domain, the majority of previous studies support a relation between Gc and broad writing or specific writing skills (Beaujean et al., 2014; Caemmerer et al., 2017; Floyd et al., 2008; McGrew & Knopik, 1991). In addition, Gsm had a significant effect: Children and adolescents who were better able to hold and manipulate information in their minds spelled more sounds and words correctly and were better able to compose sentences and essays; this finding is consistent with relations observed in several studies using the WJ (Cormier, Bulut et al., 2016; Floyd et al., 2008; McGrew & Knopik, 1993; Niileksela et al., 2016) and one WISC-V study (Caemmerer et al., 2017). Similarly, processing speed (Gs) had a significant effect on broad writing performance, which is consistent with previous research using the WJ (Cormier, Bulut et al., 2016; Floyd et al., 2008; McGrew & Knopik, 1993; Niileksela et al., 2016). A Gs-writing relation was not observed using the WISC-V and WIAT-III (Caemmerer et al., 2017). There may be two possible explanations for this inconsistency: First, the current study examined broad writing rather than specific writing skills, or second, the Gs factor as measured solely by the WISC-V is too narrowly defined in comparison to the Gs factor examined here. In addition, the significant effect shown by Glr on broad writing is consistent with previous research using the WJ (Floyd et al., 2008; McGrew & Knopik, 1991; Niileksela et al., 2016), and it suggests retrieval of words, previously learned information, and ideas has a significant influence on writing performance. Finally, no significant relation was observed between Gf and broad writing in the current study, however, some evidence using older versions of the WJ, the WISC-IV and WISC-V suggests Gf is important for writing performance (Beaujean et al., 2014; Caemmerer et al., 2017; Floyd et al., 2008; McGrew & Knopik, 1991). It is possible the influence of Gf on writing is narrowly focused and Gf influences only specific writing skills, such as more complex essay writing tasks; thus this relation was not observed when examining a broad writing factor.

Broad Math. Broad math performance was most strongly influenced by fluid reasoning (Gf); children and adolescents with stronger novel reasoning abilities scored higher on a latent broad math factor indexed by mathematic computation and multi-step math word problem tasks. The importance of Gf for math skills is highly consistent with findings based on the WJ (Floyd et al., 2003; Keith, 1999; McGrew & Hessler, 1995; Niileksela et al., 2016; Taub et al., 2008), WISC-IV, and WISC-V (Caemmerer et al., 2017; Parkin & Beaujean, 2012). In addition, Gs predicted broad math performance, which is consistent with previous WJ (Floyd et al., 2003; Keith, 1999;

McGrew & Hessler, 1995; Niileksela et al., 2016; Taub et al., 2008) and WISC-V (Caemmerer et al., 2017) studies. The importance of Gc is consistent with previous WJ studies (Floyd et al., 2003; Keith, 1999; McGrew & Hessler, 1995; Niileksela et al., 2016; Taub et al., 2008), and suggests general acquired knowledge is important for math performance. A significant relation between specific math skills and Gc was not found using the WISC-V (Caemmerer et al., 2017), however, which may be due to specific aspects of the WISC-V and WIAT-III; thus, practitioners using only the WISC-V and WIAT-III may not be concerned with such a relation, but practitioners using other tests may anticipate a significant Gc-math relation. Finally, there was no significant association between Gsm and math in the current study, which is consistent with the majority of previous studies (Niileksela et al., 2016; Parkin & Beaujean, 2012; Taub et al., 2008). Contrasting evidence suggests Gsm is important for the math performance of younger students (McGrew & Hessler, 1995; Caemmerer, et al., 2017). More research is needed to clarify these discrepancies.

In sum, the results of the current study align with previous studies which suggest children's and adolescents' Gc abilities are important for the majority of academic skills, including basic reading, broad writing, and broad math performance. Similar to previous research, Gsm and Glr have important influences on children's and adolescents' basic reading performance, but unlike previous research, there was no evidence to support a relation between Gs and basic reading. Gsm, Glr, and Gs all had important effects on broad writing performance, but some evidence supports a relation between writing skills and Gf, which was not observed here. Finally, Gs and Gf influenced children's and adolescents' broad math performance, but no relation was observed with Gsm, which was found in a couple of studies.

