Do the Kaufman Tests of Cognitive Ability and Academic Achievement Display Ethnic Bias for Students in Grades 1 through 12?

by

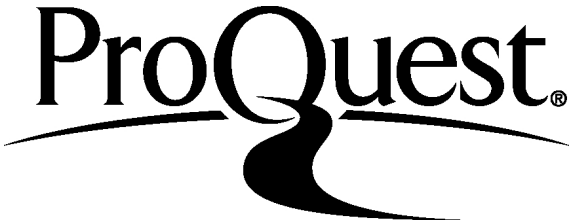
Caroline Scheiber, M.A.

A Dissertation
Presented to the Faculty of the California School of Professional Psychology at Alliant International University, San Diego

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

The California School of Professional Psychology 2015

Acknowledgement



I would like to take this opportunity and express my thankfulness and appreciation to my mentor and dissertation chair, Alan S. Kaufman, whose wisdom, patience, endless support and commitment have been invaluable in the completion of this dissertation as well as many other projects we have been working on for over three years.

In addition to Dr. Kaufman's discernible excellence and expertise in the field of intelligence and achievement testing, it were his friendship and caring that ultimately made the difference and helped to make this a constructive and encouraging experience. His high standards, generosity, and understanding were inspiring and motivating to continually improve the quality of my work. Thank you for always believing in me.

My thanks also go to Dr. Nadeen Kaufman, who always made me feel welcome in the house and always showed confidence in my ability to successfully complete this process. Her warmth and kindness are incomparable and readily apparent to everyone who meets her.

I would also like to thank my academic advisor and reader, Dr. Constance Dalenberg, who I could always rely on as supporting my best interest throughout the four years in the doctoral program, especially when things got tough. Her expertise in methodology, statistics, and research are indisputable and were instrumental in the completion of this dissertation. Dr. Dalenberg is always prepared to help whenever she can. She is an amazing professor and person.

I would also like to express my appreciation to Dr. Elizabeth Lichtenberger, who generously agreed to be one of my readers for this dissertation project. Her dedication and attention to detail not only improved the quality of the document, but also made this process comfortable for me.

And, I thank Dr. Donald Viglione, who first offered his knowledge in psychometrics and assessment during the pre-proposal meeting and ultimately at my final oral defense.

Finally, I would like to acknowledge my friends and colleagues, especially Joscelyn Rompogren, who patiently and compassionately listened to my many doubts and worries throughout the process. She has always been readily available to give advice and suggestions. Her support lasted through the end when her presence helped me get through the oral defense. I also thank Jenna VanSlyke who was so kind as to read my almost 50 page long proposal and give me feedback; as well as Arezoo Esfahani and Treniece Robinson, who supportively attended my dissertation defense.

Finally, I would like to thank my family for their support and encouragement from the far. Without them the completion of this dissertation would not have been possible.

Abstract

Cultural bias in cognitive testing has a long and controversial history. As the demographic profile in the United States continues to change and becomes ethnically more diverse, the need for culturally appropriate test instruments has become a national concern among educators, clinicians, and researchers. The Kaufman Assessment Battery for Children, – 2nd Edition (KABC-II) and the Kaufman Test of Educational Achievement – 2nd Edition (KTEA-II) are two well-known tests of intelligence and achievement. The tests' popularity for the assessment of children is not only due to the quality of their psychometric properties, but also because they appeal to an ethnically diverse client population. Although the test publishers have put great effort in ensuring the appropriate validity and reliability criteria for these tests, the issue of test bias in terms of the tests' construct and predictive validity across different ethnic groups has not been addressed.

The present study investigated construct and predictive invariance across a nationally representative sample of Caucasian (n = 1313), Black (n = 312), and Hispanic (n = 376) children in grades 1-12. Confirmatory factor analysis and structural equation modeling, using the data from the KABC-II and KTEA-II standardization samples, was used to assess whether increasing sets of equality constraints fit the tests' underlying theoretical model equally well for all three ethnic groups. Results from the construct invariance analysis showed that factorial invariance of the factor structure, based on seven Cattell-Horn-Carroll (CHC) broad abilities, was met for all three groups. Results from the predictive invariance analysis also demonstrated a lack of ethnic bias in the analysis of slopes; virtually all slopes for the five CHC-based KABC-II did not differ significantly by ethnicity. Thus, these five cognitive Indexes correlated about equally well with reading, math, and writing for Caucasians, Blacks, and Hispanics across three different grade

groups (grades 1-4: *n* = 724; grades 5-8: *n* = 743; grades 9-12: *n* = 534). However, the invariance analysis showed bias. Four of the five KABC-II CHC-based Indexes (excluding Knowledge/Gc, often considered the most culturally-loaded Index) demonstrated a persistent overprediction of the minority groups' achievement across the same three grade groups. The overprediction was especially notable in the areas of relative strength for Blacks (Sequential/Gsm and Learning/Glr) and Hispanics (Simultaneous/Gv and Planning/Gf).

One possible interpretation of this pervasive overprediction is that the educational system has failed to be flexible enough to take advantage of ethnic children's strengths when teaching reading, math, and writing. Another key finding is that the global score of Fluid-Crystallized Index emerged as the fairest predictor of achievement across the age range. Perhaps the most global score, rather than the profile of five CHC Indexes, should be featured when predicting school achievement—a contention consistent with Gary Canivez's theory and research. Outcomes of this research contribute to a scarce body of literature on ethnic test bias that goes beyond the simple comparison of mean score differences. Results of this study provide the evidence needed to justify continuous use of the KABC-II and KTEA-II in the assessment of children and adolescents for diverse ethnic groups. Furthermore, findings are generalizable beyond the Kaufman tests to other popular tests of intelligence and achievement; this is, because this study is based on the CHC factor structure, a universal theory of cognition that is used as the theoretical underpinning by many well-known tests of intelligence and achievement, including the most recent versions of the Wechsler scales.

Table of Contents

Do the Kaufman Tests of Cognitive Ability and Academic Achievement Display Ethnic Bias for Students in Grade 1 through 12?

The population in the United States has become more ethnically diverse than previous generations (U.S. Census Bureau, 2008). Former minority groups have become majority groups in various areas across the country (e.g., Blacks in Washington D.C.). The U.S. Census Bureau (2009) projects that by the year 2023 kindergarteners will consist primarily of ethnic minorities. Already today about one quarter of children in the public school systems are of Hispanic descent and the percentage of ethnic minorities in the U.S. is expected to reach 54% by 2050. By that time, 62% of children in the U.S. will be non-Caucasian. Thus, it becomes increasingly more evident that diversity within the U.S. already exists, continues to expand, and is ultimately inevitable (Llorente & Sheingold, 2010; Smith 2008).

As the population in the U.S. becomes more ethnically diverse, the need for culturally appropriate assessment measures has also become progressively more important. Cognitive ability assessments for ethnic minority groups, however, have a controversial narrative. For example, historically, differences in Intelligence Quotients (IQs) between ethnic groups have persisted for decades and continue to exist even after controlling for other variables, such as socioeconomic status (SES) (Weiss et al, 2006). Overall, mean score estimations on intelligence tests across various ethnic groups with a mean of 100 and a standard deviation (*SD*) of 15 are as follows: Caucasians 102, Blacks 92, Hispanics 94, Native Americans 90, and Asians 105 (e.g., Prifitera, Saklofske, & Weiss, 2005). Whereas the differences in scores across ethnic groups have narrowed significantly in recent years (e.g., Ceci & Kanaya, 2010; Dickens & Flynn, 2006), the disparities still endure. The implications of IQ differences are substantial, as IQ tests often

determine eligibility for special services and programs, employment, and school admission (U.S. Department of Education, 2005).

With the exception of Asians/Pacific Islanders, a remarkable overrepresentation of minority students diagnosed with a learning or intellectual disability as well as a disproportionate underrepresentation of minority students in gifted programs—has researchers, clinicians, and other scholars concerned about the fairness of assessment measures used. For example, on a national level, among 6- to 21 year olds, 7.2% of Native Americans, 5.5% of Blacks, and 4.6% of Hispanics, as compared to 3.6% of Caucasians and 1.6% of Asians/Pacific Islanders, are currently diagnosed with a specific learning disability and served under the Individuals with Disabilities Education Act (IDEA; U.S. Department of Education, 2008). Similarly, 1.7% of Blacks and 1.0% of Native Indian children and adolescents are diagnosed with an intellectual disability and served under IDEA. Alternatively, only 0.6% of Caucasians, 0.4% of Asian/Pacific Islanders, and 0.6% of Hispanics have a diagnosis of intellectual disability. However, it is important to point out that the lower representation of Hispanic students is likely to be related to cultural and language differences; many practitioners are justly hesitant to diagnose students who are still struggling with language and acculturation issues with an intellectual disability (U.S. Department of Education, 2008). In 2006, 8.0% of Caucasian students and 13.1% of Asian/Pacific Islanders in elementary and secondary school were classified as gifted, whereas only 3.6% of Blacks, 4.2% of Hispanics and 5.2% of Native Americans were placed in such programs (U.S. Department of Education, 2006).

Diagnoses such as specific learning disabilities and selection for gifted programs depend not only on IQ tests, but also on standardized measures of academic achievement. And just as there are ethnic differences in IQ, there is also ample evidence that Caucasians perform notably

better than Blacks and Hispanics on measures of reading, math, and writing (Jencks & Phillips, 1998; Lockhead, Thorpe, Brooks-Gunn, Casserly & McAloon, 1985; Naglieri, Rojahn, & Matto, 2007; Najarian, Snow, Lennon, & Kinsey, 2010). In short, the societal impact of ethnic differences in IQ and academic achievement is profound. The possibility that these differences may reflect—at least to some extent—built-in bias in the measuring instruments, therefore, is similarly of societal importance.

Goals of the Study

It is data such as the above on ethnic differences that often form the basis for educators and psychologists to express concern for tests to be fair to all ethnic groups (Weiss et al., 2006; Weiss & Prifitera, 1995). The primary purpose of this present dissertation was to investigate test bias of two individually administered, reliable, and well-normed measures of intelligence and achievement using state-of-the-art methodology. Specifically, the *Kaufman Assessment Battery for Children—Second Edition* (KABC-II; Kaufman & Kaufman, 2004a) and the Comprehensive Form of the *Kaufman Test of Educational Achievement—Second Edition* (KTEA-II; Kaufman & Kaufman, 2004b) were used to assess test bias of Caucasian, Black, and Hispanic students in grades 1 through 12. Both the KABC-II and the KTEA-II have demonstrated good convergent validity with other well-known measures of intelligence and achievement, including the *Wechsler Intelligence Scale for Children – Fourth Edition* (WISC-IV; Wechsler, 2003), the *Wechsler Individual Achievement Test – Second Edition* (WIAT-II; Wechsler, 2001), and the *Woodcock-Johnson — Third Edition* (WJ III; Woodcock, McGrew, & Mather, 2001) (Kaufman & Kaufman, 2004a, chapter 8; Kaufman & Kaufman 2004b, chapter 7). Furthermore, independent researchers (Floyd, Reynolds, Farmer, & Kranzler, 2013; Reynolds, Floyd, & Niieleksela, 2013) have found that the KABC-II measures the general intelligence factor (*g*) in

the same way as do other major tests of cognitive ability, namely the WISC-IV, WJ III, and *Differential Ability Scales — Second Edition* (DAS-II; Elliott, 2007). The *g* factor captures the positive correlations between the various components of an intelligence test, thereby demonstrating that the performance on one type of cognitive task is related to the performance on other cognitive tasks. This mental ability factor underlies all cognitive tasks and accounts for the common variance across all types of intellectual ability (Schneider & McGrew, 2012) and perhaps across academic skills as well (S. B. Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012). In that sense, findings of the present dissertation do not serve the purpose of providing evidence for the validity only of the specific two instruments used; instead the present investigation is centered on the question of whether the *constructs* of intelligence and achievement, as measured by frequently used clinical tests, are applicable not only to Caucasian students, but also to Hispanic and Black students.

To answer this broad question, three methods were used: (1) differential construct validity, (2) comparison of mean scores, and (3) differential predictive validity. Differential construct validity was established using confirmatory factor analysis (CFA) to determine whether the seven theory-based factors identified for the KABC-II and the KTEA-II (S.B. Kaufman et al., 2012) apply equally well for each of the three ethnic groups. The theory used by S.B. Kaufman and colleagues, the Cattell-Horn-Carroll (CHC) model of cognitive abilities (Schneider & McGrew, 2012), was the foundation of the present study. Once factorial invariance is established, only then can mean differences between ethnic groups be meaningfully compared (Meredith, 1993). Hence, this study evaluated test bias by determining whether the Kaufman tests measure the same theory-based constructs equally well for three ethnic groups; whether the ethnic groups differ significantly in their mean scores on these factors; and whether

the ability scores predict academic achievement equally well for Blacks, Hispanics, and Caucasians.

Socioeconomic status, as measured by parent educational attainment, was controlled. Ethnic minority children are far more likely to be from lower SES than those from nonminority Caucasian families; and lower SES creates a number of factors that have been found to negatively impact cognitive performance (e.g., less access to academic resources, increased health problems; see Weiss, et al., 2006 for review). To minimize effects related to poverty, SES was controlled. However, parent education controls only one aspect of the complex variable of SES. Further, the variable of SES becomes even *more* complex when comparing test performance of individuals from different cultural and linguistic backgrounds. In a practical sense, then, SES is only partially controlled in this investigation. Hence, mean differences, *per se*, cannot be thought of as denoting test bias but are more reasonably attributed to socioeconomic and linguistic variables that are known to impact performance on ability and achievement tests, but are not adequately controlled in scientific investigations. It is for that reason that the first and third methodologies for examining test bias (differential construct validity and differential predicative validity) are the only empirically valid methods for determining test bias in the present investigation.

Overview of the Relevant Research Literature

Test Bias Definition

A test is biased "if a test design, or the way results are interpreted …, systematically disadvantages certain groups … over others" (The Glossary of Educational Reform, 2013); it is "a systematic error in the [design of the test that results in the erroneous] measurement of a

psychological attribute as a function of membership in one or another cultural or racial subgroup" (Reynolds & Lowe, 2009, p. 333).

Traditional Method of Detecting Bias: Mean Score Differences

The detection of differences in mean scores on cognitive tests often forms the basis for arguments that tests are biased and discriminate against certain groups (e.g., Williams, 1971). In that sense, mean score differences are often used as one way of determining bias. Many studies have compared mean scores in cognition across ethnically diverse groups (e.g., Kaufman & Doppelt, 1976; Prifitera & Saklofske, 1998; Prifitera, Saklofske, & Weiss, 2005). Williams (1971) argued that "all previous research" (p. 63) that has compared the cognitive ability of Caucasians and non-Caucasians should be deemed invalid because of the tests' lack of fairness toward non-Caucasians. In that sense, Williams argues that the presence of mean score differences found on tests constitute bias with the implication being that some test items are only fair to individuals who have grown up in mainstream, middle class, Caucasian environments.

However, several other researchers (e.g., Reynolds & Lowe, 2009) have argued that mean differences alone do not necessarily prove bias. For example, mean score differences in the capacity to bench-press between men and women do not reflect bias. Instead, the different mean scores between the genders are due to inherent biological differences between males and females (e.g., males are, on average, physically stronger than females). Mean score differences denote true gender differences in bench-pressing skill; they do not imply that the bench-pressing apparatus (i.e., the instrument) is biased against females. Alternatively, mean differences in cognitive test scores by ethnicity do not reflect true differences in ability at all. Rather, as discussed previously, such discrepancies are related to SES differences that are known to be associated with growing up in the US in non-mainstream cultural and linguistic environments.

Such disparities in SES are not able to be controlled in any scientific investigation of cognitive and achievement skills.

The following section investigates the basic research findings on ethnic group differences (Blacks-Caucasians, Hispanics-Caucasians, and Blacks-Hispanic), focusing on global score differences on cognitive and achievement tests and on differences observed on CHC Broad Abilities. Research findings are evaluated across different age groups and both with and without a control for SES.

Black-Caucasian studies of cognitive and achievement tests.

Global IQ differences. Researchers generally find Black-Caucasian differences in favor of Caucasian individuals on global IQs obtained on the most frequently used clinical tests of intelligence, including the Wechsler scales, the Woodcock-Johnson tests, and the Stanford-Binet (Dickens & Flynn, 2006; J. Kaufman et al., 1995; Thorndike, Hagen, & Sattler, 1986; Tulsky et al., 2003). For school-aged children, scores usually yield Black-Caucasian differences of .8 *SD* to 1.1 *SD* (11 ½ – 16 points) in favor of Caucasians, when scores are *not* controlled for SES (e.g., Dickens & Flynn, 2006; Edwards & Oakland, 2006; Kaufman & Doppelt, 1976; Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005; Prifitera & Saklofske, 1998; Prifitera et al., 2005). When scores are controlled for SES, Black-Caucasian differences decrease to .5 *SD* – .8 *SD* (8 – 13 points) (Edwards & Oakland, 2006; Kaufman, McLean, & Kaufman, 1995; Manly, Heaton, & Taylor, 2000; Prifitera, Weiss & Saklofske, 1998; Prifitera et al., 2005; Tulsky et al., 2003; Weiss et al., 2006).

The size of the differences, however, varies depending on the test used. For example, Black-Caucasian differences have been found to be considerably smaller on some of the Kaufman tests as well as the Cognitive Assessment System (CAS; Naglieri, & Das, 1997), a

cognitive test based on the Planning, Attention, Simultaneous, and Successive (PASS) theory of intelligence. On the KABC-II, CAS, and the original Kaufman Assessment Battery for Children (K-ABC; Kaufman & Kaufman, 1983) Caucasian school-aged children outscored Black children on global intelligence scores by .4 *SD* – .6 *SD* (6–9 points), which reduced to .3 *SD*– .5 *SD* (5–8 points) when adjusted for SES (Kaufman & Kaufman, 1983, 2004; Kamphaus & Kaufman, 1986; Naglieri, Rojahn, Matto, & Aquilino, 2005; Prifitera et al., 1998).

An occasional study has demonstrated slightly larger Black-Caucasian differences for school-aged children than adults (e.g., Tulsky et al., 2003), but the bulk of research has shown similar Black-Caucasian differences for adults and school-aged children (e.g., Dickens & Flynn, 2006; Kaufman et al., 1995; Reynolds, Chastain, Kaufman, & McLean, 1988). Preschoolers, however, yield significantly smaller Black-Caucasian differences compared to school-aged children and adults across a variety of individually-administered, clinical tests of cognitive ability (Arinoldo, 1981; Hauser, 1998; Kaufman, 1973a; Kaufman & Kaufman, 1973; Kaufman, McLean, & Reynolds, 1988; Lichtenberger, Broadbooks, & Kaufman, 2000; Manly et al., 2000; Puente & Salazar, 1998; Raiford & Coalson, 2014; Reynolds et al., 1988).

In sum, Black-Caucasian mean global standard scores for school-aged children on the majority of clinical tests usually yield differences in *g* close to 14 points; differences reduce, but still remain at about 11 points, for samples matched on SES. Still, the reduction in the magnitude of the differences when SES is controlled demonstrates the importance of controlling for SES when comparing cognitive ability across ethnic groups. The KABC, the KABC-II, and the CAS produce smaller differences. Whereas differences for adults are comparable to those of school-aged children, scores for preschoolers tend to be smaller.

CHC factor differences. The present dissertation is based on the CHC theoretical model of intelligence, a theory that subsumes both cognitive abilities and academic skills (Schneider & McGrew, 2012). CHC theory is a psychometric model based on years of extensive research. It houses two influential theories of human intelligence – Cattell-Horn's Fluid-Crystallized (*G*f-*G*c) Theory and Carroll's Three-Stratum Theory (Carroll, 1993; Schneider & McGrew, 2012). Scholars identified the theory as "the most empirically grounded" and therefore one of the most "reliable" and "valid" (McGrew, 1997, p.151) classification systems of cognitive ability to date (Keith & Reynolds, 2010). And the influence of CHC theory continues to expand; for example, CHC theory has been successfully linked to a variety of popular neuropsychological assessment measures and models in a recent comprehensive investigation (Jewsbury, 2014).

Consistent with the CHC model of intelligence., the following sections on ethnic differences are organized according to the seven broad CHC factors that are the focus of this study. Five of the CHC factors are representative of traditional cognitive abilities: (1) crystallized intelligence (*G*c, which represents the knowledge one accumulates over the lifespan); (2) fluid intelligence (*G*f, which refers to one's ability to solve novel problems); (3) visual-spatial reasoning (*G*v, which refers to the ability to perceive, manipulate, and retrieve visual information); (4) short-term memory (*G*sm, which is defined as the ability to hold and manipulate small amounts of information in one's head, often requiring divided attention); and (5) long-term memory (*G*lr, which reflects the ability to store and retrieve information that has been learned previously). Two CHC factors are more associated with academic achievement than with traditional cognitive ability: (6) math (*G*q, which reflects the ability to understand and solve quantitative problems and the ability to manipulate digits); and (7) reading and writing (*G*rw, which is defined as the ability to read and spell single words. As well as to comprehend,

write, synthesize, and connect complex sentence structures) (Flanagan, Ortiz, & Alfonso, 2013; Kaufman, 2009; Schneider & McGrew, 2012).

Crystallized intelligence: Gc. Substantial *G*c differences of about 10 ½–14 ½ standard-score points between Caucasians and Blacks emerge for school-aged children on various intelligence measures, including the Wechsler scales, Kaufman scales, and Woodcock-Johnson (Edwards & Oakland, 2006; Kaufman, Chen, Kaufman, 1995; Kaufman & Kaufman, 1983, 2004; Kaufman & Lichtenberger, 2001; Kaufman et al., 1995; Kaufman & Wang, 1992; Prifitera & Saklofske, 1998; Reynolds et al., 1987; Weiss et al., 2006). Differences reduce to about 8 ½–9 points when SES is controlled (Kaufman & Kaufman, 2004; Prifitera & Saklofske, 1998; Weiss et al., 2006). Differences on *G*c are about the same magnitude for adults (Kaufman et al, 1998; Reynolds et al., 1987; Tulsky et al., 2003), but are smaller for preschoolers (Kaufman & Kaufman, 1983, 2004; Raiford & Coalson, 2014). Overall, *Gc* differences between Caucasians and Blacks mirror the differences summarized in the previous section for measures of *g*.

Nonverbal intelligence: Gf and Gv. The Wechsler's performance IQ and perceptual indexes measure *both Gf* (fluid) and *Gv* (visual-motor), as does the K-ABC Simultaneous Processing Scale . (Note: the WISC, WISC-R, and WISC-III included a performance IQ, the WISC-IV only measures Gv, and the WISC-V includes separate measures of Gv and Gf). These two CHC abilities are, therefore, often merged in the literature and are discussed together here. Some researchers have found Black-Caucasian differences on *G*f and *G*v for school-aged children that are similar in magnitude to *G*c differences (Kaufman et al., 1998; Kaufman & Kaufman, 2004a; Kaufman, et al., 1995; Kaufman & Lichtenberger, 2002; Kaufman & Wang, 1992; Prifitera et al., 2005; Prifitera & Saklofske, 1998; Tulsky et al., 2003). These differences typically range from about 10 ½-16 points (Kaufman & Kaufman, 1983, 2004; Kaufman &

Lichtenberger, 2002; Prifitera et al., 2005; Prifitera & Saklofske, 1998). When controlled for SES, these differences reduce to about 9-13 points (Kaufman & Kaufman, 2004; Kaufman & Lichtenberger, 2002; Prifitera et al., 2005; Prifitera & Saklofske, 1998). With the exception of the Naglieri Nonverbal Ability Test (NNAT; Naglieri, 1997)—which yielded differences of only about 4 points (Naglieri & Ronning, 2000)—the findings for *Gf* and *Gv* are remarkably similar to Black-Caucasian differences observed on measures of *Gc* and on most measures of *g*. Results for adult samples are essentially the same as results for school-age children: about 12 ½–14 ½ standard-score points when SES is not controlled; about 11 points when this key variable is controlled (Kaufman & Lichtenberger, 2002; Prifitera et al., 1998; Reynolds et al., 1987; Tulsky et al., 2003). And, just as was found for *Gc* and *g*, differences for preschool children are smaller, ranging from 2-6 points (Kaufman & Kaufman, 1983, 2004).

Overall, Black-Caucasian differences on traditional measures of nonverbal intelligence (measures of *Gf* and *Gv*) are about as large as differences on traditional measures of verbal intelligence (*Gc*)—even though the former are often considered to be more "culture fair" than the latter (e.g., Kaufman & Kaufman, 1983), as they minimize the language barrier (Flanagan et al., 2013).

Memory: Gsm and Glr. With the exception of the Stanford-Binet-IV (*G*sm differences = 11 standard-score points; Thorndike et al., 1986), Black-Caucasian differences on *G*sm and *G*lr are usually much smaller than differences on *Gc*, *Gf*, and *Gv*. Results for school-aged children usually yield *Gsm* differences of 7 points for unadjusted scores and 4 ½ points with SES controlled (Edwards & Oakland, 2006; Kaufman & Lichtenberger, 2002; Prifitera & Saklofske, 1998). The KABC and the KABC-II yield even smaller *Gsm* differences of about 2 ½-3 points (1 ½ when adjusted for SES) (Kaufman & Kaufman, 1983, 2004). Comparably small Black-

Caucasian differences are found for *G*lr on the KABC-II and the WJ III (Edwards & Oakland, 2006; Kaufman & Kaufman, 2004). For preschoolers, Black-Caucasians differences on *G*sm and *G*lr are even smaller, frequently not even reaching significance (Kaufman & Kaufman, 1983, 2004). Interestingly, some studies found a slight enlarging of the Black-Caucasian gap on *G*sm and *G*lr with increasing age (Kaufman & Lichtenberger, 2002; Tulsky et al., 2003).

In sum, Black-Caucasian differences on *G*sm and *G*lr are considerably smaller than differences on to *G*c, *G*v, and *G*f. The only other CHC ability that consistently produces small ethnic differences is processing speed (*Gs*), the ability to quickly perform relatively easy or overlearned tasks (Edwards & Oakland, 2006; Prifitera et al., 1998; Tulsky et al., 2003; Weiss et al., 2006). *Gs*, however, was not investigated in the present study because Gs is not included as a separate broad ability on the KABC-II. According to the authors (Kaufman & Kaufman, 2004a), "[Gs] lacked the requisite complexity for inclusion. Both of the 'speed' abilities, Gs and Gt [Decision Speed/Reaction Time], emerge consistently as weak measures of *g* in Carroll's (1993) factor-analytic survey… Gs is measured to some extent at ages 7 to 18 by the KABC-II subtests that include time points (Story Completion, Triangles, and Pattern Reasoning)" (p. 16).

Math: Gq. Studies that have compared Black and Caucasian individuals on *G*q found that Caucasians outperform Blacks on individually-administered and group-administered tests of *G*q (e.g., Lockheed et al., 1985). Results from individually-administered tests, including the Stanford-Binet-IV, three different Kaufman tests, and the Woodcock Johnson-Revised (WJ-R; Woodcock & Johnson, 1989), reveal Black-Caucasian differences in Gq of about 10 to 14 ½ points for school-aged children, favoring Caucasians (J. Kaufman et al., 1995; Naglieri et al., 2005; Thorndike et al., 1986). On the KABC, Black-Caucasian differences on Arithmetic were about 8 standard score points (Kaufman & Kaufman, 1983). Researchers using group tests, such

as the Scholastic Aptitude Test (SAT; Kobrin & Schmidt, 2005), found similar *G*q differences (e.g., Hedges & Nowell, 1998; KewalRamani, Gilbertson, Fox, Provasnik, 2002; Naglieri & Ronning, 2000; Rampey, Dion, & Donahue, 2009; Vanneman, Hamilton, Anderson, & Rahman, 2009). Furthermore, minority students, including Blacks, are less likely to be placed in advanced math and science classes (Epps 1995; Kubitschek & Hallinan, 1996) and are less likely to perform at the proficiency level in math (National Center for Education Statistics, 2009; U.S. Department of Education, 2010).

There is some evidence that the Black-Caucasian *G*q gap widens as individuals progress through school with a smaller gap observed for preschoolers and kindergarteners compared to older individuals (Careiro & Heckman, 2002; Fryer & Levitt, 2004; Jencks & Phillips, 1998; Kaufman & Kaufman, 1983). For example, one longitudinal study found that just between Fall of Kindergarten and Spring of 1st grade the Black-Caucasian gap in math and reading widened by almost 5 standard-score points, when controlling for other factors, such as SES (Fryer & Levitt, 2004).

In sum, not much research has been conducted on Black and Caucasian differences on *G*q, using individually-administered tests of achievement. Those studies that have been conducted yielded differences close to 1 *SD* for school-aged children, similar to results for group-administered tests such as the SATs or the tests used in the large-scale studies conducted by the National Assessment of Educational Progress (NAEP).

Reading and writing differences: Grw. In CHC theory reading and writing are both subsumed by a single Broad Ability (*G*rw). In practice, however, ethnic differences on reading and writing have been investigated separately. Results from individually-administered tests yield Black-Caucasian differences of about 11 standard score points on tests of reading, in favor of

Caucasians, on the WJ-R (Naglieri et al., 2005). The K-ABC produced differences of about 7-8 standard score points on Reading Decoding and Reading Understanding (Kaufman & Kaufman, 1983). Differences of 8.5 points were found on tests of writing on the WJ-R (Naglieri et al., 2005). Group-administered test also show poorer performance of Black students by about 7.5 – 10.5 standard points on standardized tests of reading and writing (Jencks & Phillips, 1998; Miller, 1995; Najarian, Snow, Lennon, & Kinsey, 2010; National Center for Educational Statistics, 2010; 2011; Vanneman, Hamilton, Anderson, & Rahman, 2009). Other group-administered achievement test data showed that a smaller percentage of Black than Caucasian school-aged children performed at the proficiency level in reading and writing (Kao, Tienda, & Schneider, 1996; U.S. Department of Education, 2007a; 2007b; 2011b). Results from group-administered achievement tests also revealed poorer performance of preschoolers in letter recognition (U.S. Department of Education, 2010). Overall, Black-Caucasian differences in *G*rw tend to be smaller than in *G*q—a 0.5 *SD* to *0*.7 *SD* Caucasian advantage in *Grw* compared to differences of close to 1 *SD* on *G*q. The KABC produced smaller ethnic group differences than the WJ. Notably, few studies have investigated Black-Caucasian differences on *Grw* or *Gq* using individually-administered, clinical tests.

Hispanic-Caucasian studies of cognitive and achievement tests.

Global IQ differences. On the global cognitive scores, school-aged Hispanic children score about midway between Caucasians and Blacks across various measures of intelligence, including the Wechsler tests, the Kaufman scales, the Woodcock-Johnson, the Stanford-Binet, and the CAS (Kaufman & Kaufman, 1983, 2004; Prifitera et al., 1998; Taylor & Richards, 1991; Thorndike et al., 1986). Hispanic-Caucasian differences are about 6–10 standard-score points, which reduce to about 3–5 ½ points when adjusted for SES (Kaufman & Kaufman, 1983, 2004;

Naglieri et al. 2007; Prifitera et al., 2005; Prifitera & Saklofske, 1995; Taylor & Richards, 1991; Tulsky et al., 2003; Weiss et al., 2006). In some instances, Hispanic-Caucasian comparisons yielded non-significant differences when adjusted for SES (e.g., Prifitera & Saklofske, 1995). When compared to differences observed for school-aged children, the Caucasian advantage over Hispanics tends to be a bit smaller for preschool children (Kaufman & Kaufman, 1983, 2004; Raiford & Coalson, 2014) and a bit larger for adults (Kaufman & Wang, 1992; Prifitera et al., 1998). Indeed, some researchers provided direct evidence that the Hispanic-Caucasian gap enlarges as a function of age (e.g., Kaufman & Wang, 1992; Taylor & Richards, 1991; Thorndike et al., 1986).

CHC factor differences. Similarly to the section on Black-Caucasian differences, this section is organized by CHC abilities.

Crystallized intelligence: Gc. Data from the most frequently used clinical tests reveal Hispanic-Caucasian differences on *G*c of about 11 ½-–12 ½ standard-score points, differences that reduce dramatically (3–6 points) when adjusted for SES (Kaufman & Kaufman, 1983, 2004; Kaufman, McLean, et al., 1995; Lichtenberger et al., 2000; Prifitera & Saklofske, 1998; Tulsky et al., 2003; Vukovich, & Figueroa, 1982). Results for preschoolers and adults are consistent with school-age results, yielding differences of 9 ½–14 points for unadjusted scores and 5 ½–9 points for SES-adjusted scores (Kaufman & Kaufman, 2004; Kaufman, McLean, et al., 1995; Raiford & Coalson, 2014). On the K-ABC, Hispanic-Caucasian differences were non-significant for preschoolers on two subtests that measure *G*c (Kaufman & Kaufman, 1983).

Overall, the unadjusted 12-point *Gc* difference in favor of Caucasians is larger than the 8-point difference observed for *g*, supporting the approach in the present study of focusing on separate CHC abilities rather than *g*. Similarly, the great reduction in *Gc* differences between

Caucasians and Hispanics when SES is controlled demonstrates emphatically that no study of ethnic differences should be conducted with taking into consideration the individuals' SES background. Finally, studies have shown that the Hispanic-Caucasian gap on verbal measures grows as a function of age (e.g., Kaufman & Lichtenberger, 2002; Prifitera et al., 1998) with an increase from 3–3 ½ standard points to 4 ½ –6 standard points between childhood and adulthood. Other studies have demonstrated an enlarging of the nonverbal versus verbal score gap of Hispanic-Caucasian individuals with increasing age, attributing the increase in the differences to poorer performance of older Hispanic individuals on verbal measures (Kaufman & Wang, 1992)

Nonverbal intelligence: Gf and Gv. Traditionally, Verbal IQ (*Gc*) has yielded much larger Hispanic-Caucasian differences than Performance IQ (*Gf* and *Gv*) across various measures of intelligence, including the Wechsler scales and the Kaufman tests (e.g., Kaufman, 1994; Valencia & Suzuki, 2001). When unadjusted for education, Caucasian school-aged children have outscored Hispanic children on measures of *Gf* and *Gv* by about 5–9 standard-score points; these differences drop to 3–4 points with SES controlled, and such differences are often not statistically significant (Kaufman & Kaufman, 1983, 2004;). Similar results have been observed for preschool children (Kaufman & Kaufman, 1983, 2004). Although the Hispanic-Caucasian differences on Performance IQ have been observed to increase slightly as a function of age, differences remain trivial for adults (Kaufman et al., 1995; Kaufman & Lichtenberger, 2002; Kaufman & Wang, 1992; Manly et al., 2001; Prifitera et al., 1998).

In sum, Hispanic-Caucasian differences on *G*f and *G*v measures are about half as large as differences observed on *Gc*, again in agreement with the emphasis in this investigation on separate CHC abilities.

Memory: Gsm and Glr. Hispanic-Caucasian differences on *G*sm and *G*lr for school-aged children are comparable to differences found of *G*v and *G*f, usually yielding differences of about 4 –8 ½ standard-score points for unmatched samples and 3–6 points for samples matched on SES (Kaufman & Kaufman, 1983, 2004; Kaufman & Lichtenberger, 2002; Kaufman et al., 1995; Manly et al., 2001; Prifitera et al., 2005; Prifitera & Saklofske, 1998; Tulsky et al., 2003). Results are similar for preschool children on *Gsm* and *Glr* (Kaufman & Kaufman, 1983, 2004). Hispanic-Caucasian differences on *G*sm have been found to be larger on the Wechsler Adult Intelligence Scale-Third Edition (WAIS-III; Wechsler, 1997) versus the Wechsler Intelligence Scale for Children-Third Edition (WISC-III; Wechsler, 1991) (Kaufman & Lichtenberger, 2002; Prifitera et al., 1998). The smallest differences between Hispanic and Caucasian school-aged children occur on *Gs* (1 ½–3 ½ points unadjusted; 2 points when controlled for SES (Prifitera et al., 2005). As noted previously, however, *Gs* is not included in the present investigation.

Math: Gq. Studies comparing Caucasian and Hispanic students' performance in math achievement (*G*q) have consistently found differences in favor of Caucasian students (e.g., Lockheed et al., 1985). On individually administered tests of *G*q, Hispanic-Caucasian differences for school-aged children have typically been about 4 –7 ½ standard points in favor of Caucasian students on the K-ABC and the WJ-R (Kaufman & Kaufman, 1983; Naglieri et al., 2007; Valencia, Rankin, & Livingstone, 1995), with differences slightly smaller for preschool children (Kaufman & Kaufman, 1983), when not adjusted for SES. No research could be found that examined *G*q differences when controlled for SES.

Results from group-administered tests of *Gq* mirror the findings for school-aged students (U.S. Department of Education, 2013). Other achievement data revealed that a smaller percentage of Hispanic students completed advanced math courses, such as geometry or

statistics, in high school and a smaller percentage of Hispanics score at the proficiency level in math compared to Caucasians (U.S. Department of Education, 2010). Furthermore, fewer Hispanic than Caucasian preschoolers were able to recognize numbers and shapes between 2005 and 2006 (U.S. Department of Education, 2010).

Reading and writing: Grw. Results from individually-administered tests of reading reveal Hispanic-Caucasian differences of about 5–9 standard score points for school-aged children on tests of reading on the K-ABC and WJ-R (Kaufman & Kaufman, 1983; Naglieri et al., 2007), not controlled for SES. Data from group-administered tests, such as the SATs reflect differences similar in size (U.S. Department of Education, 2010). Furthermore, results from other group-administered tests show that a smaller percentage of Hispanic 4th and 8th graders score at the proficiency level in tests of reading (Aud, Fox, & KewalRamani, 2010; U.S. Department of Education, 2007a; U.S. Department of Education, 2010). Hispanics were also found to perform worse in writing in 2011 at 8th and 12th grade (U.S. Department of Education, 2011). Additionally, results from the NAEP demonstrate that between 2000 and 2005, Hispanic 4-year-old children performed worse on letter recognition compared to other ethnic groups, except American-Indians, and a smaller percentage of Hispanic preschoolers was able to recognize letters at a proficiency level, compared to their Caucasian counterparts (U.S. Department of Education, 2010).

In terms of writing, results from the individually administered Woodcock-Johnson-Revised (Woodcock & Johnson, 1989) revealed Hispanic-Caucasian differences in basic writing of 8 standard-score points on the WJ R, not controlled for SES (Naglieri et al., 2007). Similar mean score differences were found on the group administered SATs in 2008 (U.S. Department of Education, 2010). Results from the NAEP revealed that a smaller percentage of Hispanic than

Caucasian school-aged children wrote at the proficient level (U.S. Department of Education, 2007b).

