This is the peer reviewed version of the following article: McGill, R. J. (2019). An instrument in search of a theory: Structural validity of the Kaufman Assessment Battery for Children-Second Edition Normative Update at school-age. *Psychology in the Schools*. Advance online publication. doi: 10.1002/pits.22304, which has been published in final form at 10.1002/pits.22304.

Please use the following citation when referencing this work:

McGill, R. J. (2019). An instrument in search of a theory: Structural validity of the Kaufman Assessment Battery for Children-Second Edition Normative Update at school-age. *Psychology in the Schools*. Advance online publication. doi: 10.1002/pits.22304

An Instrument in Search of a Theory: Structural Validity of the Kaufman Assessment Battery for Children-Second Edition Normative Update at School-Age

Ryan J. McGill

William & Mary

Author Note

Correspondence concerning this article should be addressed to Ryan J. McGill, William & Mary School of Education, P. O. Box 8795 Williamsburg, VA.

Abstract

The present study examined the factor structure of the Kaufman Assessment Battery for Children-Second Edition Normative Update with normative sample participants aged 7-18 years (*N* = 500) using confirmatory factor analysis with maximum likelihood estimation and multidimensional scaling procedures. For the 10 subtest CHC core battery configuration, an alternative hierarchical model with four group-specific factors (Crystallized Ability, Perceptual Reasoning, Short-Term Memory, and Long-Term Storage and Retrieval) provided the best fit to the normative data for both age groups. These results suggest caution for interpreting the Visual Processing and Fluid Reasoning index scores in isolation given the inability to locate those constructs independently at school-age. Implications for the clinical interpretation of the measurement instrument moving forward are discussed.

Keywords: CHC, KABC-II NU, Structural validity

Practitioner Points

1. The hypothesized CHC factor structure for the KABC-II was not replicated in the KABC-II NU.
2. Clinicians are encouraged to exercise caution when interpreting the Visual Processing and Fluid Reasoning index given questions raised about their potential redundancy.
3. Aside from Fluid Reasoning and Visual Processing, evidence from the present study was supportive of the remaining CHC dimensions suggested by the test publisher.

An Instrument in Search of a Theory: Structural Validity of the Kaufman Assessment Battery for Children-Second Edition Normative Update at School-Age

The Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004a) is an individually administered measure of the cognitive abilities for children and adolescents ages 3 to 18 years. Since its publication in 2004, the KABC-II has been cited as one of the most widely used assessments for school-aged youth with surveys of the assessment practices of school psychologists indicating that it is a preferred instrument for assessing the cognitive abilities of examinees who are culturally and linguistically diverse in school-based settings (Benson et al., 2019; Sotelo-Dynega & Dixon, 2014). Although the content and the structure of the measurement instrument has yet to be revised, the KABC-II was recently re-normed and a manual supplement (Kaufman, Kaufman, Drozdick, & Morrison, 2018) reporting updated norms and technical information has been provided to complement the original technical manual (Kaufman & Kaufman 2004b). The resulting KABC-II Normative Update (KABC-II NU; Kaufman & Kaufman, 2018) now replaces the previous version of the test. Although users are able to use existing KABC-II test kits and record forms to administer the test, they must use the updated norms provided in the manual supplement, and now available in the Q-global® system, to score and interpret the test.

Theoretical Structure of the KABC-II NU

The KABC-II NU continues to employ a dual theoretical model for test score interpretation featuring the Cattell-Horn-Carroll model of cognitive abilities (CHC; Schneider & McGrew, 2018) and Luria's neuropsychological theory of cognitive processing (1973). The CHC interpretive model features 10 core subtests that combine to form five first-order broad ability indexes (Fluid Reasoning [Gf], Crystallized Ability [Gc], Short-Term Memory [Gsm], Visual-

Spatial Processing [Gv], and Long-Term Storage and Retrieval [Glr][1]) and a full scale Fluid-Crystallized Index or FCI. The Luria interpretive model features eight core subtests that combine to form four first-order indexes (Simultaneous Processing [Gv], Successive Processing [Gsm], Learning [Glr], and Planning [Gf]) and a full scale Mental Processing Index or MPI. It is important to note that the same subtests are used to form the first-order KABC-II NU index scores regardless of the interpretive model that is employed. The only salient differences between the models are the nomenclature that is employed to interpret test scores and whether measures of Crystalized Ability are administered by the examiner. Additional supplementary subtests can be administered by examiners and the configuration of these measures varies according to the age of the child. However, these subtests do not contribute to the measurement of the KABC-II NU scales and indexes.

Even though practitioners can elect to interpret the KABC-II NU from either theoretical perspective, this decision must be made *a priori*. Though, it is made clear in available manuals and other technical resources (e.g., Drozdick et al., 2018), that the CHC model should be preferred in most clinical situations. Not surprisingly, the CHC model has been the focal point of the vast majority of KABC-II validity research (e.g., Benson, Kranzler, & Floyd, 2016; Kaufman et al., 2012; McGill, 2015; Reynolds et al., 2007; Villeneuve, Hajovsky, Mason, & Lewno, 2019) since its publication. Further, Braden and Oazts (2005) argue that it is not psychometrically plausible for a test to measure two theoretically distinct constructs simultaneously as a measure of Visual Processing does not morph into a test of Sequential Processing[2] due to the theoretical orientation of the examiner. For the sake of parsimony, the remainder of the present article will focus specifically on the validity of the CHC core battery interpretive structure at school-age (7-

18) as that is the only age range at which all intended CHC constructs were able to be located by the test publisher in previous validation studies.

Validation of the KABC-II NU

The KABC-II NU manual supplement (Kaufman, Kaufman, Drozdik, Morrison, 2018) reports a bevy of psychometric information for the updated normative sample. Updated reliability coefficients and examination of relationships between KABC-II NU variables and several external measures are provided. However, the internal structure of the KABC-II NU was not formally investigated. Instead, the *Evidence Based on Internal Structure* section (pp. 29-33) only contains a narrative description of relationships observed among the composite scores in the intercorrelation matrices reported in that section. Although it is suggested that the intercorrelations among the scores are *very* similar to those observed in the KABC-II sample, the NU correlations are noticeably weaker indicating attenuation in some of the relationships among the variables that could have implications for the structure and interpretation of the test moving forward.