Taken together, many of the previously observed cognitive-achievement relations were replicated in the current study. Inconsistencies in cognitive-achievement relations across studies, however, may be attributed to several reasons: (1) specific and unique task demands present within individual tests, (2) whether broad or specific achievement skills were examined, or (3) differences in how intelligence was modeled, such as the higher-order models used in the current study or bifactor models, which remove the influence of *g* from the broad abilities (bifactor models are discussed further below). Despite some differences between findings from the current study and previous studies, the cognitive-achievement relations observed here provide further support for the important role CHC broad abilities have on achievement skills. Results from the alternative cognitive-achievement models tested in this study support this assertion and suggest the influence of *g* on achievement skills primarily operates indirectly, through *g*'s influence on the broad abilities.

Further Implications for Practice

The cognitive-achievement patterns described in this study can be used to guide the interpretation of psychological assessment results and inform diagnostic decision-making regarding specific learning disabilities, educational recommendations, and assessment planning. Children's overall and broad cognitive abilities can explain strengths and weaknesses they exhibit in specific achievement skills. For example, assume that a child was referred to a psychologist for an evaluation because he or she was struggling in math. The psychologist might particularly focus on the child's fluid reasoning performance and would likely give the child additional fluid reasoning measures to better understand his or her abilities in this area. When interpreting the child's assessment results, findings of below average math problem solving skills accompanied with a relative weakness in the child's fluid reasoning abilities would likely suggest a potential specific learning disability in math problem solving (assuming other information was consistent with this possible diagnosis).

Another example is a child who is referred due to reading difficulties. Based on the results of this study, the psychologist might decide to give the child additional verbal comprehension subtests, given the strong relation between verbal comprehension and reading. If the child scored below average on a standardized basic reading test, and had lower verbal comprehension abilities, the psychologist may suspect a specific learning disability in basic reading. Other data would of course be considered when making these diagnostic decisions, such as the child's classroom performance, teacher and parent reports, school and medical records.

In addition to informing diagnostic decision making, the results from this study may inform educational recommendations and test planning decisions. The child's cognitive intrapersonal strengths and weaknesses may determine which accommodations are recommended. For example, if a child has a learning disability in writing and a relative weakness in Gsm, reminders regarding writing conventions or the structure of paragraphs or essays may be recommended. Another accommodation may include breaking the writing assignment into smaller steps in order to support the child's working memory abilities.

For evaluation planning purposes, if one or more specific academic skills are the referral concern, psychologists may select tests and subtests that assess the cognitive abilities relevant to the academic skill and spend less time and effort on the assessment of cognitive abilities that lack evidence to support their importance for the particular academic skill. Use of the cognitive-achievement relations found in this study to guide test planning may be particularly relevant when psychologists are using more than one intelligence test, which is encouraged in the cross-battery assessment approach. These test planning decisions may be important for initial evaluations of children, as well as re-evaluations, which are currently required every three years in schools according to Individuals with Disabilities Educational Act (IDEA) guidelines. Re-evaluations provide updated testing for children who qualify for special education services. Selective testing which focuses only on significant cognitive-achievement relations that are relevant for the children's particular disability can reduce the instructional time children miss for testing and reduce the testing demands on school psychologists.

Limitations and Future Research

The findings of this study need to be considered within the context of the study's limitations. One limitation is findings are limited to the specific tests included in this study, and may not be generalizable to other tests used by psychologists, but not included in the study. Due to the specific tests included in this study auditory processing, (Ga) was excluded from the analysis because the WJ III was the only test to include measures of Ga, and additional measures were needed from other tests for a cross-battery analysis. In order to explore the cross-battery influence of Ga on academic skills, future research may include other measures, such as the newest edition of the WJ, 4th edition, and non-cognitive measures of Ga, such as the Comprehensive Test of Phonological Processing, Second Edition (CTOPP-2). Other non-cognitive measures, such as executive functioning tests, can also be used to supplement cognitive tests in future studies in order to more comprehensively predict achievement, given findings that suggest these neuropsychological measures fit well within the CHC taxonomy (Floyd et al., 2010; Jewsbury et al., 2016; Salthouse, 2005).