Summary of Hispanic-Caucasian vs. Black-Caucasian Differences on CHC factors. In sum, the largest Hispanic-Caucasian differences—even larger than differences on the achievement-oriented *Gq* and *Grw*—are usually found on measures of *G*c, especially before controlling SES. The Hispanic-Caucasian discrepancy on measures of *G*f and *G*v is notably smaller than the discrepancy on *G*c, especially for adults due to poorer performance of older Hispanic individuals on *G*c. Whereas the Hispanic-Caucasian gap observed on *G*c is similar in size to the Black-Caucasian gap observed on this factor, the ethnic group gap on *g* and on the nonverbal factors (*G*f and *G*v) is smaller for Hispanic-Caucasian than Black-Caucasian school-aged children. Similarly, the Black-Caucasian gap on *G*q tends to be larger than the Hispanic-Caucasian gap. However, differences on *G*rw and the memory factors (*G*sm and *G*lr) tend to comparable in magnitude for Blacks and Hispanics. However, older Hispanics score lower than their younger counterparts on *G*sm, just as they did on Gc. Importantly, ethnic group differences (Black-Caucasian and Hispanic-Caucasian) on *g* and CHC abilities narrow when SES is taken into account, reinforcing the necessity of controlling for SES when studying test bias. No research could be found that explored achievement differences when controlling for SES.

The above data demonstrate that the Black-Caucasian "profile" is different from the Hispanic-Caucasian "profile" concerning the magnitude of ethnic differences on the various CHC abilities and on *g*. Consequently, the results of the bias analyses conducted in this dissertation might reveal notable differences for Blacks versus Caucasian and Hispanics versus Caucasians.

Hispanic-Black studies of cognitive and achievement tests. When it comes to Hispanics versus Blacks, the ethnic group gap seems to be smaller compared to Blacks-Caucasians and Hispanics-Caucasians; this finding has been demonstrated across various cognitive tests, including the Wechsler scales and the Kaufman tests. When *not* matched on SES, school-aged Hispanic children typically score about 1-2 IQ points higher than Blacks on global IQ scales (e.g., Kaufman & Kaufman, 1983, 2004; Kaufman, et al., 1995; Prifitera et al., 2005; Taylor, & Richards, 1991). This gap increases to 3.5 IQ points for matched samples (Kaufman & Kaufman, 2004; Prifitera et al., 2005). Hispanics also outperform Blacks on measures of *G*v and *G*f with differences up to 5 IQ points for unmatched samples and up to 6.5 IQ points for matched samples (e.g., Kaufman & Kaufman, 1983, 2004; Kaufman et al., 1995; Prifitera et al., 2005; Taylor, & Richards, 1991). Blacks outperform Hispanics on *G*c by about 1-2 IQ points (Kaufman & Kaufman, 1983, 2004; Kaufman, McLean et al., 1995); however, the differences disappear when SES is controlled (Prifitera et al., 2005). Blacks have been found to score about 2.5 IQ points above Hispanics in measures of *G*sm but, again, the differences disappear when SES is controlled (e.g., Kaufman & Kaufman, 1983, 2004; Kaufman, et al., 1995; Prifitera et al., 2005; Tulsky et al., 2003). With regards to the achievement factors, *G*q and *G*rw, the NAEP demonstrated that Hispanics and Blacks seem to score about equally well; however, there is some evidence that Hispanics outperform Blacks slightly in measures of *G*q (Aud, Fox, & KewalRamani, 2010; KewalRamani, Gilbertson, Fox, Provasnik, 2007; Lee, 2002; Lockheed, Thorpe, Brooks-Gunn, Casserly, & McAloon, 1985; Rampey, Dion, & Donahue, 2009; The Nation's Report Card., 2009; U.S. Department of Education, 2007a; 2007b; 2008; 2010; 2013).

More Sophisticated Methods of Assessing Bias: Differential Construct and Predictive Validity

In contrast to the simple question of which ethnic group earns higher mean scores on an array of intelligence or achievement tasks, more complex approaches to test bias ask (a) whether the tests measure the same constructs for different ethnic groups, and (b) whether the intelligence tests predict academic achievement equally well for diverse groups (Keith, 1999; Reynolds & Lowe, 2009).

Differential construct validity. Construct validity refers to the validity of inferences one draws from the observed scores. To warrant construct validity, measurement (factor) invariance needs to be established. Measurement invariance ensures that the assessment tool measures what it is supposed to measure across all groups. Measurement invariance denotes that the factor structure of the test is consistent and intercorrelates in the same way across different ethnic groups. Only when measurement invariance is established can inferences about the observed scores be made. Furthermore, only when measurement invariance is established, can comparisons between mean scores legitimately be made. In that sense, a test might be biased in terms of its construct if the factor structure is *not* invariant across groups. That is to say, if the assessment tool measures different constructs for one group compared to another group, such a finding would conceivably constitute bias of the instrument. Furthermore, comparisons between groups should not be made without first providing empirical evidence that the factor structure of the test is the same for all groups. In other words, mean scores cannot meaningfully be compared between groups without first establishing measurement invariance (Reynolds & Keith, 2013).

Up to this point, only a few studies have examined differential construct validity for cognitive and achievement tests across various ethnic groups; even fewer have followed the

appropriate statistical procedure of first establishing structural invariance before comparing the test scores of different ethnic groups. The following section summarizes the most important research studies that have explored construct validity of cognitive and achievement tests across ethnic groups (Black-Caucasian, Hispanic-Caucasian, and Hispanic-Black).

Studies of cognitive and achievement tests. The majority of results obtained from a variety of cognitive and achievement tests revealed similar factor structures for Black, Caucasian, and for Hispanic school-aged children and adults. This finding of similar factor structures characterizes the Woodcock-Johnson tests, the Kaufman tests, the Wechsler scales, and the Differential Ability Scales (DAS; Elliott, 1990) (e.g., Edwards & Oakland, 2006; Jensen, 1980; Kaufman 1990; Kaufman & Di Cuio, 1975; Kaufman, Kaufman, & McLean, 1995; Kaufman, McLean, & Reynolds 1991; Keith, Quirk, Schartzer, & Elliott, 1999; Nichols, 1972; Miele, 1979; Sandoval, 1982). However, those studies are few and far between, and most used simple correlation techniques (often coefficients of congruence) to compare the factor structures. Only a few researchers used the preferred method of confirmatory factor analysis (e.g., Keith, 1999; Keith et al., 1999; Kush, Watkins, Ward, Canivez, & Worrell, 2001; Trundt, 2013)

Using simple correlation techniques, several studies revealed factor invariance for school-aged Black and Caucasian children on individually-administered cognitive tests; typically, verbal and nonverbal factors identified for Caucasians resembled the verbal and nonverbal factors identified for Blacks (Kaufman & Wang, 1992; Miele, 1979; Nichols, 1972; Reschly, 1978). Invariance was also found for measures of *g* (Edwards & Oakland, 2006; Jensen, 1980; Miele, 1979). Jensen (1980) found factor invariance for the verbal and nonverbal factors for Black and Caucasian 6th through 8th graders. Using CFA methods, Kush and his colleagues (2001) found factorial invariance for *g* as well as for verbal and performance factors

on the WISC-III for Black and Caucasian school-aged children; Keith et al. (1999) established construct invariance for the DAS; and Trundt (2013), in her dissertation research, found construct invariance for the DAS-II across Black, Hispanic, and Caucasian subgroups. Results from group-administered tests also generally found factor invariance for school-aged Black and Caucasian children across the verbal factor, nonverbal factor, and *g* in national (Campell et al., 1973; Jensen, 1977; Jensen, 1980) and international (Rushton, Skuy, & Bons, 2004) samples.

Further, factor invariance of verbal, nonverbal, and *g* factors was established for a sample of adults (Kaufman & Wang, 1992) and for samples of preschoolers using individually-administered cognitive tests (Gutkin & Reynolds, 1981; Reschly, 1978; Kaufman & Hollenbeck, 1974). However, some studies revealed significantly different factor structures for Black and Caucasian school-aged children; Jensen (1980) was not able to establish factor invariance for 5th graders (Jensen, 1980), another study did not find factor invariance for Black and Caucasian adults (Kaufman, McLean, & Reynolds, 1991), and a third study did not find invariance for Black and Caucasian preschoolers tested on the McCarthy Scales of Children's Abilities (McCarthy, 1972) (Kaufman & DiCuio, 1975).

Few studies have investigated factor invariance across Black-Caucasian and Hispanic-Caucasian school-aged children using *achievement* tests. Using correlation coefficients, Nichols (1972) established measurement invariance for Caucasian and Black school-aged children for an individually-administered achievement test, and another study, using group-administered achievement measures, found factor invariance in a sample of Hispanic and Caucasian school-aged children (Hennessy & Merrifield, 1976).

Summary. In sum, the studies that have explored measurement invariance for cognitive and achievement tests have tended to find similar factor structures for Caucasians, Blacks, and

Hispanics. Nonetheless, some studies did not establish factor invariance for Black individuals for some cognitive tests (e.g., the McCarthy scales; Kaufman & DiCuio, 1975) or for some grade levels (Jensen, 1980). Perhaps most importantly, however, the majority of the previous studies lacked statistical sophistication because they were conducted before CFA and structural equation modeling were either available or very popular. They also tended to be conducted from a Wechsler-like verbal versus nonverbal factor structure rather than from a theoretical perspective. Only a few researchers have used CFA methods (e.g., Edwards & Oakland, 2006; Keith et al., 1999; Trundt, 2013); consequently, additional research on differential construct validity by ethnicity is needed.

The few studies that included both Black and Hispanic individuals in their samples generally found factorial invariance across the ethnic groups on individually-administered cognitive tests (Keith et al., 1999; Kaufman, Kaufman et al., 1995; Reschly, 1978; Sandoval, 1982) and, in one instance, on a group-administered achievement test (Hennessy & Merrifield, 1976). Using CFA methods, two other studies found that the factor structure of the DAS and the DAS-II was the same for both Hispanic and Black (as well as for Caucasian) school-aged children (Keith et al., 1999; Trundt, 2013).

Differential predictive validity. The relationship of intelligence to achievement dates back as far as the early 1900s, when E. L. Thorndike introduced the law of effect (Thorndike, 1911). According to Thorndike, the ability to learn is the most fundamental of all aptitudes; it is the capacity to learn from one's experiences (e.g., trial and error learning). Similarly, Alfred Binet, who developed the first intelligence test (Binet & Simon, 1905), recognized intelligence as the ability to acquire knowledge; for example, he tested children's accumulated knowledge, such as the ability to count from 1 to 10 or knowing the colors of the rainbow. Depending on how

much knowledge the child had acquired compared to his or her peers with the same years of school experience, the child's cognitive ability was determined (Wolf, 1973). Finally, the close link between achievement and intelligence is reflected in the name change of the Scholastic Aptitude Test to the Scholastic Achievement Test (SAT) (Mayer, 2011). The original name demonstrates that the test was initially meant to measure aptitude (ability); however, over time, researchers realized that the SATs were actually measuring the knowledge the student had accumulated in school – that is, achievement. In that sense, intellectual ability (the ability to learn) is tightly linked to achievement (what has been successfully learned). In short, intellectual ability helps the individual to obtain knowledge and thereby to learn and achieve.

Indeed, several cognitive abilities have been linked to specific achievement domains. For example, phonological awareness, a cognitive ability, is needed to perform well on word recognition/decoding (achievement) tasks (Bradley & Bryant, 1983). Similarly, research has shown that conceptual knowledge (specifically a sense for numbers, i.e., knowledge of a mental number line) is needed in order to solve arithmetic problems (Carroll, 1993). Due to the relationship between intelligence (ability to learn) and achievement (what has been learned) that intelligence tests are typically used as predictors and classifiers for academic achievement. Testing the differential predictive validity of these instruments is, therefore, one of the most important ways to examine test bias (Reynolds & Kaiser, 1990).

IQ and achievement tests usually yield correlation coefficients ranging from the mid-.60s to the mid-.80s (S.B. Kaufman et al., 2012; Naglieri & Bornstein, 2003). According to Urbina (2014), prediction is biased when (a) the magnitude of the correlation coefficient between the test scores and the outcome varies for different groups, and (b) when the test scores overestimate or underestimate (as measured by slope and intercept) the criterion performance of an individual

depending on his or her group membership. In terms of overprediction and underprediction, it is important to note that bias exists if the test puts the minority group at a disadvantage. For example, an overprediction of the minority groups' achievement would indicate that the test might not be accurate at predicting their achievement; however, it would not indicate bias, because the test would not be penalizing the minority group. As other researchers have done previously (but not recently) with the Wechsler Scales (e.g., Weiss & Prifitera, 1995), the present study explored test bias of the Kaufman tests using differential prediction bias.

Black-Caucasian studies of differential predictive validity. There are some older studies that investigated differential predictive validity as a means to assess test bias across ethnic groups, including Caucasian and Black school-aged children (e.g., Gutkin & Reynolds, 1981; Keith, 1999; Poteat, Wuensch, & Gregg, 1988; Weiss & Prifitera, 1995; Weiss et al., 1993). These studies usually used global scores of individually-administered tests of intelligence (exception: Keith, 1999) as the predictors (Wechsler scales, Woodcock-Johnson, Stanford-Binet, CAS), but sometimes used group tests (NNAT, Lorge-Thorndike). Criterion measures of achievement included the group-administered California Achievement Test (CAT; Clark & Tiggs, 1950), the Metropolitan Achievement Test (MAT, 1950), and the SATs, as well as the individually-administered WJ-R and WJ III. These studies found that IQ predicted achievement about equally well for Caucasian and Black school-aged children in terms of (a) the magnitude of the coefficients of correlation (Edwards & Oakland, 2006; Jensen, 1980; Oakland, 1983; Keith, 1999; Naglieri et al., 2005; Naglieri et al., 2000; Reynolds & Kaiser, 1990; Weiss et al., 1993; Weiss & Prifitera, 1995), and (b) the tests' slope and intercept (Reschly & Sabers, 1979; Reynolds & Kaiser, 1990; Weiss et al., 1993; Weiss & Prifitera, 1995).

Another set of differential predictive validity studies was conducted to predict performance in college. Predictors were either high school grades or group-administered tests such as the Differential Aptitude Test (DAT; Bennett, Seashore, & Wesman, 1947), School and College Ability Test (SCAT; Educational Testing Service, 1955), or California Test of Mental Maturity (CTMM; Sullivan, Clark, & Tiegs, 1963); criteria were grade point average in college. These studies produced mixed research results. Whereas some studies found that, on average, the tests predicted Black and Caucasian students' performance in college equally well in terms of the magnitude of the correlation coefficients (Centra, Linn, & Parry, 1970; Clearly, 1968; McKelpin, 1965), other studies found prediction bias (Baggaley, 1974; Boney, 1966; Maxey & Sawyer, 1981; Ramist, Lewis, & McCamley, 1994; Tracey and Sedlacek, 1984, 1985; Young, 1994).

Results from the SATs (basically an achievement test) revealed similar results when predicting academic achievement across ethnicities. Whereas one study found no bias between the Caucasian and Black correlation coefficients (Cowen & Fiori, 1991), other researchers did find differences in the magnitude of the correlation coefficients so that it put Black students at a disadvantage (Bridgeman, McCamley-Jenkins, & Ervin, 2000). For example, across all studies – cognitive tests predicting achievement as well as achievement tests (e.g., the SATs) predicting achievement – when prediction bias was found, the magnitude of the correlation was usually *weaker* for minority students. Those findings indicate that the tests were better predictors of college success for Caucasians than Blacks; in that sense, the tests were biased against Blacks.

Mattern and Patterson (2013) detected slope bias in their large sample of 475, 000 students, using the SATs and high school GPA as predictor variables and first year college GPA as the criterion. In that study, in which the authors corrected for a variety of methodological artifacts (such as population effect size, percentage of minority students included in the study,

and restriction of range), the Black regression line was found to consistently lie below the Caucasian regression line so that the performance of Black students was consistently overpredicted.

Another set of relevant studies was conducted in industrial, rather than school, settings to examine employment testing. These studies used group-administered tests as predictor measures—such as the General Aptitude Test Battery (GATB), developed by the U.S Employment Service in 1945 for personnel selection, or other paper-and-pencil aptitude tests (some of which were specifically designed and put together for a particular study, e.g., the ETS-U.S. Civil Service Commission Six-year Study; French, Ekstrom, & Price, 1963). Criteria in these studies were measures of job performance (work samples, job-knowledge tests, or supervisor rating scale). These industrial studies either found no bias (e.g., Nijenhuis & Van der Flier, 2000) or—more typically—they found bias in the intercept of the regression lines (Campell, Crooks, Mahoney, & Rock, 1973; Centra, Linn, & Parry, 1970; Chou & Huberty, 1990; Cole, 1981; Cleary, 1968; Crooks, 1972; Davis & Temp, 1971; Davis and Kerner-Hoeg, 1971; Elliott & Strenta, 1988; Farr et al. 1977; Houston & Novick, 1987; Humphreys, 1986; Hunter, Schmidt, & Rauschenberger, 1984; Jensen, 1980; Kallingal, 1971; Kuncel & Sackett, 2007; Lewis, & McCamley-Jenkins, 1994; Linn, 1978; Pfeifer & Sedlacek, 1971; Rotundo & Sackett, 1999; McCornack, 1983; Nettles, Theony, & Gosman, 1986; Rushton & Jensen, 2005; Ramist, Temp, 1971; Sackett, Schmitt, Ellingson, & Kablin, 2001; Sackett & Wilk, 1994; Schmidt & Hunter, 1981, 1998; Wilson, 1970).

For those studies that found intercept bias, the criterion outcome for Blacks (and Hispanics if they were included in the analysis) was *overpredicted*. That is to say, when significant ethnic bias was found, the ethnic minority group achieved *lower* than was predicted

for them based on their ability test scores. That finding is *opposite* to the general notion that intelligence tests are biased against Blacks or Hispanics. That general notion posits that a biased (unfair) ability test will *underestimate* ethnic minority students' ability and, therefore *underpredict* their achievement. That is not what happened, except in one study that found that the Verbal (*G*c) and Quantitative (*G*q) portion of the SCAT significantly underpredicted the performance of Black students in special education classes (Bowers, 1970).

Another set of studies was conducted in the armed forces, using aptitude tests such as the Armed Forces Qualification Test (AFQT; Brandt & Burke, 1950), the Army General Classification Test (AGCT; U.S. Army, 1945), or the Wonderlic Personnel test (Wonderlic, 1945), to assign military personnel to training programs and types of occupations within the military. These differential predictive validity studies revealed that if bias was found it was either in the intercept (e.g., Guinn, Tupes, & Alley, 1970; Foley, 1971; Thomas, 1975) or in the slope (Farr et al., 1971; Thomas, 1972). If intercept bias was found, it was again reflective of an overprediction of minority groups' ability. When slope bias was found the shape of the slope resulted in the underprediction of lower achieving Black individuals and the overprediction of higher achieving Black individuals (Fox, Taylore, & Caylor, 1969; Jensen, 1980). In that sense, whereas high achieving Black individuals performed worse than would have been expected based on their test scores, lower achieving Black individuals' performed better than the test predicted.

Hispanic-Caucasian studies of differential predictive validity. Few studies have investigated Hispanic-Caucasian differential prediction bias. Results from a variety of individually-administered and group-administered tests of intelligence and achievement revealed no significant differences in the regression equations of Caucasian and Hispanic students,

providing support for the fairness of the test in terms of the magnitude of the regression lines or the slopes and intercepts (Jensen, 1974; Keith, 1999; Naglieri et al., 2007; Naglieri et al., 2000; Reschly & Sabers, 1979; Weiss, et al., 1993; Weiss & Prifitera, 1995).

Differential validity studies aimed at predicting college success produced diverse results: (a) Some studies found no bias (Maxey & Sawyer, 1981; Pennock-Román, 1990); (b) one study found that the regression line was weaker for Hispanics compared to Caucasian students (Goldman & Richards, 1974; Ramist, Jenkins, & Lewis, 1993); and (c) some studies reported bias with regards to the intercept, so that it overpredicted Hispanics' performance (McCornack, 1983; Ramist et al., 1994). The aforementioned studies used high school GPA or group-administered tests such as the Non-Cognitive Questionnaire (NCQ; Tracey & Sedlacek, 1984) as predictor variables, with college GPA or SAT results serving as the criteria. Mattern and Patterson (2013) detected slope bias for Hispanic students. Similar to the results for Black students, the Hispanic regression line was slightly below the Caucasian students' regression line, indicating overprediction of Hispanic students' achievement. Results from the SATs revealed similar findings. Whereas two studies found no prediction bias when using the SATs as predictor variables and college success as the criteria (Bridgeman, McCamley-Jenkins, & Ervin, 2000; Cowen & Fiori, 1991), one study found that the Hispanic regression line was weaker as compared to the Caucasian regression line (Goldman, & Richards, 1974).

Summary. In sum, research that investigated the differential predictive validity of cognitive tests and the achievement-oriented SATs for Caucasians versus Blacks or Hispanics revealed mixed results. Bias was sometimes found, but such evidence of bias tended to occur in studies focused on differential prediction of college success or job success, neither of which is directly related to the questions of interest in the present dissertation. The studies that are

relevant, concerning the prediction of academic achievement for school-aged children and adolescents based on clinical tests of intelligence such as Wechsler's scales or the Woodcock-Johnson (e.g., Gutkin & Reynolds, 1981; Keith, 1999; Naglieri et al., 2007; Weiss et al., 1993), tended to find no evidence of predictive bias against Blacks or HispanicsIf prediction bias was found it was in the intercept and so that a common regression line overpredicted the achievement outcomes of minority group children. However, it is important to note that some of the studies summarized above are several decades old. Many intervention programs, such as Head Start or the Carolina Abecedarian project, were introduced in the 1960s and 1970s. Such programs have shown to improve at risk children's cognitive ability and achievement in school. Thus, it is important to note that results from older studies, when availability of intervention programs was not as readily available, might not generalize to how minority group children achieved after the introduction of such programs.

The author was unable to identify studies in the literature that have investigated differential predictive validity between Blacks and Hispanics with regards to slope and intercept bias. However, a few studies have compared the correlation coefficients between the two ethnic groups and found either (a) no differences between coefficients obtained for Blacks versus Hispanics (Bridgeman, McCamley-Jenkins, & Ervin, 2000; Elliott, 1990; Jensen, 1974; Maxey & Sawyer, 1981) or (b) the strength for the prediction for reading achievement among Hispanics was weaker compared to Blacks' correlation coefficients (Naglieri, & Ronning, 2000; Ramist, Lewis, & McCamley, 1994; Weiss, Prifitera, & Roid, 1993). One study found slightly stronger correlation coefficients for Hispanic students relative to Black students when WISC-III FSIQ predicted WIAT reading composite (Weiss & Prifitera, 1995). For example, the Hispanic regression line produced correlation coefficients of .76 compared to .69 for Blacks. The

regression line for Blacks in writing, however, was stronger compared to the Hispanic regression lines, producing a correlation coefficient of .67 versus .58 for Hispanics. And, as indicated previously, Keith (1999) found that the influence of *G*c (and *G*s) on the WJ-R was significantly stronger on reading achievement (reading comprehension) for Hispanic students grades 5-8, as compared to Black and Caucasian students.

Present Study

Statement of the Problem

The population in the U. S. continues to become more ethnically diverse. The increase in diversity is particularly prominent among younger generations, such as school-aged children. Even further, it is school-aged children from minority backgrounds that most frequently are referred for psychological testing (Weiss et al., 2006). The disproportionate overrepresentation of minority children in special education classes and underrepresentation in gifted programs, understandably, has researchers, clinicians, and scholars concerned whether tests are biased against minority groups. Many clinicians and researchers have compared mean scores between minority and Caucasian majority groups. Some scholars argue that the persistent differences in mean scores found in favor of Caucasian students provide evidence for the notion that tests are, in fact, biased and discriminate against minority groups. However, the simple detection of mean score differences alone does not necessarily constitute bias of the test (Meredith, 1993; Reynolds & Keith, 2013; Reynolds & Lowe, 2009). Gender differences studies have shown that mean score differences can sometimes reflect true differences between groups. For example, several researchers have shown that females outperform males in their writing achievement (e.g., Reynolds, Scheiber, Hajovski, Schwarz, & Kaufman, in press; Scheiber, Reynolds, Hajovski, &

Kaufman, 2015). Other explanations could be that mean score differences are the results of environmental or linguistic differences (Weiss et al., 2006).

This present dissertation aimed to explore the psychometric properties of the Kaufman achievement and intelligence tests. Instead of only comparing mean scores differences, the present dissertation used statistically more sound techniques to explore the psychometric properties of the test instruments and thereby explored test bias of the tests via the exploration of (a) differential construct validity, and (b) differential predictive validity (Reynolds, & Lowe, 2009). Those two approaches needed to be conducted alongside the comparison of mean score differences to fully understand the depth and breadth of a test's potential bias. Indeed, according to Reynolds and Keith (2013), mean score differences should not be compared without first establishing differential construct validity of the test—namely, to provide evidence that the test measures the same constructs for majority and minority samples before analyzing ethnic differences on these constructs. Surprisingly, this procedure has rarely been followed in the ethnic-difference literature, although it has been followed in recent studies of gender differences (e.g., Keith, & Reynolds, 2010; Reynolds, Keith, Ridley, & Patel, 2008). Jensen (1980) and Edwards and Oakland (2006) established measurement invariance of the *g* factor prior to comparing the global abilities of Caucasians and Blacks, but those studies are the exceptions. The present dissertation aims to fill this important gap in the literature.

Based on a strong theoretical model built on CHC theory, this dissertation explored bias of a modern, individually-administered tests of cognitive ability (KABC-II) and academic achievement (KTEA-II) by using three definitions of test bias: a) differential construct validity, b) comparison of mean scores, and c) differential predictive validity.

Using the CHC-based factor structure of the KABC-II and KTEA-II, as S.B. Kaufman and his colleagues outlined it in 2012, the present dissertation used state-of-the-art methodology to explore the construct validity of two contemporary instruments. It is important to note that with the exception of the Keith et al. (1999), Trundt (2013), and Edwards and Oakland (2006) studies, no other studies of differential construct validity by ethnicity were conducted in the past 20 years. Accordingly, most studies have used simple correlation techniques (e.g., coefficients of congruence) to show that the same set of factors emerged for Blacks, Hispanics, and Caucasians. The use of such relatively primitive analyses limits the meaningfulness of their findings; their conclusions of "no difference" in the constructs identified for Caucasians, Blacks, and Hispanics cannot be taken at face value. Few studies have used more sophisticated methods, such as structural equation modeling (e.g., Keith, 1999).

Second, not many of the previous studies of differential construct validity have established construct invariance for different CHC factors (exceptions Kaufman et al., 1995; Keith et al., 1999; Trundt, 2013); most studies that explored construct invariance tended to focus on *g* (Edwards & Oakland, 2006; Jensen, 1980) or the performance and verbal measures (e.g., Kaufman & DiCuio, 1975; Kaufman et al., 1991). Kaufman et al. (1995) examined the construct invariance of *Gf* and *Gc* for Caucasians, Blacks, and Hispanics—two of the factors in the CHC model—but those researchers were operating out of the Cattell-Horn *Gf-Gc* theory rather than the more complex model that underlies contemporary CHC theory. Studies on mean scores demonstrate that the magnitude of the ethnic group gap differs depending on CHC factor, providing evidence for the importance of investigating each factor separately. The comparison of the CHC factor structure for Caucasians, Blacks, and Hispanics demonstrates whether the components of intelligence and achievement constructs are measured equally well across

different ethnic groups. The inclusion of achievement variables (*Grw* and *Gq* in CHC terminology) is an especially important contribution of this study. Numerous previous studies have examined mean ethnic differences on *intelligence* tests and have compared the factor structure of intelligence tests; few previous studies have addressed these kinds of ethnic differences on tests of *achievement*.

Differential predictive validity was used to assess whether the cognitive variables of the KABC-II predict the achievement variables on the KTEA-II for Blacks and Hispanics equally well as they do for Caucasians. Intelligence tests have traditionally been used to assess future achievement outcome. The bulk of research on predictive validity up to this point has shown that tests are usually not biased, especially when conducted with the kind of individually-administered instruments used in this study. However, as with investigations of differential construct validity, most previous studies of differential predictive validity are more than 15 or 20 years old. Only Keith's (1999) study of the WJ-R evaluated differential predictive validity. And, Edwards and Oakland's (2006) study of the WJ III is among the very few that evaluated both structural invariance and mean differences by ethnicity (though the latter study did not include Hispanics).

The present study was the first to evaluate test bias using three different methodologies and to apply these methods to three ethnic groups; it is one of the very few studies (alongside Keith, 1999, and Weiss & Prifitera, 1995) to explore the differential predictive validity based on CHC factor scores. Cognitive variables based on CHC theory have been found to be particularly accurate at predicting future academic achievement because CHC factors provide a more detailed, comprehensive understanding of the individual's abilities (Floyd, Evans, and McGrew, 2003; Keith, 1999). Indeed Keith (1999) found that CHC factors predict achievement above and

beyond *g*. The fact that the present dissertation used CHC based factors (in the sense that they were identified by S.B. Kaufman et al., using sophisticated CFA methodology) provides clinicians and researchers with a better understanding of where interventions might be needed and where exactly the bias might lie.

Furthermore, the great majority of differential predictive validity studies on ethnic bias have been conducted with group-administered tests. There is a surprising scarcity of research conducted in the area of differential prediction across ethnic groups on individually-administered, clinical test of intelligence and achievement, although it is precisely those tests that are used to determine admission to special programs, schools, and (often) employment. Finally, there is a scarcity of studies that have compared the differential construct validity and differential predictive validity of Black versus Hispanic students.

The present dissertation addressed all of the above gaps in the literature, and thereby advanced the methodological sophistication and clinical meaning in the research on ethnic group differences in intelligence and achievement.

Research Questions

The three research questions that follow each address the larger question of whether the Kaufman tests are biased by ethnicity (Blacks, Hispanics, Caucasians). Each question focuses on a different definition of bias:

1. Using Confirmatory Factor Analysis (CFA), is the factor structure of the Kaufman tests invariant for separate groups of Blacks, Hispanics, and Caucasians in grades 1-12, using the CHC-based factor model developed by S. B. Kaufman et al. (2012) as the criterion?

2. Do the three ethnic groups differ significantly in their mean scores on the CHC latent variables that underlie the Kaufman tests, based on CFA (when factor invariance is found) and/or on the subtests that compose these factors (when invariance is *not* found)?

3. On the Kaufman tests, is there predictive validity bias across the different ethnic groups? Do the general factor (*g*) and five CHC-based cognitive factors as measured by the KABC-II (*G*f, *G*c, *G*lr, *G*sm, *G*v) predict the KTEA-II achievement composites (reading, writing, and math) equally well (magnitude of the coefficients) for Blacks, Hispanics, and Caucasians?

Theoretical Justification for the Study

The present dissertation was one of the first test bias studies to base its results on the CHC classification system of intelligence, using the factor model developed by S. B. Kaufman et al. (2012), who developed a factor model of the KABC-II and KTEA-II based on sound statistical and theoretical principles. The two measures have not been selected in order to validate their reliability as tests of intelligence and achievement, but they were selected because both instruments assess the constructs of interest (intelligence and achievement) based on CHC theory, thereby providing evidence for the technical adequacy and theoretical soundness of the instruments. Indeed, the CHC model offers researchers and clinicians statistical and theoretical dependability. CHC theory of cognitive ability is one of the most reliable and one of the most widely researched classification systems of human cognitive ability.

Perhaps most importantly, however, CHC is the first theory to bridge the gap between statistical soundness and clinical applicability. It allows researchers and clinicians alike to draw meaningful conclusions from test results, thereby merging theory with practice (McGrew, 2009). This benefit arises because CHC theory offers a standardized way of describing cognitive

ability—"a shared vocabulary" (McGrew, 1997) – which facilitates interactions between clinicians and researchers (Flanagan et al., 2007). Furthermore, CHC theory explains the constructs of intelligence and achievement above and beyond the *g* factor, thereby allowing clinicians to better target intervention strategies (Flanagan, Fiorello, & Ortiz, 2010; Keith, 1999; Wendling & Mather, 2009).

Indeed, the present dissertation bridged the gap between theoretical soundness and clinical applicability. Whereas it has been common practice to focus on *g* when it comes to the assessment of learning (McDermott, Fantuzzo, & Flutting, 1990; McDermott & Glutting, 1997), research has shown that more specific abilities (e.g., *Gc, Gv*) do, in fact, influence achievement performance above and beyond *g* (McGrew, Flanagan, Keith, & Vanderwood, 1997; Keith, 1999). For example, *G*c has been found to strongly influence reading comprehension in additional to *g*. In fact, Keith (1999) has shown that models that include specific abilities provide a better prediction of achievement outcomes for all three ethnic groups (Caucasians, Hispanics, and Blacks) for children from elementary through high school. In that sense, understanding school-aged children's performance in more detail – meaning, above and beyond the general intelligence factor *g* – allows for a more comprehensive understanding of their achievement abilities which can enhance and specify learning intervention strategies.

Clinical Justification for the Study

The present study was not only the first to explore factorial invariance and differential predictive validity based on a strong theoretical model, but its results also have important clinical implications. All across the country, clinicians and practitioners administer well-known, clinical tests of intelligence and achievement to children from ethnic minority groups. Results can have profound consequences, as scores often determine access or denial to special programs, services,

employment, and schooling. Assessments are conducted based on the assumption that tests are equally valid and reliable across different groups. However, there is virtually no empirical support, other than the proportional inclusion of ethnic minority groups in national standardization samples, that justifies the usage of those measures with minority groups.

For decades, scholars, researchers, and clinicians have debated about the fairness of cognitive and achievement tests based on mean score differences. Indeed, there is a plethora of studies that have compared mean scores between various ethnic groups; however, essentially none of those researchers has investigated construct invariance of the tests, which is an indispensable prerequisite that needs to be determined before the comparison of, and subsequent debate about, mean scores and their differences. Whereas many scholars and clinicians argue about the meaning of mean ethnic score differences on clinical tests, nobody has yet to demonstrate that the same constructs are being compared. Only once construct invariance has been established, can meaningful conclusions from mean scores be drawn. Furthermore, nobody has established factor invariance of CHC factors across ethnic groups, even though (a) research has shown that the magnitude of the differences depends heavily on the specific CHC factor measured, and (b) the most frequently used clinical tests of achievement and intelligence are either derived from CHC theory or interpreted from a CHC orientation (Flanagan & Kaufman, 2011; Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005).

Similarly, the predictive validity of the tests used in schools and other settings has not been verified to be equally valid for different ethnic groups for contemporary tests of achievement and intelligence based. Few studies have explored the differential predictive validity of individually-administered, clinical tests of intelligence, although it is those same tests that are typically used to allow students access to special classes and programs. There is no

empirical evidence that uses state-of-the-art methodology and contemporary measures of intelligence and achievement to verify that those tests actually predict what they are supposed to predict for different ethnic groups.

It is for those reasons that results of this present dissertation have important clinical implications. If those tests, which are so commonly used to determine access to special education classes, giftedness classes, and school admission, do not actually measure what they are supposed to measure and do not actually predict what they are supposed to predict for minority students, serious changes in the usage of those tests will be necessary. Alternatively, if results of this present dissertation do verify construct invariance and differential predictive validity, then researchers, clinicians, and practitioners have the empirical evidence needed to allow for continuous usage of those measures in clinical test settings with minority groups.

Methods

Participants

The data come from the group of children and adolescents in the standardization samples of the KABC-II and KTEA-II who were administered both instruments. The sample is large (N = 2,001) and stratified on key background variables (gender, parent educational attainment, ethnicity, age, geographic region) according to 2001 U.S. Census Data.

The total sample used for this dissertation includes 986 females (49.3%) and 1015 males (50.7%) and ranged in age from 6 years 0 months to 19 years 1 month (mean = 11.6, SD = 3.4). The sample comprised 312 Blacks (15.6%), 376 Hispanics (18.8%), and 1313 Caucasians (65.6%); participants from other ethnic backgrounds (e.g., Asian and Native American) were excluded from this study. Overall, 300 (15.0%) mothers completed Grade 11 or less, 661 (33.0%) obtained a High School degree or GED, 589 (29.4%) completed some College or an

Associates Degree, and 451 (22.5%) held at least a Bachelor's degree (father's education was used if maternal data were unavailable). Overall, 288 (14.4%) of the participants were from the Northeast, 509 (25.4%) came from the North Central region, 725 (36.2%) were from the South, and 479 (23.9%) resided on the West Coast.

The separate demographics for each ethnicity are shown in Table 1. The demographic subdivisions of the three ethnicities match 2001 U.S. Census Data on important background variables, including age, region, gender, and SES (parents' education). As shown, the three ethnic groups are matched almost exactly on the variables of age and sex. They differ, however, in geographic region and SES because of real differences in the U.S. Census for Blacks, Caucasians, and Hispanics. Consequently, in order to approximate U.S. Census in terms of geographic region, in this study 71% of the sample's Black participants were from the South, as compared to only 31% of Caucasians and 24% of Hispanics. Similarly, in order to approximate 2001 U.S. Census data, 59% of the sample's Hispanic participants were collected from the West, whereas only 8% of Blacks and 18% of Caucasians were from that region. The ethnic subdivision also stresses another important issue: a significantly larger percentage of Caucasian participants came from higher SES (as measured by parent educational attainment). That is to say, 27% of Caucasian participants' mothers had acquired at least a Bachelor's degree, as compared to 19% of Black participants' mothers and 10% of Hispanic participants' mothers. The demographic division again reflects 2001 U.S. Census Data, but highlights the importance of controlling for SES when conducting analyses of ethnic group differences.

In this study, SES was defined as mother's educational attainment (or father's if mother's was not available). The variable was categorical and distinguished between four educational levels: 1. Grade 11 or less; 2. High School degree or GED; 3. Some College or an Associates

Degree; 4. At least a Bachelor's degree. Per Dr. Mark Daniel, (personal communications, March 31, 2015), Project Director of the KABC-II and KTEA-II, the reason for using mother's education is the high percentage of single-parent households. He emphasized that mother's educational attainment is almost always possible to obtain, which, in turn, enabled the publishers to use U.S. Census data as a target. Dr. Daniel also highlighted that in the past, when data for both mothers and fathers were available for almost the entire sample, it was found that mother's education and father's education correlated about equally well with test scores. Even though the average of the two correlates usually slightly higher, the difference in correlation is negligible; thus, given the availability of data it was more sensible to use mother's educational attainment when collecting the data for the KTEA-II and KABC-II.

Despite the regional differences for the ethnic groups, it was not necessary to control for geographic region in this study because previous research has shown little systematic relationship between region and IQ (Kaufman, 1973b; Kaufman & Doppelt, 1976; Kaufman, McLean, & Reynolds, 1988).

The total sample was used to answer the first two questions posed in this dissertation: 1—Is the factor structure of the Kaufman tests invariant for Black, Hispanic, and Caucasian children; and 2—Do the ethnic groups differ significantly in their mean scores? For the first two questions, the sample was not divided into different age groups due to a lack of power. According to Maede and Bauer (2007) at least 200 participants per group are necessary for adequate power when conducting confirmatory factor analysis. For the third question (Is there predictive validity bias across the different ethnic groups?), participants were subdivided into the following three groups: grades 1-4 (n = 724), 5-8 (n = 743), 9-12 (n = 534). Grades 1-4 included 724 participants, 357 (49%) of whom were females, and the sample comprised 119 (16.4%)

Blacks, 150 (20.7%) Hispanics, and 455 (62.8%) Caucasians. Grades 5-8 comprised 743 participants, including 364 females (49%), 119 Blacks (16.0%), 137 (18.4%) Hispanics, and 487 (65.5%) Caucasians; at grades 9-12 the total sample comprised 534 participants, including 269 (50.4%) males, 74 (13.8%) Blacks, 89 (16.6%) Hispanics, and 371 (69.5%) Caucasians.

