

# Which of the Three KABC-II Global Scores is the Least Biased?

Caroline Scheiber<sup>1</sup> · Alan S. Kaufman<sup>2</sup>

Received: 2 April 2015 / Revised: 1 August 2015 / Accepted: 4 August 2015 / Published online: 25 August 2015  
© American Academy of Pediatric Neuropsychology 2015

**Abstract** In order to create fairer measures for the assessment of ethnic minority group children, the test authors of the Kaufman Assessment Battery for Children–Second Edition (KABC-II) created three different global indexes: the comprehensive Fluid-Crystallized Index (FCI); the Mental Processing Index (MPI), which excludes subtests that assess crystallized knowledge; and the Nonverbal Index (NVI), which excludes any subtest that requires verbal expression. The test authors encourage clinicians to use the MPI for children with cultural differences and the NVI for children with language differences. The implication is that the FCI is more biased than the MPI and NVI for assessing children from ethnic minorities. However, there has been no empirical evidence to support that hypothesis. Using structural equation modeling, we investigated the predictive validity of the three KABC-II global scores in predicting reading, writing, and math in a nationally representative sample of Caucasian, African-American, and Hispanic school-aged children across three grade groups (1–4,  $N=724$ ; 5–8,  $N=743$ ; and 9–12,  $N=534$ ). Contrary to the test authors' predictions, the FCI emerged as the fairest predictor of achievement across all age and grade groups. Alternatively, the MPI and NVI showed a persistent intercept overprediction of African-American and Hispanic students' achievements across grades 5–8. The results provide evidence that more comprehensive global scaled scores might, in fact, be fairer predictor variables than less culturally and

linguistically influenced indexes when predicting the achievement of children and adolescents from ethnic minorities.

**Keywords** Kaufman Assessment Battery for Children–Second Edition (KABC-II) · Intercept · Overprediction · Bias · Ethnicity

## Introduction

There is a strong link between intelligence and academic achievement. Historically, 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 and 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). In short, intellectual ability (the ability to learn) is tightly linked to achievement (what has been successfully learned), as intellectual ability helps the individual to obtain knowledge and thereby to learn and achieve.

Given this strong conceptual relationship between intelligence and achievement, IQ tests are often used to predict academic achievement outcomes (Naglieri and Bornstein 2003). In fact, scores on IQ tests frequently determine access or denial to special programs in school as well as college and employment eligibility, based on the notion that performance

---

✉ Caroline Scheiber  
cscheiber@alliant.edu

<sup>1</sup> California School of Professional Psychology, Alliant International University, San Diego, CA, USA

<sup>2</sup> Yale University School of Medicine, New Haven, CT, USA

on IQ tests predicts future performance in school or employment (Weiss et al. 2006). A disproportionate overrepresentation of children from ethnic minorities in special education classes, and underrepresentation in gifted programs, has researchers and neuropsychologists concerned whether IQ tests are biased against minority groups in that they do not accurately predict minority groups' achievement (U.S. Department of Education et al. 2006). According to Urbina (2014), prediction is biased when (a) the correlation coefficient between the predictor variable and the outcome varies for different groups in terms of its magnitude and (b) when the test scores consistently overpredict or underpredict the outcome of an individual depending on his or her group membership in terms of the slope or intercept. It is important to note that an overprediction of the minority groups' achievement in slope or intercept 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. An underprediction in the slope or intercept, on the other hand, would penalize the minority group and, therefore, denote bias.

The present study explored prediction bias of a popular test of intelligence in a representative sample of Caucasian, Hispanic, and African-American school-aged children across three different grade groups (grades 1–4, 5–8, and 9–12). The Kaufman Assessment Battery for Children–Second Edition (KABC-II; Kaufman and Kaufman 2004a) was used to assess whether the academic achievement domains of reading, writing, and math, as measured by the Kaufman Test of Educational Achievement–Second Edition (KTEA-II; Kaufman and Kaufman 2004b), are predicted equally well across different ethnic minority groups. Specifically, it was of interest to compare prediction bias results of the Fluid-Crystallized Index (FCI) with the Nonverbal Index (NVI) and the Mental Processing Index (MPI). All three summary scores measure global ability and are representative of the general intelligence factor (*g*) (Kaufman and Kaufman 2004a, pp. 27 and 45). The KABC-II for school-aged children comprises five scales that are founded in a dual theoretical model (Cattell-Horn-Carroll or CHC theory, and Luria's neuropsychological processing theory). The names of these scales reflect their roots in both theories: Sequential/*Gsm*, Simultaneous/*Gv*, Planning/*Gf*, Learning/*Glr*, and Knowledge/*Gc*. The abbreviations in the scale names correspond to CHC abilities, as follows: *Gsm* (short-term memory), *Gv* (visual processing), *Gf* (fluid reasoning), *Glr* (long-term storage and retrieval), and *Gc* (crystallized knowledge). Whereas the FCI offers a global measure that includes all five scales, the MPI includes only four of the scales; it excludes Knowledge/*Gc*, which includes tasks that require verbal concepts, verbal reasoning, and cultural knowledge. Knowledge/*Gc* includes subtests that are traditionally considered the most culturally loaded and are, therefore, believed to produce the

most bias against minority group children with cultural differences (Flanagan et al. 2013). The NVI takes the notion of eliminating language even further and includes only subtests that can be communicated via gestures without any need for verbal expression on the part of the child. Nonverbal ability measures have traditionally been used with individuals who have hearing difficulties or individuals with language differences (e.g., those who learned English as their second language). It is for those reasons that the test authors encouraged clinicians to use the MPI and NVI with minority groups' students; however, no empirical evidence has been provided in the test manual or in the literature to support the claim that the MPI and NVI are fairer (less biased) than the FCI as predictors of academic achievement for children from ethnic minorities.

Similarly to the Kaufman test authors, many other clinicians and neuropsychologists encourage the usage of linguistically and culturally more neutral measures in the assessment of minority group children (Flanagan et al. 2013; Weiss et al. 2015). The practice is based on the notion that ethnic minority groups are disadvantaged when their cognitive performance is, at least partially, based on tests that emphasize language and cultural knowledge. The debate stems from the idea that the verbal and linguist parts of an intelligence test are primarily based on the majority group's (Caucasians') cultural and linguistic norms and do not take into consideration the minority groups' ethnic or linguistic background. Even though this hypothesis makes theoretical and rational sense, there has not been compelling empirical evidence to support the accuracy of this hypothesis. That is the goal of this study.

Using structural equation modeling, this present study filled this gap in the literature and assessed differential predictive validity of the FCI, MPI, and NVI of the KABC-II to explore whether the culturally more neutral indexes—the NVI and MPI—are, in fact, less biased at predicting academic achievement for African-American and Hispanic school-aged children than the more comprehensive and, thus, more culturally and linguistically loaded FCI. In accordance with the test authors, it is hypothesized that the MPI and NVI would be fairer predictors than FCI of minority groups' achievement. Findings are of importance because the range of abilities measured by nonverbal or so-called culture fair tests or summary scores are generally narrower by virtue of their exclusion or limitation of tasks that measure language skills and acquired knowledge. This limited range of abilities becomes especially problematic when the referral for evaluation is based on problems in language. In fact, the majority of referrals are based on language difficulties and the exact problem cannot easily (if at all) be measured with tests that do not assess crystallized intelligence abilities (Flanagan et al. 2013). The limitations of using culture fair or language free tests (tests that do not include measures of crystallized knowledge) would only be worth accepting, if those tests are, in fact, fairer

predictors of minority groups' achievement outcomes. However, if they are not, clinicians and neuropsychologists might have reason to assess children from ethnic minorities using a more comprehensive measure of intelligence such as the FCI.