To gain better insight into the reliability and validity of the KABC-II NU, readers are encouraged to review the research literature that has accumulated for the instrument since its publication. However, it cannot be anticipated that those results, no matter how persuasive, will replicate in the NU as they were obtained from a different sample of participants that is now over 15 years old. Furthermore, as will be demonstrated below, it remains unclear what the KABC-II measures. Significant questions have been raised about the tenability of the CHC model preferred by the test publisher and numerous alternative models have been proposed in the literature with some questioning fundamental aspects of the instrument's hypothesized alignment with CHC theory. What follows is a summary of previous KABC-II structural validity research pertaining to the CHC interpretive model at school-age.

KABC-II Structural Validity Research

In the original KABC-II technical manual[3], Kaufman and Kaufman (2004b) report that "The KABC-II development process relied mostly on the technique of confirmatory factor analysis, used in an *exploratory* fashion to evaluate subtests and decide how they should be grouped into scales" (p. 103). In particular, a major focus of the factor analyses of the normative sample that followed standardization was to provide statistical evidence that Fluid Reasoning could be distinguished from Visual Processing in the normative sample. Although it is noted that exploratory factor analysis (EFA) was employed to help identify alternative structures for the data that had not been identified at this stage, the EFA results "did not make a significant contribution to the overall analysis program" (p. 104) and were not reported in the technical manual. Instead, the test publisher relied exclusively on confirmatory factor analysis (CFA) employed across two stages. In stage one, CFA was used to examine competing higher-order models ranging from a one-factor model to increasingly more complex correlated (oblique) factors models at each age. The authors report checking at each step to verify that improvement in fit was statistically significant and that there were no issues with localized strain (i.e., problematic loadings, out of bounds parameter estimates). Fit statistics for the models that were explored at this stage are not reported but the results for each age are described. It is clear from that narrative that several estimation problems were encountered when trying to locate posited CHC factors across the age range.

At age three, a one-factor (*g*) model was preferred even though a Short-Term Memory factor was located. It is not clear why this factor was not retained as it appeared to meet the inclusionary criteria for factor retention stated earlier in the technical manual. At age four, analyses supported the presence of Short-Term Memory and Long-Term Storage and Retrieval factors but a Fluid Reasoning factor could not be located so those indicators were assigned to Visual Processing. Interestingly, it was noted that the Crystallized Ability and Visual Processing factors were highly correlated suggesting that they were likely isomorphic[4]. Even so, a decision was made to retain those dimensions, even though differentiation between them was not statistically significant, on the basis that a *content perspective* suggested that they are likely sampling different abilities.

At ages 5-7, the Fluid Reasoning and Visual Processing factors were not distinguishable and those indicators were all assigned to a broad Visual Processing factor. It was later reported that the correlations between those factors, when explicated, exceeded .90 from ages 7-18 raising concern about the discriminate validity of those factors (Byrne, 2005). However, the authors noted that these abilities could be differentiated statistically from each other and a decision was made to separate Fluid Reasoning from Visual Processing across the school age range. Additionally, different configurations of the Visual Processing scale were explored to determine the combination that would *best amplify* the distinction between those abilities. Results indicated that Triangles and Rover provided optimal differentiation in the younger portion of the school age range (7-12) and Block Counting and Rover improved differentiation at ages 13 to 18.

In the second stage of the analytical plan, constrained CFAs were used to investigate hierarchical versions of the models retained at stage one containing a second-order general factor of intelligence. It was believed that these models best aligned with the scoring and interpretive structure for the test. Those models are depicted visually in Figures 8.1 and 8.2 of the technical manual (pp. 106-107). Global fit statistics indicate that all of the models provided good to excellent fit to the data. However, at ages 7-18, all of the models contain path loadings between *g* and Fluid Reasoning that meet or exceed unity.

Whereas standardized path loadings equal to 1.0 are technically permissible in CFA, they present a conceptual problem as they indicate that a variable is empirically redundant with another variable in a model. In these situations, researchers have been advised to consider the scientific law of parsimony and consider revising the model by deleting the redundant variable (Brown, 2015 [in this case Fluid Reasoning]).Values that exceed 1.0 are considered out of bounds estimates (i.e., Heywood cases) and indicate potential model misspecification. It is also worth noting that localized strain was not identified in any of the models for ages 4-6 and that those models do not contain a separate Fluid Reasoning factor which was the source of the strain at ages 7-18. In sum, the factor analytic evidence reported in the technical manual (Kaufman & Kaufman, 2004b) suggests that the CHC model preferred by the test publisher is likely not tenable at school-age.

Further complicating the matter, subsequent factor analytic research on the KABC-II has supported conflicting structures for the instrument that deviate from the theoretical structure posited by the test publisher. In one of the first independent examinations of the internal structure of the KABC-II, Reynolds and colleagues (2007) used CFA to test rival hypothesis for the normative data. Consistent with the results reported in the technical manual, specification of a baseline model consistent with publisher theory produced a Heywood case indicating an impermissible solution. After an exhaustive analysis, the authors elected to retain a final validation model containing several cross-loadings and theoretically divergent assignment of indicators on latent factors: Pattern Reasoning loaded on both Fluid Reasoning and Visual Processing, Hand Movements (supplementary subtest thought to measure Short-Term Memory) loaded on both Short-Term Memory and Fluid Reasoning, and Gestalt Closure (a supplementary measure thought to measure Visual Processing) was reassigned to load on Crystallized Ability. Although some of these parameters may be *theoretically* supported, this departure from desired simple structure creates an interpretive confound for users of the KABC-II due to the fact that the scores for the instrument do not take into account this dimensional complexity[5]. For example, in the cross-battery assessment interpretive system (XBA; Flanagan, Ortiz, & Alfonso, 2013), even though its contribution to the measurement of Visualization (Vz; a narrow dimension of Visual Processing) is acknowledged, Pattern Reasoning is assigned to Fluid Reasoning at ages 7-18[6]. Even though the final validation model converged and fit the data well, the loading coefficient between *g* and Fluid Reasoning (.97) approached unity indicating that virtually all of the variance in that construct was absorbed by general intelligence.