Another limitation concerns the planned missing design as it applied to the achievement data. Because participants only completed either the KTEA or WIAT, and no one completed both tests, it was not possible to establish a CB-CFA achievement measurement model that best fit the data. Thus, although one common linking test that all participants completed is unnecessary, it seems likely that at least partial overlap in which some portion of participants completed more than one of the measures under study is required in order to effectively use a planned missing data design. Future research can incorporate more achievement data samples in order to explore an achievement measurement model, and test potential cross-loadings of achievement subtests or relations between achievement skills. Relatedly, the non-overlapping achievement test data in the current study resulted in analysis difficulties due to convergence issues; these analysis difficulties meant it was impossible to study cognitive influences on several specific achievement skills, including reading comprehension, basic writing, written expression, basic math, and math problem solving skills. As a result children's and adolescents' writing and math performance was studied at the broader domain level. This collapsed the cognitive effects into broader influences and may have masked some of the nuanced specific influences which vary according to specific achievement skills; relations may exist for these specific skills that were not evidenced in the current study. For example, some research suggests Gf significantly predicts written expression, but not basic writing skills (Caemmerer et al., 2017; Cormier, Bulut et al., 2016); this specific relation was not examined in the current study. Future research can address this limitation in several ways. Paths from cognitive abilities to particular achievement subtests, in addition to paths to the latent achievement variable, can test potential differential relations between cognitive abilities and broad and specific achievement skills. Future research can also incorporate additional achievement data in order to address the preponderance of missing data in the current study and allow for a more focused examination of specific achievement skills.

Another limitation is related to factors that may influence cognitive-achievement relations, but were not accounted for in the current study. Previous research suggests developmental differences exist in cognitive-achievement relations (Niileksela et al., 2016; Taub et al., 2008). The strength of cognitive-achievement relations may change across development, and although a broad ability may not exert a significant effect across all ages, it may exert a significant effect at particular ages only. Future research should examine cross-battery cognitive-achievement across ages in order to address that gap in the literature. In addition to developmental differences, previous research suggests gender differences exist in some achievement skills (Scheiber, Reynolds, Hajovsky, & Kaufman, 2015), and thus potential cognitive-achievement differences between gender should be explored in future research. Finally, it is unclear whether these cognitive-achievement relations are generalizable across different racial and ethnic groups (Garcia & Stafford, 2000; Keith, 1999) and across individuals with and without disabilities (Niileksela & Reynolds, 2014). These considerations warrant further study.

Finally, this study was guided by CHC theory and thus a higher-order model of intelligence was analyzed. There is currently a debate in the school psychology literature, however, regarding whether a higher-order or bifactor model is more appropriate when modeling cognitive data. In a bifactor model *g* does not subsume the broad abilities, instead the intelligence subtests are influenced directly by both *g* and the unique effects of the broad abilities (with the effect of *g* statistically removed from the broad abilities). The magnitude of the cognitive-achievement relations would likely vary dependent on the cognitive model under study; it is likely that a bifactor model would show weaker broad ability effects than would a higher-order model because a bifactor model partials out the effects of *g* from the broad abilities. Future research is needed to clarify differences in the two models.

Summary

Cross-battery research has several implications for practice and research associated with cognitive and achievement tests. An adequately fitting cross-battery cognitive model that combines subtests and samples from six different tests supports the applicability of CHC theory to the development and interpretation of modern intelligence tests. Cross-battery cognitive-achievement relations demonstrate which relations are generalizable across tests, and which may be specific to particular tests. The cross-battery cognitive-achievement findings may inform psychologists' diagnostic decision making regarding specific learning disabilities, assessment planning, and provide empirical support for the practice of administering more than one cognitive test to children and adolescents, known as the cross-battery assessment approach. Finally, the planned missingness methodology used as part of these cross-battery analyses suggests researchers may benefit from planned missingness designs, including those that do not include a single common linking test completed by all participants.

References

Beaujean, A. A., Parkin, J., & Parker, S. (2014). Comparing Cattell–Horn–Carroll factor models: Differences between bifactor and higher order factor models in predicting language achievement. *Psychological Assessment*, *26*(3), 789–805.