### Prediction Bias Across Ethnic Groups

Studies of prediction bias are relatively rare in the IQ literature. There have been some studies, mostly in the 1990s, that investigated differential prediction bias across different ethnic groups for the Wechsler Full-Scale IQ (FSIQ) and the General Ability Index (GAI) of the Woodcock-Johnson test (Edwards and Oakland 2006; Keith 1999; Weiss and Prifitera 1996; Weiss et al. 1993). For example, using structural equation modeling, Keith (1999) established prediction invariance of the Woodcock-Johnson-Revised (WJ-R; Woodcock and Johnson 1989) GAI and six narrow CHC abilities (*Gc*, *Gs*, *Gf*, *Gc*, *Ga*, and *Gsm*) for a nationally representative sample of Hispanics, African-Americans, and Caucasians across three different grade groups (1–4, 5–8, and 9–12). Similarly, Weiss et al. (1993) and Weiss and Prifitera (1995) explored the fairness of the Wechsler Intelligence Scale for Children–Third Edition (WISC-III) FSIQ when predicting achievement outcomes for samples of Hispanic, Caucasian, and African-American children and adolescents ages 6–16 years old, using structural equation modeling. The researchers found that the FSIQ predicted reading, math, and writing equally well for all three groups in terms of (a) the magnitude of the correlation and (b) its slope and intercept. Similarly, using regression analyses, Edwards and Oakland (2006) found that the GAI of the Woodcock-Johnson III predicted reading, writing, and math achievement equally well across a representative sample of African-American and Caucasian school-aged children. Two other studies by Naglieri and colleagues (2005, 2007) used the full-scale score on the Cognitive Assessment System (CAS; Naglieri and Das 1997). The CAS does not require the retrieval of facts and knowledge or vocabulary and has, therefore, been thought of as culturally more neutral. The authors used simple correlation techniques and found that, overall, the correlations were equally strong for African-American and Caucasian school-aged children. Another study found that the Naglieri Nonverbal Ability Test (NNAT; Naglieri 1997) predicted achievement equally well for representative samples of Hispanic and Caucasian school-aged children (Naglieri and Ronning 2000).

In sum, not many studies have explored differential prediction bias of global intelligence scores across different ethnic groups using individually administered tests of cognition. The few studies that have explored this question found that achievement was predicted equally well for

Caucasian, African-American, and Hispanic children. However, the studies are few and generally old. With the exception of Keith (1999), Weiss et al. (1993), and Weiss and Prifitera (1995), all other studies used simple regression models to assess prediction bias, instead of using more sophisticated structural equation modeling techniques. Perhaps most importantly, none of these studies compared comprehensive global summary scores (such as FSIQ or FCI) with nonverbal or less culturally loaded global scaled scores. In that sense, the question has not been answered whether global scaled scores that limit language and cultural knowledge, such as KABC-II MPI and NVI, are less biased than comprehensive global scores. The present study attempted to fill this gap in the literature.

Another important contribution of this study is that it explored a possible developmental trend in the question of prediction bias, as it separated the sample into three different grade groups (grades 1–4, 5–8, and 9–12). Only Keith (1999) tested the effects of the general and specific intelligence factors to explore prediction invariance across three different age groups; he found no developmental differences on the WJ-R. Finally, it is important to note that present findings are generalizable beyond the Kaufman tests to other popular tests of cognition. This is because independent researchers (Floyd et al. 2013; Reynolds et al. 2013) have found that the KABC-II measures the general intelligence factor (*g*), as represented in the global scores, in the same way as do other major tests of cognitive ability, namely the Wechsler Intelligence Scale for Children–Fourth Edition (WISC-IV; Wechsler 2003), the Woodcock-Johnson–Third Edition (WJ III; Woodcock et al. 2001), and the Differential Ability Scales–Second Edition (DAS-II; Elliott 2007).

### Present Study

The purpose of this study was to test for differential prediction bias of the three KABC-II global scores—FCI, MPI, and NVI (all of which are representative of *g*). Analyses were conducted separately for each of the global scores. Ethnic group differences in the prediction of reading, math, and writing on the KTEA-II, using structural equation modeling, were examined. By comparing the results of the separate analyses, it was investigated (a) whether the MPI and NVI—the two culturally and linguistically more neutral global scaled scores—are, in fact, less biased and more accurate at predicting achievement for ethnic minority groups as compared to the FCI; (b) whether the slope or intercept of the prediction models was biased against one or more group(s), as would be indicated by a degradation in model fit; and (c) if the possible prediction bias was a function of *chronological age*.

## Method

### Participants

The participants ( $N=2001$  with grade range=1–12) were from the conorming sample of the KTEA-II (Kaufman and Kaufman 2004b; see pp. 110–111, Tables 7.24 and 7.25) Comprehensive Form and the KABC-II (Kaufman and Kaufman 2004a). The sample matched 2001 US Census data in terms of ethnicity, gender, age, geographic region, and parental education. Parental education within each ethnic group also closely matched the US population. The total sample used for this study included 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 African-Americans (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.

The sample was divided up into three different grade groups for sensitivity analyses: grades 1–4 ( $n=724$ ), grades 5–8 ( $n=743$ ), and grades 9–12 ( $n=534$ ).

### Measures

#### *KTEA-II Comprehensive Form*

The KTEA-II is an individually administered test of academic achievement for children and adolescents ages 4.5–25 years. The test yields several achievement composites. For this study, three composites were used: math (math computation/math concepts and applications), reading (letter and word recognition/reading comprehension), and written language (written expression/spelling) (Kaufman and Kaufman 2004b).

The KTEA-II consists of two alternate forms (forms A and B). A total of 221 children were administered both forms. Alternative form reliability ranged from the low 0.80s to the mid-0.90s (Kaufman and Kaufman 2004b, Table 7.5). Split-half reliability on the KTEA-II Composites ranged from the high 0.80s to the high 0.90s for the three achievement domains used in this study (Kaufman and Kaufman 2004b, Table 7.1). To provide support for the organization of the KTEA-II subtests into their composites, CFA was employed. The final model had good statistical fit (Comparative Fit Index (CFI)=0.992, root-mean square error of approximation (RMSEA)=0.062) (Kaufman and Kaufman 2004b, Fig. 7.1). Additionally, the KTEA-II has demonstrated good convergent validity with other tests of achievement, including the Wechsler Individual Achievement Test—Second Edition (WIAT-II; Wechsler 2005) and the WJ III, ranging from the mid-0.70 to the low 0.80s (Kaufman and Kaufman 2004b,

Tables 7.17–7.20). Overall, the KTEA-II is a reliable test of achievement with strong psychometric properties.

#### *KABC-II*

The KABC-II (Kaufman and Kaufman 2004a) is an individually administered measure of intelligence for children and adolescents ages 3–18 years. The KABC-II consists of 18 subtests (including both core and supplementary subtests). From the CHC theory standpoint, the KABC-II produces a global score, the Fluid-Crystallized Index (FCI), that is composed of five scales (Sequential/*Gsm* (short-term memory), Simultaneous/*Gv* (visual processing), Learning/*Glr* (long-term storage and retrieval), Planning/*Gf* (fluid reasoning), and Knowledge/*Gc* (crystallized knowledge). From the standpoint of the Luria model, the KABC-II produces a global score that emphasizes mental processing, the Mental Processing Index (MPI), and only includes the first four of these scales; Knowledge/*Gc* is excluded. The KABC-II also generates a Nonverbal Index (NVI) to measure cognitive and processing abilities with minimal verbal involvement. The NVI consists of five subtests, and their instructions and responses can be communicated via gestures. At ages 7–18, the subtests are hand movements, triangles, pattern reasoning, story completion, and block counting. At age 6, conceptual thinking replaces block counting. All indexes have a mean of 100 and a standard deviation of 15. For further information about the KABC-II, consult Kaufman et al. (2005).