The sample included individuals that are stratified on number of background variables, including exceptionality — a term used to identify patterns of strengths and needs common to groups of students — to truly reflect the US population (Department of Education and Early Childhood Development, 2015). Excluded were those children that were non English-speaking, had ever been institutionalized, or had physical or perceptual deficiencies that would prevent them from completing the tests.

Table 1

Demographic Characteristics of the Separate Samples of Caucasians, Blacks, and Hispanics

Demographic Characteristics	Caucasians (*n* = 1313)	Blacks (*n* = 312)	Hispanics (*n* = 376)	Total
Grade groups (number of participants)				
1 – 4	455	119	150	724
5 – 8	487	119	137	743
9 – 12	371	74	89	534
Total	1313	312	376	2001
Gender				
Male	50.2%	50.3%	52.9%	
Female	49.8%	49.7%	47.1%	
Age				
Mean	11.7	11.4	11.3	
Standard Deviation	3.4	3.5	3.5	
SES (Parent Educational Attainment)				
Grade 11 or less	6.5%	19.6%	41.0%	
High School Graduate/GED	34.4%	28.5%	31.9%	
Some College/Tech/Associates Degree	32.1%	33.0%	17.0%	
Bachelor's Degree or higher	27.0%	18.9%	10.1%	
Region				
East	18.5%	10.3%	3.5%	
North Central	32.3%	11.2%	13.3%	
South	31.6%	70.5%	23.9%	
West	17.6%	8.0%	59.3%	

Measures

The KABC-II. The Kaufman Assessment Battery for Children, Second Edition (KABC-II; Kaufman & Kaufman, 2004a) is an individually-administered test of intelligence designed for ages 3–18. The test is based on two theoretical models – Luria's (1966, 1970, 1973) neuropsychological model and CHC theory (Carroll, 1997; Flanagan, 2000; Horn & Noll, 1997). Regardless of whether the Luria model or CHC theory is used, the same subtests are administered (although the Knowledge/*G*c subtests are supplementary for the Luria model). The KABC-II comprises 18 subtests (including both core and supplementary subtests). In the present study all core subtests were used in addition to two supplementary subtests (Hand Movement and Expressive Vocabulary). From the CHC theory standpoint, the KABC-II produces a global score, the Fluid-Crystallized Index (FCI), and five CHC broad ability scores (short-term memory, visual processing, long-term storage & retrieval, and fluid reasoning). From the standpoint of the Luria model, the KABC-II produces a global score that emphasizes mental processing, the Mental Processing Index (MPI), as well as four scale scores (sequential processing, simultaneous processing, learning ability, and planning ability). The KABC-II also generates a nonverbal index (NVI) to measure cognitive and processing abilities with minimal verbal involvement. The NVI consists of four to five supplementary subtests and their instructions and responses can be communicated via gestures. All indexes have a mean of 100 and a standard deviation of 15.

Reliability. Internal-consistency reliability, as measured by split-half coefficients, is generally high for the KABC-II. For the global scales coefficients were .97 (FCI), .95 (MPI), and .92 (NVI) at ages 7-18. Similarly, on the scale level, the KABC-II also demonstrates evidence for strong internal-consistency, producing coefficients ranging from the high .80s for

the Simultaneous/*G*v, the Planning/*G*f, and the Sequential/*G*sm processing scales to the low .90s for the Learning/*G*lr and the Knowledge/*G*c scales. On the subtest level, reliability coefficients for ages 7-18 ranged from .74 for Gestalt Closure and .77 for Story Completion to .90 for Pattern Reasoning and .93 for Rebus (mean coefficient = .85). Descriptions of each KABC-II scale and subtest follow in Table 2.

Table 2

K-ABC II scale and subtest descriptions

Scale	What the scale measures	Name of KABC-II subtests	Description of the KABC-II subtest
1. Sequential/ Short-Term Memory (*G*sm)	The ability to maintain information in primary memory and immediately reproduce it either in the same sequence or after performing simple manipulations with it.	Number Recall*	Measures the child's ability to repeat previously read numbers in the same order they were read
		Word Order*	Requires the child to point to objects in the same order they were verbalized by examiner (sometimes with an interferences task)
		Hand Movement	Requires the child to copy the examiner's exact order of taps
2. Simultaneous/ Visual Processing (*G*v)	The ability to perceive complex patterns and visualize how they would look like when transformed	Rover*	Measures the child's ability to use frontal lobe executive functioning to find the quickest path for Rover to his bone
		Triangles* (core ages 6-12)	This is a visual motor construction test that requires the child to assemble triangles to match a picture
		Block Counting* (core ages 13-18)	Assesses the ability to visualize objects in 3-Dimensions by counting the exact number of blocks on a picture
3. Learning/Long-Term Storage and Retrieval (*G*lr)	The ability to store information and retrieve	Atlantis*	Requires the child to learn nonsense names for fish, plants

	it at a later point in time; the ability to learn new information		and shells and then point to each picture when named thereby measuring the child's ability to learn new information
		Rebus*	Assesses the child's ability to learn new information by learning concepts paired with each rebus and then reading sentences composed of the rebuses
4. Planning/Fluid Reasoning (*G*f)	The ability to reason logically, form concepts, and solve problems using new information	Story Completion*	Measures executive functions and planning by having the child select missing picture to complete a picture story
		Pattern Reasoning*	Measures executive functions of frontal lobe by having the child select the correct stimulus to complete a pattern
5. Knowledge/ Crystallized Knowledge (*G*c)	The accumulated breath and depth of knowledge of one's culture and the ability to understand spoken language and verbalize thoughts clearly.	Verbal Knowledge*	Assesses the child's storage of general information and vocabulary by having the child select a picture that corresponds to a vocabulary word or general knowledge question
		Riddles*	Measures fluid reasoning and crystallized ability by having the child points or names concrete/abstract verbal concept provided by examiner
		Expressive Vocabulary	Assesses crystalized and expressive language by having the child verbalizes name of pictured object

(* indicates core subtest)

Note--Data are from KABC-II manual (Kaufman & Kaufman, 2004a, Table 8.1)

Test-retest reliabilities for children and adolescents ages 7-12 (n = 82) and 13-18 (n = 61) for the global scores (MPI, FCI, and NVI) are high, ranging from .87 to .94 over a 4-week interval (Kaufman & Kaufman, 2004a, Table 8.3). At the scale level, test-retest reliabilities ranged from .76 to .88 at ages 7-12 (mean = .80) and from .78 to .95 (mean = .85) at ages 13-18.

For the two age groups combined, stability coefficients ranged from .77 for Simultaneous/*Gv* to .92 for Knowledge/*G*c (Kaufman & Kaufman, 2004a, Table 8.3).

Validity. As outlined in the manual (Kaufman & Kaufman, 2004a, chapter 8), confirmatory Factor Analysis (CFA) was employed to confirm the factor structure of the KABC-II . The procedure was used to evaluate the best groupings of subtests and scales and verify that these groupings supported the organization of subtests into the five designated scales. Separate factor models were evaluated for several age groups (age 4, 5-6, 7-12, 13-18). Hierarchical CFA was employed so that the procedure was started with a one-factor model and then additional factors were added subsequently. After each factor was added, the model was evaluated for possible improvements of the model fit. Based on the final model, core subtests were identified as those subtests with the highest loadings on their appropriate factors. All subtests' loadings on *g* were relatively high, suggesting that the abilities are strongly influenced by *g*. Sequential/*G*sm had the weakest correlation with *g* and Planning/*G*f had the strongest correlation with *g*, which is consistent with CHC theory (Kaufman & Kaufman, 2004a; Flanagan et al., 2013). Simultaneous/*G*v and Planning/*G*f produced strong intercorrelations at ages 7-18; however, separating the two factors resulted in a statistically significant improvement in model fit. The results suggest that Simultaneous/*G*v and Planning/*G*f—although strongly correlated--are, in fact, distinct factors. The final model, as noted in the manual (Kaufman & Kaufman, 2004a, Figures 8.1 & 8.2), that examined the construct validity of the core subtests, had excellent fit for all age levels (CFI = .997-.999; RMSEA = .025 - .055).

In addition to the construct validity of the KABC-II, as demonstrated by CFA, the KABC-II has been shown to correlate well with other measures of intelligence. On the global level, the FCI and the MPI of the KABC-II correlated highly with the Full Scale IQ (FSIQ) of

the WISC-IV (.89 and .88, respectively). The NVI correlated .82 with the FSIQ of the WISC-IV. Similarly, on the scale level, the KABC-II Sequential/*G*sm scale correlated .71 with the Working Memory Index of the WISC-IV; the Simultaneous/*G*v scale correlated .66 with the Perceptual Reasoning Index of the WISC-IV; and the Knowledge/*G*c scale correlated .85 with the Verbal Comprehension Index of the WISC-IV at ages 7-16 years old. In addition, the KABC-II has been shown to measure the general intelligence factor (*g*) in the same way as do other major tests of cognitive ability, namely the WISC-IV, WJ III, and *Differential Ability Scales — Second Edition* (DAS-II; Elliott, 2007) (Floyd, et al., 2013; Reynolds, et al., 2013).

The KTEA-II. The Kaufman Test of Educational Achievement, Second Edition (KTEA-II) Comprehensive Form is an individually-administered test of achievement for children, adolescents, and young adults ages 4.5 to 25 years old. The KTEA-II consists of 14 subtests and is normed both on grades 1-12 and ages 4.5 through 25 years. Eight of the 14 subtests measure achievement in four domains: Reading, Math, Writing, and Oral Language. The remaining six subtests constitute four additional achievement domains: Sound-Symbol, Decoding, Oral Fluency, and Reading Fluency. Six of the subtests that constitute the four achievement domains (Reading, Math, Writing, and Oral Language) produce the Comprehensive Achievement Composite at ages 4.5 through 25 years old.

Reliability. As the KTEA-II has two Forms (Form A and B), alternate-form reliabilities were calculated. Both Form A and Form B were administered to a total of 221 children and adolescents, divided into three grade groups (ages 4.5-grade 1; grades 2-6; and grades 7-12). The alternate-form reliabilities were substantial, ranging from the low .80s for the Oral Language composite to the high .80s and mid .90s for the Reading, Math, Writing domains for grades 7-12. Similarly, alternate-form reliabilities for grades 2-6 produced correlations ranging from the high

.60s (for Oral Language) to the high .80s and low .90 for the Reading, Writing, and Math domains (Kaufman & Kaufman, 2004b, Table 7.5).

Internal-consistency reliability (split-half) coefficients averaged .97 for the Comprehensive Achievement Composite, .96 for the Math and Reading composites, .93 for the Written Language composite, and .87 for the Oral Language composite (Kaufman & Kaufman, 2004b, Table 7.1). Table 3 presents descriptions of each KTEA-II subtest for grades 1-12.

Table 3

KTEA-II descriptions for subtests

Domain Composite	Name of KTEA-II Subtest	Description of KTEA-II Subtest
1. Reading		
	Letter Word Recognition	Requires the student to pronounce words of increasing difficulty
	Reading Comprehension	Requires the student to read single words, simple instructions, or passages and then point to the response, perform an action, or answer literal or inferential questions
2. Writing		
	Written Expression	The student is asked to complete age-appropriate writing tasks, including writing sentences from dictation, adding punctuation, completing sentences, and writing essays
	Spelling	Requires the student to write single letters that represent sound (for the younger children) or to spell orthographically (ir)regular words of increasing difficulty (for the older children)
3. Oral Language		
	Listening Comprehension	The student is asked to listen to specific parts played on a CD and then answer questions about it

	Oral Expression	The student is asked to complete specific speaking tasks in the context of a real-life scenario
4. Math		
	Math Concepts and Applications	Requires the student to respond verbally to questions that focus on real-life mathematical problem solving
	Math Computation	Assesses the student's ability to write down solutions to math problems

(Data are from the KTEA-II Comprehensive manual, Kaufman & Kaufman, 2004b, Table 7.1)

The internal-consistency of the KTEA-II is strong. Split-half reliability coefficients for grades 1-12 ranged from .73 for Associational Fluency and .78 for Oral Expression to .94 for Nonsense Word Decoding and .96 for Letter & Word Recognition (mean coefficient = .88). Overall, the KTEA-II is a reliable test of achievement with strong psychometric properties.

Validity. For Grade 1 through age 25, Confirmatory Factor Analysis (CFA) was used to verify the factor structure of the KTEA-II. CFA was employed in order to better understand the magnitude of the subtest interrelations, to determine whether the four main areas of academic achievement measured by the KTEA-II have construct validity, and to support the organization of subtests into the four scales. CFA was applied to the eight primary subtests of the KTEA-II: Letter & Word Recognition, Reading Comprehension, Math Concepts & Applications, Math Computation, Written Expression, Spelling, Listening Comprehension, and Oral Expression. The first step was to create a one-factor model and additional subsets were added subsequently. Modification indexes indicated how to group the subtests so that the best possible model fit could be created. The final model included the four achievement factors (Reading, Written Language, Math, and Oral Language) and also entailed the correlations of error variances

between the (a) Letter & Word Recognition and Spelling subtests, (b) Reading Comprehension and Listening Comprehension subtests, and (c) Written Expression and Oral Expression subtests. The final model had good statistical fit (CFI = .992, RMSEA = .062) (Kaufman & Kaufman, 2004b, Figure 7.1).

In addition, the Comprehensive Achievement Composite correlated substantially with global achievement scores on other individually administered achievement batteries: .89-.90 with WIAT-II at grades 1-11 (Kaufman & Kaufman, 2004b, Tables 7.17 & 7.18) and .84-.89 with WJ III at grades 1-10 (Kaufman & Kaufman, 2004b, Tables 7.19 & 7.20)

Strong evidence of convergent and divergent validity was provided for the KTEA-II Reading, Math, and Written Language composites. KTEA-II Reading correlated .85 with measures of reading on the WIAT-II and .76-.82 with reading scores on the WJ III (Kaufman & Kaufman, 2004b, Tables 7.17–7.20). Those coefficients support the convergent validity of the KTEA-II Reading Composite. Divergent validity for the Reading Composite was supported by lower coefficients between KTEA-II reading and scores in other areas of academic achievement—for example, .55-.72 (mean = .64) with WIAT-II math and .47-.74 (mean = .63) with WJ III math (Kaufman & Kaufman, 2004b, Tables 7.17, 7.18, 7.19, & 7.20).

The Written Language Composite of the KTEA-II correlated .62-.92 with the WIAT-II and WJ III writing subtests (Kaufman & Kaufman, 2004b, Tables 7.17, 7.18, 7.19, & 7.20), proving evidence for the convergent validity of the Written Language Composite. Evidence of divergent validity for this Composite comes from lower correlations between the KTEA-II Written Language Composite and other achievement areas, such as the correlation coefficients of .34-.48 (mean = .41) with the Oral Language Composites of the WIAT-II and the WJ III (Kaufman & Kaufman, 2004b, Tables 7.17 & 7.20).

The KTEA-II Math Composite correlated .74 - .87 with WJ III and WIAT-II math (Kaufman & Kaufman, 2004b, Tables 7.17, 7.18, 7.19, & 7.20). The strong correlation coefficients support convergent validity of the Math Composite. Weaker correlation coefficients between the KTEA-II Math Composite and other achievement composites of the WIAT-II and the WJ III provide evidence for its divergent validity. For example, the KTEA-II Math Composite produced correlation coefficients of .44-.76 (mean = .60) with the WJ III and WIAT-II Written Language Composites (Kaufman & Kaufman, 2004b, Tables 7.17 & 7.20).

Procedure

The KTEA-II and the KABC-II were co-normed. Both protocols were collected between September 2001 and May 2003. In total, data from 2,400 participants grades K through 12 were collected. The demographic information for the participants is nationally representative for the school grades. Data were collected in 39 states around the United States (as well as the District of Columbia) and participants were tested in 133 sites (Kaufman & Kaufman, 2004a, 2004b).

Site coordinators. The first step in the data collection procedure included the recruitment of site coordinators, whose responsibilities comprised the supervision and organization of the data collection at their personal locations and the recruitment of skilled examiners. The site coordinators not only recruited the examiners, but also trained and supervised them and assisted with the recruitment of examinees as well as the distribution and collection of parent consent forms. Site coordinators as well as examiners included school psychologists, diagnosticians, special education teachers, and graduate students in the field of education or psychology, who had completed the appropriate training. For each completed assessment, both examiner and site coordinator were paid. Examinees were rewarded gift

certificates and compensation was provided to schools and other organizations that assisted with the distribution and subsequent collection of consent forms.

Consent forms. A pool of potential examinees was selected from schools, preschools, daycares, churches, neighborhood organizations, and individual families. Potential examinees' parents signed consent forms that were then returned to the publisher. The consent form included an explanation regarding the testing procedure, the request for parental consent, as well as demographic characteristics of the participant. Consents were available in both English and Spanish. All information collected was entered into a confidential database and potential examinees' demographic information was matched to 2001 U.S. census data, systematically in a computer program. This procedure allowed random selection of participants. The selected participants were oversampled in order to account for incomplete or missing data. Lists of the final selected examinees were distributed to the site coordinators. The selection process was repeated, if necessary, as examinees became unavailable (Kaufman & Kaufman, 2004a; 2004b).

Quality control. In order to assure quality control, each examiner was asked to complete at least one practice examination before he/she was allowed to collect data. Before continuing with the testing, each examiner was obligated to submit his/her tryout to the AGS Publishing editorial staff and was not allowed to continue testing without prior approval. Examiners administered and scored each battery. The batteries submitted by the examiners were carefully supervised. Scoring and administration errors were closely examined and examiners received regular feedback on their administrations.

Furthermore, statistical procedures verified the accuracy of the data by identifying questionable response patterns. Protocols that seemed dubious due to obvious scoring or administration inaccuracy or as detected by the statistics program were excluded from the data.

Finally, examinees or their parents were randomly contacted to assure that testing had, in fact, taken place (Kaufman & Kaufman, 2004a, 2004b).

Counterbalancing. To reduce the influence of practice effects, the KTEA-II and KABC-II were administered in a counterbalanced order. That is to say, about half the examinees were first administered the KABC-II and the other half was first administered the KTEA-II.

KTEA-II Forms A and B. As explained in the manual (Kaufman & Kaufman, 2004b), the KTEA-II has two Forms - A and B. About half the standardization sample was administered each form. Administering each student the alternate form assessed reliability, while controlling for practice effects. Forms A and B were both standardized jointly in order to ensure that no differences between the random selection and reported reliabilities of the samples existed. Once data were collected, the items were divided with an odd-even split to create the two parallel halves. Afterwards, the compositions of the halves were adjusted to equate the level of difficulty and the representation of skill area.

S. B. Kaufman et al. (2012) conducted analyses to determine the feasibility of collapsing the two parallel forms into one data set: .

> A four-factor Grw, Gq, Oral Language, and Oral Fluency model was specified for each Form. A test of strict factorial invariance was performed. The Configural Invariance model fit was acceptable, $\chi 2$ (110) = 1084.77, CFI = .943. Next, a strict factorial invariance model was imposed, $\chi 2$ (145) = 1213.65, CFI = .938, with $\Delta\chi 2$ (35) = 128.88, $p < .001$. The ΔCFI (.005) was negligible. Given the sensitivity of the likelihood ratio test to sample size, it was deemed appropriate to collapse KTEA-II Forms A and B for the purpose of this research. (S.B. Kaufman et al., 2012; p. 130).

Based on the findings by S.B. Kaufman et al. (2012) the merging of data from Forms A and B was supported; thus, in the present dissertation, Forms A and B are used together.

Statistical Analyses

Factorial invariance using MG-MACS (Question 1 Methodology). In order to explore whether the joint factor structure of the KABC-II and KTEA-II is the same for Caucasians, Hispanics, and Blacks, factorial invariance was assessed. That is, the variable structure of the KABC-II and KTEA-II, as outlined by S.B. Kaufman et al. (2012) for the total sample, was explored to determine whether it is invariant across the three ethnic groups. The 22 subtests listed and described in Tables 2 and 3 comprised the variables entered into the CFA: (a) the 15 KABC-II subtests that are organized in Table 2 according to the CHC abilities measure: *G*c, *G*f, *G*sm, *G*lr, *G*v,; the 7 KTEA-II subtests measure *Grw* (Reading and Written Language) and *Gq* (Math). The factorial invariance of the first-order factors was assessed in each analysis. The first-order variables refer to the seven CHC latent factors (*G*c, *G*f, *G*sm, *G*lr, *G*v, *G*rw, and *G*q). Although the invariance of the second-order *g*-factor was not explored (because the invariance of *g* is irrelevant to the research questions posed in this dissertation), a single *g* factor is hypothesized to underlie all of the subtests that compose the KABC-II and KTEA-II, regardless of whether these tasks are best classified as "ability" or "achievement." From the vantage point of CHC theory, all seven first-order factors fit into the domain of intelligence, as conceptualized by Cattell, Horn, and Carroll (Schneider & McGrew, 2012). In addition, S. B. Kaufman and colleagues (2012) examined the relationship between the separate *g* factors that underlie the KABC-II (COG-*g*) and KTEA-II (ACH-*g*) and concluded from their analysis: "Although COG-*g* and ACH-*g* were not isomorphic, they correlated substantially, with an overall mean

correlation coefficient of .83, and with the correlations generally increasing with age (ranging from .77 to .94)" (p. 123).

The invariance models were explored using pairs of two ethnicities: Caucasians and Blacks, Caucasians and Hispanics, and Hispanics and Blacks.

Power Analysis. Maede and Bauer (2007) explored power for various sample size conditions in confirmatory factor analysis tests of measurement invariance. They found that sample sizes of 100 produced low power, whereas sample sizes of 400 produced high power. The researchers found acceptable power with sample sizes of around 200. The sample sizes for the differential construct validity, which is conducted on the total sample, yield n=312 (African-American group), n=376 (Hispanic sample) and n=1313 (Caucasian group). These sample sizes therefore meet criteria for acceptable to high power.

Assumptions. Statistical assumptions underlying confirmatory factor analysis, include:

1. Interval or ratio level of measurement, meaning that the distance between attributes of the variables is meaningful.
2. Independence, which can be met with random selection of the participants.
3. Linearity, which is an assumption based on the notion that all variables are related linearly with one another.
4. Normality, which refers to the normal distribution of the dependent variables for each individual variable; it refers to the skewness and kurtosis to be within normal limits and the removal or transformation of outliers.

Analytical steps. Multi-group confirmatory factor analysis (MG-CFA) serves as an excellent tool to explore factorial invariance between groups (Reynolds & Keith, 2013). Specifically, MG-CFA based on a mean and covariance structure (MG-MACS) approach was

used. Using multi-group mean and covariance structure analysis (MG-MACS), it was specified whether the factor loadings and residual variances, as well as the intercepts of the factor structure of the KABC-II and KTEA-II, were equivalent for Caucasian, Black, and Hispanic children and adolescents in grades 1-12. This method was used to assure that the observed score on a subtest is attributable to the factor that the subtest measures and not due to group membership. All analyses were completed using Amos software version 20 (Arbuckle, 1995–2011).

Testing for measurement invariance using MG-MACS required the setting of increasingly restrictive sets of equality of constraints. Meredith (1993) discusses using a hierarchy that consists of identifying configural invariance, metric invariance (weak factorial invariance), intercept invariance (strong factorial invariance), and residual invariance (strict factorial invariance).

Configural invariance. In *configural* invariance, the same factor structure is applied to all ethnic groups (Caucasians and Blacks, Caucasians and Hispanics, and Hispanics and Blacks). This procedure can be completed by either exploring each structure separately for each group or by using multi-group analysis. Configural invariance means to establish the same alignment of factors for each group. For both groups, the factors and patterns of free and fixed loadings are estimated equally. Whereas factor variances and covariances are allowed to vary freely, the reference indicator (for each first and second order factor) for both groups' factor loadings is fixed to 1. This approach balances the factors and scales them properly. The latent factor means are fixed to 0 and the observed subtest means (intercepts) can vary freely. Depending on the goodness of fit of the model for all three ethnic groups, as estimated by the size of chi-square (χ^2) as compared to the degrees of freedom and by Comparative Fit Index (CFI) and Root Mean Square Error of Approximation (RMSEA) values, the model is either contained or modified

(Reynolds & Keith, 2013). When using multi-group analysis in AMOS, the fit index RMSEA needs to be corrected, as recommended by Steiger (1998). The RMSEA has to be multiplied by its square root of two (because of multiple groups—i.e., two groups in each of the three analyses).

Metric invariance. After configural invariance has been established, *metric* invariance (weak factorial invariance) is assessed. In this step, first-order factor loadings are restricted to be equal across groups. As pointed out by Meredith (1993), the invariance of the factor loadings is a prerequisite that needs to be met before establishing measurement invariance. In configural invariance, each factor has one loading already fixed to 1. This restriction remains throughout the metric invariance step. In addition, the previously free factor loadings are now also restrained so that all corresponding factor loadings are equal across the groups. This assures that a one-unit increase in a specific factor for Caucasians results in the same unit increase for Blacks and Hispanics. The parameters are given the same letter (e.g., a) to mark the equal constraints. The model fit across the groups is assessed via $\Delta\chi^2$, ΔCFI, and the ΔRMSEA to assess whether the constraints have resulted in a statistically significant degradation in model fit (the configural model is compared to the metric model). If these added constraints have *not* resulted in a degradation of model fit, it is established that the relation between subtests and factors is the same for the ethnic groups. That is, the unit of measurement is equal across groups and latent variances and covariances may be compared across the ethnic groups (Reynolds & Keith, 2013). However, in order to determine whether the same level of CHC abilities results in the same observed scores on the KTEA-II and KABC-II for Caucasians, Blacks, and Hispanics, further investigations of the intercepts are required.

Intercept invariance. Before comparing groups means, *intercept invariance* (strong factorial invariance) needs to be established. In the metric invariance model, the unstandardized factor loadings have been fixed to be equal across the groups and the factor means have been fixed to "zero". The groups' subtest intercepts were allowed to differ. In strong factorial invariance, in addition to all previous constraints, all corresponding subtest intercepts (means) are restrained to be equal. The factor means are allowed to vary freely across groups. This specification ensures that the mean differences found on the subtests are due to the same common factor. That is, this step allows for a specific score to mean the same for one group as it does for another group. By constricting the intercepts and by allowing the factor mean scores to vary freely, one can estimate whether differences in the observed means are due to differences in the CHC factors. To allow group differences to show, the latent means of one group are constrained to 0, whereas the latent means for the other group can vary freely. Once these mean score constraints are added in addition to the subtest intercept constraints, and the fit index has not degraded, it can be concluded that differences in the observed subtest scores are due to differences in the latent means (i.e., CHC factor means) (Reynolds & Keith, 2013).

Alternatively, if strong factorial invariance is *not* established, and one group's intercept in a particular subtest is higher compared to another group's intercept, then this group would consistently score better on this subtest, despite demonstrating the same level of ability. In that sense, this group would continue to show better performance on that test, even though they do not actually have better abilities on the factor. The factor (e.g., *G*c) in that case does not completely account for the performance on that particular subtest. Such a result is an example of bias, because differences in the subtest measuring this latent factor variable (*G*c) are not solely due to *G*c ability, but due to other variables (Reynolds & Keith, 2013).

Partial strong factorial invariance. Partial strong factorial invariance means that the factor intercepts are not likely to produce group differences, but there are intercept differences found in a specific variable (e.g., a specific subtest). That is to say, measurement invariance for *G*c, for example, might not be established simply because one particular subtest was *not* invariant across the groups (e.g., Verbal Knowledge). In other words, there is something about this particular subtest (Verbal Knowledge) that cannot solely be explained by *G*c. The constraints on the Verbal Knowledge subtest intercept can then be freed, which would permit mean score comparisons on *G*c (Reynolds & Keith, 2013).

Residual invariance. Once strong factorial invariance (intercept invariance) has been established, *residual invariance (strict factorial invariance)* needs to be investigated. Strict factorial invariance refers to the equality constraints of the residual variances of the residuals (error and specific variances).

In this analysis, the residuals are constrained in addition to all previous constraints that have been made thus far. If this further restriction does *not* result in a degradation of the model fit, then latent variables can be compared across the groups - all differences in the observed scores are due to difference in the common factor means and variances/covariances, and are not due to group membership. If strict factorial invariance is established then latent mean differences can be compared within the residual invariance model (the most parsimonious model). The establishment of strict factorial invariance means that differences found on the factor means and variances account fully for all group differences found on subtest scores. It is important to note, though, that groups can be compared on factor means, even if strict factorial invariance is not met. The constraints for the specific residual variance just need to be removed and the latent factor mean differences can be compared within the intercept invariance model. In

that sense, group comparisons of observed means and variances can be made with either strong or strict factorial invariance. Professionals can then be certain that differences found on the latent factor means and observed scores are due to true differences, and not biased by the individual's ethnic background (Reynolds & Keith, 2013). Several researchers (e.g., Byrne, 2010; Reynolds & Keith, 2013), have suggested that residual invariance is NOT a necessary prerequisite in order to meaningfully compare mean factor scores or observed means. Therefore, in this dissertation, residual invariance is not evaluated.

Fit indexes. The possible degradation of model fit with increasingly restrictive sets of restraints is determined by the likelihood ratio test (χ^2), the root mean square error of approximation (RMSEA) and comparative fit index (CFI), or the standardized root mean square residual (SRMR). A "good" fit would typically result in a non-significant $\Delta\chi^2$, a RMSEA value close to .05 or less (or between .05 and .08), and a CFI value of at least .95 (Keith & Reynolds, 2012). The likelihood test ratio ($\Delta\chi^2$), ΔCFI, and ΔRMSEA are used to compare the goodness of fit for tests of factorial invariance (Cheung & Rensvold, 2002). These researchers recommend that for ΔCFI a change > .01 is considered significant and for ΔRMSEA a change > .02 is considered significant. $\Delta\chi^2$ can easily result in a significant degradation of goodness of fit because it detects minor and inconsequential differences, which are often the result of a large sample size and a large number of constraints. Given the complexity of the model, the sample size, and the number of constraints, ΔCFI as well as ΔRMSEA values are given more weight when evaluating the goodness of fit for the measurement invariance models (Byrne, 2010; Chen, Sousa, & West, 2005; Reynolds & Keith, 2013). Therefore, in this dissertation, ΔCFI (with ΔCFI > .01 considered significant change) and the ΔRMSEA (with ΔRMSEA > .02 considered significant change) are used to evaluate the degradation in model fit.

Comparisons of mean scores using Multivariate Analysis of Covariance (MANCOVA) (Question 2 Methodology). To compare mean scores of the invariant factors across the ethnic groups, multivariate analysis of covariance (MANCOVA) were employed. MANCOVA assesses whether any mean differences found in the sample represent true population mean differences, by taking the margin of error into account. MANCOVA allows for the possibility to compare several dependent variables that measure similar constructs and analyze them both simultaneously and separately, while holding other variables constant. The simultaneous comparison of various dependent variables decreases the probability of finding differences just by chance (compared to conducting several ANCOVAS). MANCOVA also has the advantage of taking into consideration the inter-dependability of the latent factors, as MANCOVA allows for the analysis of the variance-covariance matrices.

Power analysis. According to Meyers et al. (2013, p. 226-227) multivariate analysis of variance requires no more than 20 participants per group. These analyses were conducted using the total sample and, therefore, yielded samples sizes of: n=312 (African American group), n=376 (Hispanic group) and n=1313 (Caucasian group). The power criterion was met.

Assumptions. In addition to three of the assumptions that are already tested in question 1 for the confirmatory factor analysis, including (1) Independence, (2) Linearity, and (3) Normality, there are four more assumptions underlying MANCOVA and ANCOVA analyses that need to be met:

1. Multivariate normality, which refers to the normality based on measures of multivariate skewness and kurtosis, as assessed by AMOS (Meyers, Gamst, & Guarino, 2013).

2. Homogeneity of regression, which states that the slopes of the regression lines for each group have to be the same. The evaluation of an ethnicity x SES (group x covariate) interaction tests this assumption. Non-significant interactions indicate that this assumption is met.

3. Homogeneity of the variances, which says that the variances of the dependent variables across the independent variables are equally distributed for all groups. This is tested with the Levene test. A non-significant Levene indicates that this assumption is met.

4. Homogeneity of the covariance's, which stipulates that the intercorrelations between the dependent variables need to be homogenous. This assumption is tested with Box's M test. A non-significant Box's M indicates that the assumption is met.

Preliminary analysis of the homogeneity of regression assumption. It was desirable to conduct preliminary analyses to decide whether SES would be appropriate to include as a covariate in the MANCOVA analysis. Two ANOVAs were conducted, both using the same independent variables—ethnicity and SES (parent's education). In the first ANOVA, KABC-II Mental Processing Index (MPI) was the dependent variable; in the second, KTEA-II Comprehensive Achievement Composite (CAC) was the dependent variable. In order to qualify as an appropriate covariate for the MANCOVAs two conditions had to be met in these preliminary analyses: (a) SES had to yield *significant* main effects in the ANOVAs, and (b) the interactions between ethnicity and SES had to be *non-significant*. The first criterion was necessary to identify SES as a significant confound that had to be controlled in the MANCOVAs. The second criterion was necessary to ensure that the homogeneity of regression assumption was met.

Table 4 (MPI) and Table 5 (CAC) present the results of these ANOVAs. In both ANOVAs, ethnicity and SES were significant main effects (p <.001), but the ethnicity X SES interactions did not approach significance at the .05 level. Furthermore, a significant χ^2 between ethnicity and SES revealed that there is a meaningful association between these two variables (see Table 6); SES accounts for some of the differences found between the ethnic groups. Given the results of the χ^2 analysis, the significant main effect of SES in both ANOVAs, and the non-significant interactions between ethnicity and global scores on the KABC-II and KTEA-II, SES is included as a covariate in MANCOVA.

Table 4

Between-Subjects Effects (ANOVA): Ethnicity by SES (Parent's Education) with KABC-II Mental Processing Index (*MPI) as Dependent Variable*

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	64703.736[a]	11	5882.158	30.629	.001	.145
Intercept	10567664.285	1	10567664.285	55026.844	.001	.965
SES	24455.142	3	8151.714	42.447	.001	.060
Ethnicity	10031.130	2	5015.565	26.117	.001	.026
SES * Ethnicity	1033.691	6	172.282	.897	.496	.003
Error	381978.738	1989	192.046			
Total	20439886.000	2001				
Corrected Total	446682.474	2000				

a. R Squared = .145 (Adjusted R Squared = .140)

Table 5

Between-Subjects Effects (ANOVA): Ethnicity by SES (Parent's Education) with KTEA-II Comprehensive Achievement Composite as Dependent Variable

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	86095.469[a]	11	7826.861	44.024	.001	.196
Intercept	10496495.136	1	10496495.136	59039.857	.001	.967
Ethnicity	13025.098	2	6512.549	36.631	.001	.036
SES	31552.283	3	10517.428	59.158	.001	.082
Ethnicity * SES	1625.598	6	270.933	1.524	.166	.005
Error	353084.179	1986	177.787			
Total	20560027.000	1998				
Corrected Total	439179.648	1997				

a. R Squared = .196 (Adjusted R Squared = .192)

Table 6

Chi-Square Tests: SES (Parent's Education) and Ethnicity

	Value	Df	Asymp. Sig. (2-sided)
Pearson Chi-Square	303.095[a]	6	.001
Likelihood Ratio	274.969	6	.001
Linear-by-Linear Association	74.990	1	.001
N of Valid Cases	2001		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 46.78.

Analytical steps. Based on the results of Question 1, namely the invariance of all 7 CHC factors across the three ethnic groups, the decision was made to conduct two MANCOVAS. One MANCOVA was conducted with Ethnicity (Caucasian, Black, Hispanic) as the independent variable; SES (parental education) as the covariate; and the five KABC-II CHC factors as dependent variables—Simultaneous/Visual Processing-Gv, Sequential/Short-Term Memory-Gsm, Learning/Long-Term Storage and Retrieval-Glr, Planning/Fluid Reasoning-Gf, and Knowledge/Crystallized Knowledge-Gc). The second MANCOVA was conducted with the same independent variable and covariate but with the four KTEA-II composites as dependent variables—Reading, Written Language, Oral Language, Math. The mean differences between ethnic groups were examined jointly, while holding SES constant.

MANCOVA is based on the assumption that the means of each group on each dependent variable are equal. In other words, MANOVA assumes that there are no ethnic group differences on any of the factors. This *null-hypothesis* was evaluated using Pillai's trace, Wilk's lambda, Hotelling's trace, and Roy's largest root. These tests create an F-value that assesses the multivariate between and within the groups. A significant effect on any of the tests indicates that there are significantly different ethnic group means on one or more of the dependent variables. Whenever significant differences are found, follow-up planned comparisons are employed to determine which ethnic differences, on which dependent variable, are significant and which ones are not. In fact, both MANCOVAs yielded significant *F* values on the pertinent statistical indexes such as Pillai's trace and Wilk's lambda. For the KABC-II MANCOVA, 15 pair-wise planned comparisons were conducted: (a) Caucasian-Black differences between SES-adjusted mean standard scores on each of the five KABC-II CHC factors, (b) Caucasian-Hispanic differences, adjusted for SES, on the five factors, and (c) Hispanic-Black adjusted differences on

the five factors. To help minimize the possibility of Type I errors, these analyses were conducted with an adjusted alpha level based on the Bonferroni procedure. To achieve a family-wise alpha level of .05 for 15 simultaneous planned comparisons, .05 was divided by 15, yielding $p = .0033$; differences that yielded a probability of .0033 or less were interpreted as being significant at the .05 level. Using a comparable approach, $p = .00066$ was needed for a difference to be significant at the .01 level. Following the same methodology for the KTEA-II MANCOVA, 12 planned comparisons were conducted (three pair-wise ethnic group comparisons on SES-adjusted mean scores on each of the four KTEA-II composites). The Bonferroni procedure required $p = ..0042$ for a .05 family-wise alpha level and $p = .00083$ for a .01 family-wise alpha level. Such strict alpha levels could potentially lead to a Type II error.

As noted, the MG-MACS analyses conducted for Question 1 revealed invariance across ethnic groups on all seven CHC-based factors. Therefore, it was feasible to examine ethnic differences, adjusted for SES, on the latent roots measured by the KABC-II and KTEA-II. That was done by the two MANCOVAs just described. Had there NOT been invariance for one or more factors, then the specific subtests that caused the non-invariance would have been identified. For all subtests that were found to be NOT invariant, additional MANCOVAs (with SES as the covariate and the subtests as dependent variables) would have been conducted to identify significant ethnic differences on the specific subtests that produced the non-invariance. Again, an appropriate Bonferroni correction would have been applied with the precise p values determined based on the number of maverick subtests. However, based on the invariance found in Question 1 for Caucasian-Black, Caucasian-Hispanic, and Hispanic-Black CFAs, no subtest-level analyses were required for Question 2.

