Psychological Assessment
2016, Vol. 28, No. 8, 942–952

1040-3590/16/$12.00 http://dx.doi.org/10.1037/pas0000236

Do the Kaufman Tests of Cognitive Ability and Academic Achievement Display Construct Bias Across a Representative Sample of Black, Hispanic, and Caucasian School-Age Children in Grades 1 Through 12?

Caroline Scheiber
Alliant International University

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 2 popular tests of intelligence and achievement, known to appeal to an ethnically diverse client population. The present study investigated test bias in terms of the test scores' construct validity across a nationally representative sample of Caucasian, Black, and Hispanic (N = 2,001) children in Grades 1–12. Confirmatory factor analysis was used to assess whether increasing sets of equality constraints fit the test scores' underlying theoretical model equally well for all 3 ethnic groups. Results showed that factorial invariance of the factor structure, based on 7 Cattell-Horn-Carroll (CHC) model broad abilities, was met for all 3 groups. Outcomes 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 minority group children and adolescents. 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.

Keywords: construct invariance, ethnicity, bias, Kaufman Test of Education Achievement–2nd Edition, Kaufman Assessment Battery for Children–2nd Edition

Supplemental materials: http://dx.doi.org/10.1037/pas0000236.supp

The population in the United States has become more ethnically diverse than previous generations (U.S. Census Bureau, 2009). 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 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).

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, Saklofske, Holdnack, & Prifitera, 2015). For example, on the full scale IQ (FSIQ) on the WISC-V, Caucasians average a standard score of 103.5, Hispanics average 94.4, and Blacks average 91.9. Asians outperform all three groups and average 108.6. Of particular interest in this study was the score gap between Caucasian students and Hispanic and Black students, which often forms the basis for arguments that test scores are biased against minority groups students (Weiss et al., 2015). Some researchers argue that the differences in scores across ethnic groups have narrowed significantly in recent years (e.g., Ceci & Kanaya, 2010; Dickens & Flynn, 2006); however, others disagree (Rushton, 2012). Regardless, the disparities still endure and the implications of IQ differences are substantial, as test scores often determine eligibility for special services and programs, and school admission (Weiss et al., 2006). 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 (Naglieri, Rojahn, & Matto, 2007;

This article was published Online First October 26, 2015.
I would like to express my gratitude to Alan S. Kaufman as well as Constance J. Dalenberg—two fabulous mentors and wonderful persons. Without them, the completion of this project would not have been possible. I also gratefully acknowledge Pearson for providing the data set used in this study.
Correspondence concerning this article should be addressed to Caroline Scheiber, California School of Professional Psychology, Alliant International University, 10455 Pomerado Road, San Diego, CA 92131. E-mail: cscheiber@alliant.edu

Najarian, Snow, Lennon, Kinsey, & Mulligan, 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.

The primary purpose of this present study was to explore test score bias, in the form of differential construct validity, of two individually administered, reliable, and well-normed measures of intelligence and achievement, using confirmatory factor analysis (CFA). Specifically, the Kaufman Assessment Battery for Children–2nd Edition (KABC-II; Kaufman & Kaufman, 2004a) and the comprehensive form of the Kaufman Test of Educational Achievement–2nd Edition (KTEA-II; Kaufman & Kaufman, 2004b) were used to assess bias of the test scores of Caucasian, Black, and Hispanic students in Grades 1 through 12. 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.

According to Messick's (1995) Unified Theory of Construct Validity, construct validity encompasses an integrated, multilayered framework, consisting of six components: (a) consequential, (b) content, (c), substantive, (d) structural, (e) external, and (f) generalizability. The present study on differential construct validity assesses construct validity across all six components: This study tests *content* and *structural* validity of the test scores by assessing whether the test items measure the constructs of interest (intelligence and achievement) accurately and by assessing whether the test items correlate in the same way across the three ethnic groups. Furthermore, the present study is based on CHC theory and, therefore, assesses the *substantive* validity of the test scores, as results will show whether the theoretical foundation of the test scores (based on CHC theory) is accurate. Furthermore, *generalizability* of results is assessed, as CHC theory underlies basically every clinical test of intelligence and achievement and results will, thus, generalize to other popular tests, including the Wechsler scales and Woodcock Johnson tests. I am also testing the *external* validity of the test scores, as the intelligence test scores are evaluated alongside the achievement test scores, which serve as an external criterion. And, finally, potential consequences associated with a lack of construct validity found are discussed throughout the paper (*consequential*).

Both the KABC-II and the KTEA-II test scores have demonstrated good convergent validity with other well-known measures of intelligence and achievement, including the Wechsler Intelligence Scale for Children–4th Edition (WISC-IV; Wechsler, 2003), the Wechsler Individual Achievement Test–2nd Edition (WIAT-II; Wechsler, 2001), and the Woodcock-Johnson–3rd Edition (WJ III; Woodcock, McGrew, & Mather, 2001; Kaufman & Kaufman, 2004a, Chapter 8; Kaufman & Kaufman 2004b, Chapter 7). Furthermore, independent researchers (Floyd, Reynolds, Farmer, Kranzler, & Volpe, 2013; Reynolds, Floyd, & Niileksela, 2013) have found that the KABC-II scores measure the general intelligence factor (*g*) in the same way as do the test scores of other major tests of cognitive ability, namely the WISC-IV, WJ III, and Differential Ability Scales–2nd Edition (DAS-II; Elliott, 2007). The *g* 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 (Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012). In that sense, findings of the present study do not serve the purpose of providing evidence for the validity only of the test scores 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 the scores of frequently used clinical tests, are applicable not only to Caucasian students, but also to Hispanic and Black students.