Internal-consistency reliability (split-half coefficients) on the KABC-II is high. Coefficients for the global scales coefficients were 0.97 (FCI), 0.95 (MPI), and 0.92 (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 0.80s to the low 0.90s. For the global scores (MPI, FCI, and NVI), test-retest reliabilities for children and adolescents ages 7–12 ( $n=82$ ) and 13–18 ( $n=61$ ) are high, ranging from 0.87 to 0.94 (Kaufman and Kaufman 2004a, Table 8.3). At the scale level, test-retest reliabilities ranged from 0.76 to 0.95 for ages 7–12 and 13–18. Test-retest intervals occurred over a 4-week interval. To confirm the factor structure of the KABC-II, confirmatory factor analysis (CFI) was used (Kaufman and Kaufman 2004a, chapter 8). The final model for the core subtests had excellent fit for all age levels (CFI=0.997–0.999; RMSEA=0.025–0.055) (Kaufman and Kaufman 2004a, Figs. 8.1 and 8.2). The KABC-II has also demonstrated good convergent validity with other tests of intelligence, including the WISC-IV and WJ III. Global scores (FCI, NVI, and MPI) correlated in the low to high 0.80s with the global scores of WISC FSIQ. 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 (Floyd et al. 2013; Reynolds et al. 2013).

## Statistical Procedure

Multi-group path models (structural equation modeling) were used to explore whether the KABC-II global scales scores (FCI, MPI, and NVI) predict the KTEA-II achievement domains (reading, writing, and math) equally well for the three ethnic groups. Of specific interest was to measure the predictive validity of the global cognitive scales and compare whether the less culturally and linguistically loaded scores (MPI and NVI) predicted the achievement composites for Hispanics and African-Americans more accurately than the FCI. The predictive validity of the global scores was evaluated separately, and the results were subsequently compared. In order to detect any possible developmental trend, the sample was divided up into three grade groups (grades 1–4, 5–8, and 9–12). All analyses were conducted using AMOS 20 (Arbuckle 1995–2011). If the regression lines ( $Y=a+bX$ ) for any pair of variables differed across the groups, it was concluded that there was bias in the prediction. That is to say, if the slope  $b$  or intercept  $a$  differed significantly across groups, the application of the same regression line to all groups resulted in an incorrect prediction of the criterion variable (Keith and Reynolds 2003).

## Analytical Steps

Three global cognitive scales, as measured by the KABC-II, were used to 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 was employed. Paths from each cognitive scale (e.g., FCI) to the corresponding achievement composite were created. In order to assess for prediction bias, a model fit method was employed for each pair (Caucasians versus African-Americans, Caucasians versus Hispanics, and Hispanics versus African-Americans). The model fit was 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 were first compared on a baseline model, without any constraints. That way, magnitudes of the coefficients can be compared. Next, the residual variances of the achievement composites were 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 were analyzed. If the slope restriction did not result in a degradation of model fit, slope invariance was established (*weak prediction invariance*). Finally, in addition to the slope constraints, 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 prediction invariance*).

The fit of the models were evaluated with  $\Delta\chi^2$ . RMSEA and CFI were also employed as alternative fit indexes. If the

slope and intercept restrictions did not result in a significant degradation of model fit (as evaluated by  $\Delta\chi^2$ ,  $\Delta$ RMSEA, and  $\Delta$ CFI), then, prediction non-bias was concluded and the same regression lines could be used across the three ethnic groups (Keith and Reynolds 2003). However, if the slope restriction resulted in a significant degradation of model fit, that indicated that there was an interaction between ethnicity and achievement outcome. If the intercept restrictions resulted in a significant degradation of model fit, a common regression line would overpredict for one group and underpredict for the other group (Keith and Reynolds 2003). When there was slope or intercept *non*-invariance, post hoc regression analyses were conducted to better understand the direction of the bias.

## Results

### Missing Data, Means, and Standard Deviations

Before analyses could be conducted, decisions had to be made regarding how to deal with missing data and outliers. The KTEA-II subtests that were used in this study to assess reading, writing, and math had no missing data. However, there were a few missing cases on the KABC-II subtests that compose the global scaled scores. Rover (used to assess both MPI and FCI) and story completion (used to assess MPI, FCI, and NVI) each had one missing case. The two missing cases were handled using hot deck imputation (Myers 2011). Rover was scaled equal to that child's scaled score on the triangle subtest (both are on the Simultaneous/ $Gv$  scale) and story completion was scaled equal to the child's pattern reasoning scaled score (both are on the Planning/ $Gf$  scale).

Tables 1 and 2 present the means and standard deviations for the three KABC-II predictor variables and the three KTEA-II outcome variables, by ethnic subsample, separately for the three grade groups. 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 and was, therefore, far from the suggested four-point cutoff (Meyers et al. 2013). Hence, there were no problems regarding homogeneity of the variance for the present samples. And 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.

## Correlations

Table 3 presents correlations of the three KABC-II cognitive global ability factors 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

**Table 1** Means and standard deviations for each ethnic group across grade groups for each KTEA-II outcome variable

|                  | Caucasians |                    | African-Americans |                    | 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              |

Samples sizes, grades 1–4: Caucasians (455), African-Americans (119), and Hispanics (150); grades 5–8: Caucasians (487), African-Americans (119), and Hispanics (137); and grades 9–12: Caucasians (371), African-Americans (74), and Hispanics (89)

KABC-II ability factors and the KTEA-II achievement outcome variables ranged between  $r=0.45$  and  $r=0.79$  for all ethnicities across all grade groups. Correlations between the FCI and the three KTEA-II achievement composites produced correlations ranging from  $r=0.55$  to  $r=0.79$ . The MPI correlated between  $r=0.50$  and  $r=0.73$  with the three achievement domains, and the NVI correlated between  $r=0.46$  and  $r=0.73$  with reading, writing, and math. Overall, general ability accounted for about 25–60 % of the achievement variance for the present samples.

### Prediction Invariance

This section examines the differential predictive validity of the three KABC-II global cognitive scales (FCI, MPI, and NVI) across 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 three 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 FCI, MPI, and NVI.