Prediction bias using structural equation modeling (Question 3 Methodology). Lastly, an exploration of whether the KABC-II predicts the KTEA-II equally well for the three ethnic groups is conducted. Multi-group path models (structural equation modeling) is used to measure the predictive validity of the cognitive scales and compare whether they predict the achievement composites for the three ethnic groups across three different grade level groups equally well. All analyses are conducted using AMOS 20.

If the regression lines (Y = a + bX) for any pair of variables differ across the groups, it can be concluded that there is bias in the prediction. That is to say, if either the slope "*b"* or intercept "*a"* differ significantly across groups, the application of the same regression line to all groups would result in an incorrect prediction of the criterion variable. Slope bias means that the magnitude of the correlation coefficients are different and the test would be biased against the group that produces the lower correlation coefficients in terms of the magnitude of the relationship between the predictor and outcome variables. Intercept bias, on the other hand, concerns the ability of an IQ test to predict a group's correct level of achievement. A test would be biased against minority group children if it *underpredicted* their achievement. Overprediction of the minority group's achievement would indicate that the test is not entirely accurate at predicting their achievement; however, it would not indicate bias of the test against the ethnic minority group. as it would not make conceptual sense to call a test biased against a group when it overpredicts their achievement outcomes (Keith & Reynolds, 2003).

Power analysis. The power analysis using G power indicated that with an effect size of .25 (Kaufman & Kaufman, 2004, p.111), an alpha level of .01 and 3 predictor variables (e.g. *Gf* plus 2 possible age interactions), 68 participants are needed to reach a power level of .8. The

analyses are conducted using 3 different grade groups (1-4, 5-8, 9-12). The smallest sample yields $n = 74$ participants (Blacks grades 9-12). The power criterion is met.

Assumptions. In addition to previously tested assumptions of (1) multivariate normality and (2) independence, structural equation modeling also assumes (3) correct model specification, as tested by fit indexes.

Analytical steps. Five cognitive scales, as measured by the KABC-II (Planning/*G*f, Knowledge/*G*c, Learning/*G*lr, Sequential/*G*sm, and Simultaneous/*G*v) predict three KTEA-II achievement composites (Reading, Writing and Math) across three different grade groups (1-4, 5-8, and 9-12). For each prediction, a separate model is employed. Paths from each cognitive scale (e.g., *Gf* or *G*c-Lexical Knowledge) to the corresponding achievement composite, are created; The predictor variables (each CHC variable) are mean-centered, which helps with interpretation (the intercept of 0 represents the mean achievement score for someone with average intelligence). Error and unique variances for each predictor variable and subtests are also included in the model.

In order to assess for prediction bias, a model fit method is employed for each pair (Caucasians versus Blacks, Caucasians versus Hispanics, and Hispanics versus Blacks). The model fit is evaluated in a stepwise analysis, by testing the invariance of the variance, slope, and intercept of the regression lines. When using this approach, the ethnic groups are first compared on a model. Here, each group has its separate regression line (or paths) from the cognitive variable to the achievement composite, without any constraints. That way, magnitudes of the coefficients can be compared. Next, the residual variances of the achievement composites are constrained to be equal across the groups (the constriction of the residual variances does not necessarily have to be met). Following this step, the invariance of the slopes and intercepts are

analyzed. To explore the equality of the slopes, the slopes need to be constrained to be equal across the three ethnic groups. If the slope restriction does not result in a degradation of model fit, slope invariance is established (*weak* invariance). Finally, in addition to the slope constraints, the intercepts are constrained to be equal. If the slope and intercept constraints do not result in a significant degradation of model fit, prediction invariance has been established (*strong factorial invariance)*.

The fit of the models is again evaluated with $\Delta\chi^2$. RMSEA and CFI can be employed as alternative fit indexes. If the slope and intercept restrictions do not result in a significant degradation of model fit (as evaluated by $\Delta\chi^2$ and ΔRMSEA and ΔCFI), then it can be concluded that the same regression lines can be used across the three ethnic groups (that is to say, the cognitive scales or MPI impact the achievement composites equally across groups) (Keith & Reynolds, 2003).

However, if the restrictions do result in a significant degradation of model fit, then bias is present. If the slope restriction results in a significant degradation of model fit, there is an interaction between ethnicity and achievement outcome when a common regression lines is used for the ethnic groups. Depending on the shape of the slope, the regression line might overpredict achievement for one ethnic group only for those individuals that score in the lower percentile ranks, but might underpredict for those individuals that score in the higher percentile ranks. If the intercept restrictions result in a significant degradation of model fit, a common regression line results in an overprediction for the group that has a higher intercept (as compared to the common regression line intercept) and in an underprediction for the group that has a lower intercept (Keith & Reynolds, 2003).

If there is slope or intercept *non*-invariance, post-hoc analyses are conducted in order to better understand the direction of the bias. The participants are then divided depending on their ethnic background and simple regressions are conducted for each ethnic group. The output for each group is then compared and the different regression lines are created for the separate ethnic groups. The ethnic group regression lines are graphically displayed together with the common regression line to present the differences in slope and intercept (Keith, 2006).

Results

This section is organized into the following sections: (a) missing data and outliers; (b) descriptive statistics for the samples of Caucasians, Blacks, and Hispanics; (c) factorial invariance using MG-MACS (Question #1); (d) comparisons of mean scores using Multivariate Analysis of Covariance (MANCOVA) (Question #2); and (e) assessing prediction bias using structural equation modeling (Question #3).

Missing Data and Outliers

Before analyses could be conducted to answer the questions posed by this study, decisions had to be made regarding how to deal with missing data and outliers. Frequency distributions were examined to discover exactly what data were missing. On the KTEA-II, Associational Fluency (Semantic and Phonological) had missing data for 105 cases; therefore, the task was not included in the analyses. On the KABC-II, the supplementary subtests Atlantis Delayed (Glr) had missing data for 20 cases, Rebus Delayed (Glr) had missing data for 45 cases, and Gestalt Closure (Gv) had missing data for 572 cases. These three subtests were excluded from all analyses. Since all three tests were supplementary they did not impact the loadings on the CHC factors and, therefore, the validity of the results. Regarding the two KABC-II tasks of delayed recall, these two tasks had to be eliminated from the analyses for a second reason: It would have been inappropriate to include both the initial learning and then the delayed learning tasks in the same analysis due to multicollinearity.

There was also a small amount of missing data in the final dataset composed on KTEA-II and KABC-II subtests and scales. The KABC-II Rover subtest and Story Completion subtest each had 1 missing case. The two missing cases were handled using Hotdeg imputation (Myers, 2011). Rover was scaled equal to that child's scaled score on the Triangles subtest and Story

Completion was scaled equal to the child's Pattern Reasoning scaled score. This procedure takes advantage of the scale composition of the KABC-II: Rover and Triangles are both measures of Simultaneous/Gv whereas Story Completion and Pattern Reasoning are both measures of Planning/Gf.

There were also missing data on the KABC-II Planning/Gf index for all 6-year-olds in the sample (*n*=117) because this Index is computed only for children who are at least 7-years old (Kaufman & Kaufman, 2004b). However, there is empirical evidence to support a distinct fluid reasoning factor (*Gf* separate and apart from visual processing or *Gv*) for children as young as 4 years old (Raiford & Coalson, 2014). In addition, all children age 6 in the sample were administered both Planning/*Gf* subtests. It was, therefore, simple and straightforward to compute the Planning/Gf index for all of the 6-year-olds with missing data. For each of those 117 children, the sum of the scaled scores on the Pattern Reasoning and Story Completion subtests was entered into the KABC-II conversion table for age 7:0-9:11 year olds (Kaufman & Kaufman, 2004b; table D-2, p.190).

There were no outliers in the sample. All participants had previously been selected for inclusion in the standardization samples of the KABC-II and KTEA-II. According to Dr. Mark Daniel (Personal communications, September 13-14, 2014), Project Director of the KABC-II and KTEA-II, whenever apparent outliers were found and such outliers were the result of poor administration or scoring mistakes, the outliers were eliminated or corrected.

Descriptive Statistics for the Total Sample

Table 7 presents the means and *SD*s of the KTEA-II composites and subtests and Table 8 presents means and *SD*s of the KABC-II indexes and subtests for the total samples of Caucasians (*n*=1,313*),* Blacks (*n*=312), and Hispanics (*n*=376). On the KTEA-II Comprehensive

Achievement Composite, Caucasians averaged 102.8, and Blacks and Hispanics each averaged 93.8. Caucasians earned mean standard scores of 102.1-103.7 on the four KTEA-II composites (Reading, Written Language, Math, Oral Language) whereas Blacks scored between 94.6 and 95.3, and Hispanics earned means of 93.7 and 95.4. On the KTEA-II subtests, mean standard scores for Blacks ranged from 93.8 on Math Concepts & Applications to 99.9 on Associational Fluency. For Hispanics, the range was from 94.4 (Reading Comprehension) to 96.9 (Math Computation).

On the KABC-II global indexes, Caucasians scored between 102.7 and 103.2, whereas Blacks scored between 93.1 and 94.9, and Hispanics scored between 93.5 and 96.2. Despite their similar mean scores on the KABC-II global indexes, Blacks and Hispanics displayed different group profiles. Blacks performed best on Sequential/Gsm (99.6) and Learning/Glr (98.1) while earning their lowest scores on Simultaneous/Gv (92.9) and Knowledge/Gc (93.3). Hispanics scored best on Simultaneous/Gv (97.7) and Planning (96.9) and lowest on Sequential/Gsm (93.9) and Knowledge/Gc (91.9).

Table 7
Kaufman Assessment Battery for Children-2nd Edition (KABC-II) Means and Standard Deviations (SD) for each Ethnic Group (N=2011)

CHC Factor	Subtest	Caucasians (N=1313)		Blacks (N=312)		Hispanics (N=376)	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
Sequential/Gsm		**102.0**	**14.3**	**99.6**	**16.3**	**93.9**	**15.4**
	Number Recall	10.4	2.7	10.0	3.1	9.3	3.1
	Word Order	10.3	2.7	9.8	3.2	8.6	2.8
	Hand Movements	10.4	2.9	9.5	2.9	9.3	2.7
Simultaneous/Gv		**102.3**	**14.9**	**92.9**	**13.9**	**97.7**	**13.7**
	Block Counting	10.3	3.0	9.0	2.9	9.2	3.1
	Rover	10.4	2.9	9.0	2.9	9.7	2.9
	Triangles	10.3	2.9	8.3	2.8	9.7	2.7
	Gestalt Closure	10.3	2.8	9.1	3.3	9.9	2.9
Planning/Gf		**102.2**	**14.9**	**94.6**	**13.8**	**96.9**	**14.6**
	Pattern Reasoning	10.5	2.9	9.2	2.8	9.5	2.9
	Story Completion	10.4	2.9	9.1	2.9	9.6	3.0
Learning/Glr		**102.3**	**15.0**	**98.1**	**14.1**	**95.6**	**14.9**
	Atlantis	10.3	3.1	9.7	2.9	9.1	3.2
	Rebus	10.5	3.0	9.6	2.9	9.3	3.1
	Atlantis Delayed	10.1	2.8	9.8	2.7	9.2	3.1
	Rebus Delayed	10.3	2.9	9.7	3.0	9.3	3.0
Knowledge/Gc		**103.9**	**13.8**	**93.3**	**14.1**	**91.9**	**14.1**
	Expressive Vocabulary	10.8	2.6	8.3	2.5	7.6	2.7
	Riddles	10.8	2.8	8.8	2.9	8.4	2.9
	Verbal Knowledge	10.5	2.8	8.7	2.9	8.6	2.8
GLOBAL INDEXES							
Mental Processing Index (MPI)		**102.7**	**14.6**	**94.9**	**13.8**	**94.6**	**14.4**
Fluid Crystalized Index (FCI)		**103.2**	**14.4**	**94.1**	**13.6**	**93.5**	**14.2**
Nonverbal Index (NVI)		**102.7**	**14.9**	**93.1**	**13.4**	**96.2**	**13.9**

Note: The following subtests have missing data: Rebus Delayed--Caucasians (N = 1287), Blacks (N = 303), Hispanics (N = 366); Atlantis Delayed--Caucasians (N = 1302), Blacks(N = 310), Hispanics (N = 369); Gestalt Closure--Caucasians (N = 958), Blacks (N = 214), Hispanics (N = 257).
SES (Parent Educational Attainment) for Caucasians (n=1313), Blacks (n=312), and Hispanics (n=376):
Grade 11 or less: 6.5% Caucasians, 19.6% Blacks, 41.0% Hispanics;
High School/GED: 34.4% Caucasians, 28.5% Blacks, 31.9% Hispanics;
Some College/Tech/Associated Degree: 32.1% Caucasians, 33.0% Blacks, 17.0% Hispanics;
Bachelor's Degree or higher: Caucasians 27.0%, Blacks 18.9%, Hispanics 10.1%.

Table 8

Kaufman Test of Educational Achievement-2nd Edition (KTEA-II) Means and Standard Deviations (SD) for Each Ethnic Group (N=2001)

Composite	Subtest	Caucasians (n=1313)		Blacks (N=312)		Hispanics (N=376)	
		Mean	*SD*	Mean	*SD*	Mean	*SD*
Reading		**102.1**	**14.6**	**95.3**	**15.0**	**94.2**	**14.3**
	Letter Word Recognition	102.0	14.3	96.4	14.8	95.9	14.9
	Reading Comprehension	102.3	14.4	95.5	15.1	94.4	13.6
	Nonsense Word Decoding	101.5	14.2	95.1	15.3	96.7	15.7
Math		**102.5**	**14.4**	**94.6**	**13.8**	**95.4**	**13.9**
	Math Concepts & Applications	103.4	14.6	93.8	14.1	94.9	13.9
	Math Computation	101.6	14.0	96.4	13.5	96.9	13.6
Written Language		**102.1**	**14.5**	**95.2**	**14.9**	**95.2**	**14.0**
	Written Expression	102.8	14.8	94.8	14.8	95.8	14.5
	Spelling	101.3	14.4	96.3	15.6	95.4	14.2
Oral Language		**103.7**	**14.3**	**95.2**	**14.1**	**93.7**	**13.9**
	Listening Comprehension	103.3	14.1	94.9	14.2	94.9	13.6
	Oral Expression	103.3	14.3	97.1	14.5	94.8	14.1
	Letter & Word Recognition	102.0	14.3	96.4	14.8	95.9	14.9
	Associational Fluency (Semantic and Phonological)	100.8	14.2	99.9	13.9	95.4	14.4
Comprehensive Achievement Index		**102.8**	**14.5**	**93.8**	**14.3**	**93.8**	**13.6**

Note: The following subtest have missing data: Associational Fluency (Semantic and Phonological)—Caucasians (N=1305), Blacks (N=311), Hispanics (N=371).
SES (Parent Educational Attainment) for Caucasians (n=1313), Blacks (n=312), and Hispanics (n=376):
Grade 11 or less: 6.5% Caucasians, 19.6% Blacks, 41.0% Hispanics;
High School/GED: 34.4% Caucasians, 28.5% Blacks, 31.9% Hispanics;
Some College/Tech/Associated Degree: 32.1% Caucasians, 33.0% Blacks, 17.0% Hispanics;
Bachelor's Degree or higher: Caucasians 27.0%, Blacks 18.9%, Hispanics 10.1%.

Factorial invariance using MG-MACS (Question #1)

Question 1 asks: Using Confirmatory Factor Analysis (CFA), is the factor structure of the Kaufman tests invariant for separate groups of Caucasians, Blacks, and Hispanics in grade 1-12, using the CHC-based factor model developed by S.B. Kaufman et al (2012) as the criterion?

Assumptions. Preliminary analyses were conducted to determine whether the assumptions that underlie the CFA were met. These assumptions concerned the: (1) Interval or ratio level of measurement, (2) Independence, (3) Linearity, and (4) Normality (Meyers, Gamst, & Guarino, 2013).

(1) Interval level of measurement: All data are derived from raw scores on subtests that were normalized and then standardized to have a mean = 10 and *SD* = 3 (KABC-II) or a mean = 100 and *SD* = 15 (KTEA-II) and, therefore, fall at an interval level of measurement (Kaufman & Kaufman, 2004a, 2004b).

(2) Independence: this assumptions is met due to the stratified random sampling procedures used to select the participants (see Kaufman & Kaufman, 2004a, 2004b).

(3) Linearity: Scatterplots were used to visually evaluate whether the assumption of linearity was met. All variables met the assumption of linearity for the total sample (n=2001) as well as for each individual ethnic group, Caucasians (n= 1313), Blacks (n = 312), and Hispanics (n = 376). Examples illustrative of the findings: Figure 1 demonstrates the linear correlations of the KTEA-II Letter and Word Recognition subtest and the KTEA-II Math Concepts and Applications subtest for the total sample (n=2001). Figure 2 demonstrates the intercorrelation between the KABC-II Expressive Vocabulary subtest and the KABC-II Story Completion subtest for the total sample (n=2001).

Figure 1

Intercorrelation between the KTEA-II Letter and Word Recognition subtest and Math Concepts & Applications subtest for the total sample (N=2001)

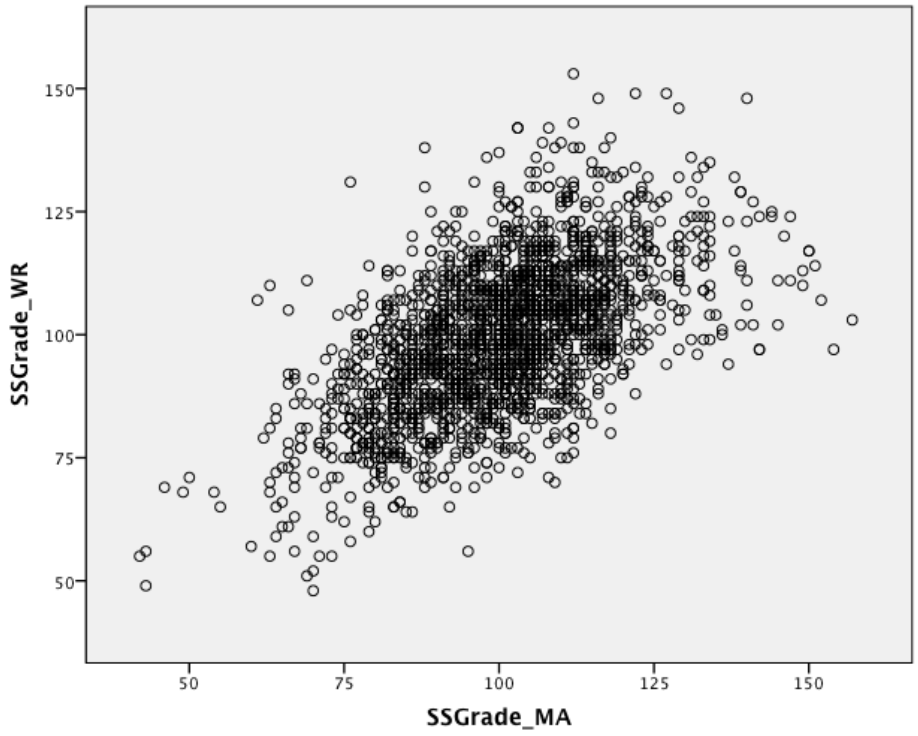
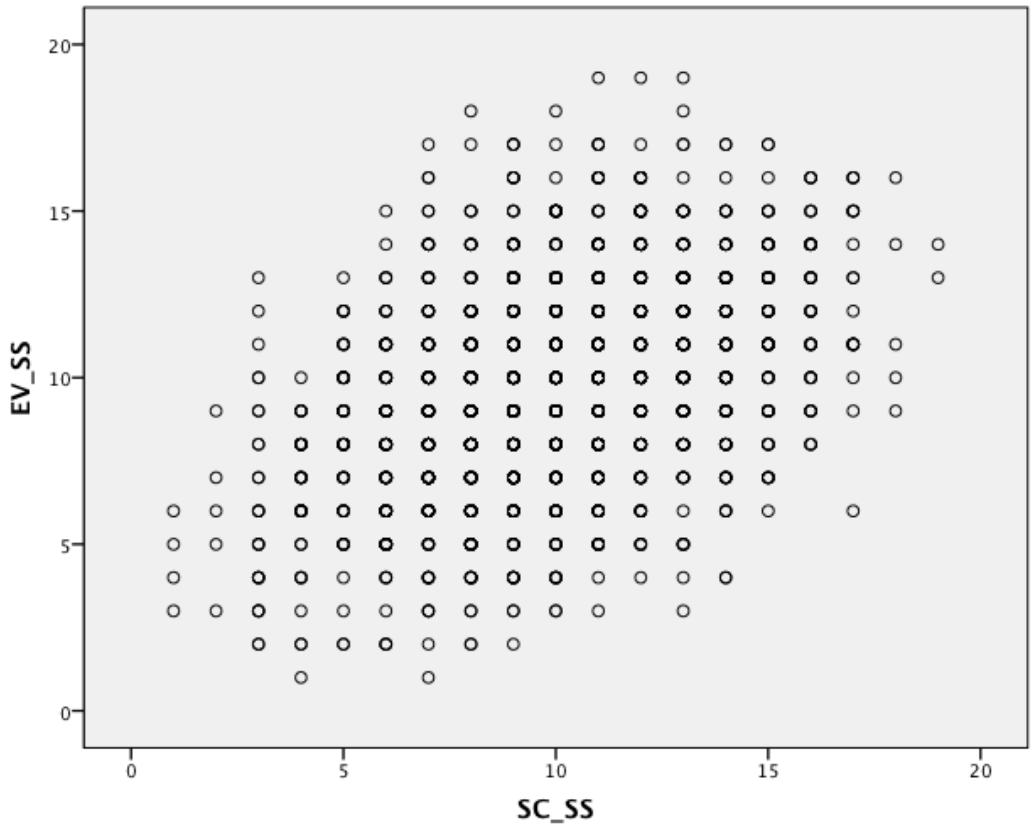


Figure 2

Distribution of the KABC-II Expressive Vocabulary subtest and Story Completion for the total sample (N=2001)



(4) Normality: The distribution of the data was explored by looking at the skewness and kurtosis values of each variable for the total sample and for each ethnic group. For the total

sample (n=2001) skewness ranged from -.152 (KABC-II Rebus) to +.148 (KABC-II Word Order) and was, therefore, far from the +/- 2.0 cutoff. Kurtosis ranged from -.399 (KABC-II Riddles) to .360 (KTEA-II Reading Comprehension) and was, therefore, far from the +/- 7 cutoff (Meyers et al., 2013). Similarly, for Blacks (n=312), skewness ranged from -.143 (KABC-II Pattern Reasoning) to .403 (KABC-II Riddles) whereas kurtosis ranged form -.273 (KABC-II Number Recall) to 1.1.05 (KTEA-II Math Concepts & Applications). Thus, skeweness and kurtosis for Blacks were within normal limits. For Hispanics (n=376), skewness ranged from -.138 (KTEA-II Math Computation) to .277 (KABC-II Riddles) and was, therefore, normal. Kurtosis ranged from -.603 (KABC-II Riddles) to .547 (KTEA-II Math Concepts & Applications) and, again, was normal. Based on skewness and kurtosis data, data for the total sample and for each ethnic group were normally distributed.

Creation of the model for the total sample and the three ethnic groups. Before invariance (or non-invariance) can be established across ethnic groups, the first step is to identify the model that best fits the data for all samples, including the total sample (N = 2,001). To do that, a model must first be hypothesized to fit the data ("original model") and then that model must be modified based on theory and research to try to improve it. For this study, the original model was based on the CHC-based factor structure developed by S.B. Kaufman et al. (2012), with two main changes: (a) in this study, only a single *g* factor was hypothesized to underlie all cognitive and achievement abilities, rather than the two used by S.B. Kaufman et al. (2012); and (b) all cross-loadings and error correlations identified by S.B. Kaufman et al. were removed. The decision to use a single *g* was made because S.B. Kaufman et al. found no evidence for separate and distinct achievement and cognitive *g* factors; also *g* played no role in this study, which was focused on seven separate CHC abilities rather than global scores or abilities. Figure

3 shows the original AMOS model, as created by S.B. Kaufman et al. (2012), without cross-loadings and error variance correlations.

Figure 3

7-Factor CHC Structure of the Kaufman Intelligence and Achievement tests as created by S.B. Kaufman et al. (2012)

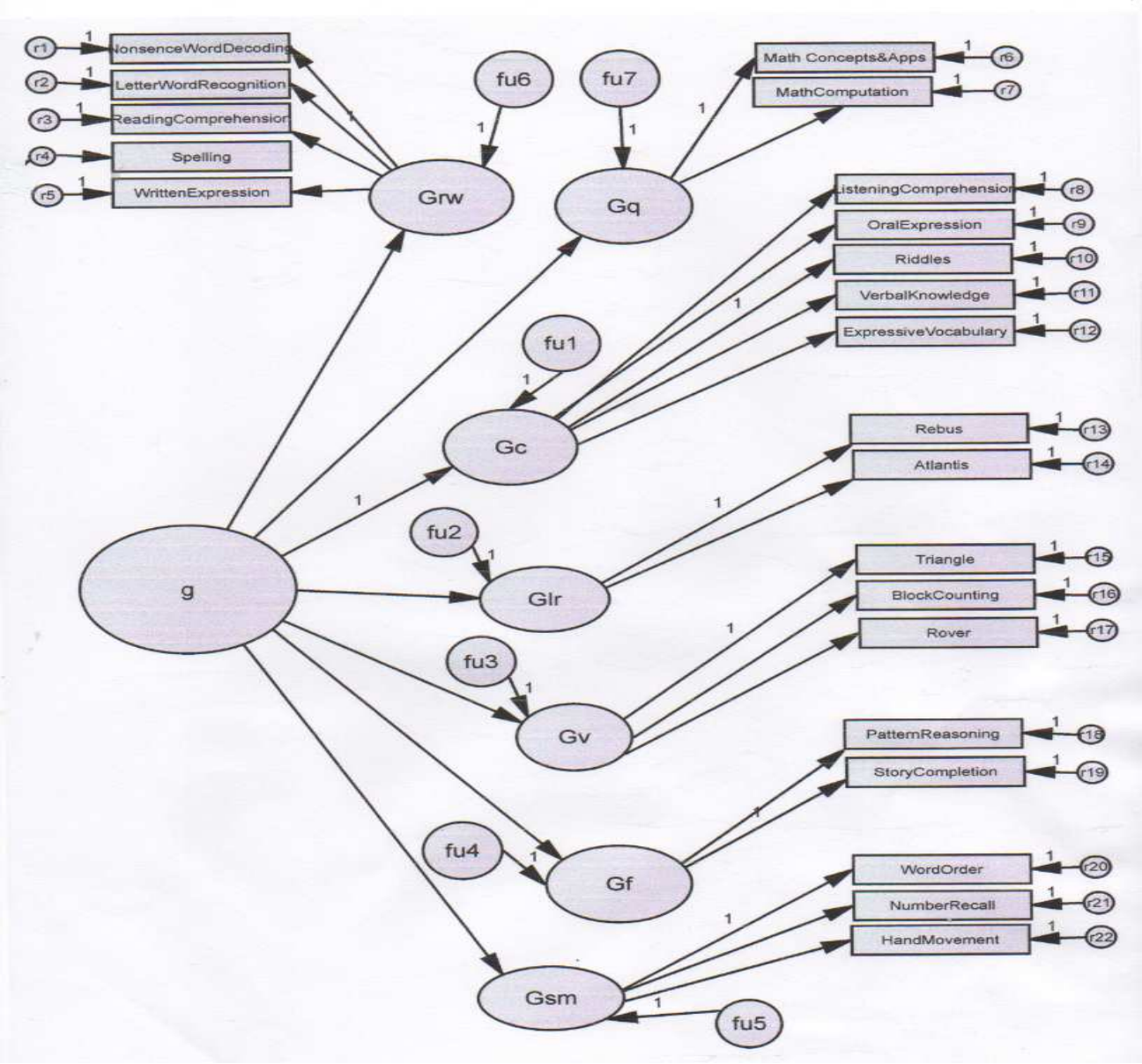


Table 9 shows the model fit for the present samples. It shows the original model fit for the total sample and for each ethnic group separately and also the final data and theory driven model for the present total sample and for each ethnic group separately.

There is no agreement among authors as to what method should be used when evaluating model fit (e.g., Chen, Sousa, & West, 2005; S.B. Kaufman et al, 2012; Reynolds & Keith, 2012). Values for model chi square ($\chi 2$), root-mean square error of approximation (RMSEA), and comparative fit index (CFI) were used to evaluate the individual models for this present study. A non-significant $\Delta\chi^2$ value, a RMSEA value close to .05, and a CFI value of .95 are considered a good model fit (Keith & Reynolds, 2012).

Based on those fit indexes, the model fit for the original model as outlined by S.B. Kaufman et al. demonstrated inadequate fit for the total sample: the values of CFI ranged from only .911-.933 and the values of RMSEA all exceeded .06.

The steps that were followed to create the final model are outlined here (the data from these intermediary steps do not appear in the tables). Based on modification indexes as well as theory, Gc was first cross-loaded with the KTEA-II Reading Comprehension subtest as well as the Written Expression subtest. This improved the fit of the model for the total sample (CFI=.944; RMSEA=.060). Reading Comprehension and Written Expression are heavily Gc loaded as both subtests measure the ability to understand ideas and have a knowledge base. For the Reading Comprehension subtest, the more knowledge one has acquired, the more one understands the content and the easier it is to make sense out of reading passages. For the Written Expression subtest, the child has to express coherent thoughts in writing in an interactive way. The child writes letters, words, and sentences as the examiner goes through a story together with the child. For example, the child is asked to complete a paragraph of a speech about time

travel. The more knowledge the child has acquired the easier it is for the child to complete the story. It is for those reasons that Gc/Knowledge heavily affects Reading Comprehension and Written Expression. The author, therefore, decided to maintain these cross-loadings, as did S. B. Kaufman et al. (2012).

Subsequently, the KABC-II Hand Movements subtest was cross-loaded with Gf, as suggested by the modification indexes. Theory and research state that Hand Movements can either cross-load on Gf (S.B. Kaufman et al., 2012) or Gv (Kaufman & Kamphaus, 1984). Cross-loadings between Gf and Hand Movements improved the model fit considerably (CFI= .949; RMSEA= .057) and was, therefore, accepted for the model. Finally, correlating the error associated with Written Expression and the error associated with Spelling improved the model fit further for the total sample (CFI=.953; RMSEA=.055). Theoretically, correlating these two error terms makes sense because both subtests require the same exact response style. They are the only two subtests that require children and adolescents to express their ideas in writing. Thus, errors that occur during the Spelling subtest are likely to occur as well during the Written Expression subtest, such as difficulties with paper-and-pencil coordination.

Based on these steps, the model was modified according to theory and data in order to improve the fit. As shown in Table 9, the final model showed good fit for the total sample with CFI values around .95 (ranging from .943 to .953) and RMSEA values below .06. Figure 4 shows the final AMOS model.

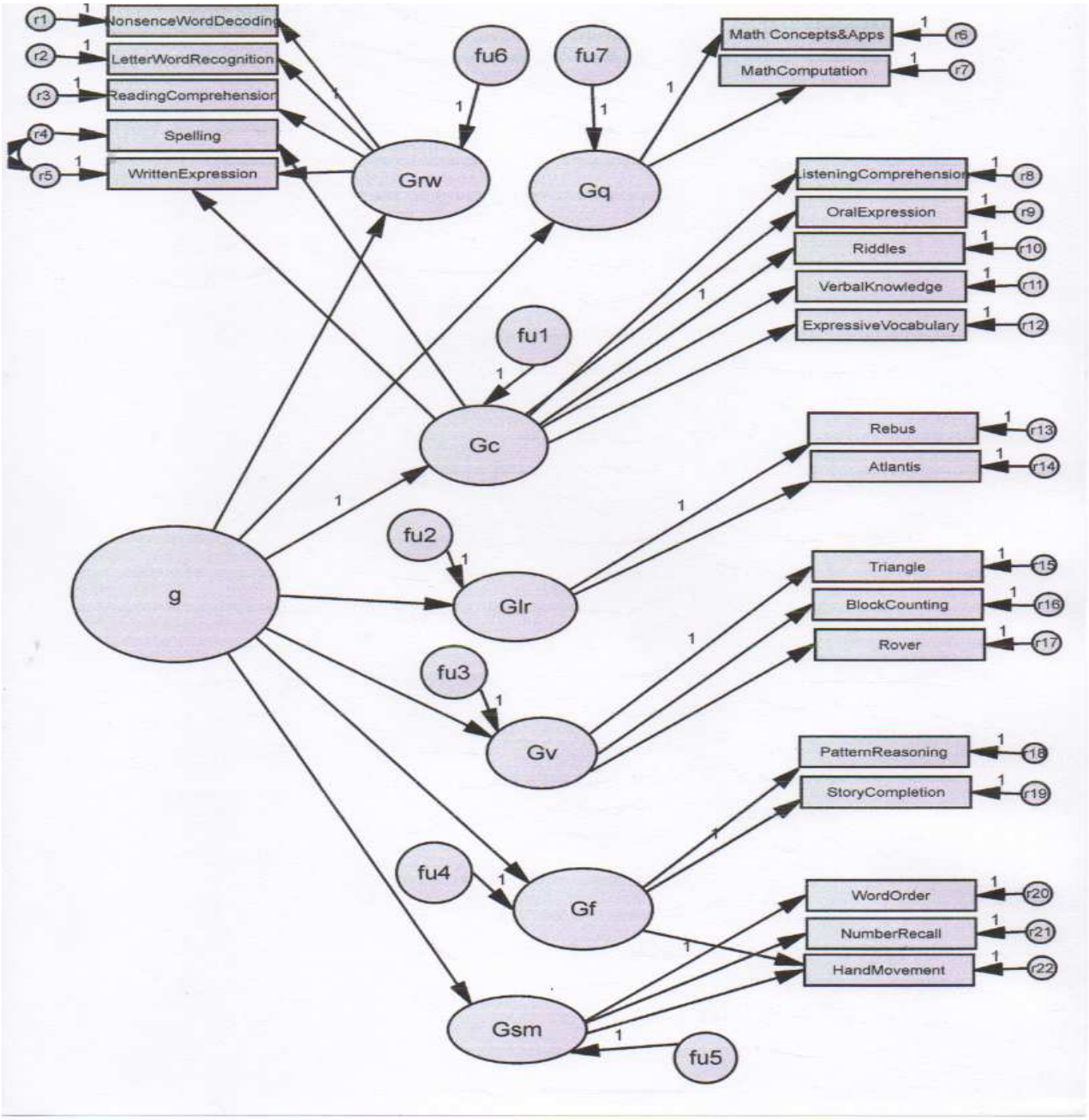
Table 9

Model Fit Indexes for Confirmatory Factor Analysis (CFA) Original and Data and Theory Driven Models For the Total Sample (N = 2001) and the three Ethnic Groups (Caucasians n = 1313; Blacks n = 312; Hispanics n = 376)

Form and Model		$\chi 2$	Df	*P*	CFI	RMSEA	RMSEA 90% Confidence Interval
Original Model							
	Total Sample	2082.2	202	< .001	0.927	0.068	.066 - .071
	Caucasians	1553.6	202	< .001	0.911	0.071	.068 - .075
	Blacks	501.2	202	< .001	0.927	0.069	.061 - .077
	Hispanics	514.3	202	< .001	0.933	0.064	.057 - .071
Data and Theory Driven Model							
	Total Sample	1397.9	198	< .001	0.953	0.055	.052 - .058
	Caucasians	1060.5	198	< .001	0.943	0.058	.054 - .061
	Blacks	412.6	198	< .001	0.948	0.059	.051 - .067
	Hispanics	437.3	198	< .001	0.949	0.057	.050 - .064

Note: CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Figure 4
Proposed 7-Factor CHC Structure of the Kaufman Intelligence and Achievement tests



Invariance analysis. Three separate invariance analyses were conducted: (1) Caucasians versus Blacks, (2) Caucasians versus Hispanics, and (3) Blacks versus Hispanics. For each invariance analysis, Meredith's (1993) hierarchy of setting increasingly restrictive sets of equality constraints was applied. That is, first configural invariance was established, then metric invariance was established, and finally intercept invariance was established.

The comparison of model fit was evaluated with ΔCFI and ΔRMSEA. According to Cheung and Rensvold (2002), ΔCFI >.01 and ΔRMSEA>.02 are considered meaningful changes in model fit and therefore indicate a significant degradation of goodness of fit. It is important to note that $\Delta\chi^2$ can easily result in a significant degradation in model fit, even if the model is actually a good fit. This is because χ^2 detects minor, insignificant differences that have no theoretical or practical consequences when (a) the sample size is large, (b) the model is complex, and (c) the number of constraints increases (Browne & Cudeck, 1993; Byrne, 2010; Reynolds & Keith, 2013). Due to the complexity, sample size, and number of constraints of this present model, ΔCFI and ΔRMSEA were given more value when evaluating the goodness of fit of the model and degradation of goodness of fit.

Caucasian-Black invariance analysis. Model fit for the Caucasian-Black comparisons are shown in table 10. As the table demonstrates, the Caucasian-Black configural model fit well (CFI=.944; RMSEA=.057). The model fit did not degrade when first order loadings were constrained to be equal ($\chi2$ (414)=3.61, p>.05). No changes in CFI or RMSEA occurred. Model change did not degrade when second factor loadings were added ($\chi2$ (420)=3.57, p>.05). Again, no changes in CFI or RMSEA were detected. Thus, metric invariance was established.

Intercept invariance had to be established next. For this analysis, the subtest intercepts were constrained to be equal across the groups. These additional constraints resulted in a

significant degradation in model fit according to $\Delta\chi 2$ ($\chi 2$ (435)=3.69, $p<.001$). However, ΔCFI was negligible and RMSEA did not change at all. Furthermore, the CFI value close to .95 and the RMSEA value of less than .06 provided further evidence of a good model fit. As discussed previously, $\Delta\chi 2$ can easily show as significant when the number of constraints increases, even if the changes are of no theoretical and practical meaning. Results based on CFI and RMSEA provide evidence for intercept invariance.