Subsequent CFA and structural equation modeling (SEM) research have provided inconsistent support for the model produced by Reynolds et al. (2007), with several studies supporting additional alternative models that fundamentally challenge KABC-II theory. Whereas Kaufman et al. (2012) adopted a KABC-II structure similar to that produced by Reynolds et al. (2007) and that model fit the data well, the loading between *g* and Fluid Reasoning was 1.0 across the age range, which the authors report was a byproduct of fixing the error variance for that dimension to zero in all of the models. Additionally, the posited Pattern Reasoning cross-loading was not retained in the final validation model for a study of the measurement invariance of the KABC-II across ethnicity (Scheiber, 2016) or in a recent cross-battery CFA by Reynolds, Keith, Flanagan, and Alfonso (2013). The cross-battery CFA by Reynolds and colleagues (2013) is noteworthy in several respects: it used the KABC-II as a target reference instrument to locate posited CHC constructs in other instruments, XBA theory was used to guide model development, and the final validation models contained numerous complex parameters (i.e., cross-loading, correlated residual terms) so it was possible for that parameter to emerge in the data.

Other validity studies have questioned whether a Fluid Reasoning factor can be located by the instrument at school-age at all. Reynolds and Keith (2007) examined the effects of Spearman's "law of diminishing returns" (SLODR) on different structural models of intelligence using the KABC-II. Briefly, SLODR refers to a decrease in *g* saturation in cognitive subtests as general ability increases—essentially *g* becomes more efficient as ability increases. For their hierarchical model, the employed a model virtually identical to the model produced by Reynolds and colleagues (2007) except Gestalt Closure was permitted to cross-load on Crystallized Ability *and* Visual Processing. Imposition of that additional parameter resulted again in a perfect loading between *g* and Fluid Reasoning indicating empirical redundancy. Interestingly, in constructing a bifactor model where *g* and the group factors have direct effects on the subtests simultaneously, the authors employed a model that did not contain a Fluid Reasoning factor at all because it was empirically under-identified.

Almost a decade later, Benson, Kranzler, and Floyd (2016) used those same models to examine KABC-II cognitive-achievement relationships using SEM. Whereas the bifactor model was identical to the one specified by Reynolds and Keith (2007), the Fluid Reasoning factor was dropped (as per Brown, 2015) from both the hierarchical and bifactor models to remedy the redundancy that was observed in the previous study. Both models fit the data well indicating that they were plausible interpretive structures for the instrument. These results were partially replicated in a similar study of the longitudinal effects of cognitive abilities on math achievement by Villeneuve et al. (2019) in which specification of a bifactor model consistent with publisher theory resulted in group factor collapse for Fluid Reasoning. It should be noted that although the authors elected to retain a Fluid Reasoning factor in the hierarchical model, its prediction paths on math achievement were constrained to zero because the latent factor contained nominal residual variance beyond general intelligence.

Finally, in one of the only direct examinations of the structural validity of the 10 subtest core battery configuration that has been conducted to date outside of the KABC-II technical manual, McGill and Dombrowski (2018) used EFA to elucidate the internal structure of those indicators at school age. Extraction test results did not support a five-factor model as posited by publisher theory. Not surprisingly, when that model was explicated for exploratory purposes, it contained a mathematically impermissible factor with no salient loadings at ages 7-12 and a single salient loading by Story Completion at ages 13-18. EFA results supported a more parsimonious four factor solution for ages 7-12 and 13-18 with the hypothesized Fluid Reasoning and Visual Processing indicators fusing together to form a complexly determined Perceptual Reasoning factor consistent with previous incarnations of the Wechsler Scales. All of the other posited CHC factors were replicated across the school age range for the core battery configuration.

Purpose

As noted by Meyer and Reynolds (2018), intelligence tests scores go through a continuous validation process and whenever a test is revised or updated, new questions are raised about the test's internal structure. The information furnished by such investigations are important because they provide the statistical rationale for the development and clinical interpretation of test scores (McGill & Dombrowski, 2017). Whereas structural validity is only one aspect of construct validity, it is often regarded as a vital first step in the test validation process as it undergirds other forms of validity (Keith & Kranzler, 1999).

Although the manual supplement contains extensive information about the psychometric characteristics of the KABC-II NU scores in the updated normative sample, no structural validity information is reported, a curious omission given questions that have been raised about the internal structure of the KABC-II since its publication and the availability of a dataset that is more than acceptable for such analyses. Without this information, users of the KABC-II NU are forced to rely on a series of inconsistent and conflicting results for the KABC-II to make inferences about how its scores should be interpreted in clinical settings moving forward.

Accordingly, the purpose of this study is to address this limitation in the literature by examining the construct validity of the KABC-II NU using confirmatory factor analysis and multidimensional scaling procedures. To this authors knowledge, this is the first time the NU has been subjected to external analysis by an independent investigator and the first time rival hypotheses suggested in the independent factor analytic literature for the KABC-II have been compared within the same analytical format. It is believed that the results furnished by this study will help to resolve questions about KABC-II NU structure and the degree to which the instrument aligns with posited CHC theory.

Method

Participants

Participants were 500 participants ages 7-18 who were included in the KABC-II NU normative sample. This sample was obtained using a stratified sampling plan designed to accord with 2015 U.S. Census estimates. Inspection of the demographic data reported in the manual supplement (Kaufman et al., 2018) reveal that the data for the normative sample was consistent with the U. S. population parameters for age, gender, race\ethnicity, parent education level (as a proxy for socioeconomic status), and geographic region. There were 50 participants in each of the 10 age groups that are the focus of the current study. It should be noted that the total normative sample for the KABC-II NU contains an additional 200 participants at ages 3-6. These participants were excluded from the present analyses because the core-battery contains at those ages contains an alternative configuration of age-specific subtests and not all posited CHC dimensions could be located prior to age 7.