Benson, N. (2007). Cattell–Horn–Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment*, *26*(1), 27–41. doi: 10.1177/0734282907301424

Breaux, K. C. (2010). *WIAT-III technical manual with adult norms*. Bloomington, MN: Pearson.

Burns, M. K., Appleton, J. J., & Stehouwer, J. D. (2005). Meta-analytic review of responsiveness-to-intervention research: Examining field-based and research-implemented models. *Journal of Psychoeducational Assessment*, *23*, 381-394.

Caemmerer, J.M., Maddocks, D.L.S., Keith, T.Z., & Reynolds, M.R. (2017). *Effects of cognitive abilities on child and youth academic achievement: Evidence from the WISC-V and WIAT-III*. Manuscript submitted for publication.

Calhoon, M. B., Emerson, R. W., Flores, M., & Houchins, D. E. (2007). Computational fluency performance profile of high school students with mathematics disabilities. *Remedial and Special Education*, *28*(5), 292-303.

Colom, R., Karama, S., Jung, R. E., & Haier, R. J. (2010). Human intelligence and brain networks. *Dialogues in Clinical Neuroscience, 12*(4), 489–501. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21319494%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3181994

Compton, D. L., Fuchs, L. S., Fuchs, D., Lambert, W., & Hamlett, C. (2012). The cognitive and academic profiles of reading and mathematics learning disabilities. *Journal of Learning Disabilities*, *45*, 79–95. http://doi.org/10.1177/0022219410393012

Dowdy, E., Mays, K.L., Kamphaus, R.W., & Reynolds, C.R. (2009). Roles of diagnosis and classification in school psychology. In C.R. Reynolds & T.B. Gutkin (Eds.), *The handbook of school psychology* (pp. 191–209). New York, NY: Wiley.

Enders, C. K. (2010). *An Introduction to Missing Data*. *Applied Missing Data Analysis*. New York, New York: Guilford

Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457.

Elliot, C. D., Hale, J.B., Fiorello, C.A., Dorvil, C., & Moldovan, J. (2010). Differential Ability Scales-II prediction of reading performance: Global scores are not enough. *Psychology in the Schools*, *47*, 698–720. doi: 10.1002/pits.20499

Evans, J. J., Floyd, R. G., McGrew, K. S., & Leforgee, M. H. (2002). The relations between measures of Cattell–Horn–Carroll (CHC) cognitive abilities and reading achievement during childhood and adolescence. *School Psychology Review, 31*, 246–262.

Flanagan, D. P. (2000). Wechsler-based CHC cross-battery assessment and reading achievement: Strengthening the validity of interpretations drawn from Wechsler test scores. *School Psychology Quarterly, 15*(3), 295–329. doi: 10.1037/h0088789

Flanagan, D. P., & McGrew, K. S. (1998). Interpreting intelligence tests from contemporary Gf-Gc theory. *Journal of School Psychology*, *36*(2), 151–182. doi: 10.1016/S0022-4405(98)00003-X

Flanagan, D.P, Ortiz, S.O., & Alfonso, V.C. (2013). *Essentials of cross-battery assessment: Third edition*. Hoboken, NJ: Wiley.

Floyd, R. G., Bergeron, R., Hamilton, G., & Parra, G. R. (2010). How do executive functions fit with the Cattell–Horn–Carroll model? Some evidence from a joint factor analysis of the Delis–Kaplan Executive Function System and The Woodcock–Johnson III Tests of Cognitive Abilities. *Psychology in the Schools, 47*, 721–738. doi: 10.1002/pits

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. doi: 10.1002/pits.10083

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

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, 132-144. doi: 10.1002/pits.20284

Floyd, R. G, Meisinger, E., Gregg, N., & Keith, T.Z. (2012). An explanation of reading comprehension across development using models from Cattell-Horn-Carroll theory: Support for integrative models of reading. *Psychology in the Schools*, *49*, 725–743. doi: 10.1002/pits.21633

Fuchs, L. S., Fuchs, D., Stuebing, K., Fletcher, J. M., Hamlett, C. L., & Lambert, W. (2008). Problem solving and computational skill: Are they shared or distinct aspects of mathematical cognition?. *Journal of educational psychology*, *100*, 30.