Table 10
Model Fit Indexes and Nested Comparisons for Confirmatory Factor Analysis (CFA) Models: Caucasian and Black Comparison (N = 1625) (Caucasians n = 1313; Blacks n = 312)

Form and Model		χ2	*df*	Δχ2	Δ*df*	*P*	CFI	ΔCFI	RMSEA	ΔRMSEA	RMSEA 90% confidence interval
Configural Invariance		1473.1	396				0.944		.057		.055 - .061
Metric Invariance											
	Model 1 (measurement weights)	1492.7	414	19.6	18	0.358	0.944	.000	.057	.000	.054 - .059
	Model 2 (structural weights)	1501.0	420	8.3	6	0.215	0.944	.000	.057	.000	.054 - .059
Intercept Invariance											
	Model 3 (measurement intercepts)	1605.4	435	104.4	15	< .001	0.939	.005	.057	.000	.055 - .061

Caucasian-Hispanics invariance analysis. Table 11 shows the Caucasian-Hispanic invariance analysis. As demonstrated in the table, the Caucasian-Hispanic configural model fit well (CFI=..944; RMSEA=.057). When factor loadings were added to the first order factors, no significant degradation in model fit was observed ($\chi 2$ (414)=3.69, p>.05). Furthermore, there were no changes in CFI or RMSEA values. When the second order factor loading was constrained, in addition to the first order factor loadings, $\Delta\chi 2$ resulted in a significant degradation in model fit ($\chi 2$ (420)=3.68, p<.05). However, ΔCFI was nearly non-existent and there was no change in RMSEA value. Again, given the complexity of the model, the increase in the number of constraints, and the large sample size, RMSEA and CFI values were weighted more than $\chi 2$.

Adding additional equality constraints to the subtest intercepts resulted in a significant $\chi 2$ value ($\chi 2$ (435)=3.92, p<.01). However, changes in CFI and RMSEA were again far from their suggested cutoff values of .01 (for CFI) and .02 (for RMSEA). Thus, intercept invariance was established.

Table 11
Model Fit Indexes and Nested Comparisons for Confirmatory Factor Analysis (CFA) Models: Caucasian and Hispanic Comparison (N = 1,689) (Caucasians n = 1313; Hispanics n = 376)

Form and Model		χ2	*df*	Δχ2	Δ*df*	*P*	CFI	ΔCFI	RMSEA	ΔRMSEA	RMSEA 90% Confidence Interval
Configural invariance		1497.8	396				0.944		.057		.054 - .061
Metric Invariance											
	Model 1 (measurement weights)	1525.6	414	27.858	18	0.064	0.944	0.00	.057	.000	.054 - .059
	Model 2 (structural weights)	1544.9	420	19.305	6	< .05	0.943	.001	.057	.000	.054 - .059
Intercept Invariance											
	Model 3 (measurement intercepts)	1704.7	435	159.799	15	< .01	0.936	.007	.059	.002	.057 - .062

Note: CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Black-Hispanic invariance analysis. The Black-Hispanic invariance analysis is demonstrated in Table 12. As the table shows, the configural model indicated good fit (CFI=.948; RMSEA=.057). Adding constraints to the second order factor loadings resulted in a significant degradation in model fit when using $\Delta\chi 2$ as the criterion ($\chi 2(414)=2.14$, $p<.05$). However, ΔCFI was again non-significant and RMSEA did not change at all. When compared to the previous model, no significant changes in $\chi 2$ were observed between the model that constrained the second order factor loadings and the model that constrained the first order factor loading ($\chi 2(420)=2.12$, $p>.05$). No changes in CFI or RMSEA were detected.

Adding additional constraints to the subtest intercepts resulted in a significant degradation in model fit according to $\chi 2$ only ($\chi 2(435)=2.20$, $p<.01$). However, negligible changes in CFI and RMSEA values provide evidence for factorial invariance on the intercept level.

Table 12

Model Fit Indexes and Nested Comparisons for Confirmatory Factor Analysis (CFA) Models: Black and Hispanic Comparison (N = 688) (Black n = 312; Hispanics n = 376)

		χ2	*Df*	Δχ2	Δ*df*	*P*	CFI	ΔCFI	RMSEA	ΔRMSEA	RMSEA 90% Confidence Interval
Configural Invariance		849.9	396				0.948		.057		.052 - .063
Metric Invariance											
	Model 1 (measurement weights)	884.2	414	34.4	18	< .05	0.946	.002	.057	.000	.052 - .063
	Model 2 (structural weights)	891.7	420	7.4	6	0.282	0.946	0.00	.057	.000	.052 - .062
Intercept Invariance											
	Model 3 (measurement intercepts)	958.0	435	66.4	15	< .01	0.940	.006	.059	.000	.054 - .064

Note: CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation

Conclusion. In sum, factorial invariance on the configural, metric, and intercept level was established for every CHC factor across all three ethnic group comparisons. Even though the changes in $\chi 2$ occasionally resulted in significant values, this finding was anticipated given the complexity of the model, the number of constraints, and the sample size. The CFI and RMSEA values were more defensible statistics for evaluating degradation of model fit in this study; the conclusion of factorial invariance across the three ethnic groups is based on the ΔCFI and ΔRMSEA values, which were negligible in each invariance analysis. Therefore, the results of these analyses provided strong evidence for good model fit for each CHC factor for Caucasians, Blacks, and Hispanics. CFI values were all around .95 and RMSEA values were all <.06. Furthermore, none of the ΔCFI and ΔRMSEA values came close to the suggested cutoff lines of .01 for CFI and .02 for RMSEA. All differences in RMSEA and CFI were non-existent or trivial.

Comparison of the CHC abilities using Multivariate Analysis of Covariance (MANCOVA) (Question #2)

Question 2 asks: Do the three ethnic groups differ significantly in their mean scores on the CHC latent variables that underlie the Kaufman tests, based on CFA (when factor invariance is found) and/or on the subtests that compose these factors (when invariance is *not* found)?

Assumptions. In addition to three of the assumptions that were already discussed and tested in question 1 for the CFA, including (1) independence, (2) linearity, and (3) normality, there were four more assumptions underlying MANCOVA analyses that needed to be met: (4) multivariate normality, (5) homogeneity of regression, (6) homogeneity of the variances, and (7) homogeneity of the covariances.

(4) Multivariate normality: For illustrative purposes, Figures 5 and 6 show normality for the KABC-II Planning/Gf scale for Blacks (N=312) ($\bar{x}$=94.1; SD=13.8) and for the KTEA-II Reading composite for Hispanics (N=376) ($\bar{x}$=94.2; SD=14.3), respectively. Separate histograms for each independent variable (for each of the four KTEA-II composites and each of the five KABC-II CHC scales) were visually explored to test the assumption of multivariate normality. Multivariate normality was met for each independent variable.

Figure 5

Distribution of KABC-II Planning/Gf for Blacks (N=312).

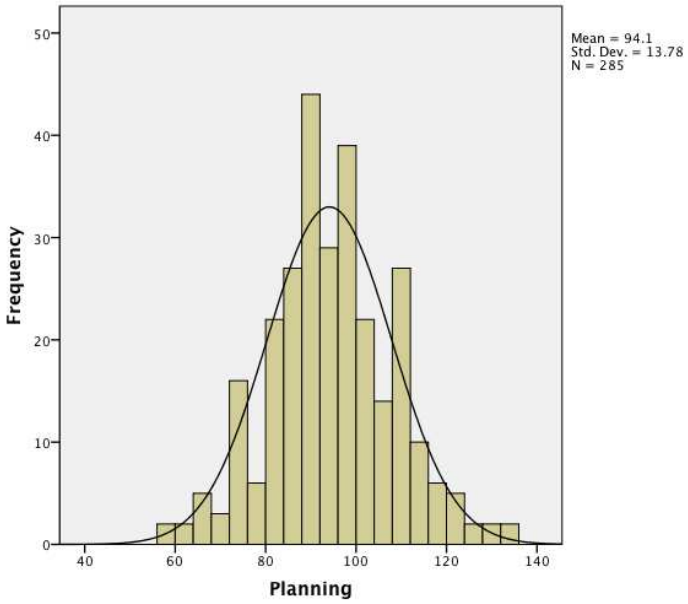
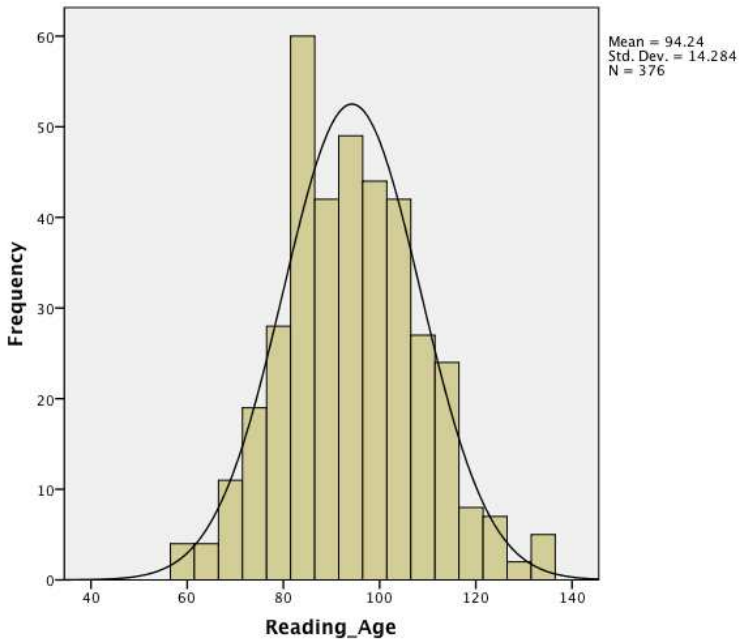


Figure 6

Distribution of KTEA-II Reading for Hispanics (N=376)



(5) Homogeneity of regression: Homogeneity of the regression line was previously tested in two preliminary analyses using ANOVAs with ethnicity and SES (parent's education) as independent variables. In the first ANOVA, KABC-II Mental Processing Index (MPI) was the dependent variable and in the second ANOVA KTEA-II Comprehensive Achievement Composite was the dependent variables (Tables 4 and 5). The results indicated that the SES X ethnicity interaction was NOT significant and, thus, SES was deemed an appropriate covariate to use for the present analyses. However, in order to fully confirm that SES was a suitable covariate, it was desirable to redo the ANOVAs based on the results of the CFA analyses conducted to answer Question 1.

These analyses demonstrated invariance across ethnic groups, permitting direct comparison of the ethnic groups' mean scores on the seven CHC-based latent roots. Had there been a *lack* of

invariance on one or more of the latent roots, then ethnic group comparisons would have been at the *subtest* level for the "not" invariant factors. So based on the invariance established in Question 1, homogeneity of regression had to be established specifically for the "final" set of dependent variables in the analyses: (a) the five KABC-II CHC-based scales, which measure *Gc, Gf, Gv, Glr,* and *Gsm*, and (b) the four KTEA-II composites that measure *Gc* (Oral Language), *Gq* (Math), and *Grw* (Reading and Written Expression). Separate KTEA-II and KABC-II ANOVAs were conducted with ethnicity and SES as the independent variables; for the KTEA-II ANOVA, the four composites served as dependent variables and for the KABC-II the five scales were the dependent variables). Tables 13 and 14 show the results. With the exception of the KABC-II Simultaneous/Gv variable (p = .041), none of the KTEA-II and KABC-II variables showed significant ethnicity x SES interactions. When the Bonferroni correction is applied to the five simultaneous analyses of dependent variables, even the value of p =.041 falls far short of the value of p = .01 that is needed to achieve a family-wise alpha level of .05. Such results confirmed that the homogeneity of regression assumption was met and that SES was a suitable covariate.

Table 13

Between-Subjects ANOVA Effects: Ethnicity by SES (Parent's Education) with KTEA-II Composites (Reading, Writing, Math, and Oral Language) as Dependent Variables

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	Reading	66373.855[a]	11	6033.987	31.110	.001
	Math	63111.688[b]	11	5737.426	31.074	.001
	Written Language	55005.461[c]	11	5000.496	25.725	.001
	Oral Language	73805.997[d]	11	6709.636	35.965	.001
Intercept	Reading	10497988.048	1	10497988.048	54126.141	.001
	Math	10513805.894	1	10513805.894	56943.814	.001
	Written Language	10555279.263	1	10555279.263	54300.791	.001
	Oral Language	10534222.872	1	10534222.872	56465.262	.001
Ethnicity	Reading	7211.035	2	3605.517	18.590	.001
	Math	8289.177	2	4144.589	22.448	.001
	Written Language	7286.778	2	3643.389	18.743	.001
	Oral Language	17003.839	2	8501.920	45.572	.001
SES	Reading	30122.279	3	10040.760	51.769	.001
	Math	24658.495	3	8219.498	44.518	.001
	Written Language	23847.884	3	7949.295	40.895	.001
	Oral Language	23721.424	3	7907.141	42.384	.001
Ethnicity * SES	Reading	919.428	6	153.238	.790	.578
	Math	1218.220	6	203.037	1.100	.360
	Written Language	1059.380	6	176.563	.908	.488
	Oral Language	962.541	6	160.424	.860	.524

Table 14

Between-Subjects ANOVA Effects: Ethnicity by SES (Parent's Education) with KABC-II Scales (Sequential, Knowledge, Planning, Learning, and, Simultaneous) as Dependent Variables

Source	Dependent Variable	Type III Sum of Squares	Df	Mean Square	F	Sig.
Corrected Model	Sequential	41193.607[a]	11	3744.873	17.926	.001
	Simultaneous	44803.146[b]	11	4073.013	20.097	.001
	Planning	48603.485[c]	11	4418.499	21.772	.001
	Learning	34945.298[d]	11	3176.845	14.962	.001
	Knowledge	105191.651[e]	11	9562.877	55.940	.001
Intercept	Sequential	10859937.943	1	10859937.943	51983.728	.001
	Simultaneous	10532686.123	1	10532686.123	51971.359	.001
	Planning	10631991.126	1	10631991.126	52388.541	.001
	Learning	10835154.166	1	10835154.166	51029.967	.001
	Knowledge	10364185.516	1	10364185.516	60627.813	.001
Race	Sequential	4843.653	2	2421.827	11.593	.001
	Simultaneous	14195.460	2	7097.730	35.022	.001
	Planning	7666.163	2	3833.082	18.887	.001
	Learning	4140.597	2	2070.299	9.750	.001
	Knowledge	23995.806	2	11997.903	70.185	.001
SES	Sequential	17106.568	3	5702.189	27.295	.001
	Simultaneous	14218.683	3	4739.561	23.386	.001
	Planning	17164.014	3	5721.338	28.192	.001
	Learning	12522.987	3	4174.329	19.660	.001
	Knowledge	30632.265	3	10210.755	59.730	.001

Race * SES	Sequential	2037.556	6	339.593	1.626	.136
	Simultaneous	2662.174	6	443.696	2.189	.041
	Planning	2006.394	6	334.399	1.648	.130
	Learning	1050.881	6	175.147	.825	.550
	Knowledge	830.256	6	138.376	.809	.562

(6) Homogeneity of the variances: This assumption was tested with the Levene test. A non-significant Levene indicated that this assumption was met. Tables 15 and 16 show the Levene tests for the KTEA-II and the KABC-II, respectively. With the exception of the KABC-II Sequential/Gsm factor (p = .01), none of the Levene tests was significant; thus, the assumption of homogeneity of the variances was met. For the Sequential/Gsm factor, a more stringent alpha level of .01 instead of .05 was used for the MANCOVA analyses to compensate for the significant Levene test (Meyers et al., 2013).

Table 15

KTEA-II Levene's Test of Equality of Error Variances

	F	df1	df2	Sig.
Reading	.325	2	1998	.723
Math	1.589	2	1998	.204
Written Language	.516	2	1998	.597
Oral Language	.270	2	1998	.763

Table 16

KABC-II Levene's Test of Equality of Error Variances

	F	Df1	df2	Sig.
Sequential/Gsm	4.588	2	1998	.010
Simultaneous/Gv	2.766	2	1998	.063
Planning/Gf	1.088	2	1998	.337
Learning/Glr	.451	2	1998	.637
Knowledge/Gc	.065	2	1998	.937

(7) Homogeneity of the covariances: This assumption was tested with Box's M test. Tables 17 and 18 represent Box's M and Bartlett's Test of Sphericity for the KTEA-II. Tables 19 and 20 represent Box's M and Bartlett's Test of Sphericity for the KABC-II. Both Box's M and Bartlett's Tests of Sphericity were significant for both the KTEA-II and the KABC-II, which indicated that the assumption of homogeneity of the convariances was not met. More stringent alpha levels of .01 had to be used for the remainder of the analyses (Meyers et al., 2013).

Table 17

KTEA-II Box's Test of Equality of Covariance Matrices

Box's M	44.483
F	2.214
df1	20
df2	3020329.551
Sig.	.001

Table 18

KTEA-II Bartlett's Test of Sphericity

Likelihood Ratio	.000
Approx. Chi-Square	4113.668
Df	9
Sig.	.001

Table 19

KABC-II Box's Test of Equality of Covariance Matrices

Box's M	65.321
F	2.165
df1	30
df2	2743527.224
Sig.	.001

Table 20

KABC-II Bartlett's Test of Sphericity

Likelihood Ratio	.000
Approx. Chi-Square	2535.496
Df	14
Sig.	.001

Multivariate analysis. As mentioned when discussing assumptions underlying the MANCOVAs. the analyses that were conducted for Question 2 were directly affected by the results of Question 1. Since factorial invariance was established for all three ethnic groups across all CHC factors, group comparisons could be made on each of the CHC factors. There was no need to compare means on the subtest level, because factorial invariance for all seven CHC factors was established in the invariance analyses for Caucasians versus Blacks, Caucasians versus Hispanics, and Hispanics versus Blacks. As a result, two separate MANCOVAS were conducted with ethnicity as the independent variable, one for the KTEA-II and one for the KABC-II (results are shown in Tables 21 and 22).

SES (parental education) was covaried in each MANCOVA. For the KTEA-II the four achievement composites (Reading/Grw, Written Language/Grw, Math/Gq, and Oral Language/Gc) served as the dependent variables and for the KABC-II the five ability scales (Sequential/Gsm, Simultaneous/Gv, Planning/Gf, Learning/Glr, and Knowledge/Gc) were the dependent variables. Therefore Gc was evaluated in both of these MANCOVAS (KTEA-II: Oral Language and KABC-II: Knowledge) and the latent trait of Grw was evaluated by using the Reading and Written Language composites of the KTEA-II.

Tables 21 and 22 demonstrate the MANCOVA results for the KTEA-II and the KABC-II, respectively. The two MANCOVAs conducted for both the KTEA-II and the KABC-II showed significant mean score differences on all KTEA-II and KABC-II variables among the three ethnic groups. This significance at $p = .001$, regarding the Intercept, the main effect of Ethnicity, and the covariate of SES, was demonstrated in both MANCOVAs by all multivariate indexes (Pillai's Trace, Wilks' Lambda, Hotelling's Trace, Roy's Largest Root).

Table 21
KTEA-II Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.876	3516.622[b]	4.000	1994.000	.001
	Wilks' Lambda	.124	3516.622[b]	4.000	1994.000	.001
	Hotelling's Trace	7.054	3516.622[b]	4.000	1994.000	.001
	Roy's Largest Root	7.054	3516.622[b]	4.000	1994.000	.001
SES	Pillai's Trace	.116	65.341[b]	4.000	1994.000	.001
	Wilks' Lambda	.884	65.341[b]	4.000	1994.000	.001
	Hotelling's Trace	.131	65.341[b]	4.000	1994.000	.001
	Roy's Largest Root	.131	65.341[b]	4.000	1994.000	.001
Ethnicity	Pillai's Trace	.059	15.257	8.000	3990.000	.001
	Wilks' Lambda	.941	15.419[b]	8.000	3988.000	.001
	Hotelling's Trace	.063	15.581	8.000	3986.000	.001
	Roy's Largest Root	.058	28.816[c]	4.000	1995.000	.001

Table 22
KABC-II Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.904	3759.483[b]	5.000	1993.000	.001
	Wilks' Lambda	.096	3759.483[b]	5.000	1993.000	.001
	Hotelling's Trace	9.432	3759.483[b]	5.000	1993.000	.001
	Roy's Largest Root	9.432	3759.483[b]	5.000	1993.000	.001
SES	Pillai's Trace	.126	57.302[b]	5.000	1993.000	.001
	Wilks' Lambda	.874	57.302[b]	5.000	1993.000	.001
	Hotelling's Trace	.144	57.302[b]	5.000	1993.000	.001
	Roy's Largest Root	.144	57.302[b]	5.000	1993.000	.001
Ethnicity	Pillai's Trace	.124	26.371	10.000	3988.000	.001
	Wilks' Lambda	.879	26.527[b]	10.000	3986.000	.001
	Hotelling's Trace	.134	26.684	10.000	3984.000	.001
	Roy's Largest Root	.097	38.745[c]	5.000	1994.000	.001

Following the significant MANCOVAs, planned comparisons were conducted in order to explore which ethnic group pairs differed significantly on which CHC factors, and which variables yielded the largest effect sizes. Table 23 portrays the planned comparison results. All results reflect adjustment for the covariate of SES; the Bonferroni correction, at both a .05 and .01 level, was applied to control for the errors that occur when many comparisons are made simultaneously. In order to evaluate the magnitude of the differences (effect size) Cohen's *d* was used. The following terminology best describes different levels of Cohen's *d*: values of about 0.2 indicate small differences, values of 0.5 translate into moderate differences, and values of 0.8 or larger reflect large differences (Meyer et al., 2013).

Caucasians and Blacks did not differ significantly at the .01 (or at the .05) level on the *Gsm*, as measured by the KABC-II Sequential/Gsm scale, but Caucasians scored significantly higher ($p < .01$) than Blacks on all other CHC factors on both the KABC-II and KTEA-II. Significant differences were small to moderate in magnitude, ranging from 0.22 on *Glr* to .55-56 on *Gc* and *Gv*. (The value for *Gc* is the mid-point of the Caucasian-Black differences on the separate measures of *Gc*.)

Caucasians and Hispanics did not differ significantly at the .01 (or at the .05) level on measures of *Gf* and *Gv*, both of which minimize the role of language. Caucasians scored significantly higher than Hispanics ($p <.01$) on all other CHC abilities, but except for the language-oriented *Gc* scales ($d = 0.48$), which produced a small-to-moderate effect size, all other effect sizes were small (about .20-.30).

Hispanics and Blacks did not differ significantly on *Gc*, *Glr*, or *Grw*. Hispanics scored significantly higher on *Gv* ($p <.01$) and *Gf* ($p <.01$).with moderate effect sizes ($d = 0.3$-0.4). The Black-Hispanic differences in favor of Hispanics on *G*q was significant at $p < .05$ but not $p < .01$.

Similarly, Blacks outperformed Hispanics on *Gsm* ($p < .05$, $d = .25$). Both G*q* and G*sm* results can only be interpreted tentatively because Box'M and Bartlett's Test of Sphericity were significant; strictly, speaking, the differences on *Gsm* and *G*q should be considered non-significant.

Table 23

Ethnic differences on seven CHC factors for Caucasians, Blacks, and Hispanics, adjusted for SES (parental education)

	Mean Standard Score Adjusted for SES			Mean Ethnic Group Differences					
	Caucasian (n=1313)	Black (n=312)	Hispanic (n=376)	Caucasian-Black		Caucasian-Hispanic		Hispanic-Black	
CHC Factor				SS	SD units (d)	SS	SD units (d)	SS	SD units (d)
Gc-Crystallized Ability									
KABC-II Knowledge/Gc	102.9	93.8	95.1	9.2**	.61	7.9**	.52	1.3	.09
KTEA-II Oral Language	102.9	95.5	96.3	7.3**	.49	6.5**	.44	0.8	.05
Gf-Fluid Reasoning									
KABC-II Planning/Gf	101.5	94.9	99.3	6.5**	.44	2.1	.14	4.4**	.29
Gv-Visual Processing									
KABC-II Simultaneous/Gv	101.6	93.2	99.6	8.4**	.56	1.9	.13	6.5**	.43
Glr-Long Term Storage & Retrieval									
KABC-II Learning Ability/Glr	101.7	98.4	97.6	3.3**	.22	4.0**	.27	-0.7	.05
Gsm-Short Term Memory									
KABC-II Sequential Processing/Gsm	101.4	99.8	96.1	1.5	.10	5.3**	.35	-3.8*+	.25
Grw-Reading and Writing Ability									
KTEA-II Written Language	101.5	95.5	97.9	5.7**	.38	3.4**	.23	2.3	.16
KTEA-II Reading	101.2	95.6	97.2	5.5**	.37	3.9**	.26	1.6	.11
Gq-Quantitative Reasoning									
KTEA-II Math	101.6	94.9	98.3	6.7**	.44	3.3**	.22	3.4*	.22

Note: SS = standard score; SD units = standard deviation units.

*$p < .05$ based on Bonferroni correction

**$p < .01$ based on Bonferroni correction

Caucasian-Black differences = mean for Caucasians minus means for Blacks. Caucasian-Hispanic differences = mean for Caucasians minus means for Hispanics. Hispanic Black differences = mean for Hispanics minus means for Blacks (e.g. ,+). In the MANCOVA for Hispanic-Black differences, the Levene Test of Equality of Error was violated for the KABC-II Sequential Processing/Gsm index. That violation required a more stringent alpha level of .01 for that comparison. Therefore, even though the difference of -3.8 points in favor of Blacks is asterisked to denote significance at the .05 alpha level, that difference should not be interpreted as meaningful because it fell short of the .01 level

Conclusion. Mean score differences were compared using two MANCOVAs, one MANCOVA using the four KTEA-II CHC variables as dependent variables (Reading/Grw, Writing/Grw, Math/Gq, and Oral Language/Gc) and the other MANCOVA using the five KABC-II CHC scales as dependent variables (Sequential/Gsm, Simultaneous/Gv, Planning/Gf, Learning/Glr, and Knowledge/Gc). All analyses were adjusted for SES (parent education attainment). MANCOVA results revealed significant ethnic group differences on all KTEA-II and KABC-II variables. Despite the strict Bonferroni corrected alpha level, differences as small as 3 standard score points were significant. Thus, it is unlikely that Type II errors were committed. In general, Caucasians scored significantly higher than Blacks and Hispanics on most CHC factors, with effect sizes in the small to moderate range. Important exceptions—no significant *Gsm* difference versus Blacks and no significant *Gf* and *Gv* differences versus Hispanics. Blacks and Hispanics differed significantly on some CHC factors with the most notable differences being a small-to-moderate Hispanic advantage on *Gf*, *Gv*, and *Gq*.

SES as a covariate. It is of interest to determine whether SES was effective as a covariate in the MANCOVAs. Did it reduce ethnic differences? Was it equally effective in reducing differences in each of the three analyses? Did it reduce ethnic differences on scales with a cultural component (*Gc*, *Grw*, *Gq*) to the same extent as on scales with "culturally-neutral" stimuli (*Gf*, *Gv*, *Gsm*, *Glr*)? Logically, controlling for SES should make a bigger difference on scales that are especially sensitive to cultural opportunities, such as general information and math ability, than on scales that measure nonverbal problem solving.

Table 24 shows the effect size (Cohen's *d*) for each ethnic comparison, both with and without the SES covariate. Clearly, covarying SES was effective in reducing the size of the ethnic difference in each analysis. However, SES functioned differently in each of the three

MANCOVAs. It had a substantial effect in the Caucasian-Hispanic analysis: the Caucasian advantage over Hispanics reduced by .18-.28 *SD* units (about 2 ½-4 standard-score points) when SES was controlled. In contrast, it had almost a trivial effect in the Caucasian-Black analysis as the Caucasian advantage reduced only slightly (.06-.10 *SD* units = 1-1½ points). In the Hispanic-Black MANCOVA, covarying SES reduced ethnic differences moderately (.11-.18 *SD* units = 1½-2½ points). Covarying SES in the latter analysis had the effect of increasing scores earned by Hispanics and decreasing scores earned by Blacks.

In each of the three analyses, SES reduced ethnic differences to a greater extent on "cultural" scales than on "culturally-neutral" scales. In the Caucasian-Hispanic analysis, for example, covarying SES reduced the Caucasian advantage by 3½-4 points on measures of *Gc*, *Gq*, and *Grw*, versus 3 points on *Gf*, *Gv*, and the memory scales.

Table 24
KABC-II and KTEA-II ethnic group differences with and without SES adjustments

	Caucasian-Black			**Caucasian-Hispanic**			**Hispanic-Black**		
CHC Factor	Obtained	SES-Adjusted	Difference in *d*	Obtained	SES-Adjusted	Difference in *d*	Obtained	SES-Adjusted	Difference in *d*
Gc-Crystallized Ability									
KABC-II Knowledge/Gc	.71	.61	.10	.80	.52	.28	-.09	.09	.18
KTEA-II Oral Language	.57	.49	.08	.67	.44	.23	-.10	.05	.15
Gf-Fluid Reasoning									
KABC-II Planning/Gf	.51	.44	.07	.35	.14	.21	.15	.29	.14
Gv-Visual Processing									
KABC-II Simultaneous/Gv	.63	.56	.07	.31	.13	.18	.32	.43	.11
Glr-Long Term Storage & Retrieval									
KABC-II Learning Ability/Glr	.28	.22	.06	.45	.27	.18	-.17	-.05	.12
Gsm-Short Term Memory									
KABC-II Sequential Processing/Gsm	.16	.10	.06	.54	.35	.19	-.38	-.25	.13
Grw-Reading and Writing Ability									
KTEA-II Written Language	.46	.38	.08	.46	.23	.23	.00	.16	.16
KTEA-II Reading	.45	.37	.08	.53	.26	.27	-.07	.11	.18
Gq-Quantitative Reasoning									
KTEA-II Math	.53	.44	.09	.47	.22	,25	.05	.22	.17

Assessing prediction bias using structural equation modeling (Question #3).

Question 3 asks: On the Kaufman tests, is there predictive validity bias across the different ethnic groups? Do the general factor (g) and the five CHC-based cognitive factors as measured by the KABC-II (Sequential/Gsm, Simultaneous/Gv, Planning/Gf, Learning/Glr, and Knowledge/Gc) predict the KTEA-II achievement composites (reading, math, and written language) equally well (magnitude of the coefficients) for Blacks, Hispanics, and Caucasians? The results of the structural equation modeling were evaluated and summarized in the following sections: a) means and standard deviations, b) correlations, c) prediction invariance.

In order to assess prediction invariance, a model fit method was employed for each pair (Caucasians versus Blacks, Caucasians versus Hispanics, and Hispanics versus Blacks). The model fit was evaluated in a stepwise analysis, by testing the invariance of the variance, slope, and intercept of the regression lines. Residual invariance was not a necessary prerequisite to determine prediction invariance (Reynolds & Keith, 2012). However, slope and intercept invariance needed to be met in order to determine prediction NON-bias. If the slope restriction did not result in a degradation of model fit, slope invariance was established (*weak* invariance). Next, the intercepts were constrained to be equal. If the slope and intercept constraints did not result in a significant degradation of model fit, prediction invariance was established (*strong factorial invariance)*. The fit of the models was evaluated with $\Delta\chi^2$. RMSEA and CFI were employed as alternative fit indexes.

Means and Standard Deviations. Table 25 presents the means and standard deviations for the five KABC-II predictor variables as well as the global Fluid-Crystallized Index (FCI) and the three KTEA-II outcome variables, by ethnic subsample, separately for the three grade groups. On both the KABC-II and the KTEA-II, Caucasians produced means ranging from 101 to 103

standard score points. Mean scores for Blacks ranged from 93 to 98, with the exception of the KABC-II Sequential/Gsm ability factor, which produced means of 97 to 103. Hispanics' mean scores ranged from 92 to 97. In general, *SD*s were in the expected range of 14-15 for each ethnicity at each grade level on all cognitive and achievement variables (even though there were a few instances were SDs exceeded 15 by 0.5 to 1.5 standard score point. In one occasion the SD reached 17.3 points). Homogeneity of the variance was tested by using Hartley's Fmax statistic. As indicated by Meyers et al. (2013), Fmax is not problematic as long as it does not exceed 4 points. Fmax ranged from 1.1 to 1.3 on the KABC-II variables and from 1.1 to 1.2 on the KTEA-II variables. Hence, there were no problems regarding homogeneity of the variance for the present samples.

Information about variability is important for the interpretation of validity coefficients. Other things being equal, coefficients are spuriously inflated when samples are unusually heterogeneous and are spuriously low for samples that are restricted in range (Urbina, 2014).

Table 25

Means and Standard Deviations for each Ethnic Group across Grade Groups for each KABC-II Predictor Variable

	Caucasians		Blacks		Hispanics	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Grade 1-4						
Sequential/Gsm	101.91	14.42	103.08	16.55	95.39	16.46
Simultaneous/Gv	101.84	15.03	93.81	13.86	97.37	14.58
Planning/Gf	101.40	14.86	96.10	13.39	97.49	14.98
Learning/Glr	102.84	15.39	99.42	13.81	96.81	14.99
Knowledge/Gc	103.53	14.47	94.05	13.13	91.92	14.85
Fluid-Crystallized Index	102.71	14.93	95.80	13.34	94.01	15.42
Grade 5-8						
Sequential/Gsm	102.33	14.55	97.65	15.55	93.37	12.57
Simultaneous/Gv	101.75	15.36	93.46	13.82	98.16	12.78
Planning/Gf	102.48	15.02	93.80	13.28	95.99	14.34
Learning/Glr	102.13	14.77	98.92	15.29	94.73	14.99
Knowledge/Gc	104.37	13.32	92.24	14.51	90.77	13.23
Fluid-Crystallized Index	103.43	14.31	93.64	13.91	92.71	13.33
Grade 9-12						
Sequential/Gsm	101.74	13.67	96.97	16.19	92.55	17.29
Simultaneous/Gv	103.27	14.29	90.53	13.97	97.36	13.66
Planning/Gf	102.97	14.98	93.51	15.13	97.20	14.63
Learning/Glr	101.85	14.97	94.68	12.04	94.91	14.83
Knowledge/Gc	103.85	13.60	93.95	15.08	93.46	13.95
Fluid-Crystallized Index	103.73	13.94	92.33	13.54	93.74	13.44

Note: Samples Sizes Grades 1-4: Caucasians (455), Blacks (119), Hispanics (150);
Grades 5-8: Caucasians (487), Blacks (119), Hispanics (137);
Grades 9-12: Caucasians (371), Blacks (74), Hispanics (89).

Table 25 continued

Means and Standard Deviations for each Ethnic Group across Grade Groups for each KTEA-II Outcome Variable

	Caucasians		Blacks		Hispanics	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
			Grade 1-4			
Math	101.03	14.40	96.38	13.76	96.30	15.03
Reading	100.55	15.15	96.68	13.89	94.05	14.88
Written Language	100.93	15.00	97.06	15.22	95.81	15.48
			Grade 5-8			
Math	102.86	14.14	93.99	13.63	94.31	13.78
Reading	103.35	14.21	95.08	15.55	94.01	13.27
Written Language	103.28	14.29	94.47	14.99	93.50	12.72
			Grade 9-12			
Math	103.88	14.71	92.61	14.06	95.71	12.03
Reading	102.29	14.36	93.26	15.93	94.90	14.89
Written Language	102.11	14.17	93.46	14.13	96.90	13.28

Note: Samples Sizes Grades 1-4: Caucasians (455), Blacks (119), Hispanics (150);
Grades 5-8: Caucasians (487), Blacks (119), Hispanics (137);
Grades 9-12: Caucasians (371), Blacks (74), Hispanics (89).

Correlations. Table 26 presents correlations of the five KABC-II cognitive ability factors and global FCI with the three KTEA-II achievement outcome variables for the three ethnic groups at grades 1-4, 5-8, and 9-12. Correlations between the KABC-II ability factors and the KTEA-II achievement outcome variables ranged between r=.40 and r=.80 for all ethnicities across all grade groups. The KABC-II ability factors that produced coefficients less than r=.40 on occasion were Sequential/Gsm and Simultaneous/Gv (and Planning/Gf and Learning/Glr once each). Among the five ability factors, Knowledge/Gc produced the strongest ability-achievement relationships, correlating .r=50 to r=.80 with Math, Reading, and Written Language. Correlations between the FCI and the three KTEA-II achievement composites produced correlations ranging from r=.65 to r=.75, indicating that general ability accounted for about half the achievement variance for the present samples.

Table 26
Correlation between KABC-II Predictors and the three KTEA-II Achievement Composites across Age and Ethnicity

Predictor	Math			Reading			Written Language		
	Caucasians	Blacks	Hispanics	Caucasians	Blacks	Hispanics	Caucasians	Blacks	Hispanics
Grades 1-4									
Sequential/Gsm	.491	.335	.533	.479	.347	.576	.432	.485	.540
Simultaneous/Gv	.552	.477	.572	.440	.497	.500	.424	.455	.485
Planning/Gf	.598	.640	.671	.557	.493	.580	.529	.472	.548
Learning/Glr	.455	.512	.599	.539	.599	.615	.543	.585	.608
Knowledge/Gc	.564	.617	.556	.580	.608	.601	.532	.473	.537
Fluid-Crystallized Index	.698	.687	.733	.696	.683	.738	.657	.670	.690
Grades 5-8									
Sequential/Gsm	.386	.471	.401	.433	.383	.470	.357	.415	.508
Simultaneous/Gv	.544	.472	.499	.477	.461	.423	.335	.359	.445
Planning/Gf	.497	.682	.559	.488	.570	.641	.395	.521	.571
Learning/Glr	.385	.509	.551	.489	.585	.581	.429	.606	.551
Knowledge/Gc	.560	.642	.565	.701	.728	.771	.509	.568	.673
Fluid-Crystallized Index	.642	.740	.685	.711	.735	.768	.551	.665	.727
Grades 9-12									
Sequential/Gsm	.401	.388	.440	.470	.512	.639	.378	.397	.599
Simultaneous/Gv	.488	.499	.416	.427	.474	.267	.317	.422	.296
Planning/Gf	.572	.642	.476	.516	.588	.566	.480	.599	.429
Learning/Glr	.489	.567	.416	.494	.581	.512	.461	.548	.494
Knowledge/Gc	.591	.693	.559	.717	.815	.703	.602	.713	.556
Fluid-Crystallized Index	.696	.743	.621	.716	.790	.745	.613	.704	.648

Note: Samples Sizes Grades 1-4: Caucasians (455), Blacks (119), Hispanics (150); Grades 5-8: Caucasians (487), Blacks (119), Hispanics (137); Grades 9-12: Caucasians (371), Blacks (74), Hispanics (89).

Prediction Invariance. This section examines the differential predictive validity of the five KABC-II cognitive scales (and FCI) for the three ethnic groups for the subsamples--grades 1-4, 5-8, and 9-12. The approach to interpretation was first (a) to explain the evaluation of model fit used for the present analyses; then (b) to examine slope bias of the five ability factors, including the degree to which the ability-achievement relationships were similar or different across ethnic groups; and then (c) to explore intercept bias of the five ability factors. Subsequently, slope and intercept bias for FCI was explored.

Evaluation of model fit. Equality constraints across the groups were applied to the parameters in sequential fashion--1. Restriction of the residuals, 2. Restriction of the Slope, and 3. Restriction of the Intercept. Homogeneity of the residuals was not an absolutely necessary prerequisite (e.g., Reynolds & Keith, 2012). If this assumption was *not* met, the constraint was simply released. Residual invariance, slope invariance, and intercept invariance were each evaluated with model chi square ($\chi 2$), root-mean square error of approximation (RMSEA), and comparative fit index (CFI). Given that these models were considerably less complex than the model used for question #1 (factorial invariance), $\chi 2$ values were weighted more than CFI and RMSEA for these analyses.