**Table 2** Means and standard deviations for each ethnic group across grade groups for each KABC-II predictor variable

|                          | Caucasians |                    | African-Americans |                    | Hispanics |                    |
|--------------------------|------------|--------------------|-------------------|--------------------|-----------|--------------------|
|                          | Mean       | Standard deviation | Mean              | Standard deviation | Mean      | Standard deviation |
| Grade 1–4                |            |                    |                   |                    |           |                    |
| Fluid-Crystallized Index | 102.71     | 14.93              | 95.80             | 13.34              | 94.01     | 15.42              |
| Mental Processing Index  | 102.12     | 14.95              | 96.80             | 13.75              | 95.28     | 15.68              |
| Nonverbal Index          | 101.86     | 14.877             | 94.46             | 13.24              | 97.07     | 15.04              |
| Grade 5–8                |            |                    |                   |                    |           |                    |
| Fluid-Crystallized Index | 103.43     | 14.31              | 93.64             | 13.91              | 92.71     | 13.33              |
| Mental Processing Index  | 102.77     | 14.83              | 94.66             | 13.82              | 93.94     | 13.34              |
| Nonverbal Index          | 102.72     | 15.03              | 93.32             | 13.23              | 95.67     | 13.53              |
| Grade 9–12               |            |                    |                   |                    |           |                    |
| Fluid-Crystallized Index | 103.73     | 13.94              | 92.33             | 13.54              | 93.74     | 13.44              |
| Mental Processing Index  | 103.29     | 14.02              | 92.39             | 13.49              | 94.39     | 13.89              |
| Nonverbal Index          | 103.71     | 14.81              | 90.68             | 13.94              | 95.84     | 12.77              |

Samples sizes, grades 1–4: Caucasians (455), African-Americans (119), and Hispanics (150); grades 5–8: Caucasians (487), African-Americans (119), and Hispanics (137); grades 9–12: Caucasians (371), African-Americans (74), and Hispanics (89)

**Table 3** Correlation between KABC-II predictors and the three KTEA-II achievement composites across age and ethnicity

| Predictor                | Math       |                   |           | Reading    |                   |           | Written language |                   |           |
|--------------------------|------------|-------------------|-----------|------------|-------------------|-----------|------------------|-------------------|-----------|
|                          | Caucasians | African-Americans | Hispanics | Caucasians | African-Americans | Hispanics | Caucasians       | African-Americans | Hispanics |
| <b>Grades 1–4</b>        |            |                   |           |            |                   |           |                  |                   |           |
| Fluid-Crystallized Index | 0.698      | 0.687             | 0.733     | 0.696      | 0.683             | 0.738     | 0.657            | 0.670             | 0.690     |
| Mental Processing Index  | 0.602      | 0.554             | 0.653     | 0.627      | 0.603             | 0.654     | 0.596            | 0.631             | 0.629     |
| Nonverbal Index          | 0.652      | 0.543             | 0.653     | 0.578      | 0.495             | 0.533     | 0.552            | 0.458             | 0.519     |
| <b>Grades 5–8</b>        |            |                   |           |            |                   |           |                  |                   |           |
| Fluid-Crystallized Index | 0.642      | 0.740             | 0.685     | 0.711      | 0.735             | 0.768     | 0.551            | 0.665             | 0.727     |
| Mental Processing Index  | 0.601      | 0.710             | 0.683     | 0.636      | 0.672             | 0.716     | 0.506            | 0.643             | 0.700     |
| Nonverbal Index          | 0.602      | 0.732             | 0.654     | 0.556      | 0.610             | 0.651     | 0.455            | 0.585             | 0.620     |
| <b>Grades 9–12</b>       |            |                   |           |            |                   |           |                  |                   |           |
| Fluid-Crystallized Index | 0.696      | 0.743             | 0.621     | 0.716      | 0.790             | 0.745     | 0.613            | 0.704             | 0.648     |
| Mental Processing Index  | 0.665      | 0.679             | 0.573     | 0.646      | 0.696             | 0.677     | 0.556            | 0.630             | 0.614     |
| Nonverbal Index          | 0.661      | 0.708             | 0.562     | 0.584      | 0.671             | 0.560     | 0.529            | 0.656             | 0.467     |

Samples sizes, grades 1–4: Caucasians (455), African-Americans (119), and Hispanics (150); grades 5–8: Caucasians (487), African-Americans (119), and Hispanics (137); grades 9–12: Caucasians (371), African-Americans (74), and Hispanics (89)

*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 and Keith 2013). 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).

*Slope Bias*

Alpha levels of 0.01 and 0.001 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 81 comparisons were completed (3 (grade groups) × 3 (ethnic groups) × 3 (predictor variables) × 3 (outcome variables)). No evidence of slope bias was found. 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 3). The correlation table shows that the coefficients between global intelligence factors and achievement composites are substantial for all three ethnicities across all grade groups. For example, the FCI correlated in the mid-0.60s to the mid-

0.70s with math across all three ethnicities across grades 1–12. Indeed, all correlation coefficients between the three KABC-II global ability factors and the three KTEA-II achievement outcome composites were moderate to high for all three ethnic groups.

*Intercept Differences*

Tables 4 and 5 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  $p < 0.01$  and  $p < 0.001$  to protect against multiple comparisons, results indicated that intercept differences were present such that a common regression line would overpredict performance on particular aspects of achievement for African-Americans and Hispanics and underpredict performance for Caucasians.

Table 4 shows the Caucasian-African-American comparisons. Using  $p < 0.01$ , 5/27 (18.5 %) produced significant intercept differences between African-Americans and Caucasians (and only 3/27 were significant at the  $p < 0.001$  level). Interestingly, FCI produced no intercept bias between Caucasians and African-Americans. In other words, FCI was the most accurate at predicting African-American’s achievement in math, reading, and writing. MPI and NVI, on the other hand, produced frequent intercept bias at grades 5–8 for all three achievement domains. The bias was so that MPI and NVI tended to overpredict achievement for African-

**Table 4** Significant intercept fit indexes and nested comparisons for confirmatory factor analysis (CFA) models for Caucasians and African-Americans across the grade groups

|                         | $\chi^2$ | <i>Df</i> | $\Delta\chi^2$ | $\Delta df$ | <i>P</i> | CFI   | $\Delta CFI$ | RMSEA |
|-------------------------|----------|-----------|----------------|-------------|----------|-------|--------------|-------|
| Grades 5–8              |          |           |                |             |          |       |              |       |
| Predictor variable: MPI |          |           |                |             |          |       |              |       |
| Math                    | 19.596   | 3         | 12.140         | 1           | 0.000**  | 0.982 | 0.005        | 0.096 |
| Reading                 | 10.787   | 3         | 7.143          | 1           | 0.008*   | 0.976 | 0.019        | 0.066 |
| Writing                 | 19.041   | 3         | 12.536         | 1           | 0.000**  | 0.922 | 0.056        | 0.094 |
| Predictor variable: NVI |          |           |                |             |          |       |              |       |
| Math                    | 15.594   | 2         | 9.034          | 1           | 0.003*   | 0.956 | 0.076        | 0.106 |
| Written language        | 16.940   | 3         | 10.711         | 1           | 0.001**  | 0.913 | 0.061        | 0.088 |

Samples sizes, grades 1–4: Caucasians (455) and African-Americans (119); grades 5–8: Caucasians (487) and African-Americans (119); and grades 9–12: Caucasians (371) and African-Americans (74)

\* $p < 0.01$ ; \*\* $p < 0.001$

American students and underpredict Caucasian students' achievement.

Table 5 presents the significant Caucasian-Hispanic comparisons, and the results mirror the results of the Caucasian-African-American analyses. 8/27 (29.6 %) comparisons were significant at  $p < 0.01$ , and 5/27 (18.5 %) were significant at the  $< 0.001$  level. Every one produced overprediction for the ethnic minority group (in this case Hispanics). Intercept bias was most prevalent at grades 5–8. Similar to the African-American-Caucasian comparison, with the exception of written language at grades 5–8 (only at the  $p < 0.01$  level), FCI did not produce any intercept bias for Hispanics. MPI and NVI, on the other hand, produced consistent overprediction for Hispanics at grades 5–8 (and once at grades 1–4 when NVI predicted reading for Hispanics).

The African-American-Hispanic comparisons are not shown in any tables because they produced no significant results at any grade level. Such results indicate that a common regression line can be used for both Hispanics and African-Americans when predicting achievement, as no strong evidence for intercept bias was found.