Fuchs, L. S., Geary, D. C., Fuchs, D., Hamlett, C. L., & Bryant, J. D. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abitilies? *Developmental Psychology*, *46*, 1731–1746. http://doi.org/10.1037/a0020662.Do

Garcia, G. M., & Stafford, M. E. (2000). Prediction of reading by Ga and Gc specific cognitive abilities for low-SES white and hispanic English-speaking children. *Psychology in the Schools*, *37*, 227–235.

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology*, *74*, 213–239. http://doi.org/10.1006/jecp.1999.2515

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, *78*, 1343–1359. http://doi.org/10.1111/j.1467-8624.2007.01069.x

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, *13*(1), 1–4. doi:10.1111/j.0963-7214.2004.01301001.x

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343. doi:10.1037/1082-989X.11.4.323

Gustafsson, J., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28 (4)*, 407-434. doi: 10.1207/s15327906mbr2804

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J.-A. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly*, *16*, 31–55. doi: 10.1521/scpq.16.1.31.19158

Hajovsky, D., Reynolds, M. R., Floyd, R. G., Turek, J. J., & Keith, T. Z. (2014). A multigroup investigation of latent cognitive abilities and reading achievement relations. *School Psychology Review*, *43*(4), 385–406.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

Jewsbury, P. A., Bowden, S. C., & Duff, K. (2016). The Cattell–Horn–Carroll model of cognition for clinical assessment. *Journal of Psychoeducational Assessment*, *1*, 1–21. http://doi.org/10.1177/0734282916651360

Kamphaus, R.W. (2009). Assessment of intelligence and achievement. In C.R. Reynolds & T.B. Gutkin (Eds.), *The handbook of school psychology* (pp. 230–246). New York, NY: Wiley.

Kaufman, A. S., & Kaufman, & N. L. (2004a). Kaufman assessment battery for children—Second edition: Technical manual. Circle Pines, MN: American Guidance Service.

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock–Johnson and Kaufman tests. *Intelligence*, *40*(2), 123–138. doi: 10.1016/j.intell.2012.01.009

Keith, T. Z. (1999). Effects of general and specific abilities on student achievement: Similarities and differences across ethnic groups. *School Psychology Quarterly*, *14*(3), 239–262. doi: 10.1037/h0089008

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). New York, NY: Routledge.

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

Keith, T., Kranzler, J., & Flanagan, D. (2001). What does the Cognitive Assessment System (CAS) measure? Joint confirmatory factor analysis of the CAS and the Woodcock-Johnson Tests of Cognitive Ability (3rd Edition). *School Psychology Review*, *30*(1), 89-119.

Keith, T. Z., Low, J. A., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2010). Higher-order factor structure of the Differential Ability Scales–II: Consistency across ages 4 to 17. *Psychology in the Schools, 47*, 676-697.

Keith, T.Z. & Novak, C.G. (1987). Joint factor structure of the WISC-R and K-ABC for referred school children. *Journal of Psychoeducational Assessment, 4*, 370–386.

Keith, T. Z., & Reynolds, M. R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools, 47*(7), 635-650. doi: 10.1002/pits

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

Mayes, S.D., & Calhoun, S. L. (2007). Wechsler Intelligence Scale for Children–Third and – Fourth Edition Predictors of Academic Achievement in Children With Attention-Deficit/ Hyperactivity Disorder. *School Psychology Quarterly*, *22*, 234–249. http://doi.org/10.1037/1045-3830.22.2.234

McBride-Chang, C., & Manis, F. R. (1996). Structural invariance in the associations of naming speed, phonological awareness, and verbal reasoning in good and poor readers: A test of the double deficit hypothesis. *Reading and Writing*, *8*, 323–339. http://doi.org/10.1007/BF00395112

McGrew, K. S. (1993). The relationship between the WJ-R Gf-Gc cognitive clusters and reading achievement across the lifespan. *Journal of Psychoeducational Assessment*, Monograph Series: WJ R Monograph, 39–53.

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

McGrew, K. S., Flanagan, D. P., Keith, T. Z., & Vanderwood, M. (1997). Beyond g: The impact of Gf-Gc specific cognitive abilities research on the future use and interpretation of intelligence tests in the schools. *School Psychology Review*, *26*(2), 189–210.