Construct Invariance

Definition of Test Bias

A test is biased "if a test design, or the way results are interpreted . . . , systematically disadvantages certain groups . . . over others. . . . [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).

The most common way of determining whether test scores are biased is to compare mean scores across different groups. The idea is that if there are statistically significant differences found in mean scores, the test scores are biased against the group that scores, on average, lower. However, the simple comparison of mean scores is not a statistically sound way of determining whether a test's psychometric properties are flawed and the test scores are, therefore, biased against a specific group. A more sophisticated and statistically appropriate way of determining bias is through the assessment of construct invariance. From a psychometric point of view, test scores might be biased in terms of their construct if the factor structure is *not* invariant across groups. That is to say, if the test scores of the assessment tool measure different constructs for one group as compared with another group, such a finding would conceivably constitute bias of the instrument (Reynolds & Keith, 2013).

Up to this point, only a few studies have examined differential construct validity for cognitive and achievement test scores across various ethnic groups. The following section summarizes the most important research studies that have explored construct validity of cognitive and achievement test scores across ethnic groups.

Studies of Cognitive and Achievement Tests

The majority of results obtained from a variety of cognitive and achievement test scores revealed similar factor structures for Black, Caucasian, and Hispanic school-age children and adults. This finding of similar factor structures characterizes the Woodcock-Johnson, the Wechsler, and the Differential Ability test scores (DAS; Elliott, 1990; e.g., Edwards & Oakland, 2006; Jensen, 1980; Kaufman & Di Cuio, 1975; Kaufman, Kaufman, & McLean, 1995; Kaufman, McLean, & Reynolds, 1991; Keith, Quirk, Schartzer, & Elliott, 1999; Miele, 1979; Nichols, 1972). 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 et al., 1999; Kush et al., 2001; Trundt, 2013).

Using simple correlation techniques, several studies revealed factor invariance for school-aged Black and Caucasian children on individually administered cognitive test scores; typically, verbal and nonverbal factor scores identified for Caucasians resembled the verbal and nonverbal factor scores identified for Blacks (Kaufman & Wang, 1992; Miele, 1979; Nichols, 1972). 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 factor scores for Black and Caucasian sixth through eighth graders. With the exception of Wicherts and Dolan (2010), who found noninvariance of 7-year-old Moroccan and Turkish children on the scores on the Revised Amsterdam Child Intelligence Test scores (RACIT; Bleichrodt, Drenth, Zaal, & Resing, 1984), other researchers were able to establish factorial invariance of individually administered test scores of cognition using CFA methods (e.g., Keith et al., 1999; Kush et al., 2001; Trundt, 2013).

For example, using CFA methods, Kush et al. (2001) found factorial invariance for *g* as well as for verbal and performance factor scores on the WISC–III for Black and Caucasian school-age children. Additionally, Keith, Quirk, Schartzer, and Elliott (1999) established construct invariance for the DAS scores; and Trundt (2013), in her dissertation research, found construct invariance for the DAS-II scores across Black, Hispanic, and Caucasian subgroups. Results from group-administered test scores also generally found factor invariance for school-aged Black and Caucasian children across the verbal factor, nonverbal factor, and *g* factor scores in national (Jensen, 1977; Jensen, 1980) and international (Rushton, Skuy, & Bons, 2004) samples.

Further, factor invariance of verbal, nonverbal, and *g* factor scores was established for a sample of adults (Kaufman & Wang, 1992) and for samples of preschoolers using individually administered cognitive test scores (Gutkin & Reynolds, 1981; Kaufman & Hollenbeck, 1974). However, some studies revealed significantly different factor structures for Black and Caucasian school-age children; Jensen (1980) was not able to establish factor invariance for fifth 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 of the scores for Black and Caucasian preschoolers tested on the McCarthy Scales of Children's Abilities (Kaufman & DiCuio, 1975; McCarthy, 1972).

Few studies have investigated factor invariance across Black–Caucasian and Hispanic–Caucasian school-age children using *achievement* test scores. Using correlation coefficients, Nichols (1972) established measurement invariance for Caucasian and Black school-age children for the scores of an individually administered achievement test, and another study, using group-administered achievement measures, found factor invariance of test scores in a sample of Hispanic and Caucasian school-age children (Hennessy & Merrifield, 1976). No studies established invariance for achievement test scores, using sophisticated methodology, such as CFA.

Summary

In sum, the studies that have explored measurement invariance for cognitive and achievement test scores have tended to find similar factor structures for Caucasians, Blacks, and Hispanics. Nonetheless, some studies did not establish factor invariance for Black individuals for the test scores of 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 was 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), even though, as pointed out by several researchers (e.g., Lubke, Dolan, Kelderman, & Mellenbergh, 2003), there are numerous statistical advantages associated with applying multi-Group CFA. The advantages of CFA, relative to simple methodologies such as regression analysis or analysis of variance, are especially true when investigating construct invariance of test scores across different ethnic groups. Additional research using CFA on differential construct validity by ethnicity is needed.

Present Study

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 study explored construct invariance of the scores of two individually administered, contemporary instruments of cognition and achievement using state-of-the-art methodology. Not many studies have studied construct invariance across various ethnic groups. With the exception of the Keith et al. (1999), Trundt (2013), Kush et al. (2001), 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. The present study filled this gap in the literature.