### Summary

As demonstrated in summary Tables 6, 7, and 8, overall, for all grade levels and for all ethnic groups, there was no evidence for slope bias. The magnitudes of the path from global 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

**Table 5** Significant intercept fit indexes and nested comparisons for confirmatory factor analysis (CFA) models for Caucasians and Hispanics across the grade groups

|                         | $\chi^2$ | <i>Df</i> | $\Delta\chi^2$ | $\Delta df$ | <i>P</i> | CFI   | $\Delta CFI$ | RMSEA |
|-------------------------|----------|-----------|----------------|-------------|----------|-------|--------------|-------|
| Grades 1–4              |          |           |                |             |          |       |              |       |
| Predictor variable: NVI |          |           |                |             |          |       |              |       |
| Reading                 | 10.338   | 3         | 9.826          | 1           | 0.002*   | 0.972 | 0.028        | 0.064 |
| Grades 5–8              |          |           |                |             |          |       |              |       |
| Predictor variable: NVI |          |           |                |             |          |       |              |       |
| Math                    | 19.064   | 3         | 16.043         | 1           | 0.000**  | 0.945 | 0.055        | 0.093 |
| Reading                 | 30.299   | 3         | 23.024         | 1           | 0.000**  | 0.892 | 0.087        | 0.121 |
| Written language        | 39.018   | 2         | 34.870         | 1           | 0.000**  | 0.791 | 0.191        | 0.173 |
| Predictor variable: MPI |          |           |                |             |          |       |              |       |
| Math                    | 14.466   | 3         | 8.712          | 1           | 0.003*   | 0.962 | 0.026        | 0.078 |
| Reading                 | 20.235   | 3         | 12.597         | 1           | 0.000**  | 0.950 | 0.034        | 0.096 |
| Written language        | 30.785   | 2         | 24.108         | 1           | 0.000**  | 0.877 | 0.089        | 0.152 |
| Predictor variable: FCI |          |           |                |             |          |       |              |       |
| Written language        | 15.883   | 2         | 11.584         | 1           | 0.001*   | 0.950 | 0.038        | 0.106 |

Samples sizes, grades 1–4: Caucasians (455) and Hispanics (150); grades 5–8: Caucasians (487) and Hispanics (137); and grades 9–12: Caucasians (371) and Hispanics (89)

\* $p < 0.01$ ; \*\* $p < 0.001$



**Table 6** Specificity of bias by predictor and achievement across age: Caucasians and African-Americans; slope bias and underpredicted achievement

| Predictor                      | Math                         | Reading                      | Written language             |
|--------------------------------|------------------------------|------------------------------|------------------------------|
| Grades 1 through 4             |                              |                              |                              |
| Fluid-Crystallized Index (FCI) | No bias                      | No bias                      | No bias                      |
| Mental Processing Index (MPI)  | No bias                      | No bias                      | No bias                      |
| Nonverbal Index (NVI)          | No bias                      | No bias                      | No bias                      |
| Grades 5 through 8             |                              |                              |                              |
| Fluid-Crystallized Index (FCI) | No bias                      | No bias                      | No bias                      |
| Mental Processing Index (MPI)  | Intercept: Caucasians (−0.8) | No bias                      | Intercept: Caucasians (−0.9) |
| Nonverbal Index (NVI)          | Intercept: Caucasians (−0.6) | Intercept: Caucasians (−0.5) | Intercept: Caucasians (−0.8) |
| Grades 9 through 12            |                              |                              |                              |
| Fluid-Crystallized Index (FCI) | No bias                      | No bias                      | No bias                      |
| Mental Processing Index (MPI)  | No bias                      | No bias                      | No bias                      |
| Nonverbal Index (NVI)          | No bias                      | No bias                      | No bias                      |

Numerical values represent the underpredicted achievement for Caucasians and African-Americans at the 0.01 alpha level

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. That conclusion is not supported in the analyses of intercepts.

The results of the Caucasian-African-American and Caucasian-Hispanic analyses did show evidence for intercept differences between ethnic minority groups and Caucasians (Tables 6, 7, and 8). The bias was such that a common regression line consistently overpredicted achievement for African-Americans and Hispanics and underpredicted achievement for Caucasians. Most importantly, it was NVI as well as MPI, which showed consistent overprediction for Hispanics and African-Americans, especially at grades 5–8. FCI, on the other hand, did not show any evidence for prediction bias, except in the Hispanic-Caucasian comparison when predicting

written expression at grades 5–8; FCI significantly overpredicted Hispanic achievement ( $p < 0.01$ ).

Whereas the underprediction for Caucasians is of small effect size (about one standard-score point, usually  $< 0.10$  SD), the amount of overprediction is moderate to large (2–5 points, typically  $> 0.3$  SD) for African-Americans and Hispanics (see Table 9, which summarizes the amount of overprediction for all significant intercepts). Overall, MPI and NVI produced the strongest evidence of overprediction for both African-Americans and Hispanics at grades 5–8. Such findings are opposite to common beliefs that nonverbal or less culturally loaded indexes, such as MPI and NVI, are fairer predictors of minority group’s achievement outcomes. Whereas the FCI includes all five ability scales, the MPI excludes Knowledge/*Gc* and NVI also excludes tasks that

**Table 7** Specificity of bias by predictor and achievement across age: Caucasians and Hispanics; slope bias and underpredicted achievement

| Predictor                      | Math                         | Reading                      | Written language             |
|--------------------------------|------------------------------|------------------------------|------------------------------|
| Grades 1 through 4             |                              |                              |                              |
| Fluid-Crystallized Index (FCI) | No bias                      | No bias                      | No bias                      |
| Mental Processing Index (MPI)  | No bias                      | No bias                      | No bias                      |
| Nonverbal Index (NVI)          | No bias                      | Intercept: Caucasians (−0.9) | No bias                      |
| Grades 5 through 8             |                              |                              |                              |
| Fluid-Crystallized Index (FCI) | No bias                      | No bias                      | Intercept: Caucasians (−1.0) |
| Mental processing index (MPI)  | Intercept: Caucasians (−0.6) | Intercept: Caucasians (−0.7) | Intercept: Caucasians (−1.1) |
| Nonverbal index (NVI)          | Intercept: Caucasians (−1.0) | Intercept: Caucasians (−1.2) | Intercept: Caucasians (−1.4) |
| Grades 9 through 12            |                              |                              |                              |
| Fluid-Crystallized Index (FCI) | No bias                      | No bias                      | No bias                      |
| Mental Processing Index (MPI)  | No bias                      | No bias                      | No bias                      |
| Nonverbal Index (NVI)          | No bias                      | No bias                      | No bias                      |

Numerical values represent the underpredicted achievement for Caucasians and Hispanics at the 0.01 alpha level

**Table 8** Specificity of bias by predictor and achievement across age: African-Americans and Hispanics; slope bias and underpredicted achievement

| Predictor                      | Math    | Reading | Written language |
|--------------------------------|---------|---------|------------------|
| Grades 1 through 4             |         |         |                  |
| Fluid-Crystallized Index (FCI) | No bias | No bias | No bias          |
| Mental Processing Index (MPI)  | No bias | No bias | No bias          |
| Nonverbal Index (NVI)          | No bias | No bias | No bias          |
| Grades 5 through 8             |         |         |                  |
| Fluid-Crystallized Index (FCI) | No bias | No bias | No bias          |
| Mental Processing Index (MPI)  | No bias | No bias | No bias          |
| Nonverbal Index (NVI)          | No bias | No bias | No bias          |
| Grades 9 through 12            |         |         |                  |
| Fluid-Crystallized Index (FCI) | No bias | No bias | No bias          |
| Mental Processing Index (MPI)  | No bias | No bias | No bias          |
| Nonverbal Index (NVI)          | No bias | No bias | No bias          |

Numerical values represent the underpredicted achievement for Caucasians and African-Americans at the 0.01 alpha level

require language ability or acquired knowledge. Even though the MPI and NVI are recommended as the global index of choice for ethnic minority children (Kaufman and Kaufman 2004a), the unbiased nature of FCI makes this global index a better choice when evaluating ethnic minority group children. The overprediction does not denote bias against African-Americans and Hispanics; rather, *underprediction* of achievement would have been indicative of ethnic bias. Nonetheless, the unanticipated overprediction for both African-Americans and Hispanics by the MPI and FCI indicates that these two global indexes are less accurate than the FCI in estimating math, reading, and writing for the two ethnic minorities.