Slope bias. alpha levels of .01 and .001 were used to report significant findings in an attempt to control for the chance findings that are known to occur when many statistical comparisons are made simultaneously. In these analyses the slopes were constrained to be equal across groups (*weak prediction invariance*). Three ethnic group comparisons were conducted across the three grade groups. A total of 135 comparisons were completed. Virtually no evidence of slope bias was found.

Table 27 presents the one significant result from the slope invariance analyses. Of the 45 comparisons between Caucasians and Blacks, only 1 produced significant slope bias ($p < .01$): KABC-II Sequential/Gsm was biased against Blacks at grades 1-4 when predicting the KTEA-II math achievement score. This significant slope bias is evident in Table 27 by the significant $\chi 2$ value ($\chi 2$ (2)=6.898, p<.01) and the relatively large RMSEA value of .09. Furthermore, as demonstrated by Table 26, Sequential/Gsm correlated .49 with Math for Caucasians versus .34 for Blacks. None of the 45 comparisons in the Caucasian-Hispanic analyses yielded significant slope bias, nor did any of the 45 comparisons in the Hispanic-Black analyses. Thus, over all three sets of analyses, 134 out of 135 comparisons (99.3%) resulted in no significant slope bias.

The lack of slope bias is easiest understood by examining the magnitude of the correlation coefficients between ability and achievement across the ethnic groups (Table 26). The correlations table shows that the coefficients between intelligence factors and achievement composites are substantial for all three ethnicities across all grade groups. For example, Planning/Gf correlated r=.50 to r=.70 with math across all three ethnicities across grades 1-12. Indeed, all correlation coefficients between the five KABC-II ability factors and the three KTEA-II achievement outcome composites were moderate to high for all three ethnic groups. Finding only one significantly different slope at $p < .01$ (and none at $p < .001$) suggests that the result is probably due to chance. Indeed, when 135 comparisons are made, one would expect one or two to be statistically significant at $p < .01$ simply due to chance alone.

Table 27

Significant Slope Fit Indexes and Nested Comparisons for Confirmatory Factor Analysis (CFA) Models for Caucasians and Black across the Grade Groups

	χ2	*Df*	Δχ2	Δ*df*	*P*	CFI	ΔCFI	RMSEA	Biased Against
				Grades 1-4					
Predictor Variable: Sequential/Gsm									
Math	6.898	2	6.730	1	.009*	.964	.036	.091	Blacks
				Grades 5-8					
				No bias					
				Grades 9-12					
				No bias					

Note: Samples Sizes Grades 1-4: Caucasians (455) and Blacks (119); Grades 5-8: Caucasians (487) and Blacks (119); Grades 9-12: Caucasians (371) and Blacks (74).
Only 1/45 slope analyses were significant, which means that 97.8% of the slopes did not produce a significant bias. In the Caucasian-Hispanic and Hispanic-Black analyses, none of the slopes reached significance at p < .01.
** p < .01*

Intercept differences. Tables 28, 29, and 30 present the significant results from the intercept invariance analyses. If slopes were *not* statistically significantly different from each other, the intercepts were constrained to be equal across groups (differences in intercepts, with slope invariance, suggests *strong prediction invariance*). Again using $p < .01$ and $p < .001$ to protect against multiple comparisons, results indicated that intercept differences were present such that a common regression line would over-predict performance on particular aspects of achievement for Blacks and Hispanics and under-predict performance for Caucasians.

Table 28 shows the Caucasian-Black comparisons. Altogether, 44 comparisons were completed (the significant slope bias for Sequential/Gsm predicting math at grades 1-4 eliminated the need to examine the intercept in that instance). Using $p < .01$, a majority of the comparisons (24 of 44 or 55%) produced significant intercept differences between Blacks and Caucasians; even using the .001 level yielded numerous significant intercept differences (20 of 44 or 36%). In every instance, the common regression would over-predict achievement for Blacks and under-predict achievement for Caucasians. All scales produced intercept bias for at least one grade group on at least one outcome variable. For example, Sequential/Gsm over-predicted Blacks' Reading and Written Language at grades 1-4 and over-predicted their achievement in all three academic areas at grades 5-12. Simultaneous/Gv, Planning/Gf, and Learning/Glr tended to over-predict Blacks' achievement in nearly all academic areas in grades 5-12 (but not 1-4). The only KABC-II variable that did not produce strong evidence for intercept bias was Knowledge/Gc. In other words, Knowledge/Gc was the most accurate of the five KABC-II scales at predicting Blacks' achievement in Math, Reading, and Written Language.

Table 28
Significant Intercept Fit Indexes and Nested Comparisons for Confirmatory Factor Analysis (CFA) Models for Caucasians and Blacks across the Grade Groups

	χ2	*Df*	Δχ2	Δ*df*	*P*	CFI	ΔCFI	RMSEA
			Grades 1-4					
Predictor Variable: Sequential/Gsm								
Reading	16.449	3	10.219	1	.001**	.898	.070	.126
Written Language	10.011	3	9.937	1	.007*	.943	.057	.091
			Grades 5-8					
Predictor Variable: Sequential/Gsm								
Math	29.392	3	27.846	1	.001**	.750	.025	.171
Reading	24.144	3	21.509	1	.001**	.819	.176	.153
Written Language	26.212	3	25.887	1	.001**	.738	.262	.159
Predictor Variable: Simultaneous/Gv								
Math	14.693	3	14.684	1	.001**	.940	.006	.113
Reading	13.793	3	11.249	1	.001**	.929	.067	.109
Written Language	19.012	3	18.184	1	.001**	.781	.219	.133
Predictor Variable: Planning/Gf								
Math	23.203	2	14.67	1	.001**	.899	.065	.187
Reading	13.436	3	8.862	1	.003*	.943	.043	.107
Written Language	18.654	3	14.182	1	.001**	.871	.109	.132
Predictor Variable: Learning/Glr								
Math	35.953	3	33.740	1	.001**	.704	.285	.191
Reading	28.789	3	26.770	1	.001**	.860	.140	.168
Written Language	36.105	3	30.778	1	.001**	.784	.194	.191
			Grades 9-12					
Predictor Variable: Sequential/Gsm								
Math	29.384	3	28.398	1	.001**	.647	.353	.199
Reading	16.708	3	16.053	1	.001**	.878	.122	.144
Written Language	16.416	3	16.202	1	.001**	.801	.199	.141
Predictor Variable: Simultaneous/Gv								
Math	8.441	3	8.051	1	.005*	.955	.045	.091

Predictor Variable: Planning/Gf								
Math	15.966	3	14.247	1	.001**	.929	.071	.140
Written Language	8.948	3	6.765	1	.009*	.953	.046	.095
Predictor Variable: Learning/Glr								
Math	24.731	3	21.440	1	.001**	.829	.161	.181
Reading	16.011	3	10.733	1	.001**	.901	.074	.140
Written Language	13.942	3	10.819	1	.001**	.903	.087	.129
Predictor Variable: Knowledge/Gc								
Math	13.343	3	10.30	1	.001**	.950	.045	.124

Note: Samples Sizes Grades 1-4: Caucasians (455) and Blacks (119); Grades 5-8: Caucasians (487) and Blacks (119); Grades 9-12: Caucasians (371) and Blacks (74).
** p < .01*
*** p < .001*

Table 29 presents the significant Caucasian-Hispanic comparisons, and the results mirror the results of the Caucasian-Black analyses. About half of the comparisons were significant at both $p < .01$ $p < .001$ and every one produced over-prediction for the ethnic minority group (in this case Hispanics). Intercept bias was more prevalent at grades 5-12 than 1-4 and Knowledge/Gc showed the least intercept bias among the five KABC-II scales.

Table 30 represents the Black-Hispanic comparisons across the grade groups. Out of the 45 comparisons only 2 resulted in significant intercept bias at the .01 alpha level. Simultaneous/Gv under-predicted reading achievement for Blacks at grades 1-4 and Sequential/Gsm under-predicted Written Language for Hispanics at grades 9-12. None of the comparisons were significant *at* $p < .001$. Such results indicate that a common regression line can be used for both Hispanics and Blacks when predicting achievement, as no strong evidence for intercept bias was found.

Table 29
Significant Intercept Fit Indexes and Nested Comparisons for Confirmatory Factor Analysis (CFA) Models for Caucasians and Hispanics across the Grade Groups

	χ2	*Df*	Δχ2	Δ*df*	*P*	CFI	ΔCFI	RMSEA
			Grades 1-4					
Predictor Variable: Simultaneous/Gv								
Reading	13.391	3	12.10	1	.001**	.925	.075	.107
Predictor Variable: Planning/Gf								
Reading	13.357	3	12.99	1	.001**	.955	.045	.107
Predictor Variable: Learning/Glr								
Reading	9.761	3	7.091		.008*	.970	.027	.086
			Grades 5-8					
Predictor Variable: Sequential/Gsm								
Math	15.958	3	15.237	1	.001**	.870	.130	.117
Reading	20.801	3	18.388	1	.001**	.865	.132	.139
Written Language	31.042	2	27.617	1	.001**	.729	.248	.216
Predictor Variable: Simultaneous/Gv								
Math	33.203	3	32.989	1	.001**	.853	.147	.179
Reading	40.466	3	40.139	1	.001**	.753	.247	.201
Written Language	51.678	3	44.175	1	.001**	.441	.496	.229
Predictor Variable: Planning/Gf								
Math	22.239	3	20.399	1	.001**	.897	.103	.144
Reading	33.903	2	31.015	1	.001**	.847	.144	.226
Written Language	41.622	2	38.801	1	.001**	.712	.275	.252
Predictor Variable: Learning/Glr								
Math	25.560	3	19.391	1	.001**	.820	.147	.156
Reading	26.900	3	22.955	1	.001**	.875	.115	.159
Written Language	33.721	2	33.270	1	.001**	.786	.214	.226
			Grades 9-12					
Predictor Variable: Sequential/Gsm								
Math	12.775	2	10.548	1	.001**	.868	.117	.153
Predictor Variable: Simultaneous/Gv								
Math	17.801	3	12.587	1	.001**	.872	.100	.147

Reading	12.638	3	9.792	1	.002*	.878	.111	.119
Predictor Variable: Planning/Gf								
Math	18.620	3	12.921	1	.001**	.907	.071	.151
Reading	9.743	3	9.071	1	.003*	.954	.046	.099
Predictor Variable: Learning/Glr								
Math	16.727	3	11.082	1	.001**	.881	.087	.141
Reading	7.393	3	7.171	1	.007*	.966	.034	.081

Note: Samples Sizes Grades 1-4: Caucasians (455) and Hispanics (150); Grades 5-8: Caucasians (487) and Hispanics (137); Grades 9-12: Caucasians (371) and Hispanics (89).
** p < .01*
*** p < .001*

Table 30
Significant Intercept Fit Indexes and Nested Comparisons for Confirmatory Factor Analysis (CFA) Models for Blacks and Hispanics across the Grade Groups:

	χ2	*Df*	Δχ2	Δ*df*	*P*	CFI	ΔCFI	RMSEA	Underprediction
				Grades 1-4					
Predictor Variable: Simultaneous/Gv									
Reading	8.661	3	8.066	1	.005*	.924	.076	.119	Blacks
				Grades 5-8					
No significant intercept fit indexes									
				Grades 9-12					
Predictor Variable: Sequential/Gsm									
Written Language	11.881	3	7.726	1	.005*	.821	.0136	.192	Hispanics

Note: Samples Sizes Grades 1-4: Blacks (119) and Hispanics (150); Grades 5-8: Blacks (119) and Hispanics (137); Grades 9-12: Caucasians (371) and Hispanics (89).
** p < .01*
*** p < .001*

Summary. As demonstrated in summary Table 31, overall, for all grade levels and for all ethnic groups, there was no evidence for slope bias. The magnitudes of the path from ability factors to achievement factors were the same across all three ethnic groups (ranging from moderate to high in terms of effect size). The finding means that an individual's ethnic background does not interact with the effect of cognitive abilities on predicting achievement outcomes when the coefficient of correlation (i.e., slope) is the focus of the analyses. However, that conclusion is not supported in the analyses of intercepts.

The results of the Caucasian-Black and Caucasian-Hispanic analyses did show substantial evidence for intercept differences between ethnic minority groups and Caucasians (Table 31). The bias was such that a common regression line consistently over-predicted achievement for Blacks and Hispanics and under-predicted achievement for Caucasians. A common regression line would, therefore, favor the selection of minority group individuals in, for example, the selection of students for giftedness programs in school. Such findings are contrary to the common belief that cognitive test scores underestimate minority groups' performance in school or college selection.

Whereas the under-prediction for Caucasians is of small effect size (about 1 standard-score point, usually <.10 *SD*), the amount of over-prediction is moderate to large (2-7 points, typically > .25 *SD*) for Blacks and Hispanics (see Table 32, which summarizes the amount of over-prediction for all significant intercepts). Overall, Sequential/Gsm and Learning/Glr produced the strongest evidence of over-prediction for Blacks and Simultaneous/Gv, Planning/Gf, (and at times Learning/Glr) produced the strongest evidence for over-prediction for Hispanics across the grade groups. The most consistent and largest over-predictions were generally found for grade groups 5-8 for both Hispanics and Blacks.

Global score analysis. As most of the KABC-II separate ability factors consistently demonstrated intercept bias, especially at grades 5-12, it was sensible to explore bias on a more global level, to see if the over-prediction also characterized global intelligence. Knowledge/Gc emerged as the least biased ability factor among the five abilities. Whereas the Fluid-Crystallized Index (FCI) includes all five ability scales, the Mental Processing Index includes only four (excluding Knowledge/Gc). Even though the MPI is recommended as the global index of choice for ethnic minority children (Kaufman & Kaufman, 2004a), the unbiased nature of Knowledge/Gc made FCI a better choice for this dissertation.

As demonstrated in Table 26, the magnitude of the correlation coefficients between FCI and achievement were high, ranging from r=.70 to r=.80 for all three ethnicities across grade groups. Accordingly, no slope bias was present for FCI for any of the comparisons. Furthermore, there was virtually no evidence of intercept bias (as demonstrated in Table 32). Only once FCI over-predicted Hispanics' achievement for Written Language by two standard score points in the Caucasian-Hispanic comparison for grades 5-8. Such results are solid indications for the fact that FCI is an excellent predictor for achievement for Caucasians, Blacks, and Hispanics across grade groups 1-12.

Table 31

Specificity of Bias by Predictor and Achievement across Age: Caucasians and Blacks; Slope Bias and Underpredicted Achievement

	Math	**Reading**	**Written Language**
Predictor		*Grades 1 through 4*	
Sequential/Gsm	Slope Bias (Blacks)	Intercept: Caucasians (-0.9)	Intercept: Caucasians (- 0.9)
Simultaneous/Gv	No bias	No bias	No bias
Planning/Gf	No bias	No bias	No bias
Learning/Glr	No bias	No bias	No bias
Knowledge/Gc	No bias	No bias	No bias
Fluid-Crystallized Index (FCI)	No bias	No bias	No bias
Predictor		*Grades 5 through 8*	
Sequential/Gsm	Intercept: Caucasians (-1.5)	Intercept: Caucasians (-1.3)	Intercept: Caucasians (-1.4)
Simultaneous/Gv	Intercept: Caucasians (-1.0)	Intercept: Caucasians (-0.9)	Intercept: Caucasians (-1.2)
Planning/Gf	Slope Bias: Caucasians	Intercept: Caucasians (- 0.8)	Intercept: Caucasians (-0.9)
Learning/Glr	Intercept: Caucasians (-1.6)	Intercept: Caucasians (-1.3)	Intercept: Caucasians (-1.4)
Knowledge/Gc	No bias	No bias	No bias
Fluid-Crystallized Index (FCI)	No bias	No bias	No bias
Predictor		*Grades 9 through 12*	
Sequential/Gsm	Intercept: Caucasians (-1.5)	Intercept: Caucasians (-1.1)	Intercept: Caucasians (-1.2)
Simultaneous/Gv	Intercept: Caucasians (-0.7)	No bias	No bias
Planning/Gf	Intercept: Caucasians (-1.0)	No bias	Intercept: Caucasians (-0.7)
Learning/Glr	Intercept: Caucasians (-1.2)	Intercept: Caucasians (-0.9)	Intercept: Caucasians (-0.8)
Knowledge/Gc	Intercept: Caucasians (-0.7)	No bias	No bias
Fluid-Crystallized Index (FCI)	No bias	No bias	No bias

Note: Numerical values represent the underpredicted achievement for Caucasians and Blacks at the .01 alphas level.

Table 31 continued

Specificity of Bias by Predictor and Achievement across Age: Caucasians and Hispanics; Slope Bias and Underpredicted Achievement

	Math	**Reading**	**Written Language**
Predictor		*Grades 1 through 4*	
Sequential/Gsm	No bias	No bias	No bias
Simultaneous/Gv	No bias	Intercept: Caucasians (-1.1)	No bias
Planning/Gf	No bias	Intercept: Caucasians (-1.1)	No bias
Learning/Glr	No bias	Intercept: Caucasians (-0.8)	No bias
Knowledge/Gc	No bias	No bias	No bias
Fluid-Crystallized Index (FCI)	No bias	No bias	No bias
Predictor		*Grades 5 through 8*	
Sequential/Gsm	Intercept: Caucasians (-1.1)	Intercept: Caucasians (-1.2)	Intercept: Caucasians (-1.3)
Simultaneous/Gv	Intercept: Caucasians (-1.5)	Intercept: Caucasians (-1.8)	Intercept: Caucasians (-1.9)
Planning/Gf	Intercept: Caucasians (-0.9)	Intercept: Caucasians (-1.3)	Intercept: Caucasians (-1.6)
Learning/Glr	Intercept: Caucasians (-1.3)	Intercept: Caucasians (-1.3)	Intercept: Caucasians (-1.4)
Knowledge/Gc	No bias	No bias	No bias
Fluid-Crystallized Index (FCI)	No bias	No bias	Intercept: Caucasians (-1.0)
Predictor		*Grades 9 through 12*	
Sequential/Gsm	Intercept: Caucasians (-0.9)	No bias	No bias
Simultaneous /Gv	Intercept: Caucasians (-1.1)	Intercept: Caucasians (-1.0)	No bias
Planning/Gf	Intercept: Caucasians (-1.0)	Intercept: Caucasians (-0.9)	No bias
Learning/Glr	Intercept: Caucasians (-1.0)	Intercept: Caucasians (-0.8)	No bias
Knowledge/Gc	No bias	No bias	No bias
Fluid-Crystallized Index	No bias	No bias	No bias

Note: Numerical values represent the underpredicted achievement for Caucasians and Hispanics at the .01 alphas level.

Table 31 continued

Specificity of Bias by Predictor and Achievement across Age: Blacks and Hispanics; Slope Bias and Underpredicted Achievement

	Math	**Reading**	**Written Language**
Predictor		*Grades 1 through 4*	
Sequential/Gsm	No bias	No bias	No bias
Simultaneous/Gv	No bias	Intercept: Blacks (-2.5)	No bias
Planning/Gf	No bias	No bias	No bias
Learning/Glr	No bias	No bias	No bias
Knowledge/Gc	No bias	No bias	No bias
Fluid-Crystallized Index (FCI)	No bias	No bias	No bias
Predictor		*Grades 5 through 8*	
Sequential/Gsm	No bias	No bias	No bias
Simultaneous/Gv	No bias	No bias	No bias
Planning/Gf	No bias	No bias	No bias
Learning/Glr	No bias	No bias	No bias
Knowledge/Gc	No bias	No bias	No bias
Fluid-Crystallized Index (FCI)	No bias	No bias	No bais
Predictor		*Grades 9 through 12*	
Sequential/Gsm	No bias	No bias	Intercept: Hispanics (-0.4)
Simultaneous /Gv	No bias	No bias	No bias
Planning/Gf	No bias	No bias	No bias
Learning/Glr	No bias	No bias	No bias
Knowledge /Gc	No bias	No bias	No bias
Fluid-Crystallized Index (FCI)	No bias	No bias	No bias

Note: Numerical values represent the underpredicted achievement for Blacks and Hispanics at the .01 alphas level.

Table 32
Significant Intercept Over-predictions for Blacks and Hispanics as compared to Caucasians across Age Groups

Predictor	Math		Reading		Written Language	
	Blacks	Hispanics	Blacks	Hispanics	Blacks	Hispanics
Grades 1-4						
Sequential/Gsm			+3.3		+3.4	
Simultaneous/Gv				+3.3		
Planning/Gf				+3.1		
Learning/Glr				+2.3		
Knowledge/Gc						
Fluid-Crystallized Index						
Grades 5-8						
Sequential/Gsm	+5.6	+3.7	+5.0	+3.9	+5.6	+4.7
Simultaneous/Gv	+3.8	+5.3	+3.5	+5.9	+4.7	+6.7
Planning/Gf		+4.1	+2.9	+4.6	+4.0	+5.3
Learning/Glr	+6.2	+4.1	+5.3	+4.3	+5.0	+4.9
Knowledge/Gc						
Fluid-Crystallized Index						+1.9
Grades 9-12						
Sequential/Gsm	+7.6	3.4	+5.4		+5.4	
Simultaneous/Gv	+3.7	4.1		+3.9		
Planning/Gf	+4.7	4.1		+3.5	+3.3	
Learning/Glr	+6.3	4.0	+4.3	+3.1	+4.3	
Knowledge/Gc	+3.9					
Fluid-Crystallized Index						

Note: Samples Sizes Grades 1-4: Caucasians (455), Blacks (119), Hispanics (150); Grades 5-8: Caucasians (487), Blacks (119), Hispanics (137); Grades 9-12: Caucasians (371), Blacks (74), Hispanics (89).

Discussion

Aims of the Study

The demographic profile of the United States is changing. The population is increasing in size and age and is becoming more ethnically diverse. The U.S. census bureau predicts that by 2050 more than half of the U.S. population will be composed of ethnic minorities. Given the changing demographic profile in the United States, the need for culturally and ethnically fair test instruments has become increasingly more important. Specifically, the exploration of test bias of cognitive and achievement tests is important because of the disproportionate representation of ethnic minority group children diagnosed with a learning disability and the significant underrepresentation of those children in giftedness classes. Indeed, the implications of scores on cognitive test instruments are vast, as scores often determine access or denial to special programs and services.

The most common way of determining test bias is through the comparison of mean score differences. However, the detection of mean score differences between two groups (e.g., a minority group and a Caucasian majority group) is not a statistically sound way of determining bias; clinicians can easily draw erroneous conclusions from such results (e.g., Reynolds & Lowe, 2009). This is because, sometimes there can exist true differences between groups. For example, it is easier for males to lift a 50-pound weight than it is for females, but the differences in performance do not indicate that the weight tool is biased, they just demonstrate the [true differences in strength between males and females] he said it sounds like my study determined TRUE differences, but he said that he doesn't think my study did this. But Connie wanted me to give such an example – what should we do?. Instead of defining ethnic bias as the presence of mean score differences, more sophisticated and appropriate methodology was used in this

dissertation to determine test bias of two individually administered, well-known tests of intelligence and achievement. In this study, ethnic group bias of the KABC-II and KTEA-II was examined for a representative sample of Caucasian, Hispanic, and Black children and adolescents in grades 1 through 12. More specifically, construct and predictive invariance of the Kaufman tests were addressed by asking three research questions: (1) Is the factor structure of the Kaufman tests construct invariant across all three ethnic groups? (2) Do the three ethnic groups significantly differ in their mean scores? (3) Is there predictive validity bias across the different ethnic groups?

In order to answer those research questions, CFA was used to explore construct invariance of the test structure and SEM was used to measure predictive invariance. The methodology applies increasingly restrictive sets of equality constraints in order to incrementally test whether the different levels of equality were met across the groups (Meredith, 1993).

Major Findings of the Study

The most important findings of this study are as follows:

Construct and prediction invariance

- Using CFA and SEM, factorial invariance and predictive invariance of the KABC-II and KTEA-II test structures were established for Caucasian, Hispanic, and Black school-aged children.

Construct invariance

- To the author's knowledge, this is the first investigation to establish differential construct validity of a comprehensive achievement battery alongside a comprehensive measure of cognitive ability.

- Construct invariance of seven CHC factors was established as a prerequisite for comparing mean score differences on the separate ability factors (the appropriate research methodology that is not usually conducted in "mean difference" studies).

Mean score differences

- Mean score differences were explored on measures of academic achievement across ethnic groups. Results indicated that mean score differences found on the KTEA-II were comparable to the differences found on the KABC-II. This is an important contribution to the literature on academic achievement because of the lack of previous studies that have reported ethnic differences in academic achievement on well-standardized, individually-administered, clinical tests of achievement.

Differential prediction invariance

- The differential prediction results demonstrate consistent findings that measures of cognitive ability *overpredict* the reading, math, and written expression of Hispanic and Black students across all grade levels from elementary school through high school, especially at grades 5-8.
- The one cognitive ability that did *not* show signs of overprediction of achievement in reading, writing, and math for Blacks and Hispanics was crystallized intelligence (*Gc*); that is, the variable that is traditionally considered the most culturally loaded showed the least bias.
- Despite evidence that CHC broad ability factors are good predictors of achievement, results of this study suggest that FCI (*g*) is the "fairest" (least biased) predictor of achievement for Caucasian, Hispanic, and Black school-aged children.

Elaboration of Major Findings

Overall, the results of CFA and SEM analyses established both the factorial and predictive invariance of the Kaufman achievement and intelligence tests. This is because there was no slope bias detected and the only intercept bias that was found was in favor of minority group children (i.e., bias was against Caucasians, whose achievement was underpredicted). Test bias against the minority groups would only have been concluded if either the slope or the intercept had *underpredicted* minority groups' achievement. This sophisticated approach to test bias has rarely been done in the past for standardized tests of achievement and intelligence. When researchers did follow the appropriate methodology, they established either construct or predictive validity, but not both. For example, Keith et al. (1999) and Trundt (2013) used CFA to demonstrate the differential construct validity of two versions of the DAS for Caucasians, Hispanics, and Blacks (Trundt, 2013, also included Asians). Additionally, Kush et al. (2001) provided evidence for the invariance of the four-factor structure of the WISC-III in a sample of Black and Caucasian students, using exploratory and confirmatory factor analysis (although there were some inconsistencies observed for the Freedom from Distractibility and Processing Speed factors for the Black sample). And Keith (1999) used SEM to demonstrate the differential predictive validity of the WJ-R when predicting the achievement outcomes for Caucasians, Hispanics, and Blacks. In sum, although a few researchers have used state-of-the-art methodology to investigate construct and predictive validity of other intelligence tests, no one has investigated both types of bias for the same samples.

Factorial invariance. This study established factorial invariance of comprehensive intelligence and achievement tests based on the theoretical CHC model of human intelligence. Five of the seven CHC factors were representative of cognitive abilities and two were associated

with achievement (*G*rw and *G*q). This is one of the few studies that established construct invariance of CHC broad cognitive abilities, by ethnicity, and the *only* study that established construct invariance of achievement factors. Keith et al. (1999) established factorial invariance of *G*f, *G*v, and *G*c on the original version of the DAS and Trundt (2013) established factorial invariance of *G*c, *G*f, *G*v, *G*lr, *G*s, and partial invariance for *G*sm for the DAS-II. Additionally, Rush et al. (2001) established invariance for the four-factor structure of the WISC-III in a non-representative sample of Black and Caucasian students. The factors explored by those authors were cognitive-based CHC factors. Even though contemporary CHC theory includes achievement factors as part of their broad ability spectrum, construct validity across ethnic groups on measures of academic achievement has been under-studied. Thus, CHC theory recognizes that achievement ability constitutes an important part of human intelligence and findings of this study provide evidence that Gq and Grw - two broad achievement factors - are invariant across ethnic groups.

Additionally, this is the first study to establish factorial invariance of CHC factors *before* comparing mean score differences across ethnic groups. *If* researchers followed the procedure for the examination of ethnic differences, they only established invariance for the *g* factor (e.g., Edwards & Oakland, 2006), but not for CHC broad ability factors. Thus, this is the first study to follow the correct statistical procedure and first established invariance by ethnicity for seven CHC broad factors and then examined mean score differences on those separate abilities. Note, however, that the correct methodological procedure has been followed in several investigations of *gender* differences on the KABC-II (Reynolds, Keith, Ridley, & Patel, 2007), the DAS-II (Keith, Reynolds, Roberts, Winter, & Austin, 2011), WJ III (Keith, Reynolds, Patel, & Ridley,

2008), and the joint structure of the KABC-II and KTEA-II (Reynolds, Scheiber, Hajovsky, & Kaufman, 2015).

Mean score differences. Generally speaking, results from the mean score differences analysis suggest that both the KABC-II and the KTEA-II produce ethnic group differences in the small to moderate range in terms of their effect size (.15-.61 *SD*). Even though the Black-Caucasian and Hispanic-Caucasian difference profiles are consistent with results from other popular tests of cognition, such as the Wechsler or Woodcock-Johnson tests, the magnitude of the differences are smaller than ethnic group differences found on other intelligence tests. Tests such as the Wechsler and Woodcock-Johnson usually produce large differences ranging from .66 *SD* (controlled for SES) to 1 *SD* (not controlled for SES) (Edwards & Oakland, 2006; Prifitera et al., 2005; Tulsky et al., 2003; Weiss et al., 2006).

Consistent with previous studies, this study found small Black-Caucasian differences on Sequential Processing/*G*sm and Learning /*G*lr (.22 and .10 *SD*); Knowledge/*G*c, Simultaneous/*G*v, and Planning/*G*f produced moderate differences (.44 –.61 *SD*). In the Hispanic-Caucasian comparisons, the largest difference was found on Knowledge/*G*c (.52 *SD*). Differences on Simultaneous/*G*v and Planning/*G*f were small (.13 and .14 *SD*) and differences on Sequential Processing/*G*sm and Learning/*G*lr were moderate (.35 and .27 *SD*). Differences were slightly larger (around .66 *SD*) when not controlled for SES. This profile is consistent with previous findings on the Woodcock Johnson and Wechsler scales in terms of relative cognitive strengths and weaknesses of each ethnic group (e.g., Edwards & Oakland, 2006; Prifitera et al., 2005; Tulsky et al., 2003; Weiss et al., 2006); however, the KABC-II generates smaller ethnic differences in terms of magnitude than other tests of cognition.

Whereas there have been a plethora of studies that compared ethnic mean score differences on various tests of cognition, including the K-ABC (Kaufman & Kaufman, 1983) and the KABC-II (Kaufman & Kaufman, 2004a), not many studies have compared mean score differences on individually-administered tests of achievement. With the exception of the original K-ABC (Kaufman & Kaufman, 1983), which includes the achievement domains of reading decoding, reading understanding, and arithmetic, and two studies conducted by Naglieri and his colleagues (Naglieri, Rojan, & Matto, 2007; Naglieri, Rojahn, Matto, & Aquilino, 2005), there has been a dearth of research comparing mean score differences, by ethnicity, on individually-administered tests of achievement.

Results of this present study found that Black-Caucasian differences on the KTEA-II are in the moderate range, producing differences of .45-.60 *SD*s for reading, writing, and math, not controlled for SES. Such results are consistent with findings on the K-ABC achievement subtests of arithmetic, reading decoding and reading understanding (Kaufman & Kaufman, 1983); however, the differences are smaller than differences reported on the WJ R (Naglieri et al. 2005), which ranged from .66-1.0 *SD*. Group-administered tests, such as the Stanford Achievement Test (SAT; Naglieri & Ronning, 2000) also produced larger differences of about 1.0 *SD*. No previous study was found that explored ethnic mean score differences on individual or group administered tests of achievement that controlled SES. Present finding demonstrate that the Black-Caucasian achievement gap decreases to .40 *SD* when controlling for SES, reinforcing the necessity of controlling for SES when comparing achievement mean score differences across different ethnic groups.

Results from the Caucasian-Hispanic analysis showed moderate differences ranging from .46-.53 *SD* for reading, writing and math, not controlled for SES. Results are consistent with

findings on the K-ABC achievement subtests (Kaufman & Kaufman, 1983) and with results on the WJ-R (Naglieri et al., 2007), both of which produced differences ranging from .26-.60 *SD*, not controlled for SES. Results from group-administered tests such as the SAT produced slightly larger differences ranging from .46 -.67 *SD*. Again, no studies that controlled for SES could be found. The present study showed that the Caucasian-Hispanic achievement gap reduced from moderate in terms of its magnitude to small-moderate differences (.20-.44 *SD*) when controlling for SES. Such results again stress the importance of taking SES into consideration when comparing ethnic groups on achievement.

In sum, results of this study demonstrated that ethnic achievement mean score differences on the KTEA-II are as large in terms of the magnitude as the differences found on the KABC-II cognitive test, and smaller than differences found on other popular tests of cognition and achievement, especially when it comes to Black-Caucasian comparisons. Results of this present study contribute to a sparse literature on ethnic mean score differences in achievement. Those studies that were conducted are about 10-30 years old, and none have taken the necessary step of controlling for SES before comparing ethnic mean scores on achievement domains. The reduction in mean score differences between the Caucasian group and the two ethnic minority samples in this study highlights the influence that SES has on both cognitive and achievement tests.

In this study, SES was defined as mother's educational attainment. However, many other factors that contribute to SES, such as disparities in income, in availability of resources, in schooling system, and in nutrition, have been found to also have a profound impact on cognitive and achievement scores (Weiss et al., 2006). It is not feasible to control for all of these factors. Thus, even though this study reported ethnic mean score differences on both the KABC-II and

the KTEA-II, the differences cannot be interpreted as being the result of ethnic group membership. On the contrary, because of the vast influence of environmental factors that cannot all be controlled, ethnic mean score differences should generally not be interpreted as evidence for or against test bias.

Predictive invariance. This is one of the few studies to investigate predictive invariance of individually-administered tests of both achievement and cognition across three ethnic and grade groups, using SEM. Results provided evidence for a lack of slope bias. This finding is consistent with Keith (1999), who found that the magnitude of the path coefficients between the WJ-R ability factors and the achievement domains of reading and math were, in general, equally strong for Black, Hispanic, and Caucasian school-aged children. Other studies found that IQ predicted achievement about equally well for Caucasians and Blacks (and Hispanics if included in the analysis) in terms of (a) the magnitude of the coefficients of correlation (Edwards & Oakland, 2006; Keith, 1999; Naglieri et al., 2005; Naglieri et al., 2000), and (b) the tests' slopes (Weiss et al., 1993; Weiss & Prifitera, 1995). Present results are consistent with those findings. To the author's knowledge, Mattern and Patterson (2013) were the only researchers who detected small slope differences in their large sample of 475,000 students after adjusting for statistical artifacts that they believed could have prevented other studies from detecting slope bias. The researchers found very minimal slope bias; the Black and Hispanic regression lines were consistently below the Caucasian regression line so that the performance of Black and Hispanic students was consistently overpredicted.

Although results of this study did not find any evidence for slope bias, pervasive and persistent evidence for intercept bias was found, such that cognitive scores consistently

overpredicted Black and Hispanic academic achievement, especially at grades 5-8. Jensen (1980, p. 513) provides a detailed explanation of how intercept overprediction emerges:

> "… overprediction of the minor group's performance from the major group's regression equation (or the common regression equation) is a consequence of a positive difference in intercepts, that is, the major group's regression line being above the minor group's. This positive difference in intercepts, or levels of the regression lines, of the two groups comes about when (1) the group mean difference on the criterion is greater than zero and in the same direction as the group mean difference on the predictor and also (2) the correlation (i.e., validity coefficient) between predictor and criterion multiplied by the group mean difference on the predictor in standard score units is less than the groups' mean difference on the criterion in standard score units [...] what it means, in so many words, is that the predictor variable does not account for enough of the variance in the criterion variable to account for the major-minor groups' mean difference on the criterion."

The presence of intercept overprediction is consistent with results from previous studies (primarily employment studies), which found that the criterion outcome for Blacks (and Hispanics if they were included in the analysis) was overpredicted (e.g., Jensen, 1980; Kuncel & Sackett, 2007; Rotundo & Sackett, 1999; Rushton & Jensen, 2005; Sackett, Schmitt, Ellingson, & Kablin, 2001). While such results do not indicate bias against any of the minority groups (Hispanics or Blacks), results do suggest that four of the five KABC-II CHC factors did not accurately predict the achievement outcomes of Hispanic and Black school-aged children. However, Knowledge/Gc and the most global score, FCI, are good and accurate predictors of Black and Hispanic students' achievement in reading, writing, and math.

Indeed, remarkable was the fact that the KABC-II's measure of crystallized intelligence (Knowledge/*G*c) did not show any evidence of slope bias and essentially no evidence of intercept bias. In other words, Knowledge/*G*c was the one broad ability that showed the least bias and

was, therefore, the most accurate at predicting achievement for all three ethnic groups across all three grade levels. Indeed, coefficients of correlation ranged from *r*=.60 to *r*=.80 between Knowledge/*G*c and the three achievement outcome criteria. It is not surprising that Knowledge/Gc was such a good predictor of reading and writing because Gc assesses language abilities that are similar the language-related abilities of reading and writing. However, it is noteworthy that Knowledge/Gc was also the most accurate predictor for Hispanic and Black students' math achievement. Such findings, opposite to common beliefs, indicate that Knowledge/*G*c -- the variable that is supposedly most influenced by culture and language-- is the best predictor of achievement among the broad factors not only for Caucasians, but also for Black and Hispanic students.

Finally, despite evidence that CHC variables are good predictors of achievement, results of this study suggest that FCI (*g*), like *Gc*, is the fairest predictor of achievement. Again, contrary to common beliefs, results demonstrated that a more global score, such as FCI, might be a better predictor when estimating achievement for minority group students than the separate abilities -- including ability factors that are theoretically culturally and linguistically neutral, such as Sequential/*G*sm and Simultaneous/*G*v. Such findings support Gary Canivez's (2013) theory, which postulates that interpretation of an individual's cognitive level should be based "primarily (if not exclusively)" on the most global score on the Wechsler scale, the Full Scale IQ (*g*) (p. 93). He supports his argument psychometrically, stressing the fact that global IQs have been found to have the strongest internal consistency, short and long-term temporal stability, and predictive validity coefficients. He also highlights that the more global scores have less error variance and, thus, are more likely to reflect true scores, and account for the largest portion of the variance in cognitive tests. He therefore believes that more global scores are more valid and reliable as

compared to broad factors when it comes to the interpretation of an individual's cognitive capacity. Results of this study support his argument--namely that the broad ability factors, as opposed to FCI, demonstrated pervasive patterns of achievement overprediction for Hispanics and Blacks. In that sense, if the goal is to accurately predict achievement of Hispanic and Black school-aged children using the KABC-II, it is recommended to use the comprehensive FCI as the primary predictor variable.

Possible Explanations for Overprediction

The persistent overprediction of the broad cognitive ability factors when estimating Black' and Hispanics' achievement in reading, writing, and math was unanticipated. Such results are surprising as they are contrary to the common belief that cognitive tests put ethnic minority children at a disadvantage. What follows are possible explanations for this overprediction in minority groups' achievement.

Is the KABC-II biased? One possible explanation for the overprediction of the achievement scores is that the KABC-II does not accurately measure cognitive ability. However, as outlined in the manual (Kaufman & Kaufman, 2004a), the KABC-II was carefully developed and rigorous procedures were employed to ensure adequate content and construct validity of the test. Evidence of content validity indicates that the KABC-II includes important facets of the construct – intelligence – and construct validity ensures that the intended construct is actually being measured by the test.