Measurement Instrument

At school age (ages 7-18), and interpreted from the CHC perspective, the KABC-II NU contains 16 subtests (10 core and 6 supplemental) that combine to form five first-order index scores thought to measure various broad CHC abilities and a global full scale composite score as a proxy for psychometric *g*. Average internal consistency coefficients range from .84 to .96 for the subtests and from .91 to .98 for the index and composite scores. Extensive normative and psychometric data can be found in the KABC-II NU manual supplement (Kaufman et al., 2018). As previously mentioned, to maintain consistency with previous KABC-II validity research, the focus of the present study is limited to the CHC core battery interpretive structure at school age.

Procedure and Data Analyses

Confirmatory Factor Analysis. The KABC-II NU subtest correlation matrices for normative participants ages 7-18 were extracted from the manual supplement (pp. 32-33, Tables 3.6 and 3.7) and subjected to CFA using Mplus, Version 8.0 (Muthen & Muthen, 2017) with maximum likelihood estimation. Separate analyses were conducted at ages 7-12 and 13-18 due to the different configurations of the hypothesized Visual Processing Index for those age groups.

Consistent with previous KABC-II structural analyses, four first-order models were specified and examined: (a) a one factor baseline model; (b) four oblique Crystallized Ability, Perceptual Reasoning, Short-Term Memory, and Long-Term Storage and Retrieval factors (i.e, McGill & Dombrowski, 2018); (c) five oblique CHC factors (i.e., Kaufman & Kaufman, 2004b) and (d) a five oblique factors CHC model with Pattern Reasoning permitted to load on both Fluid Reasoning and Visual Processing (i.e., Reynolds et al. 2007). Consistent with CFA best practice, theory and data driven hierarchical and bifactor expressions of the first-order models were also examined (e.g., Benson, Kranzler, & Floyd, 2016). Specification of an oblique factors model with the alternative solution proposed by Benson et al. (2016) was not possible as it contains an indicator with no group factor loading. In the bifactor models, factors with only two subtest indicators were constrained to be equal to ensure specification. Detailed descriptions of the salient differences between bifactor and hierarchical measurement models are provided by Beaujean (2015) and Canivez (2016). It has been argued that the bifactor model is more consistent with how ability tests are *actually* interpreted by practitioners in clinical settings (Rodriguez, Reise, & Haviland, 2016). However, its viability as a structure for intelligence continues to be debated (Murray & Johnson, 2013).

Multiple fit indices were examined to evaluate the adequacy of global fit. Specifically, the (a) chi-square (χ^2), (b) comparative fit index (CFI), (c) root mean square error of approximation (RMSEA), (d) standardized root mean square residual (SRMR), and (e) Akaike's information criterion (AIC) were used. Although there are no golden rules for evaluating model fit indices, the following guidelines were used for good-model fit criteria: (a) CFI $\geq$ 0.95; (b) SRMR and RMSEA $\leq$ 0.06 (Hu & Bentler, 1999). There are no specific criteria for information-based indices like the AIC, but smaller values may indicate better approximations of the true measurement model after accounting for model complexity (Vrieze, 2012). Meaningful differences between well-fitting models were evaluated based upon the following criteria: (a) exhibit good fit according to CFI, RMSEA, and SRMR indices; (b) demonstrate a ΔCFI value

≤0.01 for nested models ; and/or (c) display the smallest AIC value. In addition to global fit, each model was also inspected for the presence of local strain.

Multidimensional Scaling. At its heart, factor analysis is about reducing complexity in data to better facilitate theory development. Although its use in psychology and education to determine which structures best explain cognitive data is ubiquitous, some scholars (e.g., McGrew, 2018) contend that EFA/CFA are presently given to much weight in academic school psychology. One multivariate alternative to factor analysis is multidimensional scaling (MDS; Kruskal, 1964). MDS is visualization technique that maps information about the theoretical distances between variables into a Cartesian space. Within the assessment literature, MDS has been used to evaluate the structure of intelligence test scores for well over 50 years (e.g., Cohen, Fiorello, & Farley, 2006; Guttman & Levy, 1991; Meyer & Reynolds, 2018), and it was incorporated here to supplement CFA results. This is the first time that the KABC-II has been subjected to this procedure.

In MDS, a correlation matrix is converted into a distance or proximity matrix and an MDS program plots the coordinates among the variables contained in that matrix into a geometric space. It is important to note that the purpose of MDS is try and construct a visual map that preserves the observed distances between the variables in the most efficient way possible. Thus, the global fit of an MDS solution is evaluated by a loss function (Stress) which evaluates the fit between the resulting MDS map and the distance matrix. A Stress value of ≤ 15 is considered acceptable although ≤ 10 is preferred (Kruskal & Wish, 1978). Once global fit is established, the map is visually inspected to identify relationships between the variables. Variables that are expected to have a relationship should cluster together within the same relative proximity or area of the map. For example, Cohen et al. (2006) used MDS to examine the WISC-

IV standardization data and found that the subtests clustered together in a pattern resembling the hypothesized group factors. Additionally, Guttman & Levy (1991) demonstrated that the positioning of tests in psychological space was not arbitrary and that variables closer to the center of the map can be interpreted as being more “cognitively complex” than variables positioned further from the center.

Proximity matrices were derived from the same KABC-II NU normative data correlation matrices that were the target of CFAs. The resulting matrices were then subjected to nonmetric MDS analyses. All data analyses were conducted using the *MASS* and *Vegan* packages in the R Statistical System (R Core Development Team, 2019).

Results

Descriptive statistics reported for the KABC-II NU subtest scores for participants ages 7-18 in the manual supplement illustrate univariate normality and Mardia’s standardized multivariate kurtosis estimate for these data was well within the criterion of |5.0| suggesting the data meet assumptions for multivariate normality (Byrne, 2006). As a result, use of maximum likelihood estimation for CFA in the present analysis was deemed appropriate.