**Table 9** Significant intercept overpredictions for African-Americans and Hispanics as compared to Caucasians across age groups

| Predictor                      | Math              |           | Reading           |           | Written language  |           |
|--------------------------------|-------------------|-----------|-------------------|-----------|-------------------|-----------|
|                                | African-Americans | Hispanics | African-Americans | Hispanics | African-Americans | Hispanics |
| Grades 1–4                     |                   |           |                   |           |                   |           |
| Fluid-Crystallized Index (FCI) |                   |           |                   |           |                   |           |
| Mental Processing Index (MPI)  |                   |           |                   |           |                   |           |
| Nonverbal Index (NVI)          |                   |           |                   | +2.7      |                   |           |
| Grades 5–8                     |                   |           |                   |           |                   |           |
| Fluid-Crystallized Index (FCI) |                   |           |                   |           |                   | +2.0      |
| Mental Processing Index (MPI)  | +3.2              | +2.4      |                   | +2.8      | +3.5              | +3.8      |
| Nonverbal Index (NVI)          | +2.5              | +3.3      | +2.3              | +4.1      | +3.3              | +4.9      |
| Grades 9–12                    |                   |           |                   |           |                   |           |
| Fluid-Crystallized Index (FCI) |                   |           |                   |           |                   |           |
| Mental Processing Index (MPI)  |                   |           |                   |           |                   |           |
| Nonverbal Index (NVI)          |                   |           |                   |           |                   |           |

Samples sizes, grades 1–4: Caucasians (455), African-Americans (119), and Hispanics (150); grades 5–8: Caucasians (487), African-Americans (119), and Hispanics (137); grades 9–12: Caucasians (371), African-Americans (74), and Hispanics (89)

## Discussion

In this study, ethnic group bias of the KABC-II global scores (FCI, MPI, and NVI) was examined for a representative sample of Caucasian, Hispanic, and African-American children and adolescents in grades 1 through 12. More specifically, it was explored whether the less culturally and linguistically loaded global scores, MPI and NVI of the KABC-II, were fairer and more accurate at predicting minority group's achievement outcomes than the traditionally more culturally and linguistically loaded FCI. In order to answer this research question, structural equation modeling was used to measure predictive invariance of the FCI, MPI, and NVI separately. The methodology applied increasingly restrictive sets of equality constraints in order to incrementally test whether the different levels of equality were met across the groups—residual, slope, and intercept invariance (Meredith 1993). Despite the firm belief by many neuropsychologists and educators that less culturally loaded scales, such as MPI and NVI, are the fairest (least biased) predictors of achievement for ethnic minority group children, results of this study suggest that FCI is the “fairest” predictor of achievement for Caucasian, Hispanic, and African-American school-aged children.

### Predictive Invariance

This is undoubtedly the first study to compare prediction invariance of three global ability measures of an individually administered test of cognition across three ethnic and grade groups, using structural equation modeling. Comparison of the FCI, MPI, and NVI results demonstrated that the FCI, the most comprehensive global score, emerged as the least

biased predictor variable for achievement, not only for Caucasian school-aged children but also, most importantly, for Hispanics and African-Americans. This finding is contrary to the KABC-II test authors' predictions and contrary to what many clinicians and neuropsychologists believe, based on the inclusion of the language-oriented and fact-oriented Knowledge/*Gc* scale in the FCI. Additionally, the MPI and NVI, the two global indexes that are linguistically and culturally more neutral, did not accurately predict the level of the reading, math, and writing abilities of children from the two ethnic minority groups. These indexes were not biased against African-Americans and Hispanics—they correlated as highly with achievement for the two ethnic minorities as they did for Caucasians, and they did not *underpredict* the achievement of African-American and Hispanic children—but they did not do a good job of identifying their level of achievement. The MPI and NVI overpredicted their actual levels of academic achievement, especially at grades 5–8.

In sum, the results of this present study show that the MPI and NVI produced consistent overprediction in terms of their intercept when assessing African-American and Hispanic minority group children's achievement, especially at grades 5–8. The more comprehensive FCI, on the other hand, was not biased in terms of its slope or intercept for Caucasian, Hispanic, and African-American school-aged children. The FCI findings are consistent with previous studies. Keith (1999), Weiss et al. (1993), and Weiss and Prifitera (1995) found no bias in terms of the slope and intercept when assessing prediction invariance of the WISC-III FSIQ and the GAI, both of which are comparable to the FCI in terms of content. Not many studies have assessed psychometric test bias in terms of prediction invariance, and those that did are more than 15 years old (Keith 1999; Weiss et al. 1993; Weiss and Prifitera 1995). No study has previously investigated prediction bias in terms of slope and intercept bias of culturally and linguistically free global scaled scores using structural equation modeling. Naglieri and colleagues did evaluate prediction bias of the CAS and NNAT, both with culturally reduced content, but they used simple coefficients of correlation for their analytic approach rather than structural equation modeling; therefore, it is not possible to determine whether the Naglieri studies also found overprediction of achievement for the ethnic minority children. The results of the present study have important implications for neuropsychologists.

### Clinical Implications for Neuropsychologists

The results demonstrate several important findings for neuropsychologists. First of all, some neuropsychologists believe that global scores should not be interpreted, or even used at all, as summary scores are often thought not to be reflective of an individual's neuropsychological status and profile (e.g., Kaplan 1988; Luria 1979; Lezak 1988). Data of this present

study, however, suggest that global scores have value. The most comprehensive global KABC-II score, FCI, was, in fact, very accurate at predicting achievement not only for Caucasians but also, most importantly, for Hispanics and African-Americans. FCI demonstrated no slope bias and virtually no intercept bias. The results of this present study suggest that the FCI, apart from being reliable and valid, is an unbiased and accurate predictor of academic achievement for Caucasian, Hispanic, and African-American school-aged children. Such results suggest that global scores can be useful indexes for clinical neuropsychologists to interpret, even though such scores mask the interpretation of the multiple individual characteristics (processes) that are more truly reflective of an individual's cognitive functioning. Further, findings of this study support Canivez's (2013) argument that global scores are valid and reliable when it comes to the interpretation of an individual's cognitive capacity. He supports his argument statistically, stating that global IQs have been found to have the strongest internal consistency, short- and long-term temporal stability, and predictive validity coefficients; they produce less error variance and account for the largest portion of the variance with a variety of criteria. The results of this study support his argument—namely that FCI is a fair predictor variable to use in the evaluation of Caucasian, Hispanic, and African-American school-aged children.