The following research question was addressed to capture the larger question of whether the Kaufman test scores are biased by ethnicity in terms of their construct (Blacks, Hispanics, Caucasians):

Using CFA, is the factor structure of the Kaufman test scores invariant for separate groups of Blacks, Hispanics, and Caucasians in Grades 1–12, using the CHC-based factor model developed by Kaufman, Reynolds, Liu, Kaufman, and McGrew (2012) as the criterion?

Method

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. Supplementary Table 1 provides more details regarding the demographic breakdown of the sample.

Measures

The KABC-II. The KABC-II (Kaufman & Kaufman, 2004a) is an individually administered test of intelligence designed for ages 3–18. 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 can also be interpreted from a Luria's neuropsychological model), the KABC-II produces two global scores, the Fluid-Crystallized Index (FCI) and the Nonverbal Index (NVI), as well as five CHC broad ability scores (short-term memory [STM], visual processing, long-term storage and retrieval, crystallized ability, and fluid reasoning). 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 test scores. For the global scaled scores, coefficients ranged from the low to high .90s and for the index level scores coefficients ranged from the high .80s to the low .90s. Test–retest reliabilities for children and adolescents for the global scores 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 for the scores ranged from .77 to .92 (Kaufman & Kaufman, 2004a, Table 8.3).

Validity. CFA was employed to confirm the factor structure of the KABC-II scores (Kaufman & Kaufman, 2004a, Chapter 8). The final model, that examined the construct validity of the core subtest scores, had excellent fit for all age levels (CFI = .997–.999; RMSEA = .025–.055; Kaufman & Kaufman, 2004a, Figures 8.1 and 8.2). Core subtests were identified as those subtests with the highest loadings on their appropriate factors.

In addition to the construct validity of the KABC-II scores, as demonstrated by CFA, the KABC-II scores have been shown to correlate well with other scores of measures of intelligence (e.g., producing correlation coefficients ranging from .66 to .89 with the WISC-IV global and scaled scores). In addition, as mentioned previously, the KABC-II scores have 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 DAS-II (Elliott, 2007; Floyd et al., 2013; Reynolds et al., 2013). For more information regarding the KABC-II, please consult the manual (Kaufman & Kaufman, 2004a).

The KTEA-II. The 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 (eight core and six supplementary subtests) and is normed both on Grades 1–12 and ages 4.5 through 25 years. In this present study all eight core achievement subtests in addition to Nonsense Word Decoding (because of its important in the reading domain) were used. The KTEA-II scores also produce the comprehensive achievement composite.

Reliability. The alternate-form reliabilities (as measured by Form A and Form B) were substantial for the test score interpretations, ranging from the low .80s to the low .90s for the reading, math, and writing domains. Oral language scores produced correlations ranging from the high .60s to the low .90s. The comprehensive achievement composite scores produced alternative-form reliabilities in the mid .90s (Kaufman & Kaufman, 2004b, Table 7.5). Internal-consistency reliability (split-half) coefficients ranged from the mid .70s to the high .90s for both the test scores on both forms A and B (Kaufman & Kaufman, 2004b, Table 7.1).

Validity. For Grade 1 through age 25, CFA was used to verify the factor structure of the KTEA-II scores. The final model had good statistical fit (CFI = .992, RMSEA = .062; Kaufman & Kaufman, 2004b, Figure 7.1). In addition, the comprehensive achievement composite scores correlated substantially with global achievement scores on other individually administered achievement batteries: The KABC-II scores correlated in the low .60s to the low .90s with the corresponding WIAT-II (Kaufman & Kaufman, 2004b, Tables 7.17 and 7.18) and the WJ III scores (Kaufman & Kaufman, 2004b, Tables 7.19 and 7.20). For further detail regarding psychometric properties of the KTEA-II please refer to the manual (Kaufman & Kaufman, 2004b).

Statistical Analysis

Factorial invariance using MG-MACS. In order to explore whether the joint factor structure of the KABC-II and KTEA-II scores is the same for Caucasians, Hispanics, and Blacks, factorial invariance was assessed. That is, the variable structure of the KABC-II and KTEA-II scores, as outlined by Kaufman et al. (2012) for the total sample, was explored to determine whether it is invariant across the three ethnic groups. The 22 subtests used comprised the variables entered into the CFA: (a) the 15 KABC-II subtest scores are organized according to the CHC abilities measure: *G*c, *G*f, *G*sm, *G*lr, *G*v; (b) the seven KTEA-II subtest scores measure *Grw* (reading and written language) and *Gq* (math). The factorial invariance of the first-order factor scores 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). The invariance of the second-order *g*-factor, which is hypothesized to underlie all of the subtests that compose the KABC-II and KTEA-II, was also explored. Kaufman et al. (2012) examined the relationship between the separate *g* factors that underlie the KABC-II (COG-*g*) and KTEA-II (ACH-*g*) scores 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). In that sense, Kaufman et al. (2012) provided evidence that: (a) some of the subtests on the achievement test are best classified as measures of cognitive ability, and (b) the *g* underlying cognitive test scores and achievement test scores is essentially the same *g* across a representative sample of school-age children in K–12. Based on these findings, in this present study, the measurement invariance analysis was conducted jointly for the KTEA-II and the KABC-II test scores. The KABC-II and KTEA-II test scores together produce a factor structure—composed of seven CHC-based factors and an underlying *g*—that is valid both from a theoretical CHC perspective (e.g., Flanagan, Ortiz, Alfonso, & Dynda, 2014) and a research perspective (Kaufman et al., 2012). Other researchers have also demonstrated that the g underlying academic achievement and cognitive measures are essentially the same (e.g., Frey & Detterman, 2004; Koenig, Frey, & Detterman, 2008).