Insert Table 1

Confirmatory Factor Analysis

Ages 7-12. Model fit statistics presented in Table 1 illustrate increasingly better fit from one to four factors. Fit statistics indicate that a one factor model was inadequate due to low CFI and RMSEA values. Fit statistics indicate that the four oblique factors model (Model 2) with the Fluid Reasoning and Visual Processing indicators combining to form a single Perceptual

Reasoning factor, provided excellent fit to the data. Subsequent higher-order models consistent with publisher theory (i.e., that the KABC-II measures five CHC related abilities at the first-order level) yielded a relatively equivalent fit however, they were both statistically undistinguishable from Model 2 and had higher AIC values. Additionally, inspection of their model parameters revealed localized strain. In Model 3, the correlation between Fluid Reasoning and Visual Processing (.92) indicates those variables are likely indistinguishable and in Model 4, specification of the cross-loading for Pattern Reasoning suggested by Reynolds and colleagues (2007), resulted in a statistically non-significant loading for Pattern Reasoning on Fluid Reasoning. As a result, those models were deemed inadequate for the data. Because the four KABC-II NU factors were highly correlated (.50 to .77), a higher-order dimension may be present (Gignac & Kretzschmar, 2017). As a result, alternative models with a general factor were also examined.

Insert Figure 1

A series of hierarchical and bifactor models consistent with previous KABC-II research and publisher theory (e.g., Kaufman & Kaufman, 2004b; McGill & Dombrowski, 2018; Reynolds et al., 2007) were specified and examined. Additional consideration was given to data driven models proposed by Benson, Kranzler, and Floyd (2016) which eliminate the Fluid Reasoning factor but retain a Visual Processing dimension. All of the eight models that were examined fit the data well. Initially, a hierarchical four factor solution (Model 5) consistent with Model 2 was examined. Subsequent models were statistically indistinguishable from Model 5. Even though Model 12 (a bifactor version of Model 2) provided a nominal improvement in CFI and SRMR to Model 5, those values did not exceed *a priori* criteria for model retention. After giving consideration to theory and parsimony (as per Pedhazur, 1997), a decision was made to retain the four factor hierarchical model (Model 5). It had the lowest AIC value of any of the models that were examined and evaluation of differences in ΔAIC values (e.g., Wagenmakers & Farrell, 2004) indicate that Model 5 is two to four times more likely to replicate than competing models. As illustrated in Figure 1, the final validation model for ages 7-12 is an elegant solution for the data. All of the path loadings were strong (i.e., > .50) and statistically significant and no localized strain was identified.

Ages 13-18. Similar results were obtained for ages 13-18. Of the four higher-order models that were examined, only the four factors model and the five factors model fit the data well. The alternative five factors model containing the Pattern Reasoning cross-loading failed to converge indicating it was not a tenable solution to the data. Model four was statistically undistinguishable from Model three and the more parsimonious four factors model produced lower SRMR, RMSEA, and AIC values. Again, as observed in the previous sample, the correlation between Fluid Reasoning and Visual Processing (.91) was large calling into question the discriminate validity of those factors (Byrne, 2005). Correlations between the factors in the four factors model were moderate to strong (.59 to .72) indicating that an oblique factors model likely does not capture all of the dimensions measured by the instrument (Caretta & Ree, 2001).

The fit statistics produced by the hierarchical and bifactor models were univocal in demonstrating the superiority of the four factor hierarchical model. Model 5 provided excellent fit to the data on all of the indices that are reported in Table 1. CFA results for that model did not reveal any issues of localized strain. No subsequent model provided a statistically significant improvement in fit and the differences in the strength of fit between Model 5 and other models approached meaningful levels on several indices. Furthermore, Model 5 also had the lowest AIC value of the all of the models that were examined. Evidence ratios calculated from the ΔAIC values for the most plausible models for that age group (as per, Wagenmakers & Farrell, 2004) indicate that Model 5 is two times more likely to replicate than Model 2 and ~10 times more likely to replicate than Model 7. On that basis, Model 5 was selected as the best model for the data. It's standardized solution is represented graphically in Figure 2. Interestingly, model misspecification was observed in several of the bifactor models due to negative residual variance for several KABC-II NU indicators. The cause of the misspecification was consistently sourced to Word Order and Rebus Learning. The fact that these same models converged for one age group and not another adds to the growing literature raising questions about the tenability of the bifactor model as a viable structure for some psychological dimensions (e.g., Bonifay, Lane, & Reise, 2017).

Insert Figure 2

Multidimensional Scaling. A 2-D nonmetric MDS map was fit to the KABC-II NU normative data. The resulting Stress values of 9.4 and 6.8 respectively for the 7-12 and 13-18 age ranges indicate that the MDS solutions provided good fits to the data. Visual inspection of the maps in Figures 3 and 4 reveal results that are consistent with those produced from the CFAs. For ages 7-12, the Short-Term Memory and Crystallized Ability indicators cluster together in theoretically expected alignment. Rebus Learning and Atlantis are more diffuse owing to the strong *g*-loading in the former; however, they are located on the same half of the map suggesting alignment. As expected, the Fluid Reasoning and Visual Processing measures largely cluster together in the same quadrant. The location of Story Completion in the bottom half of the map suggests it is measuring similar content to measures located in that dimension which may be why it fails to align with Pattern Reasoning (which is more closely aligned with Rover and Triangles).

For ages 13-18, the Short-Term Memory, Long-Term Storage and Retrieval, and Crystallized Ability factors are clearly defined. However, Pattern Reasoning straddles the border between dimension 1 and dimension 2 suggesting it aligns weakly with Story Completion and more strongly with Block Counting and Rover. Given their location at the centers of the maps at both ages, Riddles, Verbal Knowledge, Pattern Reasoning, and Rebus Learning are among the more cognitively complex (i.e., *g*-loaded) indicators on the KABC-II NU. Interestingly, at ages 13-18, Story Completion was located the furthest away from the center of the map indicating it is likely not a very cognitively complex task at that age range. Given the fact that *g* and Fluid Reasoning are frequently statistically indistinguishable in validity studies of commercial ability measures and scholars continue to conflate these two dimensions (e.g., Hajovsky et al., 2018), it may very well be that Fluid Reasoning is simply not measured well by the KABC-II NU in adolescence.