Some researchers might argue that the FCI is the fairest predictor variable of achievement due to criterion contamination. However, it is important to note that such an argument would have been true for earlier versions of the Wechsler scales, which only consisted of the Performance and Verbal IQ scales; naturally, the Verbal IQ scale overlapped greatly with achievement variables. The FCI is composed of five indexes, only one of which—Knowledge/*Gc*—is akin to academic achievement. The other four indexes measure visual-spatial ability, short-term memory, long-term retrieval, and fluid reasoning; none of these abilities (or cognitive processes) are taught in school.

Secondly, outcomes of this study showed persistent evidence for intercept bias on the MPI and NVI, such that the global cognitive scales consistently overpredicted African-American and Hispanic academic achievement at grades 5–8. Even though Kaufman and Kaufman (2004a) suggest using the MPI in preference to the more comprehensive FCI when assessing children from non-mainstream backgrounds, such as Hispanics and African-Americans, the present findings suggest otherwise. Kaufman and Kaufman generally suggest using the FCI as the index of choice in most neuropsychological evaluations, for example, for the diagnosis of learning disabilities or brain damage; however, they make the exception of suggesting the MPI for ethnic minorities. The present findings suggest that the comprehensive FCI should be the index of choice for neuropsychological evaluations, even if the referred child or adolescent is African-American or

Hispanic. By comparing the prediction invariance results of the three global scores, the present study showed that the FCI emerged as the least biased global index on the KABC-II—and that includes the NVI, which reduces language skills to an even greater extent than the MPI. Such findings are important because both the NVI and MPI measure a limited range of abilities, as they exclude language- and fact-oriented subtests. This limitation becomes especially problematic when the referral for evaluation is based on problems in language (Flanagan et al. 2013).

Thus, overall, results of this present study suggest that neuropsychologists should opt to use the FCI for all children, ethnic minority, or otherwise, when the goal is simply to predict their current level of achievement. As the MPI and NVI consistently overpredicted minority students' achievement, these two global indexes are likely to prove less accurate as predictors of their reading, math, and writing. However, we encourage examiners to use the MPI or NVI to identify minority students' capability to achieve higher than their current level; these indexes are less language and achievement oriented and are better suited at identifying their cognitive potential. For example, the MPI and NVI would be better choices than the FCI for Black and Hispanic students when the KABC-II is used for gifted placement as well as for the assessment of intellectual impairment or disability. It is also important to note that even though the KABC-II global scores are valuable predictors of current achievement or future potential, results do not imply that global scores are especially useful for planning interventions. Intelligent testing demands that clinicians rely on children's patterns of strengths and weaknesses for selecting the best educational interventions for each student (Kaufman et al. 2015; Lichtenberger and Kaufman 2013). Here, schools have the responsibility to make use of existing cognitive strengths in order to allow students to achieve to their fullest potential.

Finally, it is important to recognize that findings of this present study not only pertain to the Kaufman tests but also generalize to 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 is essentially the same *g* that is measured by the WJ III. Similarly, Reynolds et al. (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 and, by extension, to the current versions (WISC-V and WJ IV). Thus, neuropsychologists can be reasonably confident that results of the present study generalize to other popular tests of cognition and achievement.

### *Possible Explanations for the Overprediction at Grades 5–8*

It was interesting to see the persistent intercept overprediction of NVI and MPI for African-American and Hispanic achievement outcomes at grades 5–8. Such findings indicate the overprediction might depend on the developmental age of the students. For some reason, ethnic minority group children have the cognitive capacity to achieve higher in middle school (as evidence by their higher KABC-II scores) than they actually achieve. One explanation for the overprediction at grades 5–8 could be that verbal skills become extremely important for academic success in middle school, more so than in the primary grades when the goal is to learn the basics of reading, math, and writing. For example, solving math problems in middle school not only require moving letters and numbers but also require the student to read the problem, sketch the situation, and solve the problem both verbally and quantitatively (Wendling and Mather 2009). Other explanations for the overpredictions in middle school include the fact that the early adolescence is a critical time for brain development. For example, students are moving from more concrete problem solving to abstract, analytical thinking as the prefrontal cortex is undergoing rapid development (Luria 1979). However, the early adolescent years are also marked by difficulties paying attention to several stimuli at the same time (related to short-term memory limitations). Behaviors reflect this stage of brain development when adolescents engage intensively, but briefly, in a specific activity. Also, interaction with peers and active, experimental learning is preferred. In order to assist struggling students to achieve to their fullest potential in middle school, teachers should try to focus on experimental, group learning techniques and possibly limit the amount of stimuli students at these ages are presented with (Wendling and Mather 2009; Ryan and Patrick 2001). Furthermore, young adolescents are more emotionally driven (this is related to the fact that the prefrontal cortex is still developing and the amygdala is more easily activated) (Somerville et al. 2010). It is also possible that some individuals from ethnic minorities might perceive increased awareness over their minority status; for example, some pre-adolescents and adolescents might feel socially rejected or perceive the differences between them and their Caucasian school-teachers, all of which can impact their psychological health and, therefore, their ability to succeed in school (e.g., Parkhurst and Asher 1992; Weiss et al. 2006). It is important for neuropsychologists and school psychologists to take these suggestions into consideration, especially when evaluating middle school-aged minority group students.

It is also important to note that whereas reliability of the intelligence test would have affected the *slope* of the regression line, *intercept* bias arises due to omitted variables that are separate from the predictor variable (Meade and Fetzer 2009). In this study, the NVI and MPI produced consistent intercept overprediction, but no slope bias. Such results strongly

suggest that the minority group children have the cognitive capacity to achieve substantially higher than their current level of achievement, especially at grades 5–8. Intercept overprediction means that there are other variables, independent of the cognitive variables, that influence the minority groups' ability to achieve to their fullest potential. Many of these independent variables that impact achievement are likely related to socioeconomic disparities, such as differences in income and percent of single-parent households, as well as differences in nutrition and physical health (Weiss et al. 2015). Undoubtedly, there are many socioeconomic variables that contribute to differences in test results and it is impossible to account for all of those disparities. It is for those reasons that differences in mean scores between Caucasian and minority group students should not be taken at face value and should not be interpreted as meaningful. Finally, another variable that likely contributes to the "overprediction" is the failure of the American educational system to capitalize on minority children's strengths.

### Limitations

The results need to be understood in the context of the study's limitations. First of all, there is disagreement among researchers who have published on measures of ability and promoted theories of intelligence. Whereas some accept the notion that a test, which requires knowledge and skills taught in school, can be used to measure ability, others disagree due to criterion contamination (Dumont, Willis, and Elliott 2009). Furthermore, it is 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. Additionally, keep in mind that the term "Hispanic" was used to classify a very broad and heterogeneous group of individuals who differ in terms of their cultural and historical background. Unfortunately, no representative subsamples of Hispanics were available. In order to generalize present findings, future researchers need to replicate the analyses with different ethnic subsamples. Furthermore, it is important to note that the sample was not large enough to permit ethnic bias analysis for students from different socioeconomic backgrounds (as measured by mother's educational attainment). Future studies should address this limitation. For example, future studies could split their groups by parental education or other socioeconomic variables to evaluate whether results maintain for different SES groups. Other limitations include the fact the standardization sample used in this study is representative of 2001 US Census data. The demographic profile in the USA has undoubtedly changed since 2001; thus, the stratification of the sample

does not exactly reflect the current US population. Furthermore, even though we examined a developmental trend by dividing the sample into three grade groups, it is important to consider that this was a cross-sectional sample. Thus, just as there are drawbacks with regard to using longitudinal data sets, such as practice effects, there are also limitations to using cross-sectional data sets, such as cohort effects (Kaufman 2009; Kaufman and Weiss 2010). Future studies may want to replicate this present study using longitudinal data sets.