McGrew, K.S. & Hessler, G.L. (1995). The relationship between the WJ-R Gf-Gc cognitive clusters and mathematics achievement across the life-span. *Journal of Psychoeducational Assessment, 13*, 21–38.

McGrew, K. S. & Knopik, S. N. (1991). The relationship between the WJ-R Gf-Gc cognitive clusters and writing achievement across the life-span. *School Psychology Review, 22*(4), 687-695.

McGrew, K., & Wendling, B. (2010). Cattell–Horn–Carroll cognitive-achievement relations: What we have learned from the past 20 years of research. *Psychology in the Schools, 47*(7), 651–675. doi: 10.1002/pits

Muthén, L. K., & Muthén, B. O. (2012). Mplus user's guide (Seventh ed.). Los Angeles, CA: Muthén & Muthén.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Ceci, S. J., Loehlin, J. C., & Sternberg, R. J. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77–101.

Niileksela, C. R., & Reynolds, M. R. (2014). Global, broad, or specific cognitive differences? Using a MIMIC model to examine differences in CHC abilities in children with learning disabilities. *Journal of Learning Disabilities*, *47*(3), 224–36. http://doi.org/10.1177/0022219412453172

Niileksela, C. R., & Reynolds, M. R., Keith, T.Z., & McGrew, K.S. (2016). A special validity study of the WJ IV: Acting on evidence for specific abilities. In D.P. Flanagan & V.C. Alfonso (Eds.), *WJ IV clinical use and interpretation: Scientist-practitioner perspectives.*

Parkin, J. R., & Beaujean, A.A. (2012). The effects of Wechsler Intelligence Scale for Children-Fourth Edition cognitive abilities on math achievement. *Journal of School Psychology, 50*(1), 113–28. doi: 10.1016/j.jsp.2011.08.003

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

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51*(4), 535–55. doi: 10.1016/j.jsp.2013.02.003

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63, 581–592.

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

Salthouse, T.A. (2005). Relations between cognitive abilities and measures of executive functioning. *Neuropsychology*, *19*, 532–45. doi: 10.1037/0894-4105.19.4.532

Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological Methods*, *7*(2), 147–177. doi: 10.1037/1082-989X.7.2.147

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

Stone, B. J. (1992). Joint factor analysis of the DAS and WISC-R. *Journal of School Psychology, 30*, 185–195.

Swanson, H. L., & Alexander, J. E. (1997). Cognitive processes as predictors of word recognition and reading comprehension in learning-disabled and skilled readers: Revisiting the specificity hypothesis. *Journal of Educational Psychology, 89*, 128–158. http://doi.org/10.1037/0022-0663.89.1.128

Taub, G. E., Keith, T. Z., Floyd, R. G., & McGrew, K. S. (2008). Effects of general and broad cognitive abilities on mathematics achievement. *School Psychology Quarterly, 23*(2), 187–198. doi: 10.1037/1045-3830.23.2.187

Thorsen, C., Gustafsson, J.-E., & Cliffordson, C. (2014). The influence of fluid and crystallized intelligence on the development of knowledge and skills. *The British Journal of Educational Psychology*, *84*(4), 556–70. doi: 10.1111/bjep.12041

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*, 4-70.

Vanderwood, M. L., McGrew, K. S., Flanagan, D. P., & Keith, T. Z. (2002). The contribution of general and specific cognitive abilities to reading achievement. *Learning and Individual Differences*, *13*, 159–188.

Vellutino, F. R., Scanlon, D. M., Sipay, E. R., Small, S. G., Pratt, A., Chen, R., & Denckla, M. B. (1996). Cognitive profiles of difficult-to-remediate and readily remediated poor readers: Early intervention as a vehicle for distinguishing between cognitive and experiential deficits as basic causes of specific reading disability. *Journal of Educational Psychology*, *88*(4), 601–638. http://doi.org/10.1037/0022-0663.88.4.601

Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. *The science of prevention: Methodological advances from alcohol and substance abuse research*, 281-324.

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children – Fifth Edition*. Bloomington, MN: Pearson.

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