Analytical steps. In order to control for the effects of age, age standardized scores were used to conduct this analysis. Even though it is preferred to use raw scores when conducting CFA analysis in AMOS (Cudeck, 1989), it is common practice and, in fact, recommended to use age-corrected scores when it comes to the use of clinical test scores with variables that have quite different variances (Cudeck, 1989). It has been accepted methodological practice to use the CFA procedure with age-corrected standard scores (rather than mean raw scores) when analyzing data from clinical tests of intelligence (e.g., Floyd et al., 2013; Kaufman et al., 2012; Keith, 1999; Keith et al., 1999; Reynolds, Keith, Fine, Fisher, & Low, 2007; Reynolds, Keith, Ridley, & Patel, 2008; Reynolds et al., 2013).

Multigroup confirmatory factor analysis (MG-CFA) served as an excellent tool to explore factorial invariance of test scores between groups (Reynolds & Keith, 2013). Specifically, MG-CFA based on a mean and covariance structure (MG-MACS) approach was used. Using MG-MACS, it was specified whether the factor loadings as well as the intercepts of the factor structure of the KABC-II and KTEA-II test scores were equivalent for Caucasian, Black, and Hispanic children and adolescents in Grades 1–12. This method was used to explore whether the same construct was being measured across different groups. 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) discussed 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).

In *configural* invariance, the same factor structure was applied to all ethnic groups. For all three groups, the factors and patterns of free and fixed loadings were estimated equally. Whereas factor variances and covariances were allowed to vary freely, the reference indicator (for each first- and second-order factor) for the groups' factor loadings was fixed to 1. This approach balanced the factors and scales them properly. The latent factor means were fixed to 0 and the observed subtest means (intercepts) could vary freely.

After configural invariance was established, *metric* invariance (weak factorial invariance) was assessed. In this step, first-order factor loadings were restricted to be equal across groups. The previously free factor loadings were now also restrained so that all corresponding factor loadings were equal across the groups. This assured that a one-unit increase in a specific factor for Caucasians resulted in the same unit increase for Blacks and Hispanics. If these added constraints did *not* result in a degradation of model fit, it was established that the relation between subtests and factors was the same for the ethnic groups. That is, the unit of measurement was equal across groups (Reynolds & Keith, 2013). However, in order to determine whether the same CHC abilities and the same *g* on the KTEA-II and KABC-II were measured for Caucasians, Blacks, and Hispanics, further investigations of the intercepts were required.

The next step included the establishment of *intercept invariance* (strong factorial invariance). In addition to all previous constraints, all corresponding subtest intercepts (means) were restrained to be equal. The factor means were allowed to vary freely across groups. This specification ensured that the same constructs were being measured across all three groups. That is, this step allowed for a specific score to mean the same for one group as it would for another group. By constricting the intercepts and by allowing the factor mean scores to vary freely, one could estimate whether the same CHC factor abilities—and *g*—were being measured for Caucasians, Hispanics, and Blacks. Once these mean score constraints were added and the fit index did not degrade, it was concluded that the test scores measured the same underlying construct across all three ethnic groups (Reynolds & Keith, 2013).

Alternatively, if strong factorial invariance was *not* established either for the latent CHC factors or *g*, and the groups' intercepts on particular subtests were not equal, then this was an example of bias, because different constructs were being measured; in such a case, observed scores on the tests would not be based on ability, but would be the result of psychometric test score bias in terms of a lack of construct validity (Reynolds & Keith, 2013).

Residual invariance. Once strong factorial invariance (intercept invariance) was established, there was an option to investigate *residual invariance* (strict factorial invariance). Strict factorial invariance refers to the equality constraints of the residual variances of the residuals (error and specific variances). However, several researchers (e.g., Byrne, 2010; Reynolds & Keith, 2013) have suggested that residual invariance is NOT a necessary prerequisite in order to establish construct invariance, and is, therefore, not a necessary prerequisite to establish NON-bias. Therefore, in this study, residual invariance was not evaluated.

Fit indexes. The possible degradation of model fit with increasingly restrictive sets of restraints was determined by the likelihood ratio test (χ^2), the root mean square error of approximation (RMSEA), and comparative fit index (CFI). A "good" fit would typically result in a nonsignificant $\Delta\chi^2$, a RMSEA value close to .05 or less (or between .05 and .08), and a CFI value of at least .95 (Reynolds & Keith, 2013). The likelihood test ratio ($\Delta\chi^2$), ΔCFI, and ΔRMSEA were used to compare the goodness of fit for tests of factorial invariance (Cheung & Rensvold, 2002). Several researchers (Cheung & Rensvold, 2002) recommend that for ΔCFI a change $>$.01 would be considered significant and for ΔRMSEA a change $>$.02 would be 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 were 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 study, ΔCFI (with ΔCFI $>$.01 considered significant change) and the ΔRMSEA (with ΔRMSEA $>$.02 considered significant change) were used to evaluate the degradation in model fit. The fit index RMSEA needed to be corrected, as recommended by Steiger (1998). That is to say, RMSEA had to be multiplied by its square root of three (because of multiple groups).