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. Future researchers ought to address these limitations.

### Conclusions

The results of the present study provide evidence for differential predictive validity of the KABC-II global scaled score, specifically FCI, across a representative sample of Caucasian, African-American, and Hispanic school-aged children in grades 1–12. Findings indicate that the comprehensive, global summary score, FCI, is more accurate at predicting achievement outcomes across ethnic minority groups than the culturally and linguistically reduced MPI and NVI. Indeed, the FCI was the one global index that showed the least bias and was, therefore, the most accurate at predicting achievement for all three ethnicities across all three grade groups. Neuropsychologists might, therefore, keep in mind giving the FCI more consideration than other KABC-II global indexes when evaluating African-American and Hispanic children. Using this scale has many advantages because of its comprehensiveness. The use of the MPI and NVI becomes especially problematic when the referral for evaluation is based on problems in language; in fact, the majority of referrals are based on language difficulties (e.g., Figueroa 1990; Flanagan et al. 2013; Sattler 1992). The results of the present study endorse the notion that clinical neuropsychologists and other clinicians can use the FCI (and by extension, other comparable global indexes such as Wechsler's Full-Scale IQ), even when evaluating children from ethnic minorities. It was found to be unbiased at grades 1–12, using the rigorous techniques of structural equation modeling. Furthermore, the present study adds to a sparse and outdated literature on the evaluation of differential predictive validity across ethnic groups.

**Ethical Standards** All procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of

1975, as revised in 2000 (5). Informed consent was obtained from all patients for being included in the study.

**Conflict of Interest** Dr. Kaufman is author of the instruments and receives royalties from the sales.

## References

- Arbuckle, J. L. (1995–2011). *AMOS 20.0 User's Guide*. Crawfordville, FL: AMOS Development Corporation.
- Binet, A., & Simon, T. (1905). New methods for the diagnosis of the intellectual level of subnormals. *L'Année Psychologique*, *12*, 191–244.
- Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. *The Oxford Handbook of Child Psychological Assessments*, 84–112.
- Dumont, R., Willis, J. O., & Elliott, C. D. (2009). *Essentials of das-II assessment*. Hoboken, NJ: Wiley & Sons.
- 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. (2007). *Administration and scoring manual differential abilities scale 2nd edition (DAS-II)*. San Antonio, TX: The Psychological Corporation.
- Figuroa, R.A. (1990). Best practices in the assessment of bilingual children. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology-II* (pp. 93–106) Washington, DC: National Association of School Psychologists
- Flanagan, D. P., Ortiz, S. O., & Alfonso, V. C. (2013). *Essentials of cross-battery assessment* (3rd ed.). Hoboken, NJ: Wiley & Sons, Inc.
- 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.
- Kaplan, E. (1988). The process approach to neuropsychological assessment. *Aphasiology*, *2*(3–4), 309–311.
- Kaufman, A. S. (2009). *IQ testing 101*. New York, NY: Springer.
- 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., & Weiss, L. G. (Eds.) (2010). Special issue on the Flynn Effect. *Journal of Psychoeducational Assessment* (28, Whole No. 5) Thousand Oaks, CA: Sage
- Kaufman, A. S., Lichtenberger, E. O., Fletcher-Janzen, E., & Kaufman, N. L. (2005). *Essentials of KABC-II assessment*. Hoboken, NJ: Wiley.
- Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. (2012). Are cognitive g and academic g one and the same g?: an exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence*, *40*, 123–138.
- Kaufman, A. S., Raiford, S. E., & Coalson, D. L. (2015). *Intelligent testing with the WISC-V*. Hoboken, NJ: Wiley.
- 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.
- Keith, T. Z., & Reynolds, C. R. (2003). Measurement and design issues in child assessment research. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: intelligence, aptitude, and achievement* (2nd ed., pp. 79–111). New York, NY: Guilford Press.
- Lezak, M. D. (1988). Iq: rip. *Journal of Clinical and Experimental Neuropsychology*, *10*(3), 351–361.
- Lichtenberger, E. O., & Kaufman, A. S. (2013). *Essentials of WAIS-IV assessment* (2nd ed.). Hoboken, NJ: Wiley.
- Luria, A. R. (1979). *The making of mind: A personal account of Soviet psychology*. Cambridge, MA: Harvard University Press.
- Meade, A. W., & Fetzer, M. (2009). Test bias, differential prediction, and a revised approach for determining the suitability of a predictor in a selection context. *Organizational Research Methods*, *12*, 738–761.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*(4), 525–543.
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2013). *Applied multivariate research: design and interpretation*. Thousand Oaks, CA: SAGE Publications Inc.
- 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.
- Naglieri, J. A. (1997). *The Naglieri nonverbal ability test*. San Antonio, TX: The Psychological Cooperation.
- Naglieri, J. A., & Bornstein, B. T. (2003). Intelligence and achievement: just how correlated are they? *Journal of Psychoeducational Assessment*, *21*, 244–260.
- Naglieri, J. A., & Das, J. P. (1997). *Cognitive assessment system interpretive manual*. Itasca, IL: Riverside.
- 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.
- 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., Rojahn, J., & Matto, H. C. (2007). Hispanic and non-Hispanic children's performance on PASS cognitive processes and achievement. *Intelligence*, *35*(6), 568–579.
- Parkhurst, J. T., & Asher, S. R. (1992). Peer rejection in middle school: Subgroup differences in behavior, loneliness, and interpersonal concerns. *Developmental Psychology*, *28*(2), 231.
- Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment research. In *Oxford handbook of psychological assessment of children and adolescents*, ed. C. R. Reynolds. New York, NY: Oxford University.
- 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.
- Ryan, A. M., & Patrick, H. (2001). The classroom social environment and changes in adolescents' motivation and engagement during middle school. *American Educational Research Journal*, *38*(2), 437–460.
- Sattler, J. M. (1992). *Assessment of children - 3rd ed.* San Diego, CA: Publisher, Inc.
- Somerville, L. H., Jones, R. M., & Casey, B. J. (2010). A time of change: behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain and Cognition*, *72*(1), 124–133.
- Thorndike, E. L. (1911). *Animal intelligence*. New York, NY: Hafner.
- 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/>
- Urbina, S. (2014). *Essentials of psychological testing* (2nd ed.). Hoboken, NJ: Wiley.
- Wechsler, D. (2003). *WISC-IV: administration and scoring manual*. San Antonio, TX: Psychological Corporation.

- Wechsler, D. (2005). *The Wechsler Individual Achievement Test-Second Edition (WIAT-II)*. San Antonio, Tx: The Psychological Cooperation.
- Weiss, L. G., & Prifitera, A. (1996). 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), 297–304.
- Weiss, L. G., Saklofske, D. H., Prifitera, A., & Holdnack, J. A. (2006). *WISC-IV advanced clinical interpretation*. Burlington, MA: Academic.
- Weiss, L. G., Saklofske, D. H., Holdnack, J. A. & Prifitera, A., (2015). *WISC-V assessment and interpretation: Scientist-Practitioner Perspectives*. Waltham, MA: Academic Press.
- Wendling, B. J., & Mather, N. (2009). *Essentials of evidence-based academic interventions*. Hoboken, NJ: Wiley.
- Wolf, T. H. (1973). *Alfred Binet*. Chicago, IL: University of Chicago Press.
- Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock Johnson-Revised (WJ-R) tests of cognitive ability*. Itasca, IL: Riverside Publishing.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson Third edition (WJ III)*. Itaska, IL: Riverside Publishing.