Content validity. Appropriate content validity of the KABC-II was warranted in several ways: as discussed previously, the KABC-II is based on two theoretical models – Luria's (1966, 1970, 1973) neuropsychological model and CHC theory (Carroll, 1997; Flanagan, 2000; Horn & Noll, 1997) in order to create a link between intelligence and neuropsychology. Regardless of

which theoretical underpinning is elected, Luria's as well as CHC-based interpretations both use the same subtests. In this present study, the CHC based interpretation was used. To ensure content validity, the KABC-II (when interpreted from a CHC-theory point of view) measures five different facets (constructs) of intelligence, which provide the examiner with a comprehensive description of a child's main cognitive strengths and weaknesses. The facets measured are in accordance with the Individuals with Disabilities Education Act (IDEA, 2004), as they can be used to assess all eight types of learning disabilities. The five constructs measure five CHC broad ability factors. Each subtest was carefully selected by consulting appropriate experimental, neuropsychological, and psychometric literature, by consulting experts in the field, and by reviewing the original K-ABC (Kaufman & Kaufman, 1983) and other popular tests of intelligence. The subtests were selected to measure certain CHC narrow abilities, which loaded on the corresponding broad ability factors (as evidenced by results from the CFA), in accordance with CHC theory. The content of the KABC-II was finalized through pilot studies and the tryout phase. For example, each new subtest was piloted with a small number of children to ensure adequacy of each test with the target population. The tryout phase included 696 children ages 3-18 who were tested by trained professionals in 50 testing sites around the country. Final subtest decisions were made based on results from several statistical procedures (e.g., EFA, CFA, item fit statistics, and item discrimination) and expert opinions (Kaufman & Kaufman, 2004a, chapters 6 & 7). Such rigorous procedures allowed for adequate content validity of the test.

Construct validity. Confirmatory Factor Analysis (CFA) was employed to provide evidence for the test's construct validity. This present dissertation as well as chapter 8 of the test's manual provide more detail regarding the procedures (Kaufman & Kaufman, 2004a). CFA

was used to evaluate the best groupings of subtests and scales and verify that these groupings supported the organization of subtests into the five designated scales. The final model provided excellent fit for the data (CFI = .997-.999; RMSEA = .025 - .055) and the subtests' loadings on *g* and the corresponding broad factors were consistent with CHC theory (Kaufman & Kaufman, 2004a; Flanagan et al., 2013). Furthermore, several decisions were made in order to ensure fairness among different ethnic groups. For example, the Rasch (1961) method was employed to ensure ethnic non-bias. That is, after initial tryout, the subtests were evaluated separately for each ethnic group. Indeed, results of this present study and the manual (Kaufman & Kaufman, 2004) show that the KABC-II produces smaller differences between ethnic groups as compared to other intelligence tests, such as the Wechsler scales. Additionally, findings of construct validity between Blacks, Hispanics, and Caucasians in this present study provide evidence that the KABC-II measures the same construct across all three groups, which provides further support of ethnic non-bias.

Another layer of construct validity refers to convergent validity. As outlined in detail in the KABC-II manual (Kaufman & Kaufman, 2004a, chapter 8), the correlations between the KABC-II and other popular tests of intelligence, specifically the WISC-IV and the WJ III Cognitive, provide evidence of strong convergent validity. That is to say, the KABC-II measures intelligence in the same way as those tests do. For example, correlations between the KABC-II FCI (as well as MPI and NVI) and the WISC-IV Full Scale IQ were high, producing correlations of .89 (and .88 and .82, respectively). Similarly, the KABC-II scales produced moderate to high correlations with the corresponding WISC-IV index measures, ranging from the low .60s to the mid .80s. The strongest correlation was found between the KABC-II Knowledge/*Gc* factor and the WISC-IV Verbal Comprehension Index (.85) (Kaufman & Kaufman, 2004). Comparable

results were demonstrated with the WJ III Cognitive. For example, the KABC-II global indexes (FCI, MPI, and NVI) produced correlations in the mid to high .70s with the WJ III Cognitive – GIA. Correlations between the WJ III Cognitive indexes and the corresponding KABC-II scales ranged from the low .50s to the mid .80s; and, the strongest correlation was again found between the KABC-II Knowledge/*G*c factor and the WJ III Cognitive Comprehension Knowledge Composite (.84) (Kaufman & Kaufman, 2004). Such results provide evidence for the convergent validity of the KABC-II with the WISC-IV and the WJ III Cognitive. Thus, if the KABC-II does not measure intelligence accurately then neither do these other popular tests of intelligence.

Further evidence of convergent validity stems from independent researchers. Using confirmatory factor analysis, two studies provided evidence that the same *g* underlies the KABC-II, the DAS-II, the WJ III and the WISC-IV (Floyd, Reynolds, Farmer, & Kranzler, 2013; Reynolds, Floyd, and Niileksela, (2013). Similarly, S.B. Kaufman et al. (2012) found that the g's measured on the KABC-II and the WJ III are essential the same. Results by Floyd et al., Reynolds et al., and S.B. Kaufman et al. demonstrate that the same *g* factor is measured across the different batteries. Even further, the *g* factor explained over 80% of total test score variance. Such results provide robust evidence that the same underlying construct is measured by the KABC-II, the WISC-IV, the WJ III, and the DAS-II. Convergent validity results as outlined in the manual in addition to the findings by Reynolds et al. (2013) and S.B. Kaufman et al. (2012) provide strong evidence of generalizability. In other words, any results found with the KABC-II (e.g., that FCI and Knowledge/Gc were the best predictors of achievement) are most likely generalizable and applicable to other popular tests of intelligence.

In sum, the KABC-II is a strong measure of intelligence as evidenced by both content and construct validity. Thus, it is unlikely that the overprediction is the test's fault. If the KABC-II measures intelligence inaccurately so that it would result in an overprediction of achievement across ethnic groups then so do all other popular tests of intelligence.

Is the KTEA-II biased? Another possible explanation for the overprediction in achievement found among Black and Hispanic children is that the KTEA-II might be biased against those groups and not accurately measuring their achievement ability. In order to investigate this question both content and construct validity of the KTEA-II were explored.

Content validity. To provide evidence of content validity, as outlined in detail in the manual (Kaufman & Kaufman, 2004b, chapter 5), it is important to highlight that the KTEA-II measures all eight specific learning disabilities identified by the Individuals with Disabilities Education Act Amendments of 1997) (IDEA, 1997): written expression, basic reading skill and reading comprehension, reading fluency, listening comprehension, math calculation and math reasoning, and oral expression. Furthermore, the authors were careful to ensure that the appropriate skills would be measured in each achievement domain (reading, written expression, math, and oral language). The content of every subtest was based on current research and curriculum blueprints that matched the achievement topics in all academic areas taught in U.S. schools. Panels of experts were consulted and literature reviews conducted to ensure selection of items that operationalized each achievement domain appropriately. The KTEA-II manual (Kaufman & Kaufman, 2004b, pp. 57-69) explains in detail how each subtest and the corresponding items were chosen and developed. For example, the math subtest - math concepts & applications - was carefully developed by first consulting current school textbooks and curricula of several large school districts and state guidelines. Based on this research, a large list

of possible items was created, which was then compared and contrasted against the items on the original K-TEA as well as against the content of other popular achievement tests. After the items on the list were reduced to a reasonable number, tryout and standardization analyses of item discrimination were used to reduce the size of the items further and to keep a balance with regards to the skills that were measured. Selection and development of the other KTEA-II subtests followed a similar careful and systematic procedure. The thorough procedures that were employed in the development of the KTEA-II provide strong evidence that the KTEA-II is not a biased measure of achievement, but reflects the content of school curricula. In other words, reading, writing, math, and oral language skills are measured in the same way as a child would be evaluated in a school setting.

Construct validity. Evidence of construct validity is provided in several ways, as previously outlined in more detail in chapter 2 of this present study and in the test's manual (Kaufman & Kaufman, 2004b, chapter 7). Results of the confirmatory factor analysis demonstrated good statistical fit of the structure underlying the KTEA-II (CFI = .992; RMSEA = .062) and also showed that the subtests loaded high on each of their corresponding factors (Kaufman & Kaufman, 2004b). Such findings provide evidence that the test measure the intended content. In addition, present findings of construct invariance across different ethnic groups provide further evidence for the test's construct validity.

In order to investigate the question of construct validity further, the convergent validity of the KTEA-II was explored. As outlined in the manual (Kaufman & Kaufman, 2004b, chapter 7), correlations between the KTEA-II and other popular achievement tests, including the WIAT-III and the WJ III Achievement, provide strong evidence for convergent validity. The KTEA-II Comprehensive Achievement score correlated .90 with the WIAT-II Total score and .89 with the

WJ III Total Achievement score. The KTEA-II Reading, Math, and Written Language composites correlated between the mid- to high-.80s with the corresponding composites on the WIAT-II and from the low .60s to mid-.80s with the corresponding composites on the WJ III Achievement (Kaufman & Kaufman, 2004b). Such results indicate that the KTEA-II measures achievement in the same way as the WIAT-II and the WJ III Achievement do. Thus, if the KTEA-II is biased against minority group children then so are other popular tests of achievement, including the WIAT-II and the WJ III Achievement.

In sum, evidence from content and construct validity studies demonstrate that both the KABC-II and the KTEA-II measure all facets of the theoretical constructs of intelligence and achievement correctly, based on CHC theory, and they measure it in the same manner as do other well-established tests of intelligence and achievement. Indeed, essentially all modern tests of cognition, including the Wechsler, the Woodcock Jonson, and the DAS tests, are founded on or readily interpretable from CHC theory (e.g., Flanagan et al., 2013; Flanagan& Kaufman, 2009), a theory that is well validated with more than 75 years of research. Thus, it is unlikely that the overprediction of Hispanic and Black children's' achievement is due to test bias or psychometric flaws of the KABC-II or KTEA-II.

Is it our school system? An alternative explanation for the overprediction could be that our schools are not effectively teaching math, reading, and writing. It is possible that Black and Hispanic children have the cognitive capacity to achieve higher; however, the schools might not be taking advantage of their aptitude. For example, Tables 23 and 25 of this study demonstrate the ethnic score differences on the KABC-II indexes; the tables show that Black and Hispanic students scored fairly high on the Learning/Glr index. Such findings demonstrate that the students do have the cognitive capacity to learn. In fact, the scores are close to 100 and only

about 3-4 points below the Learning/Glr scores for Caucasians. Such results indicate that the children have the ability to learn in school; however, the schools have not utilized their intellectual capacity. That is to say, the problem may not be that the tests are biased, but that the schools do not avail of the existing cognitive capacities. The overprediction could be the result of the schools' difficulties to effectively teach ethnically students. Schools, therefore, need to find methods to teach reading, writing, and math to ensure that each child's aptitude is fully taken advantage of. What follows are educational recommendations that might assist in utilizing the Blacks' and Hispanics' cognitive capacities more effectively.

However, before proceeding to suggesting new and creative ways of learning, it is important to highlight that any type of overgeneralization needs to be considered with caution. Even though results of this study suggest that Black and Hispanic students have certain relative cognitive strengths and weaknesses, each student's learning needs differ and by no means are the following suggestions meant to categorize students according to their ethnic group membership. Differences in home environments, nutrition, a lack of availability of resources, and dissimilarities in the quality of schools attended are all plausible factors that impact students' success and their achievement outcomes. Similarly, differences in parenting style and the experiencing of distressing events (e.g., neighborhood crimes) can affect the students' self-confidence, self-esteem, and mental-health – all of which can have direct or indirect effects on their ability to learn and achieve. It is those difficult situations that can make learning in school much more difficult and challenging for minority group students, especially when they come from lower socioeconomic environments (Weiss et al., 2006).

Ideally, every learning intervention would be tailored specifically to each student's pattern of cognitive strengths and weaknesses in order to meet personal needs and aspirations.

However, in light of limited resources, it is sometimes not feasible to personalize learning environments. Below strategies are meant to provide educators with a general understanding of what some members of certain groups' might benefit from, without disregarding each and every student's individual strengths.

Learning strategies. One way for schools to teach reading, writing, and math more effectively to Black and Hispanic students is to target existing learning strengths. The idea is that understanding how a child naturally solves problems or thinks can help determine the most effective academic interventions (Kaufman & Kaufman, 1983). For example, referring back to Tables 23 and 25, on average, Black students showed relative cognitive strengths in Sequential/Gsm and Learning/Glr. Hispanic students, on the hand, demonstrated relative strengths in Planning/Gf and Simultaneous/Gv. Results of the overprediction of achievement indicate that the schools are not taking advantage of these relative strengths in cognitive ability, because these students, as a group, achieve lower than those scores predict. That is to say, Black and Hispanic students do have the cognitive aptitude to perform better in reading, writing, and math, but the schools may not be effectively tailoring their intervention techniques to assist Black and Hispanic students to effectively utilize their cognitive strengths.

Learning approaches focused on using existing cognitive strengths. Kaufman and Kaufman (1983) proposed a strength-based learning approach that targets already existing cognitive processing abilities. The authors argue that children either process information 'simultaneously' or 'sequentially'. Whereas simultaneous processing refers to fluid and visual spatial intelligence (the capacity to reason and solve novel problems that requires the integration of different stimuli), sequential processing pertains to short-term memory and learning capacity

(the capacity to manipulate stimuli and learn in an orderly fashion) (Ashaman & Das, 1980; Das, Kirby, & Jarman, 1975).

As mentioned previously, Black students, as a group, demonstrated a relative strength in sequential processing skills or short-term memory and learning. Hispanic students had relative strengths in fluid reasoning and simultaneous processing or visual-spatial abilities, nonverbal strengths that have been identified in numerous previous studies of Hispanics on Wechsler's scales (Kaufman & Lichtenberger, 2006). Educators are encouraged to use the group's particular strength in processing and emphasize teaching that utilizes their type of learning pattern. What follows are teaching and intervention suggestions for reading, spelling, and math, which specifically target either sequential processing (and short-term memory and learning ability) or simultaneous processing (and fluid reasoning) as previously articulated (Gunnison, Kaufman, & Kaufman, 1982; Kaufman & Kaufman, 1983).

Teaching approaches that utilize sequential and memory learning strategies focus on breaking down stimuli into single parts and components to then solve a problem in an orderly way. For example, in order to improve reading techniques, the breath and depth of vocabulary knowledge needs to be enhanced. Using sequential processing and memory and learning skills, teachers should practice matching words with common synonyms (e.g., favorite = special) and distinguishing between the meanings of words that have the same root (e.g., like, likely, likewise). More sophisticated reading involves the ability to comprehend texts. For the sequential learner (a person who finds it relatively easy to memorize single step) it would be easier to read a sentence in parts, identify the main ideas, and then combine the parts to understand the full text (Gunnison, 1982). Other strategies include organizing scrambled sentences into coherent ones, and learning to distinguish between topic sentence and supporting

details in a paragraph. In order to enhance spelling skills using simultaneous processing or memory ability, it is recommended to separate different syllabi of a word and have the child practice pronouncing these syllabi. Such a strategy enhances a student's decoding ability, which is used both in reading and writing tasks. Math skills can be improved by copying numbers, practicing the recognition of equivalence of number combinations arranged in different orders (e.g., 5+2 and 2+5 both equal 7), verbalizing mathematical problems, and practice solving addition and subtraction problems by breaking down the steps (Gunnison, Kaufman, & Kaufman, 1982).

In contrast to a sequential processing/memory approach, which focuses on breaking down a task and providing a step-by-step solution, simultaneous processing/fluid reasoning approaches address the ability to integrate different stimuli. As suggested by Tables 23 and 25 and a host of research on Wechsler's scales (e.g., Kaufman & Lichtenberger, 2006), Hispanic students have strong simultaneous processing and fluid reasoning abilities (e.g., typical Wechsler profiles for Hispanic individuals show high nonverbal-low verbal). Learning strategies that focus on simultaneous processing and reasoning include the ability to organize and connect parts into a single whole often using visualization (Ashman, & Das, 1980). For example, in order to improve decoding ability - an essential skill when it comes to reading, especially in younger age groups - it is advisable to have students picture letter features that can help them remember and recognize letters (e.g., T looks like a cross) or practice filling in missing letters to complete a word. Other strategies include matching words or sentences with pictures. When students are older and they need to start building their comprehension skills a teacher might read a paragraph and encourage the student to visualize what is being read; the instructor could also ask the student to fill in missing parts of a story or have them organize an entire story, using series of

pictures, in order to enhance text organization skills (McRae, 1981). Spelling skill strategies center on recognizing patterns that word families have in common (e.g., what do words that end in *ion* have in common?). And Math strategies that target a child's ability to process information simultaneously or the ability to reason (e.g., to identity patterns and relationships between different stimuli) focus on visualization techniques (e.g., picture the steps needed to solve a certain math problem) (Kaufman & Kaufman, 1983).

Overall, given the above research findings, teachers should expect that students with sequential processing and short-term memory strengths might perform better academically when tasks are broken down into their components and the problem is solved in a step-by-step manner that takes advantage of the student's good memory. Students with better simultaneous processing and reasoning skills, on the other hand, might benefit from learning strategies that involve visualization, pattern recognition, task completion, and the application of reasoning.

Learning approaches focused on developing cognitive skills. An alternative to the strength-based approach of learning is to focus on developing certain cognitive skills that a child might be lacking. Here, the concept postulates that knowing what cognitive processing factors lead to the child's academic problems can help educators to use creative way to target those areas (Mascolo et al., 2014; Naglieri & Pickering, 2003; Wendling & Mather, 2009). Data presented in Tables 23 and 25 suggest that Black students, as a group, demonstrate relative weaknesses on Gf, Gv, and Gc abilities and Hispanic students demonstrate relative weaknesses in Gsm, Glr, and Gc skills. What follows are teaching approaches that target cognitive processing in creative ways with the aim to improve achievement outcomes, as suggested by several researchers (Mascolo et al., 2014; Naglieri & Pickering, 2003; Wendling & Mather, 2009). These learning methods are non-conventional, alternative approaches to teaching children, who do or do not

experience learning problems in school. The use of non-traditional teaching methods, especially ones that take into account the group strengths and weaknesses, might reduce the overprediction that was observed in this study.

For example, students that struggle with Gf or higher-level thinking and reasoning, and planning can have difficulties with math, as it might be challenging for them to reason quantitatively (Jiglesias-Sarmiento, & Deaño, 2011). Fluid intelligence can also influence the students' ability to draw inferences from texts, as might be required in a reading task; and, it might be difficult for them to generalize concepts in writing (McGrew & Wendling, 2010). To assist with learning, teachers can encourage students to sort, categorize, and, thereby, plan certain tasks in a step-by-step, sequential manner. For example, teachers could use graphic organizers, and, perhaps most importantly, model meta-cognition and quantitative and verbal problem solving by explaining aloud how one has reached a certain conclusion (Feifer & DeFina, 2002; Naglieri, & Pickering, 2003; Naglieri & Johnson, 2000). Iseman and Naglieri (2011) emphasized the importance of having the child self-reflect and self-evaluate out-loud upon prompting by instructors on how a certain conclusion was reached, and showed that this intervention Improved children's math abilities. A weakness in Gv can also impact a student's ability to solve abstract problems, specifically those that involve visual stimuli. For example, it can be challenging for the student to visualize the spatial orientation of objects and to read graphs or charts, as would be required in more advanced math tasks. Gv also impacts the student's ability to comprehend what is being read when it requires the student to think spatially, and to plan spatially during writing tasks (McGrew & Wendling, 2010). It is thus recommended to practice discriminating between different visual features, to graphically record information,

and to use teaching methods that target various senses (e.g., combining visualization with kinesthetic or tactile learning) (Naglieri & Pickering, 2003; Pressley & Woloshyn, 1995)

For students that have weaker Gc or crystallized verbal ability it can be challenging to use prior knowledge, and to acquire new information, especially when verbally communicated. *G*c most drastically influences the students' reading abilities. For example, it can have an impact on the students' decoding and comprehending skills as well as their ability to retell or paraphrase when reading texts. However, it can also be challenging to understand math vocabulary and concepts, and to express oneself in writing (McGrew & Wendling, 2010). Strategies that target crystallized intelligence include creating a learning environment that is rich in both language and experience. It is also highly recommended to continue exposing students to a variety of vocabulary, having the children read out loud, and increasing their time spend reading (Mather, Lynch, & Richards, 2001).

Limitations in *G*sm and *G*lr are highly related, as both factors are linked to memory capacity. A weakness in *G*lr, for example, can make it difficult to learn new concepts, to retrieve and recall information, and to learn and generate ideas quickly. Such problems predominantly impact reading and writing ability. It can make it more challenging to access previously learned information, as might be required during reading tasks, or to paraphrase read passages. Also, accessing and memorizing vocabulary, generating new ideas and putting them into writing can be challenging. Math skill can also be impacted, as it is difficult for this group to recall certain facts or rules (Glisky, & Glisky, 2002; Mascolo et al., 2014; Naglieri & Pickering, 2003). Similarly, a weakness in *G*sm can make it challenging to hold information in memory and manipulate and use it. This would influence the students' abilities to understand what is actually being read, to memorize math facts and procedures, and to identify and express main ideas in

writing (Baddeley, Eysenck, & Anderson, 2009; Dehn, 2008). In order to target *G*lr and *G*sm, it is recommended to focus interventions on the rehearsal of information and the teaching of memorization techniques, such as chunking, association and clustering strategies, and mnemonics (Elliott, Gathercole, Alloway, Holems, & Kirkwood, 2010; Schneider, 2010; Wendling & Mather, 2009).

Generally speaking, when instructors elect to focus their interventions on learning strategies that target fluid (Gf) and visual (Gv) processing, it is recommended to model step-by-step meta-cognition and problem solving, which will help the students learn to discriminate between visual information by including kinesthetic and tactile learning techniques. Exposure to vocabulary and reading tasks target crystallized (Gc) processing, and practicing memorization techniques might be especially useful for students who would like to enhance their learning (Glr) and short-term memory (Gsm) abilities. Although above learning suggestions provide educators with a list of creative, innovative, and, most importantly, evidence-based learning strategies that have been found to be effective in improving reading, writing, and math outcomes, using cognitive processing, it is important to consider that above strategies provide educators only with examples. Needless to say, individual learning needs always have to be prioritized over group generalizations.

Results from Tables 23 and 25 strongly suggest that Black and Hispanic students have the ability to learn (as indicated by their Learning/Glr scores) and to apply that learning to academic subjects in school. Such findings demonstrate that there is massive potential that has been untapped by the home environment or educational system. These students have the capacity to read, write, and compute at a higher level than they have achieved, but do not, for some reason that is unclear. Possible explanations include differences in learning environments, working

parents, and responsibility to take care of younger siblings and the household. Whatever those factors are that have prevented minority group students to avail of their cognitive resources, generally speaking, it seems that experimental and engaging learning environments and creative ways of teaching can be helpful in enhancing reading, writing, and math outcomes for students who struggle.

Clinical Implications

Building upon the suggestions for educators that can help reduce the overprediction of academic achievement for Blacks and Hispanics, the following section includes the key findings of this study as pertaining to the most important implications for clinical psychologists, neuropsychologists, neuropsychologists, and others who assess ethnic minority children's cognitive ability and achievement.

Results from the construct invariance analysis demonstrate that when using the Kaufman intelligence and achievement tests, the same CHC factor structure is measured for Hispanic and Black as well as Caucasian children. For clinicians this is an important finding as they can now be confident that the profile of strengths and weaknesses based on the results from the Kaufman tests, in fact, accurately reflects the child's abilities, regardless of ethnic origin (Caucasian, Black or Hispanic).

Even further, there is strong evidence that suggests that findings of this present study not only pertain to the Kaufman tests, but generalize to other popular tests of cognition and achievement. For example, the study conducted by S.B. Kaufman et al. (2012) demonstrated that the *g* measured by the KABC-II is essentially the same *g* that is measured by the WJ III. Similarly, Reynolds et al. and (2013) and Floyd et al. (2013) demonstrated that the same g underlies the KABC-II, the WISC-IV, the WJ III, and the DAS-II. Such findings provide strong

evidence for the fact that the same global construct that is being measured by the KABC-II is also measured by the WISC-IV, the DAS-II, and the WJ III. Any findings pertaining to the KABC-II are therefore likely to be generalizable to those other tests.

Additionally, the strong correlations between the KABC-II scales and the corresponding WISC-IV and WJ III Cog indexes as well as the strong correlations between the KTEA-II and corresponding WIAT-II and WJ III Achievement indexes indicate that the same factors measured by the Kaufman tests are also measured by the Wechsler and Woodcock Johnson tests (due to evidence of strong convergent validity). Thus, clinicians can be reasonably confident that results of the present study generalize to other popular tests of cognition and achievement. Such findings are even more important considering that the most recent versions of the Wechsler scales offer scales that increasingly resemble the theoretical framework of CHC theory. The WPPSI-IV (Wechsler, 2012) and WISC-V (Wechsler, 2014) each yield scores on five scales that measure *Gc, Gf, Gsm, Gv,* and *Gs*. The WISC-V also yields a sixth (supplementary) scale, Symbol Translation, that measures *Glr*. All of these CHC factors, except *Gs*, were validated as invariant across ethnic groups in this present dissertation.

Results from the predictive invariance analysis also demonstrate several important findings for clinicians. Perhaps most importantly, two myths that clinicians have followed without adequate data supporting them have been disconfirmed by this study. First of all, there is a general belief that global scores should not be interpreted or even used with ethnic minority groups, because those scores are thought to be biased and unreliable for those children (e.g., Lezak, 1988). Data of this preset study, however, suggest otherwise. The most global KABC-II score, FCI, was, in fact, the most accurate at predicting achievement not only for Caucasians, but, most importantly, for Hispanics and Blacks as well. Results of this present study suggest

that the FCI is a highly reliable and fair scale for Hispanic and Black school-aged children. Even further, the FCI scale is fairer than other scales that are thought to be linguistically and culturally neutral, such as Sequential/*G*sm and Simultaneous/*G*v. Such findings are important, because, as recommended in the manual (Kaufman & Kaufman, 2004), users are encouraged to use the FCI whenever possible, as this scale provides clinicians with the most comprehensive ability profile. The authors had originally recommended to use the MPI based on the belief that the MPI is a fairer scale for ethnic minority groups, as it excludes those items that are generally considered to be more culturally and linguistically loaded (e.g., the Knowledge/Gc subtests). However, the data of this study suggest that the FCI is a fair predictor for ethnic minority group children. Thus, results of this study provide the necessary evidence that allow clinicians to use the FCI with Black and Hispanic children, which will provide clinicians with a more comprehensive and complete profile of cognitive strengths and weaknesses.

The second, and related, myth that clinicians follow without the necessary empirical evidence is the idea that less lexical and culturally more neutral scales, such as Sequential/*G*sm and Simultaneous/*G*v, are fairer scales to use with ethnic minority groups as compared to more culturally and linguistically loaded scales, such as Knowledge/Gc. Results of this study showed that Knowledge/Gc was the only scale that demonstrated basically no predictive bias both in terms of its slope and its intercept. In fact, Knowledge/Gc was the most accurate predictor variable for Black and Hispanic achievement outcomes, in addition to FCI. Such results are not entirely surprising because the subtests that comprise Knowledge/Gc (Expressive Vocabulary, Riddles, Verbal Knowledge) are clearly related to the knowledge required in completing a reading and writing task successfully.

Such findings are important for clinicians to consider when evaluating Black or Hispanic children. Even though it is the common belief to abstain from administering culturally and linguistically loaded subtests to minority group students, as those are thought to be biased, the data suggest that Knowledge/Gc on the KABC-II is, in fact, the fairest predictor of Black and Hispanic achievement. Even further, the strong correlations between the KABC-II Knowledge/*G*c scale and the corresponding WJ III Cognitive and the WISC-IV scales (.84-.85) suggest that findings pertaining to the *G*c scales might also be generalizable beyond the Kaufman tests to the Woodcock-Johnson and Wechsler scales.

Theoretical Implications

In addition to providing evidence necessary to allow for the continuous valid clinical use of the Kaufman tests with ethnic minority group children, results from the study also provide important theoretical implications for researchers and clinicians. The construct invariance analysis validates the theoretical CHC model of intelligence for Blacks and Hispanics. Even though researchers and clinicians seem to assume that the CHC model of intelligence is applicable to everybody, regardless of ethnic origin, there has hardly been any data supporting the hypothesis that this is, in fact, the case (exceptions are Keith et al., 1999; Rush et al., 2003; Trundt, 2013). Results of this study provide the necessary empirical evidence that confirms CHC theory as valid for Hispanic and Black children and adolescents. Such findings have important implications also for other popular tests of intelligence and achievement as many tests, such as the Woodcock Johnson, the DAS, and the most recent versions of the Wechsler tests, use CHC theory as their theoretical underpinnings. Even further, Jewsbury (2014) established CHC theory as an appropriate structure underlying a variety of popular neuropsychological assessment measures. Findings of this study provide the necessary evidence needed for the continuous use

of CHC theory as an appropriate interpretation model of the cognitive abilities and academic skills for ethnic minority group children.

Results from the predictive invariance analysis also have important theoretical implications. Practitioners and school psychologists often use cognitive tests to provide further understanding of a child's academic strengths and weaknesses; thus, the relationship between cognitive ability factors and academic outcome, which was important when Alfred Binet first developed the Stanford-Binet (Binet & Simon, 1905) remains important more than a century later. There is an increasing amount of literature on the successful linkage of several CHC cognitive ability factors with specific achievement skills (see summary by Flanagan, Ortiz, Alfonso, and Dynda, 2014); however, virtually nobody has provided evidence that this existing research generalizes to different ethnic groups. Keith (1999) explored the path coefficients between several WJ-R cognitive and achievement variables and found that generally the same CHC factors predicted reading and math across Caucasian, Hispanic, and Black students (with the exception of reading comprehension, which was more strongly influenced by Gs and Gc for Hispanic students as compared to Black and Caucasian students). Along with Keith's study, this present study is among the first to investigate the relationships between CHC cognitive abilities and specific academic skills across ethnic groups.

Results from the regression analysis provide important information regarding the relationship of cognitive CHC factors and achievement skills. [Findings demonstrate that the CHC ability factors relate essentially in the same way to the specific academic skills across all three ethnic groups.] Dr. V. says this is misleading. Such results are of key theoretical importance as they provide evidence that the correlations between the theory-based CHC ability factors and achievement outcomes are generalizable and universal. Whereas there is research that

has related specific cognitive CHC variables to specific academic skills, with the exception of Keith (1999), nobody has validated those relationships for different ethnic groups.

A review of the results section suggested that the CHC ability factors Planning/Gf and Knowledge/Gc demonstrate the strongest relations with math achievement for Caucasians (correlating in the mid to high .50s) and, likewise, for Hispanic and Black students (correlating in the low .50s to mid-.60s). Such findings are consistent with Keith's results (in addition to Gs) and with the findings of other researchers (e.g., Flanagan et al., 2006; McGrew & Hessler; 1995; McGrew & Wendling, 2010) in samples that were not separated by ethnicity. The relationship between Gv and math has not been consistently reported in previous studies (e.g., Flanagan et al., 2014). This study suggests that Gv might be a moderate predictor for math achievement for Caucasians, Blacks and Hispanics (correlating between the mid-.40s and mid -50s for the three ethnic groups across all ages). Thus, the pattern of influence of CHC cognitive abilities on the math domain is essentially the same across the three ethnic groups.

Similar results were found for the reading domain. For example, results demonstrate a consistent and strong relationship with Knowledge/Gc across all three ethnic groups (producing correlation coefficients ranging from the high .50s to the low .80s). Such findings are not surprising and consistent with previous results from Keith (1999) and from studies that did not take ethnic differences into consideration (e.g., Flanagan et al., 2014). Present findings provide further evidence that the relationship between *G*c and reading is also important for Black and Hispanic school-aged children. Learning/*G*lr, another CHC factor that has been previously linked to reading, especially in the early school years, has been found to be a particularly strong predictor variable for Hispanic children in grades 1-4 (correlating about .60 with reading).

Finally, previous research findings demonstrated an important relationship between writing and Gc (although research on writing has generally been scarce). Knowledge/*G*c was also found to be the strongest predictor for Written Expression for all three ethnic groups in this study, especially for grades 5-12 (producing correlations ranging from the .50s to .70s). At the youngest age group, grades 1-4, Learning/*G*lr was also an important predictor for Written Expression across all ethnic groups (producing correlations in the mid .50s to low .60s). Only a few previous studies were able to successfully link *G*lr to writing for predominantly Caucasian samples (Mascolo et al., 2014). *G*sm (as well as *G*a) is another factor that has previously been linked to writing; however, present findings found that *G*c produced the most consistent and strongest relationship with writing across all three ethnic groups.

In sum, such results demonstrate that essentially the same CHC factors that predicted Reading, Math, and Written Expression for Caucasians were also the best predictors for Blacks and Hispanics. Such findings are consistent with Keith's (1999) results (who, however, did not investigate the relationships between CHC broad ability factors and writing) and provide further evidence that CHC theory, in fact, is a universal theory of intelligence and achievement and applicable to other ethnic groups.

Limitations of the Present Study

Even though present findings provide support for the lack of construct and predictive bias against Black and Hispanic school-aged children on the KABC-II and KTEA-II, results need to be understood in the context of the study's limitations. Probably the most important limitation of the present study that needs to be taken into consideration pertains to the measure of SES. To begin with, results from the ANOVA (Tables 4, 5, and 6) demonstrated that the variable SES produced a significant main effect with the KABC-II MPI and the KTEA-II CAC. Those results

indicated that SES was a confound that needed to be controlled. Furthermore, no significant interactions between ethnicity and SES were detected, which made this variable an appropriate confound to be used across all three ethnic groups (SES had the same effect on each of the three ethnicities). In this study, SES was defined as the mother's (or father's) educational attainment. One important limitation pertains to the fact that not both parents' educational attainment was used to determine SES. Using a combination of mother's and father's educational attainment would have been a slightly more accurate determinant of the child's level of SES. However, given the archival nature of this data set, the author was unable to remedy this shortcoming. Furthermore, implications pertaining to the fact that SES for some children was based on the father's level of education, instead of the mother's (if mother's was not obtainable), cannot be determined. Future research might want to address this issue. Furthermore, SES was a categorical variable, which was also limiting, as exact differences between educational attainments could not be made (e.g., whether they had a Bachelor's or Master's Degree).

Whereas parental educational attainment is a common variable used to define SES (Weiss et al., 2006), many other factors beyond parental educational attainment significantly influence a child's cognitive development. As outlined in detail by Weiss et al., factors such as mental and physical health differences, differences in income, differences in home environments, and quality of schools attended all impact a child's cognitive and achievement scores. Given the archival nature of the present data set, no further information on the above variables could be obtained, which, therefore, is another limitation of the present study. Indeed, if the data set had not been archival, the author of this present study would have had the opportunity to experimentally manipulate the design and include questionnaires about home environment, mental and physical health, and so forth.

Other limitations pertain to the study's measures and methodology. First, in terms of the study's measures, it is important to consider that the KABC-II and KTEA-II were limited in that the tests only measured seven of the CHC broad abilities, but did not include *Gs*, which is a factor commonly used in other tests of cognition and which shows strong relationships to different aspects of academic achievement (Flanagan et al., 2014). Future researchers need to validate the use of Gs with ethnic minority groups. Furthermore, only children grades 1-12 and the corresponding subtests appropriate for school-aged children were included in the study. Younger children and the subtests designed specifically for preschool children (e.g., Face Recognition, Conceptual Thinking) were excluded from the study due to limitations in sample size. Also, the KTEA-II Associational Fluency and the KABC-II Atlantis Delayed, Rebus Delayed, and Gestalt Closure subtests all had substantial amounts of missing data. Thus, those subtests were excluded from the analysis. In order to ensure that findings are generalizable to other age groups and subtests, future research needs to include those subtests not included in this study as well as children who have not yet entered first grade. Second, in terms of the study's methodology, the lack of power also prevented the author from studying possible developmental differences when examining construct invariance. The present study could have potentially missed detecting an important developmental trend in terms of the structure of the construct. In order to ensure that present findings maintain even if a sample is divided up into different age groups, future research needs to replicate current findings with children at different age groups.

It is also important to take into consideration limitations pertaining to the sample's demographics. Only three broad ethnic groups were included in the sample. Due to a lack of sample size, other ethnic groups, such as Asians, Pacific Islanders, and Native Americans, could not be included in the analysis. In that sense, the evidence of nonbias found in this study might

not necessarily be generalizable to other ethnic groups. Future research ought to address this question. Additionally, it is important to keep in mind that the term 'Hispanic,' used in order to classify the standardization sample, is a broad term and encompasses many heterogeneous populations that differ in terms of their culture and histories. The heterogeneity of the group labeled 'Hispanic' was not taken into consideration, as no representative subsamples from each country were available. The issue of labeling heterogeneous groups as if they were one and the same is a limitation that not only pertains to this present study, but captures a larger problem with research on the topic area in general (Weiss et al., 2006). Furthermore, within the Hispanic culture there is wide variability in fluency with the English language. These findings may not be applicable to students who are less proficient in English.

Other limitations relate to the generalizability of present findings. The data for this present sample stemmed from the standardization samples of the KABC-II and KTEA-II, which were stratified according to 2001 U.S. Census data. Even though the fact that the sample was stratified on important background variables is a strength of the study, the U.S. census has undoubtedly changed since 2001. Thus, the stratification might not be representative of the current U.S. census. Further and importantly, even if no major changes have occurred in the U.S. census, the present results are only generalizable to the U.S. and not to countries in Europe or Asia, where numerous adaptations and translations of the K-ABC and KABC-II have been in use for years; indeed, the German KABC-II was recently published in March, 2015.

Future research should replicate present findings in other countries to ensure generalizability of the findings beyond the United States. Finally, it is crucial to take into consideration that the sample was composed of normally developing children. However, the children that are most commonly referred for psychological testing are those who struggle with

learning disabilities or other developmental disorders. In order to ensure the generalizability of results, future research should replicate present findings using special populations.

Conclusions

Results of the present study provide evidence of differential construct and predictive validity on the KABC-II and KTEA-II across a representative sample of Caucasian, Black, and Hispanic school-aged children. Such findings provide the evidence needed to justify the continuous use of those measures with ethnic minority group children when assessing intellectual and achievement ability. Educators and clinicians can feel confident that the two tests measure the same construct across different ethnic groups. Furthermore, the KABC-II has been found either to predict achievement outcome equally well for minority group children or to overpredict their achievement. The persistent overprediction of achievement outcome was remarkable and could be indicative of the fact that schools do not adequately utilize Blacks' and Hispanics' cognitive abilities based on contemporary methodologies for teaching academic subjects in U.S. schools. This study provided an overview of teaching strategies that could potentially help to improve achievement outcomes for Hispanic and Black children and reduce the overprediction.