Results

Descriptives

Means and standard deviations for the KABC-II scales and KTEA-II CHC factor indexes are reported in Table 1. The means and standard deviations (*SD*) by ethnicity have previously been reported in the KABC-II and KTEA-II manuals (Kaufman & Kaufman, 2004a, 2004b). Assumptions of normality, interval or

Table 1
Means and Standard Deviations (SD) for Each CHC-Factor and KABC-II and KTEA-II Subtest

	Caucasians (N = 1,313)		Blacks (N = 312)		Hispanics (N = 376)	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
KABC-II indexes						
Short-term memory/Gsm	102.0	14.3	99.6	16.3	93.9	15.4
Visualization/Gv	102.3	14.9	92.9	13.9	97.7	13.7
Fluid-intelligence/Gf	102.2	14.9	94.6	13.8	96.9	14.6
Learning/Glr	102.3	15.0	98.1	14.1	95.6	14.9
Crystallized intelligence/Gc	103.9	13.8	93.3	14.1	91.9	14.1
Fluid-crystallized index	103.2	14.4	94.1	13.6	93.5	14.2
KTEA-II indexes						
Reading	102.1	14.6	95.3	15.0	94.2	14.3
Written expression	102.1	14.5	95.2	14.9	95.2	14.0
Math	102.5	14.4	94.6	13.8	95.4	13.9
Oral language	103.7	14.3	95.2	14.1	93.7	13.9
Comprehensive achievement composite	102.8	14.5	93.8	14.3	93.8	13.6

Note. KABC-II = Kaufman Assessment Battery for Children–2nd Edition; KTEA-II = Kaufman Test of Educational Assessment–2nd Edition; CHC = Cattell-Horn-Carroll model. Data copyright Pearson. All rights reserved.

ration level of measurement, independence, and linearity underlie CFA. For the total sample (n = 2,001) 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). Skeweness and kurtosis was also evaluated for Black and Hispanic data points separately and was normal. Based on skewness and kurtosis data, data points for the total sample and for each ethnic group were normally distributed.

Furthermore, all data are derived from raw scores on subtests that were normalized and then standardized to have a M = 10 and SD = 3 (KABC-II) or a M = 100 and SD = 15 (KTEA-II) and, therefore, fall at an interval level of measurement (Kaufman & Kaufman, 2004a, 2004b). Independence was met due to the stratified random sampling procedures used to select the participants (see Kaufman & Kaufman, 2004a, 2004b). And, finally, 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 = 2,001) as well as for each individual ethnic group, Caucasians (n = 1,313), Blacks (n = 312), and Hispanics (n = 376).

Frequency distributions were examined to discover missing data. Three subtests had to be eliminated due to too many missing cases (on the KTEA-II: Associational Fluency both Semantic and Phonological; on the KABC-II: Atlantis Delayed, Rebus Delayed, and Gestalt Closure). 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 one missing case. The two missing cases were handled using Hotdeg imputation (Myers, 2011). 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 index scores 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- to 9.11-year-olds (Kaufman & Kaufman, 2004b; Table D-2, p.190).

Creation of the model for the total sample and the three ethnic groups. Before invariance (or noninvariance) 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 data to try to improve it. For this study, the original model was based on the CHC-based factor structure developed by Kaufman et al. (2012), with two main changes: (a) based on the findings by Kaufman et al. (2012) in this study, a single *g* factor was hypothesized to underlie all cognitive and achievement ability scores; and (b) all cross-loadings and error correlations identified by Kaufman et al. (2012) were removed.

Table 2 shows the model fit for the present sample. 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. Based on those fit indexes, the model fit for the original model as outlined by Kaufman et al. (2012) demonstrated inadequate fit for the total sample: the values of CFI ranged from only .91–.93 and the values of RMSEA all exceeded .06.

The steps that were followed to create the final model are outlined next (the data from these intermediary steps do not appear in the tables). Based on modification indexes as well as theory and research findings, *G*c 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 = .94; RMSEA = .06). Reading comprehension and written expression are heavily *G*c loaded as both subtests measure the ability to understand ideas and have a knowledge base. It was, therefore, decided to maintain these cross-loadings, as did Kaufman et al. (2012).

Subsequently, the KABC-II hand movements subtest was cross-loaded with *G*f, as suggested by the modification indexes. Theory and research state that hand movements can either cross-load on *G*f (Kaufman et al., 2012) or *G*v (Kaufman & Kamphaus, 1984). Cross-loadings between *G*f and hand movements improved the model fit considerably (CFI = .95; RMSEA = .06) 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 = .95; RMSEA = .05). Theoretically, correlating these two error terms made 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. As shown in Table 2, the final model showed

Table 2
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 = 1,313; Blacks n = 312; Hispanics n = 376)

Form and model	χ^2	*df*	*p*	CFI	RMSEA	RMSEA 90% CI
Original model						
Total sample	2082.2	202	<.001	.93	.068	[.066, .071]
Caucasians	1553.6	202	<.001	.91	.071	[.068, .075]
Blacks	501.2	202	<.001	.93	.069	[.061, .077]
Hispanics	514.3	202	<.001	.93	.064	[.057, .071]
Data and theory driven model						
Total sample	1397.9	198	<.001	.95	.055	[.052, .058]
Caucasians	1060.5	198	<.001	.94	.058	[.054, .061]
Blacks	412.6	198	<.001	.95	.059	[.051, .067]
Hispanics	437.3	198	<.001	.95	.057	[.050, .064]

Note. CFI = Comparative Fit Index; RMSEA = Root Mean Square Error of Approximation. Data copyright Pearson. All rights reserved.

good fit for the total sample with CFI values around .95 (ranging from .94 to .95) and RMSEA values around <.06. Figure 1 shows the final proposed seven-factor CHC structure of the Kaufman Intelligence and Achievement test scores.