Insert Figure 3

Discussion

The present study examined the internal structure of CHC core battery configurations for participants in the recent normative update for the KABC-II (ages 7-18) using CFA and MDS procedures. To this authors' knowledge, this is the first independent investigation of the validity of the KABC-II NU since its publication. Formal analyses to verify the theoretical structure posited for the instrument were not conducted, or if so, not reported by the test publisher in the manual supplement. Instead, users are encouraged to consult the psychometric literature that has accumulated for the KABC-II since its publication in 2004 to better understand what the instrument measures. Given the questions that have been raised about the posited CHC structure for the KABC-II since its publication (e.g., McGill & Dombrowski, 2018; Reynolds et al., 2007) and the half dozen alternative models that have been reported in this literature, which model should practitioners prefer? Should they interpret the test based on the publisher suggested scoring configuration and disregard (a) that CFA results in the technical manual (Kaufman & Kaufman, 2004b) suggest potential model misspecification or at a minimum empirical redundancy between posited CHC constructs, and (b) the fact that independent factor analytic studies have largely supported alternative models that deviate from posited CHC theory in non-trivial ways (e.g., Benson, Kranzler, & Floyd, 2016; McGill & Dombrowski, 2018; Reynolds et al., 2007)? To shed insight additional insight on these matters, the current study evaluated all possible rival models within the same analytical framework.

Insert Figure 4

Results from the current study suggest that a more parsimonious, theoretically divergent, four-factor hierarchical model appears to fit the normative data better at ages 7-12 and 13-18. At both age groups, the hypothesized Fluid Reasoning and Visual Processing indicators were found to load on a complexly determined Perceptual Reasoning factor and subsequent models, (consistent with publisher theory) where those dimensions were specified as unique factors, failed to substantively improve model fit or, in some cases, provided a worse fit to the data.

Whereas in the first-order models that did not contain a general factor of intelligence these differences were trivial, the differences in fit were meaningful for both age groups in the hierarchical models that were examined. As the KABC-II NU is inherently a hierarchical instrument, these results are compelling and cast doubt on the hypotheses that Fluid Reasoning failed to emerge in younger age groups because of developmental differences as the fit for the five-factors hierarchical model at ages 13-18 worsened.

While some practitioners and scholars may find these results surprising, they are consistent with the results furnished by a recent EFA of the KABC-II by McGill and Dombrowski (2018) and presaged by the numerous disclosures in the technical manual (Kaufman & Kaufman, 2004b) suggesting Fluid Reasoning and Visual Processing were difficult to separate across the age range of the test. To wit, prior to age seven, attempts to explicate a five-factor model consistent with publisher theory were not successful. Instead, the Fluid Reasoning indicators were assigned to load on the same factor as the Visual Processing indicators. Furthermore, descriptions of the narrow dimensions measured by the hypothesized Fluid Reasoning measures suggest they are both measures of Induction *and* Visualization (Kaufman & Kaufman, 2004b, pp. 68-69).

Is Fluid Reasoning measured by the KABC-II? Perhaps. However, MDS results compliment the results that were obtained by the CFAs and show that Story Completion is not aligned with Pattern Reasoning in a way that suggests they combine to form their own factor. Instead, Pattern Reasoning appears to have more in common with measures of Visual Processing at both age groups. It should be noted that these results are remarkably consistent with previous results from factor analytic research on rival CHC instruments such as the Woodcock-Johnson IV Tests of Cognitive Abilities and the Wechsler Intelligence Scale for Children-Fifth Edition (e.g., Canivez, Watkins, & Dombrowski, 2017; Dombrowski, McGill, & Canivez, 2018) where evidence for a Perceptual Reasoning factor was also found. These results do not appear to be an artifact of empirical redundancy between *g* and Fluid Reasoning as the path loadings between those indicators in all of the five factor models that were examined did not approach unity as in previous studies. In a study of the WISC-IV using Bayesian SEM, Golay and colleagues (2013), demonstrated that these constructs are separable in a hierarchical measurement model; thus, standardized path loadings between these variables that approach or exceed unity should be regarded as problematic.

Implications for the Clinical Interpretation of the KABC-II NU

In many ways the KABC-II NU appears to be an exemplary instrument. It measures four first-order abilities and psychometric *g* well and the four-factor hierarchical solution obtained in the present study provides a coherent and elegant solution for interpreting the instrument. And the authors deserve kudos for updating aspects of interpretive guidance in professional resources (e.g., Drozdick et al., 2018) to accord with new research on the issue of cognitive scatter analysis. Nevertheless, the results obtained in the current study suggest practitioners should employ caution when interpreting the KABC-II NU using the CHC interpretive structure suggested by the test publisher. As compelling evidence for distinct Visual Processing and Fluid Reasoning factors was not obtained across the CFA and MDS analyses that were employed, those index scores should be interpreted cautiously, if at all, until such evidence is furnished. In closing, whereas previous validity studies of the KABC-II suggest that those dimensions can be distinguished statistically (e.g., Potvin, Keith, Caemmerer, & Tundt, 2015; Reynolds et al., 2007), those results were not replicated here.

Limitations

As with any study, the current investigation is not without limitations that should be taken into consideration by readers when interpreting the results. First, given the sample size for normative participants ages 7-18 in the KABC-II NU (*N* = 500) is less than a quarter of the size of that same sample in the KABC-II (*N* = 2,025), it is possible that posited CHC constructs failed to emerge as an artifact of sampling error. That is, the Fluid Reasoning and Visual Processing factors failed to emerge in this particular sample but may emerge in others. Even so, it is worth nothing that the sample used in the current study is *the* reference sample for the normative scores developed for the instrument. Given that there are presently no comparable CFA studies examining the structural validity of the CHC core battery at school-age for the KABC-II, it is also possible that the present results are due to longitudinal changes in how those variables are measured by the instrument (Ferrer & McArdle, 2010). Whereas previous research suggests that the KABC-II measurement model is invariant across age (e.g., Reynolds et al., 2007), future research is needed to determine if these effects replicate for the NU core battery.

Even when posited latent dimensions (i.e., CHC broad abilities) are located by a factor analysis, additional evidence is needed to determine if those variables are useful for diagnostic or treatment decisions (Borsboom, Mellenbergh, & van Heerden, 2004). Recent incremental and diagnostic validity studies for the KABC-II suggest that the CHC-based factor scores account for trivial portions of achievement variance after the effects of the general factor are controlled for and that the cognitive profiles generated by these scores are of limited value for specific learning disability identification (Benson, Kranzler, & Floyd, 2016; McGill, 2015, 2018). Future extensions of this research would benefit the field given the extended lifespan of the instrument.