Notable was the fact that Knowledge/*G*c did not show any evidence of slope bias and essentially no evidence of intercept bias. In other words, Knowledge/*G*c was the one ability that showed the least bias and was therefore most accurate at predicting achievement for all three ethnicities across all three grade groups. Similarly, a more global score, such as FCI, appeared to be a better predictor when estimating achievement for minority group students than the separate abilities (including ability factors that are considered culturally and linguistically neutral, such as Sequential/*G*sm and Simultaneous/*G*v). Clinicians might, therefore, keep in mind giving the global intelligence scales more consideration when evaluating the prediction of achievement for

Black and Hispanic ethnic minority group children. In sum, results demonstrate that the KABC-II and the KTEA-II can provide meaningful and accurate information about minority group children's cognitive and achievement ability profiles.

Even further, as present analyses were based on the CHC construct of intelligence and achievement, a theoretical framework that underlies many popular tests of intelligence and achievement (e.g., the WISC-V, the WJ IV), results of this present study are likely to be generalizable to other tests. Given present findings, clinicians can be reasonably confident that the evaluations of Hispanic and Black minority group children with tests that use CHC theory as their interpretation are likely to be non-biased. However, it is important to remember that even if the KABC-II and the KTEA-II, as well as other cognitive or achievement tests, are not biased in terms of their psychometric properties and theoretical interpretations, such findings do, by no means, imply that group mean differences found are not biased. Cognitive and achievement scores are impacted by many different factors (e.g., income, home environment, quality of school) that are impossible to control effectively, but have been found to correlate highly with lower scores on cognitive ability measures (Nisbett, 2009). Thus, it is strongly recommended not to draw meaningful conclusions from differences in mean scores found between different ethnic groups in this study or in other studies reported in the literature.

References

Arbuckle, James L. (1995-2011). *AMOS 20.0 User's Guide.* Crawfordville, FL: AMOS Development Corporation.

Arinoldo, C. G. (1981). Black–white differences in the general cognitive index of the McCarthy scales and in the full scale IQs of Wechsler's scales. *Journal of Clinical Psychology*, *37*(3), 630–638. doi: 10.1002/1097-4679(198107)37:3<630::AIDJCLP2270370331>3.0.CO;2-#

Army, U. S. (1945). Personnel research section. The army general classification test. *Psychology Bulleting*, *42*, 760-768.

Ashman, A.F., & Das, J.P. (980) Relations between planning and simultaneous-successive processing. *Perceptual and Motor Skills, 51,* 371-381.

Aud, S., Fox, M. A., & KewalRamani, A. (2010). U.S. Department of Education, Institute of Education Sciences, National Center for Educational Statistics (2007). *Status and trends in the education of racial and ethnic groups.* Retrieved from: http://nces.ed.gov/pubs2010/2010015.pdf

Baddeley, A.D., Eysenck, M.W., & Anderson, M.C. (2009). Memory. New York: NY: Psychology Press.

Baggaley, A. R. (1974). Academic prediction at an Ivy League college, moderated by demographic variables. *Measurement & Evaluation in Guidance, 6*(4), 232–235.

Bane, M. J., & Ellwood, D. T. (1989). One fifth of the nation's children: Why are they poor? *Science*, *245*(4922), 1047-1053. doi: 10.1126/science.245.4922.1047

Bayley, N., & Jones, H. E. (1937). Environmental correlates of mental and motor development: A cumulative study from infancy to six years. *Child Development*, *8*(4), 329-341.

Bennet, G. K., Seashore, H. G., & Wesman, A. G. (1947). *Differential aptitude test.* New York, NY: Psychological Corporation

Boney, J. D. (1966). Predicting the academic achievement of secondary school Negro students. *The Personnel and Guidance Journal*, *44*(7), 700-703.

Bradley, L., & Bryant, P. (1985). *Rhyme and reason in reading and spelling.* Ann Arbor: University of Michigan Press.

Brandt, H., & Burke, L. K. (1950). Standardization of the Armed Forces Qualification Test AFQT-1 and 2, *American Psychologist*, *5*, 285.

Bridgeman, B., McCamley-Jenkins, L., & Ervin, N. (2000). *Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning Test* (College Board Report 2000–2001). New York: College Entrance Examination Board.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & R. A. Stine (Eds.), Testing structural equation models (pp.136-145). Newbury Park, CA: Sage Publications

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*(3), 456–66. doi: 10.1037/0033-2909.105.3.456

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. *The Oxford Handbook of Child Psychological Assessments*, 84-112.

Campbell, J.T., Crooks, L.A., Mahoney, M.H., & Rock, D.A. (1973). *An investigation of sources of bias in the prediction of job performance – A six-year study.* Princeton, NJ: Educational Testing Service.

Carneiro, P., & Heckman, J. J. (2002). The evidence on/ credit constraints in post-secondary schooling. *The Economic Journal*, *112*(482), 705–734. doi: 10.1111/1468- 0297.00075

Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge, MA: Cambridge University Press.

Ceci, S. J., & Kanaya, T. (2010). "Apples and oranges are both round": Furthering the discussion on the Flynn Effect. *Journal of Psychoeducational Assessment*, *28*(5), 441-447. doi: 10.1177/0734282910373339.

Centra, J. A., Linn, R. L., & Parry, M. E. (1970). Academic growth in predominantly Negro and predominantly white colleges. *American Educational Research Journal, 7*(1), 83–98.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural EquationMmodeling*, *9*(2), 233-255. doi: 10.1207/S15328007SEM0902_5.

Chen, T., Kaufman, A. S., & Kaufman, J. C. (1994). Examining the interaction of age X race pertaining to black□white differences at ages 15 to 93 on six Horn abilities assessed by K□FAST, K□SNAP and KAIT subtests. *Perceptual and Motor Skills, 79*, 1683−1690. doi: 10.2466/pms.1994.79.3f.1683.

Chou, T., & Huberty, C.J. (1990). *A freshman admissions prediction equation: An evaluation and recommendation.* Athens, GA: University of Georgia.

Clark, W. W., & Tiggs, E. W. (1950). *California achievement tests, primary battery.* Los Angeles, CA: California Test Bureau.

Clearly, T.A. (1968). Test bias: Prediction of grades of Negro and white students in integrated colleges. *Journal of Educational Measurement, 5*(2), 115–124. doi: 10.1111/j.1745-3984.1968.tb00613.x

Cole, N. S. (1981). Bias in testing. *American Psychologist*, *36*(10), 1067–1077. doi: 10.1037/0003-066X.36.10.1067

Cowen, S., & Fiori, S. J. (1991, November). *Appropriateness of the SAT in selecting students for admission to California State University, Hayward.* Paper presented at the annual meeting of the California Educational Research Association, San Diego, CA.

Crooks, L. A. (1972). *An investigation of sources of bias in the prediction of job performance. A six-year study*. Princeton, NJ: Educational Testing Service.

Das, J.P., Kirby, J.R., & Jarman, R.F., (1975). Simultaneous and successive syntheses: An alternative model for cognitive abilities. *Psychological Bulletin, 82*, 87-103.

Davis, J. A., & Temp, G. (1971). Is the SAT Biased Against Black Students? *College Board Review, 81*, 4-9.

Dehn, M.J. (2008). Working memory and academic learning: Assessment and intervention. Hoboken, NJ: Wiley.

Department of Education and Early Childhood Development (2015). Exceptionalities. Retrieved from: http://www.ed.gov.nl.ca/edu/k12/studentsupportservices/exceptionalities.html.

Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap evidence from standardization samples. *Psychological Science*, *17*(10), 913–920. doi: 10.1111/j.1467-9280.2006.01802.x.

Edwards, O. W., & Oakland, T. D. (2006). Factorial invariance of Woodcock-Johnson III scores for African Americans and Caucasian Americans. *Journal of Psychoeducational Assessment*, *24*(4), 358–366. doi: 10.1177/0734282906289595.

Elliott, C.D. (1990). *Differential Ability Scales: Introductory and technical manual.* San Antonio, TX: The Psychological Corporation.

Elliott, C. D. (2007). *Administration and scoring manual differential abilities scale 2nd edition (DAS-II).* San Antonio, TX: The Psychological Corporation.

Elliott, J.G., Gathercole, S.E., Alloway, T.P., Holems, J., & Kirkwood, H. (2010). An evaluation of a classroom-based intervention to help overcome working memory difficulties and improve long-term academic achievement. *Journal of Cognitive Education and Psychology, 9*, 227-250.

Epps, E. G. (1995). Race, class, and educational opportunity: Trends in the sociology of education. *In Sociological Forum, 10*(4), 593–608.

Feifer, S.G., & DeFina, P.D. (2002a). The neuropsychology of mathematics: Diagnosis and intervention. Middletown, MD: School of Neuropsych Press.

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., Fiorello, C. A., & Ortiz, S. O. (2010). Enhancing practice through application of Cattell–Horn–Carroll theory and research: A "third method" approach to specific learning disability identification. *Psychology in the Schools*, *47*(7), 739–760. doi: 10.1002/pits.20501

Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2007). Use of the cross-battery approach in the assessment of diverse individuals. In A. S. Kaufman & N. L. Kaufman (Series Ed.), *Essentials of cross-battery assessment second edition* (2nd ed., pp. 146 –205). Hoboken, NJ: Wiley.

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

Flanagan, D. P., Ortiz, S. O., Alfonso, V. C. & Dynda, A. (2014). Cognitive Assessment: Progress in Psychometric Theories, the Structure of Cognitive Tests, and Approaches to Test Interpretation. In D. Saklofske, V. Schwean, & C. Reynolds (Eds.), *Oxford handbook of psychological assessment of children and adolescents*. New York: Oxford University Press.

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

Floyd, R. G., Reynolds, M. R., Farmer, R. L., Kranzler, J. H., & Volpe, R. (2013). Are the general factors from different child and adolescent intelligence tests the same? Results from a five-sample, six-test analysis. *School Psychology Review*, *42*(4), 383–401.

Foley, P. P. (1971). *Validity of the officer qualification test for minority group applicants to officer candidate school*. Washington, DC: Naval Personnel Research and Development Lab.

Fox, W. L., Taylor, J. E., & Caylor, J. S. (1969). Aptitude level and the acquisition of skills and knowledge in a variety of military training tasks. Research Organization Technical Report. Washington, DC: Chief of Research and Development, Department of the Army.

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Kit of reference tests for cognitive factors.* Princeton, NJ: Educational Testing Service.

Fryer Jr, R. G., & Levitt, S. D. (2004). Understanding the black-white test score gap in the first two years of school. *Review of Economics and Statistics*, *86*(2), 447–464. doi: doi:10.1162/003465304323031049.

Glisky, E.L., & Glisky, M.L. (2002). Learning and memory impairments. In PJ. Eslinger quantitative and qualitative analysis. *Journal of Communication Disorders, 29,* 79-93.

Goldman, R. D., & Richards, R. (1974). Tbe SAT prediction of grades for Mexican-American versus Anglo-American students at the University of California, Riverside. *Journal of Educational Measurement*, *11*(2), 129-135.

Goodenough, F. L. (1927). The consistency of sex differences in mental traits at various ages. *Psychological Review*, *34*(6), 440-462. doi: 10.1037/h0075869.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*(11), S78-S94. doi: 10.1097/01.mlr.0000245454.12228.8f.

Guinn, N., Tupes, E.C., & Alley, W.E. (1970). *Cultural subgroup differences in the relationships between air force aptitude composites and training criteria* (Technical Report 70-35). Brooks Air Force Base, Texas: Air Force Human Resource Laboratory.

Gunnison, J. A. (1982). Remediation strategies based on the roles of simultaneous and successive processing in reading. *Journal of Educational Neuropsychology*, *2*, 36-69.

Gunnison, J.A., Kaufman, N.L., & Kaufman, A.S. (1982). Sequential and simultaneous processing applied to remediation. *Academic Therapy, 17,* 297-307.

Gutkin, T. B., & Reynolds, C. R. (1981). Factorial similarity of the WISC-R for White and Black children from the standardization sample. *Journal of Educational Psychology*, *73*(2), 227-231. doi: 10.1037/0022-0663.73.2.227.

Hauser, R.M. (1998). Trends in Black-White test-score differentials I: use and misuse of NAEP/SAT data. In U. Neisser (Ed.). *The rising curve: Long-term gains in IQ and related measures* (pp. 219-250). Washington, DC: American Psychological Association.

Hedges, L V., & Nowell, A. (1998). Group differences in mental test scores: Mean differences variability, and talent. In C.S. Jencks and M. Phillips (eds.) *The black-white test score gap.* Washington, DC: Brookings Institution.

Hennessy, J.J., & Merrifield, P.R. (1976). A comparison of the factor structures of mental abilities in four ethnic groups. *Journal of Educational psychology, 68,* 754-759. doi: 10.1037/0022-0663.68.6.754.

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

Iglesias-Sarmiento, V., & Deaño, M. (2011). Cognitive processing and mathematical achievement: A Study with schoolchildren between fourth and sixth grade of primary education. *Journal of learning disabilities*, *5*(37), doi: 0022219411400749.

Iseman, J. S., & Naglieri, J. A. (2011). A cognitive strategy instruction to improve math calculation for children with ADHD and LD: A randomized controlled study. *Journal of Learning Disabilities*, *44*(2), 184-195. doi: 10.1177/0022219410391190

Jencks, C. and Phillips, M., (1998). *The black–white test score gap*. Washington, DC: Brookings Institution Press.

Jensen, A.R. (1974). Ethnicity and scholastic achievement. *Psychological Reports, 34*, 659-668.

Jensen, A.R. (1977). An examination of cultural bias in the Wonderlic Personnel Test. *Intelligence*, 1, 51-64.

Jensen, A. R. (1980). *Bias in mental testing.* New York, NY: The Free Press.

Jewsbury, P. A. (2014). *The Cattell-Horn-Carroll model of cognition as an account of diverse clinical assessment batteries for neuropsychological assessment (*Unpublished doctoral dissertation). Melbourne School of Graduate Research, Australia.

Kamphaus, R. W., & Kaufman, A. S. (1986). Factor analysis of the Kaufman Assessment Battery for Children (K□ABC) for separate groups of boys and girls. *Journal of Clinical Child Psychology, 15*(3), 210□213. doi: 10.1207/s15374424jccp1503_2.

Kao, G., Tienda, M., & Schneider, B. (1996). Racial and ethnic variation in academic performance. *Research in sociology of education and socialization*, *11*, 263-297.

Kaufman, A. S. (1973a). Comparison of the performance of matched groups of black children and white children on the Wechsler Preschool and Primary Scale of Intelligence. *Journal of Consulting and Clinical Psychology, 41*(2), 186□191. doi: 10.1037/h0035095.

Kaufman, A. S. (1973b). The relationship of WPPSI IQs to SES and other background variables. *Journal of Clinical Psychology, 39*, 354-357.

Kaufman, A.S. (2009). *IQ testing 101.* New York, NY: Springer Publishing Company.

Kaufman, J. C., Chen, T., & Kaufman, A. S. (1995). Ethnic group, education, and gender differences on six Horn abilities for adolescents and adults. *Journal of Psychoeducational Assessment, 13*(1), 49□65. doi: 10.1177/073428299501300104.

Kaufman, A. S., Daramola, S.F., & Di Cuio, R.F. (1977). Interpretation of the separate WPPSI tests for boys and girls at three age levels. *Contemporary Educational Psychology, 2*(3), 232□238.

Kaufman, A. S., & Di Cuio, R. F. (1975). Separate factor analyses of the McCarthy Scales for groups of black and white children. *Journal of School Psychology, 13*(1), 10□18.

Kaufman, A. S., & Doppelt, J. E. (1976). Analysis of WISC□R standardization data in terms of the stratification variables. *Child Development, 47*(1), 165□171.

Kaufman, A. S., Harrison, P. L., & Ittenbach, R. F. (1990). Intelligence testing in the schools. In: T. B. Gutkin & C. R. Reynolds (Eds), *Handbook of school psychology* (pp. 289–327). New York: Wiley.

Kaufman, A. S**.,** & Hollenbeck, G. P. (1974). Comparative structure of the WPPSI for blacks and whites. *Journal of Clinical Psychology, 30*(3), 316□319. doi: 10.1002/1097-4679(197407)30:3<316::AID-JCLP2270300329>3.0.CO;2-B

Kaufman, A. S**.**, & Kamphaus, R. W. (1984). Factor analysis of the Kaufman Assessment Battery for Children (K-ABC) for ages 2 ½ through 12 ½ years. *Journal of Educational Psychology, 76*, 623-637.

Kaufman, A. S., & Kaufman, N. L. (1973). Black□white differences at ages 2□1/2 to 8□1/2 on the McCarthy Scales. *Journal of School Psychology, 11*(3), 194□204.

Kaufman, A. S., & Kaufman, N.L. (1983). *K-ABC: Kaufman assessment battery for children: Interpretive manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A.S., & Kaufman, N.L. (2004a). *Kaufman Assessment Battery for Children–Second Edition.* Circle Pines, MN: American Guidance Service.

Kaufman, A.S., & Kaufman, N. L. (2004b). *Kaufman Test of Educational Achievement–Second Edition (KTEA-II).* Circle Pines, MN: American Guidance Service.

Kaufman, A. S., Kaufman, J. C., & McLean, J. E. (1995). Factor structure of the Kaufman Adolescent and Adult Intelligence Test (KAIT) for whites, African Americans, and Hispanics. *Educational and Psychological Measurement, 55*(3), 365□376. doi: 10.1177/0013164495055003001.

Kaufman, A.S., & Lichtenberger, E.O. (2001). *Assessing adolescent and adults intelligence-Second edition.* Boston, MA: Allyn & Bacon.

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

Kaufman, A. S., McLean J. E., & Kaufman, J. C. (1995). The fluid and crystallized abilities of white, black, and Hispanic adolescents and adults, both with and without an education covariate. *Journal of Clinical Psychology, 51*(5), 637□647. doi:10.1002/1097-4679(199509)51:5<636::AID-JCLP2270510509>3.0.CO;2-8.

Kaufman, J. C., McLean J. E., Kaufman, A. S., & Kaufman, N. L. (1994). White□black and white Hispanic differences on fluid and crystallized abilities by age across the 11□ to 94□year range. *Psychological Reports, 75*(3), 1279□1288. doi: 10.2466/pr0.1994.75.3.1279.

Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region, and education differences of the eleven WAIS□R subtests. *Journal of Clinical Psychology, 44*, 231□248.

Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1991). Analysis of WAIS□R factor patterns by sex and race. *Journal of Clinical Psychology, 47*, 548□557.

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.

Kaufman, A. S., & Wang, J. (1992). Gender, race, and education differences on the K□BIT at ages 4 to 90 years. *Journal of Psychoeducational Assessment, 10*(3), 219−229. doi: 10.1177/073428299201000302.

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. 10.1037/h0089008.

Keith, T. (2006). *Multiple regression and beyond*. Boston, MA: Pearson Education.

Keith, T.Z. & Reynolds, C.R. (2003). Measurement and design issues in child assessment research.In Reynolds, C. R. and Kamphaus, R. W. (Ed), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed.), (pp. 79-111). New York, NY: Guilford Press.

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.20496.

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence*, *36*(6), 502-525. doi: 10.1016/j.intell.2007.11.001.

Keith, T. Z., Reynolds, M. R., Roberts, L. G., Winter, A. L., & Austin, C. A. (2011). Sex differences in latent cognitive abilities ages 5 to 17: Evidence from the Differential

Ability Scales—Second Edition. *Intelligence*, *39*(5), 389-404. doi: 10.1016/j.intell.2011.06.008.

Keith, T. Z., Quirk, K. J., Schartzer, C., & Elliott, C. D. (1999). Construct bias in the Differential Ability Scales? Confirmatory and hierarchical factor structure across three ethnic groups. *Journal of Psychoeducational Assessment*, *17*(3), 249-268. doi: 10.1177/073428299901700305.

Kewal Ramani A., Gilbertson L., Fox M. A., Provasnik S. (2007). U.S. Department of Education, Institute of Education Sciences, National Center for Educational Statistics (2007). *Status and trends in the education of racial and ethnic minorities* (NCES Publication No. 2007-039). Retrieved from: http://nces.ed.gov/pubs2007/minoritytrends/ind_4_17.asp

Kobrin, J. L., & Schmidt, A. E. (2005). The research behind the new SAT (College Board Research Summary RS-11). New York: College Entrance Examination Board.

Kubitschek, W. N., & Hallinan, M. T. (1996). Race, gender, and inequity in track assignments. *Research in Sociology of Education and Socialization*, *63*, 178-–93.

Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F. C. (2001). Construct validity of the WISC-III for White and Black students from the WISC-III standardization sample and for Black students referred for psychological evaluation. *School Psychology Review*, *30*(1), 70-88.

Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity? *Educational researcher*, *31*(1), 3–12. doi: 10.3102/0013189X031001003.

Lezak, M.D. (1988). IQ: RIP. *Journal of Clinical and Experimental Neuropsychology, 10,* 351-361.

Lichtenberger, E.O., Broadbooks, D.Y., & Kaufman, A.S. (2000). *Essentials of cognitive assessment with KAIT and other Kaufman measures.* New York, NY: Wiley.

Llorente, E., & Sheingold, D. (2010, September 15). Minorities now nearly half of Bergen County children, and just over half in the state. *Passaic County News.*

Lockheed, M. E., Thorpe, M., Brooks-Gunn, J., Casserly, P., & McAloon, A. (1985). *Understanding sex/ethnic related differences in mathematics, science, and computer science for students in grades four to eight.* Princeton, NJ: Educational Testing Service.

Luria, A. R. (1979). *The making of mind: A personal account of Soviet psychology.* Cambridge, MA: Harvard University Press.

Manly, J., Heaton, R. K., & Taylor, M. (2000, January). The effects of demographic variables and the development of demographically adjusted norms for the WAIS-III and WMS-III. In D. S. Tusky & D. Saklofske (Chairs), *The clinical interpretation of the WAIS-II and WMS-II: New research findings.* Symposium presented at the meeting of he American Psychological Association, Washington, DC.

Mascolo, J. T., Flanagan, D. P., & Alfonso, V. C. (2014). The term intervention is one that is familiar to anyone working in a school. *Essentials of Planning, Selecting, and Tailoring Interventions for Unique Learners*, 3.

Mather, N., Lynch, K., & Richards, A.M., (2001). The thinking blocks: language, images and strategies. In S. Godlstein & N. Mather (Eds.) Learning disabilities and challenging behaviors. A guide to intervention and classroom management. Baltimore, MD: Brookes.

Mattern, K. D., & Patterson, B. F. (2013). Test of slope and intercept bias in college admissions: A response to Aguinis, Culpepper, and Pierce (2010). *Journal of Applied Psychology, 98*(1)*,* 134–147. doi: 10.1037/a0030610.

Maxey, E. J., & Sawyer, R. (1981). Predictive validity of the ACT assessment for Afro-American/Black, Mexican-American/Chicano, and Caucasian-American/White students (ACT Research Bulletin 81-1). *Iowa City, IA: The American College Testing Program*.

McCarthy, D. (1972). *Manual for the McCarthy scales of children's abilities*. New York, NY: Psychological Corporation.

McCornack, R. L. (1983). Bias in the validity of predicted college grades in four ethnic minority groups. *Educational and Psychological Measurement*, *43*(2), 517–522. doi: 10.1177/001316448304300220.

McDermott, P. A., Fantuzzo, J. W., & Glutting, J. J. (1990). Just say no to subtest analysis: A critique on Wechsler theory and practice. *Journal of Psychoeducational Assessment*, *8*(3), 290–302. doi: 10.1177/073428299000800307.

McDermott, P. A., & Glutting, J. J. (1997). Informing stylistic learning behavior, disposition, and achievement through ability subtests: Or, more illusions of meaning? *School Psychology Review, 26*(2), 163–175.

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

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.

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., & Wendling, B.J. (2010). Cattell-Horn-Carroll cognitive-achievement relations: What have we learned from the past 20 years of research? *Psychology in the Schools, 47*(7), 651-675. doi: 10.1002/pits.20497

McKelpin, J.P. (1965). Some implications of the intellectual characteristics of freshman entering a liberal arts college. *Journal of Educational Measurement, 2*(2), 161–166. doi: 10.1111/j.1745-3984.1965.tb00411.x.

McRae, S.G (1981). *Simultaneous and sequential processing as related to reading and comprehension skills of second and third grade readers*. Unpublished master's thesis, National College of Education.

Meyers, L.S., Gamst, G., & Guarino, A.J. (2013). Applied multivariate research: Design and interpretation. Thousand Oaks, CA: SAGE Publications Inc.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*(4), 525–543.

Miele, (1979). Cultural bias in the WISC. *Intelligence, 3*(2), 149–164.

Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures*, *5*(4), 297-310.

Mullis, R. L., Rathge, R., & Mullis, A. K. (2003). Predictors of academic performance during early adolescence: A contextual view. *International Journal of Behavioral Development*, *27*(6), 541–548.

Naglieri, J.A. (1997). *The Naglieri nonverbal ability test.* San Antonia, TX: The Psychological Cooperation.

Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: Just how correlated are they? *Journal of Psychoeducational Assessment*, *21*(3), 244–260. doi: 10.1177/073428290302100302.

Naglieri, J. A. & Das, J. P. (1997). *Cognitive assessment system interpretive manual.* Itasca, IL: Riverside.

Naglieri, J.A. & Johnson, D. (2000). Effectiveness of a cognitive strategy intervention in improving arithmetic computation based on the PASS theory. *Journal of Learning Disabilities*, *33*(6), 591-597. doi: 10.1177/002221940003300607.

Naglieri, J. A., Rojahn, J., & Matto, H. C. (2007). Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence*, *35*(6), 568–579.

Naglieri, J. A., Rojahn, J., Matto, H. C., & Aquilino, S. A. (2005). Black-white differences in cognitive processing: A study of the planning, attention, simultaneous, and successive theory of intelligence. *Journal of Psychoeducational Assessment*, *23*(2), 146–160. doi:10.1177/073428290502300204.

Naglieri, J. A., & Ronning, M. E. (2000). Comparison of White, African American, Hispanic, and Asian children on the Naglieri Nonverbal Ability Test. *Psychological Assessment*, *12*(3), 328–334. doi: 10.1037/1040-3590.12.3.328.

Najarian, M., Snow, K., Lennon, J., & Kinsey, S. (2010). *Early childhood longitudinal study, birth cohort (ECLS-B), preschool–kindergarten 2007 psychometric report* (NCES 2010-009). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, US Department of Education.

Nichols, P.L. (1972). *The effects of heredity and environment on intelligence test performance in 4-and 7-year-old white and Negro sibling pairs (*Unpublished doctoral dissertation). University of Minnesota, MN.

Nijenhuis, T. J., & Van Der Flier, H. (2000). Differential prediction of immigrant versus majority group training performance using cognitive ability and personality measures. *International Journal of Selection and Assessment*, *8*(2), 54–60. doi: 10.1111/1468-2389.00133.

Oakland, T. (1983). Joint use of adaptive behavior and IQ to predict achievement. *Journal of Consulting and Clinical Psychology, 51*(2), 298–301. doi: 10.1037/0022-006X.51.2.298.

Pennock-Román, M. (1990). *Test validity and language background: A study of Hispanic American students at six universities*. New York, NY: College Entrance Examination Board.

Poteat, G. M., Wuensch, K. L., & Gregg, N. B. (1988). An investigation of differential prediction with the WISC-R. *Journal of School Psychology*, *26*(1), 59–68.

Pressley, M., & Woloshyn, V. (eds.) (1995). Cognitive strategy instruction that really improves children's academic performance. Cambridge, MA: Brookline Books.

Prifitera, A., & Saklofske, D. (1998). *WISC-III clinical use and interpretation: Science-practitioner perspectives.* San Diego, CA: Academic Press

Prifitera, A., Saklofske, D.H., & Weiss, L.G. (2005). *WISC-IV clinical use and interpretation.* Burlington, MA: Elsevier Academic Press.

Puente, A. E., & Salazar, G. D. (1998). Assessment of minority and culturally diverse children. In A. Prifitera and D. H. Saklofske (Eds.), *WISC–III clinical use and interpretation: Scientist–practitioner perspectives* (pp. 227–248). San Diego, CA: Academic Press.

Raiford, S.E., & Coalson, D.L. (2014). *Essentials of WPPSI-IV assessment.* Hoboken, NJ: Wiley.

Ramist, L., Lewis, C., & McCamley, L. (1994). *Student group differences in predicting college grades.* New York, NY: College Entrance Examination Board.

Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). *NAEP 2008 trends in academic progress (NCES 2009-479).* Washington, DC: National Center for Education Statistics, U. S. Department of Education.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology, pp. 321–334 in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, IV. Berkeley, California: University of California Press.

Raven, J. C. (1938). *Manual for progressive matrices.* HK Lewis, London: University of London.

Reschley, D. (1978). WISCR-R factor structures among Anglos, Blacks, Chicanos, and Native American Papagos. *Journal of Consulting and Clinical Psychology, 46*(3)*,* 417–422. doi: 10.1037/0022-006X.46.3.417.

Reschly, D. J., & Sabers, D. L. (1979). An examination of bias in predicting MAT scores from WISC-R scores for four ethnic-racial groups. *Journal of Educational Measurement*, *16*, 1–9.

Reynolds, C. R., Chastain, R. L., Kaufman, A. S., & McLean, J. E. (1988). Demographic characteristics and IQ among adults: Analysis of the WAIS□R standardization sample as a function of the stratification variables. *Journal of School Psychology, 25*(4)*,* 323–342.

Reynolds, M. R., Floyd, R. G., & Niileksela, C. R. (2013). How well is psychometric *g* indexed by global composites? Evidence from three popular intelligence tests. *Psychological Assessment*, *25*(4), 1314–1321. doi: 10.1037/a0034102.

Reynolds, C. R., & Kaiser, S. M. (1990). Bias in assessment of aptitude. In C. R. Reynolds and R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children* (pp. 611–653). New York, NY: Guilford.

Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment research. In C. R. Reynolds (Ed.), *Oxford handbook of psychological assessment of children and adolescents*. New York, NY: Oxford University.

Reynolds, M. R., Keith, T. Z., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence*, *36*(3), 236–260.

Reynolds, C. R., & Lowe, P.A. (2009). The problem of bias in psychological assessment. In T.B. Gutkin and C.R. Reynolds (Eds.), *The handbook of school psychology* (pp. 332–374). (4th ed.). Hoboken, NJ: Wiley.

Reynolds, M. R., **Scheiber, C.,** Hajovsky, D. B., Schwartz, B., & Kaufman, A. S. (in press). Gender differences in academic achievement: Is writing an exception to the gender similarities hypothesis? *Journal of Genetic Psychology.*

Rushton, J. P., Skuy, M., & Bons, T. A. (2004). Construct Validity of Raven's Advanced Progressive Matrices for African and Non□African Engineering Students in South Africa. *International Journal of Selection and Assessment*, *12*(3), 220-229.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, *56*(4), 302–318. doi: 10.1037/0003-066X.56.4.302.

Sackett, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in preemployment testing. *American Psychologist*, *49*(11), 929–954. doi: 10.1037/0003-066X.49.11.929.

Sandoval-Martinez, S. (1982). Findings from the Head Start bilingual curriculum development and evaluation effort. *NABE Journal*, *7*(1), 1–12. doi: 10.1080/08855072.1982.10668431.

Scheiber, C., Reynolds, M.R., Hajovski, D., & Kaufman, A.S. (2014). Evidence of a gender difference in writing in a large, nationally-representative sample of children and adolescents. *Psychology In The Schools*, *52*(4), 335-348.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, *36*(10), 1128–1137. doi: 10.1037/0003-066X.36.10.1128.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*(2), 262–274. doi: 10.1037/0033-2909.124.2.262.

Schneider, W. (2010). Metacognition and memory development in childhood and adolescents. In H.S. Waters & W. Schneider (Eds.), *Metacognition, strategy use, and instruction* (54-84). New York, NY: Guilford Press.

Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. *Contemporary intellectual assessment: Theories, tests, and*, (3rd), 99-144.

Sellers, A. H., Burns, W. J., & Guyrke, J. S. (1996). Prediction of premorbid intellectual functioning of young children using demographic information. *Applied Neuropsychology*, *3*(1), 21–27. doi: 10.1207/s15324826an0301_4.

Smith, D. (2008, August 7). Minority children become majority. *Washington Post.*

Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling: A Multidisciplinary Journal, 5*(4), 411-419. doi: 10.1080/10705519809540115

Sullivan, E. T., Clark, W. W., & Tiegs, E. W. (1963). *California test of mental maturity*. Monterey, CA: California Test Bureau.

Taylor, R. L., & Richards, S. B. (1991). Patterns of intellectual differences of Black, Hispanic, and White children. *Psychology in the Schools*, *28*(1), 5–9. doi: 10.1002/1520-6807(199101)28:1<5::AID-PITS2310280102>3.0.CO;2-6.

The Glossary of Educational Reform (2013). *Test bias.* Retrieved from: http://edglossary.org/test-bias/

The Nation's Report Card. (2009). *2013 mathematics and reading*. Retrieved from http://nationsreportcard.gov/reading_math_2013/#/

Thomas, P.J. (1972). *An investigation of possible test bias in the Navy basic test battery.* San Diego, CA: Naval Personnel and Training Research Laboratory.

Thomas, P.J. (1975). *Racial differences in the prediction of class "A" school grades (technical bulletin NPR-TR-75-39).* San Diego, CA: Navy personnel Research and Development Center.

Thorndike, E. L. (1911). Edward Lee Thorndike. *Animal intelligence*. New York, NY: Hafner.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet intelligence scale: Fourth edition (technical manual).* Chicago, IL: Riverside.

Tracey, T. J., & Sedlacek, W. E. (1984). Noncognitive variables in predicting academic success by race. *Measurement and Evaluation in Guidance, 16*(4), 171–178.

Tracey, T. J., & Sedlacek, W. E. (1985). The relationship of noncognitive variables to academic success: A longitudinal comparison by race. *Journal of College Student Personnel, 26,* 405–410.

Trundt, K. M. (2013). *Construct bias in the differential ability scales, (DAS-II): a comparison among African American, Asian, Hispanic, and White ethnic groups* (Published doctoral dissertation). University of Texas, Austin, Tx.

Tulsky, D. S., Saklofske, D. H., Chelune, G. J., Heaton, R. K., Ivnik, R. J., Bornstein, R., et al. (2003). *Clinical interpretation of the WAIS–III and WMS–III.* San Diego, CA: Academic Press.

Urbina, S. (2014). *Essentials of psychological testing* (2nd ed.). Hoboken, NJ: Wiley.

U.S. Census Bureau (2008). *An older and more diverse nation by midcentury*. Retrieved from http://www.census.gov/newsroom/releases/archives/population/cb08-123.html

U.S. Census Bureau. (2009). *Current population survey.* Retrieved from: http://www.census.gov/cps/

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics U.S. Department of Education. (2006). *Digest of education statistics: 2012.* Retrieved from: http://nces.ed.gov/programs/digest/d12/

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2007a). *National assessment of educational progress. The nation's report*

card: Reading 2007. Retrieved from http://nces.ed.gov/nationsreportcard/pubs/main2007/2007496.asp.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2007b). *The nation's report card: Writing 2007. National assessment of educational progress at grades 8 and 12.* Retrieved from http://nces.ed.gov/nationsreportcard/pdf/main2007/2008468.pdf

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics U.S. Department of Education. (2008). *The condition of education 2008.* Retrieved from: http://nces.ed.gov/pubs2008/2008031.pdf

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics U.S. Department of Education. (2010). *Fast facts. SAT scores.* Retrieved from: https://nces.ed.gov/fastfacts/display.asp?id=171.

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2011a). *The nation's report card: Writing 2011.* Retrieved from: http://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2011b). *Digest of education statistics*. Retrieved from: http://nces.ed.gov/programs/digest/d11/tables/dt11_126.asp

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics (2013). *National assessment of educational progress: Mathematics assessment.* Retrieved from http://nces.ed.gov/nationsreportcard/mathematics/

U.S. Employment Service. (1945). *The general aptitude test battery.* Berkley, CA: U.S. Employment Service (USES), Division of Testing.

Valencia, R. R., Rankin, R. J., & Livingston, R. (1995). K-ABC content bias: Comparisons between Mexican American and White children. *Psychology in the Schools*, *32*(3), 153–169. doi: 10.1002/1520-6807(199507)32:3<153::AID-PITS2310320302>3.0.CO;2-G.

Valencia, R.R., & Suzuki, L.A. (2011). *Intelligence testing and minority students.* Thousand Oaks, CA: Sage.

Vanneman, A., Hamilton, L., Baldwin Anderson, J., & Rahman, T. (2009). *Achievement gaps: How Black and White students in public schools perform in mathematics and reading on the National Assessment of Educational Progress* (NCES 2009–455). Washington, DC: U.S. Department of Education, National Center for Education Statistics, Institute of Education Sciences.

Vukovich, D., & Figueroa, R. A. (1982). *The validation of the system of multicultural pluralistic assessment: 1980-1982.* Unpublished manuscript, University of California at Davis, Department of Education, Davis, CA.

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

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

Wechsler, D. (2001). *Individual Achievement Test–Second Edition (WIAT-II).* San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *WISC-IV: Administration and scoring manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2012). The Wechsler Preschool and Primary Scale of Intelligence – 4th edition. Psych Cooperation. Bloomington, MN

Wechsler, D. (2014). Wechsler Intelligence Scale for Children – 5th edition. PsychCooperation. Bloomington, MN.

Weiss, L. G., & Prifitera, A. (1995). An evaluation of differential prediction of WIAT achievement scores from WISC-III FSIQ across ethnic and gender groups. *Journal of School Psychology, 33*(4), 297–304.

Weiss, L.G., Prifitera, A., Roid, G. (1993). The WISC-III and the fairness of predicting achievement across ethnic and gender groups. *Journal of Psychoeducational Assessment, 33*(4), 397–304.

Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (2006). *WISC-IV advanced clinical interpretation*. Burlington, MA: Academic Press.

Wendling, B.J., & Mather, N. (2009). *Essentials of evidence-based academic interventions.* Hoboken, NJ: John Wiley & Sons, Inc.

Williams, R. L. (1971). Abuses and misuses in testing Black children. *The Counseling Psychologist, 2*(3), 62−73. doi: 10.1177/001100007100200314.

Wilson, K. M. (1970). *Contribution of SAT's to prediction of freshman grades at CRC-member colleges (women).* Poughkeepsie, NY: College Research Center.

Wolf, T.H. (1973). *Alfred Binet.* Chicago, IL: University of Chicago Press.

Wonderlic, E. F. (1945). *Manual for the Wonderlic Personnel Test.* Northfield, IL: Wonderlic and Associates, INC.

Woodcock, R. W., & Johnson, M. B. (1989). *WJ-R tests of cognitive ability.* Itasca, IL: Riverside Publishing Company.

Woodcock, R., McGrew, K., & Mather, N. (2001). *Woodcock–Johnson III.* Itasca, IL: Riverside Publishing Company.

Young, J. W. (1994). Differential prediction of college grades by gender and by ethnicity: A replication study. *Educational and Psychological Measurement, 54*(4), 1022–1029. doi: 10.1177/0013164494054004019.