Invariance analysis. Model fit for the multigroup comparisons are shown in Table 3. As the table demonstrates, the configural model fit well (CFI = .95; RMSEA = .057). As expected, the model fit did degrade when first order loadings were constrained to be equal, $\chi^2(630) = 3.115$, $p < .05$. But, because only trivial changes in CFI (<.01) and no changes in RMSEA occurred it was concluded that metric invariance for the first order factor model was established. As expected, model change did also degrade when the second factor loadings were added, $\chi^2(642) = 3.096$, $p < .05$. However, no changes in CFI were detected and only minimal changes in RMSEA values were detected (<.02). 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 $\Delta\chi^2$, $\chi^2(672) = 3.31$, $p < .001$. However, ΔCFI was nonexistent and ΔRMSEA was negligible. Finally, the second order intercept constraints were added, which again resulted in a significant degradation according to $\Delta\chi^2$, $\chi^2(684) = 3.64$, $p < .001$. However, due to the trivial ΔCFI (<.01) and the minimal ΔRMSEA (<.02), it was concluded that intercept invariance for both the first order factor loadings and second-order factor loadings were established.

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 χ^2 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 and *g* 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 nonexistent or trivial.

Discussion

Aims of the Study

Based on the analyses, construct invariance of the KABC-II and KTEA-II test score structures were established for Caucasian, Hispanic, and Black school-age children. This study established factorial invariance of comprehensive intelligence and achievement test scores 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 ability scores, by ethnicity, in addition to the *g* factor, and the *only* study that established construct invariance of achievement factor scores. Of the handful of researchers, who have investigated construct invariance using sophisticated methodology, some researchers established factorial invariance of the *g* factor (e.g., Edwards & Oakland. 2006). Keith et al. (1999) established factorial invariance of *G*f, *G*v, and *G*c factor scores 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 factor scores for Hispanics, Blacks, and Caucasians (Trundt also included Asians). Additionally, Kush and colleagues (2001) established invariance for the four-factor structure of the WISC–III scores in a nonrepresentative sample of Black and Caucasian students. However, the factor scores 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 of scores across ethnic groups on measures of academic achievement has been understudied. Thus, CHC theory recognizes that achievement ability constitutes an important part of human intelligence and findings of this study provide evidence that *G*q and *G*rw factor scores—two broad achievement factors—are invariant across ethnic groups.