Finally, given the acrimony surrounding these debates, the current study chose to forgo discussion regarding the relative importance of the first-order dimensions in comparison to the higher-order *g* factor. Whereas there is an ample body of literature suggesting that commercial ability measures such as the KABC-II NU are largely composed of reliable variance attributable to *g* with little residual variance attributable to various broad abilities (Canivez, 2013), elaborating on this literature often elicits partisan reactions. Reynolds and Keith (2013) describe procedures by which residualized loadings approximating the Schmid-Leiman (1957) procedure can be calculated from the hierarchical models contained in Figures 1 and 2. To enable readers to calculate these indices and come to their own conclusions on these matters, the R^2 values extracted from those CFA models have been made available as a supplemental resource at the following link (link masked for review).

Conclusion

The results furnished by this investigation have direct implications for the clinical interpretation of the KABC-II NU and supplant previous results reported in the KABC-II literature. Current and future users of the KABC-II NU may use these results to compliment the information that is contained in the manual supplement. Put simply, the current study was not able to locate posited CHC constructs (Visual Processing and Fluid Reasoning) at ages 7-12 or 13-18. As an instrument that purports to measure CHC theory, the results furnished by this investigation suggest that the instrument either needs to be revised to better comport with the theory, or the theory needs to be revised to comport with what is measured by the instrument.

References

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3,* 121-136. doi: 10.3390/jintelligence3040121

Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 national survey. *Journal of School Psychology, 72,* 29-48. doi: 10.1016/j.jsp.2018.12.004

Benson, N. F., Kranzler, J. H., & Floyd, R. G. (2016). Examining the integrity of measurement of cognitive abilities in the prediction of achievement: Comparisons and contrasts across variables from higher-order and bifactor models. *Journal of School Psychology, 58,* 1-19. doi: 10.1016/j.jsp.2016.06.001

Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science, 5,* 184-186. doi: 10.1177/2167702616657069

Boorsboom, D., Mellenbergh, G. J., & van Heerden, J. (2004). The concept of validity. *Psychological Review, 111,* 1061-1071. doi: 10.1037/0033-295X.111.4.1061

Braden, J. P. & Ouzts, S. M. (2005). Review of Kaufman Assessment Battery for Children, Second Edition. In R. A. Spies & B. S. Plake (Eds.), *The sixteenth mental measurements yearbook* (pp. 517-520). Lincoln: University of Nebraska.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). New York: Guilford.

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment, 85,* 17-32. doi:

10.1207/s15327752jpa8501_02

Bryne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). New York: Routledge.

Canivez, G. L. (2013). Psychometric versus actuarial interpretation of intelligence and related aptitude batteries. In D. H. Saklofske, C. R. Reynolds, & V. L. Schwean (Eds.). *The Oxford handbook of child psychological assessment* (pp. 84-112). New York: Oxford University Press.

Canivez, G. L. (2016). Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation. In K. Schweizer & C. DiStefano (Eds.), *Principles and methods of test construction: Standards and recent advancements* (pp. 247-271). Gottingen, Germany: Hogrefe.

Canivez, G. L., Watkins, M. W., & Dombrowski, S. C. (2017). Structural validity of the Wechsler Intelligence Scale for Children-Fifth Edition: Confirmatory factor analyses with the 16 primary and secondary subtests. *Psychological Assessment, 29,* 458-472. doi: 10.1037/pas0000358

Carretta, T. R., & Ree, J. J. (2001). Pitfalls of ability research: *International Journal of Selection and Assessment, 9,* 325-335. doi: 10.1111/1468-2389.00184

Cohen, A., Fiorello, C. H., & Farley, F. H. (2006). The cylindrical structure of the Wechsler Intelligence Scale for Children-IV: A retest of the Guttman model of intelligence. *Intelligence, 34,* 587-591. doi: 10.1016/j.intell.2006.05.003

Dombrowski, S. C., McGill, R. J., & Canivez, G. L. (2018). An alternative conceptualization of the theoretical structure of the WJ IV Cognitive at school age? A Confirmatory factor analytic investigation. *Archives of Scientific Psychology, 6,* 1-13. doi:

10.1037/arc0000039

Drozdick, L. W., Singer, J. K., Lichtenberger, E. O., Kaufman, J. C., Kaufman, A. S., & Kaufman, N. L. (2018). The Kaufman Assessment Battery for Children-Second Edition and KABC-II Normative Update. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 333-359). New York: Guilford.

Ferrer, E., & McArdle, J. J. (2010). Longitudinal modeling of developmental changes in psychology research. *Current Directions in Psychological Science, 19,* 149-154. doi: 10.1177/0963721410370300

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

Gignac, G. E., & Kretzschmar, A. (2017). Evaluating dimensional distinctness with correlated-factor models: Limitations and suggestions. *Intelligence, 62,* 138-147. doi: 10.1016/j.intell.2017.04.001

Golay, P., Reverte, I., Rossier, J., Favez, N., & Lecerf, T. (2013). Further insights on the French WISC-IV factor structure through Bayesian structural equation modeling. *Psychological Assessment, 25,* 496-508. doi: 10.1037/a0030676

Guttman, L. & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence, 15,* 79-103. doi: 10.1016/0160-2896(91)90023-7

Hajovsky, D. B., Villeneuve, E. F., Reynolds, M. R., Niileksela, C. R., Mason, B. A., & Shudak, N. J. (2018). Cognitive ability influences on written expression: Evidence for developmental and sex-based differences in school-age children. *Journal of School Psychology, 67,* 104-118. doi: 10.1016/j.jsp.2017.09.001

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1-55. doi:10.1080/ 10705519909540118

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 Assessment Battery for Children-Second Edition manual*. Circle Pines, MN: American Guidance Service.

Kaufman, A. S., & Kaufman, N. L. (2018). *Kaufman Assessment Battery for Children-Second Edition Normative Update*. Bloomington, MN: NCS Pearson.