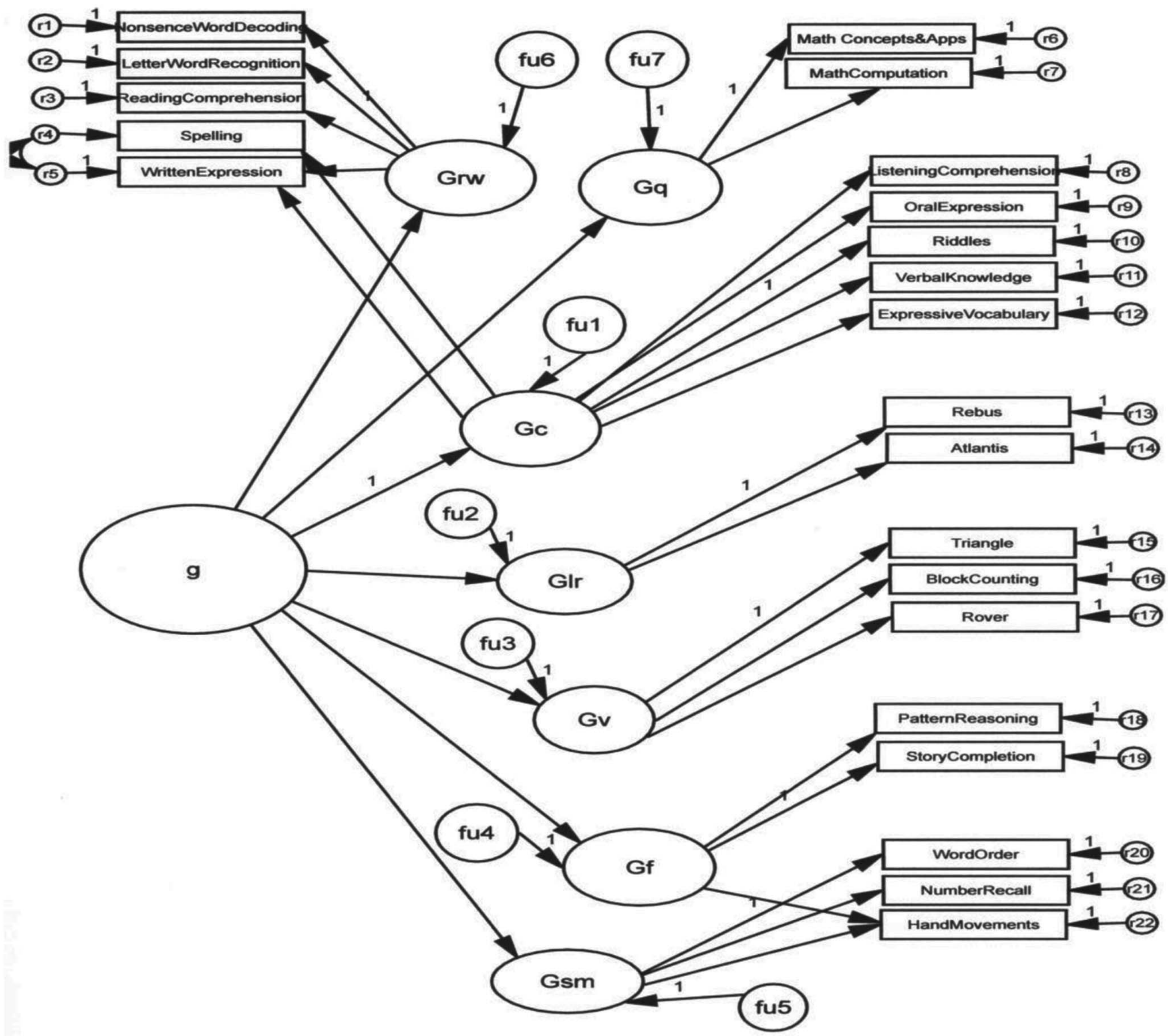

Figure 1. Proposed 7-Factor Cattell-Horn-Carroll Structure of the Kaufman Intelligence and Achievement tests. Data copyright Pearson. All rights reserved.

Clinical Implications

The following section includes the key findings of this study as pertaining to the most important implications for clinical psychologists, neuropsychologists, and others who assess ethnic minority children's cognitive ability and achievement. All across the country, clinicians and practitioners administer well-known, clinical tests of intelligence and achievement to children from ethnic minority groups and assessments are conducted based on the assumption that test scores 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. Results from this study demonstrate that when using the Kaufman intelligence and achievement test scores 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 Kaufman test scores, in fact, measure what they are supposed to measure for each child, 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 test scores, but generalize to scores of other popular tests of cognition and achievement. For example, the study conducted by Kaufman et al. (2012) demonstrated that the *g* measured by the KABC-II and KTEA-II scores is essentially the same *g* that is measured by the WJ III scores. Similarly, Reynolds et al. (2013) and Floyd et al.

Table 3
Model Fit Indexes and Nested Comparisons for Confirmatory Factor Analysis (CFA) Models: Caucasian, Hispanic, and Black Comparison (N = 2,001) (Caucasians n = 1,313; Blacks n = 312; Hispanics n = 376)

Form and model	χ^2	*df*	$\Delta\chi^2$	Δdf	*p*	CFI	ΔCFI	RMSEA	ΔRMSEA	RMSEA 90% CI
Configural invariance	1910.4	594				.95		.057		[.055, .060]
Metric invariance										
Measurement	1962.5	630	52.1	36	.040	.94	.001	.057	.000	[.054, .059]
Structural	1987.7	642	25.2	12	.014	.94	.000	.055	.002	[.054, .059]
Intercept invariance										
Measurement	2223.8	672	236.1	30	<.001	.94	.000	.058	.003	[.055, .062]
Structural	2491.9	684	268.1	12	<.001	.93	.001	.062	.004	[.060, .065]

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

(2013) demonstrated that the same *g* underlies the KABC-II, the WISC-IV, the WJ III, and the DAS-II scores. Such findings provide strong evidence for the fact that the same global construct that is being measured by the KABC-II scores is also measured by the WISC-IV, the DAS-II, and the WJ III scores. Any results pertaining to the Kaufman test scores are therefore likely to be generalizable to those other test scores. Such findings are even more important considering that the most recent versions of the Wechsler tests 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 factor scores, except *Gs* factor scores, were validated as invariant across ethnic groups in this present study.

Theoretical Implications

In addition to providing evidence necessary to allow for the continuous valid clinical use of the Kaufman test scores 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; Trundt, 2013). Results of this study provide the necessary empirical evidence that supports the validity of CHC theory for Hispanic and Black children and adolescents. 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 in addition to investigating g (e.g., Kaufman & Kaufman, 1983, 2004a, 2004b; Weiss et al., 2006). The inclusion of achievement variables (*Grw* and *Gq* in CHC terminology) is an especially important contribution of this study.

Findings have important implications also for other popular tests of intelligence and achievement as most frequently used clinical 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.

Study Limitations

Even though present findings provide support for the lack of construct bias against Black and Hispanic school-age children on the KABC-II and KTEA-II scores, results need to be understood in the context of the study's limitations. Perhaps most importantly, in this present study, factorial invariance analyses were not conducted on the subtest level. However, in order to allow comparisons of mean scores on the different subtests, a differential item functioning analysis needs to be conducted. The focus of this present study was the invariance analysis of the latent factor variables, but future research should investigate the question of invariance at the subtest and item level.

Furthermore, in terms of the study's measures and methodology, the KABC-II and KTEA-II were limited in that the tests did not include *Gs* factor scores, which is a factor commonly used in other tests of cognition (Flanagan et al., 2014). Furthermore, the lack of power prevented me from studying possible developmental differences when examining construct invariance. Future research needs to replicate current findings with children at different age groups. It is also important to take into consideration that the evidence of nonbias found in this study might not necessarily be generalizable to ethnic groups other than Hispanics, Blacks, and Caucasians. 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. Furthermore, although 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 exactly representative of the current U.S. census. 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 validity on the KABC-II and KTEA-II test scores across a representative sample of Caucasian, Black, and Hispanic school-age 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 scores on the two tests measure the same construct across different ethnic groups. 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 nonbiased. However, it is important to remember that even if the KABC-II and the KTEA-II scores, as well as other scores of cognitive or achievement tests, are not biased in terms of their construct 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.

References

Arbuckle, J. L. (1995–2011). *AMOS 20.0 user's guide*. Crawfordville, FL: AMOS Development Corporation.

Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1984). *Revisie Amsterdamse kinder intelligentie test* [Revised Amsterdam Child Intelligence Test]. Lisse, The Netherlands: Swets & Zeitlinger.

Byrne, B. A. (2010). *Structural equation modeling with amos. Basic concepts, applications, and programming* (2nd edition). New York, NY: Taylor & Francis Group.

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

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12,* 471–492.

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

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327. http://dx.doi.org/10.1037/0033-2909.105.2.317

Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science, 17,* 913–920. http://dx.doi.org/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,* 358–366. http://dx.doi.org/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 ed. (DAS-II)*. San Antonio, TX: The Psychological Corporation.

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* (pp. 239–285). New York, NY: Oxford University Press.

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,* 383–401.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the Scholastic Assessment Test and general cognitive ability. *Psychological Science, 15,* 373–378. http://dx.doi.org/10.1111/j.0956-7976.2004.00687.x

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,* 227–231. http://dx.doi.org/10.1037/0022-0663.73.2.227

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. http://dx.doi.org/10.1037/0022-0663.68.6.754

Jensen, A. R. (1977). An examination of cultural bias in the Wonderlic Personnel Test. *Intelligence, 1,* 51–64. http://dx.doi.org/10.1016/0160-2896(77)90026-5

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.

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,* 10–18. http://dx.doi.org/10.1016/0022-4405(75)90032-1

Kaufman, A. S., & Hollenbeck, G. P. (1974). Comparative structure of the WPPSI for Blacks and Whites. *Journal of Clinical Psychology, 30,* 316–319. http://dx.doi.org/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 1/2 through 12 1/2 years. *Journal of Educational Psychology, 76,* 623–637. http://dx.doi.org/10.1037/0022-0663.76.4.623

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,* 365–376. http://dx.doi.org/10.1177/0013164495055003001

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* (2nd ed.). Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Kaufman test of educational achievement–2nd ed. (KTEA-II)*. Circle Pines, MN: American Guidance Service.

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, 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,* 219–229. http://dx.doi.org/10.1177/073428299201000302

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,* 123–138. http://dx.doi.org/10.1016/j.intell.2012.01.009

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

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,* 249–268. http://dx.doi.org/10.1177/073428299901700305

Koenig, K. A., Frey, M. C., & Detterman, D. K. (2008). ACT and general cognitive ability. *Intelligence, 36,* 153–160. http://dx.doi.org/10.1016/j.intell.2007.03.005

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,* 70–88.

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*.

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within-and between-group differences and measurement invariance in the common factor model. *Intelligence, 31,* 543–566. http://dx.doi.org/10.1016/S0160-2896(03)00051-5

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

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543. http://dx.doi.org/10.1007/BF02294825

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749. http://dx.doi.org/10.1037/0003-066X.50.9.741

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

Miele. (1979). Cultural bias in the WISC. *Intelligence, 3,* 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,* 297–310. http://dx.doi.org/10.1080/19312458.2011.624490

Naglieri, J. A., Rojahn, J., & Matto, H. C. (2007). Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence, 35,* 568–579. http://dx.doi.org/10.1016/j.intell.2006.11.001

Najarian, M., Snow, K., Lennon, J., Kinsey, S., & Mulligan, G. (2010). *Early childhood longitudinal study, birth cohort (ECLS-B). Preschool-kindergarten 2007 psychometric report*. Washington, DC: U. S. 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.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. New York, NY: W. W. Norton & Company Inc.

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

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

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,* 1314–1321. http://dx.doi.org/10.1037/a0034102

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* (pp. 48–85). New York, NY: Oxford University.

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. A. (2007). Confirmatory factor structure of the Kaufman assessment battery for children: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly, 22,* 511–539. http://dx.doi.org/10.1037/1045-3830.22.4.511

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,* 236–260. http://dx.doi.org/10.1016/j.intell.2007.06.003

Rushton, J. P. (2012). No narrowing in mean Black–White IQ differences—Predicted by heritable g. *American Psychologist, 67,* 500–501. http://dx.doi.org/10.1037/a0029614

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,* 220–229. http://dx.doi.org/10.1111/j.0965-075X.2004.00276.x

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

Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling, 5,* 411–419. http://dx.doi.org/10.1080/10705519809540115

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.

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

Wechsler, D. (2001). *Individual achievement test–2nd ed. (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 ed.). Bloomington, MN: Psych Cooperation.

Wechsler, D. (2014). *Wechsler intelligence scale for children* (5th ed.). Bloomington, MN: PsychCooperation.

Weiss, L. G., Saklofske, D. H., Holdnack, J. A., & Prifitera, A. (2016). *WISC-V assessment and interpretation*. Waltham, MA: Academic Press.

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

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice, 29,* 39–47. http://dx.doi.org/10.1111/j.1745-3992.2010.00182.x

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

Received March 20, 2015
Revision received August 17, 2015
Accepted August 31, 2015 ■