Kaufman, A. S., Kaufman, N. L., Drozdik, L. W., & Morrison, J. (2018). *Kaufman Assessment Battery for Children-Second Edition Normative Update manual supplement*. Bloomington, MN: NCS Pearson.

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. doi: 10.1016/j.intell.2012.01.009

Keith, T. Z., & Kranzler, J. H. (1999). The absence of structural fidelity precludes construct validity: Rejoinder to Naglieri on what the cognitive assessment system does and does not measure. *School Psychology Review, 28,* 303-321. Retrieved from http://www.nasponline.org

Kruskal, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika, 29,* 1-27. Doi: 10.1007/BF02289565

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Thousand Oaks, CA: Sage.

McGill, R. J. (2015). Interpretation of KABC-II scores: An evaluation of the incremental validity of CHC factor scores in predicting achievement. *Psychological Assessment, 27*, 1417-1426. doi: 10.1037/pas0000127

McGill, R. J. (2018). Confronting the base rate problem: More ups and downs for cognitive scatter analysis. *Contemporary School Psychology, 22,* 384-393. doi: 10.1007/s40688-017-0168-4

McGill, R. J., & Dombrowski, S. C. (2017). School psychologists as consumers of research: What school psychologists need to know about factor analysis. *Communiqué, 46 (1),* 16-18.

McGill, R. J., & Dombrowski, S. C. (2018). Factor structure of the CHC model for the KABC-II: Exploratory factor analyses with the 16 core and supplementary subtests. *Contemporary School Psychology, 22*, 279-293. doi: 10.1007/s40688-017-0152-z

McGrew, K. S. (2018, April 12). *Dr. Kevin McGrew and Updates to CHC Theory* [Video webcast]. Invited podcast presentation for School Psyched! Podcast presented 12 April 2018. Retrieved from https://itunes.apple.com/us/podcast/episode-64-dr-kevin-mcgrew-and-updates-to-chc-theory/id1090744241?i=1000408728620&mt=2

Meyer, E. M., & Reynolds, M. R. (2018). Scores in space: Multidimensional scaling of the WISC-V. *Journal of Psychoeducational Assessment, 36,* 562-575. doi: 10.1177/0734282917696935

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence, 37,* 487-494. doi: 10.1016/j.intell.2013.06.004

Muthen, L. K., & Muthen, B. O. (2017). *Mplus* [Version 8.0]. Los Angeles, CA: Muthen &

Muthen.

Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). New York: Holt, Rinehart, & Winston.

Potvin, D. C. H., Keith, T. Z., Caemmerer, J. M., & Trundt, K. M. (2015). Confirmatory factor structure of the Kaufman Assessment Battery for Children-Second Edition with preschool children: Too young for differentiation? *Journal of Psychoeducational Assessment, 33,* 522-533. doi: 10.1177/0734282914568538

R Core Team (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria.

Reynolds, M. R., & Keith, T. Z. (2007). Spearman's law of diminishing returns in hierarchical models of intelligence for children and adolescents. *Intelligence, 35,* 267-281. doi: 10.1016/j.intell.2006.08.002

Reynolds, M. R., & Keith, T. Z. (2013). Measurement and statistical issues in child assessment research. In D. H. Saklofske, C. R., Reynolds, & V. L. Schwean (Eds.), *The Oxford handbook of child psychological assessment* (pp. 48-83). New York: Oxford University Press.

Reynolds, M. R., Keith, T. Z., Fine, J. G., Fisher, M. E., & Low, J. (2007). Confirmatory factor structure of the Kaufman Assessment Battery for Children-Second Edition: Consistency with Cattell-Horn-Carroll theory. *School Psychology Quarterly, 22,* 511-539. doi: 10.1037/1045-3830.22.4.511

Reynolds, M. R., Keith, T. Z., Flanagan, D. P., & Alfonso, V. C. (2013). A cross-battery, reference variable, confirmatory factor analytic investigation of the CHC taxonomy. *Journal of School Psychology, 51,* 535-555. doi: 10.1016/j.jsp.2013.02.003

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*, 223-237. doi: 10.1080/00223891.2015.1089249

Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53-61. doi: 10.1007/BF02289209

Schneider, W. J., & McGrew, K. S. (2018). The Cattell-Horn-Carroll Theory of Cognitive Abilities. In D. P. Flanagan & E. M. McDonough (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (4th ed., pp. 73-163). New York: Guilford.

Scheiber, C. (2016). 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? *Psychological Assessment, 28,* 942-952. doi: 10.1037/pas0000236

Sotelo-Dynega, M., & Dixon, S. G. (2014). Cognitive assessment practices: A survey of school psychologists. *Psychology in the Schools, 51*, 1031-1045. doi: 10.1002/pits.21800

Villeneuve, E. F., Hajovsky, D. B., Mason, B. A., & Lewno, B. M. (2019). Cognitive ability and math computation developmental relations with math problem solving: An integrated, multigroup approach. *School Psychology Quarterly, 34,* 96-108. doi: 10.1037/spq0000267

Vrieze, S. I. (2012). Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17,* 228-243. doi: 10.1037/a0027127

Wagenmakers, E-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin and Review, 11*, 192-196. doi: 10.3758/BF03206482

Footnotes

[1] Although the names of the KABC-II scores have not been changed in the NU, report updated nomenclature for broad and narrow test alignment consistent with recent changes to the CHC model in Table 1.2 (pp. 6-11). From this perspective, Long-Term Storage and Retrieval tasks are now thought to measure Learning Efficiency (Gl).

[2] Ironically, General Sequential Reasoning (RG) has been identified as a *well-supported* narrow ability within Gf in the CHC model, raising concern about potential jingle-jangle fallacies when using the nomenclature associated with the Luria model.

[3] Only the CHC model was explicated in the structural validity studies reported in the technical manual.

[4] It was suggested that this correlation was likely due to a problematic indicator (Conceptual Thinking) which was found to have strong loadings on both Gv and Gc. It is not clear how this cross-loading was discovered given the loading was not *theoretically* consistent. Curiously, this cross-loading was not modeled in the final validation models for that age.

[5] Pattern Reasoning is used to calculate the Fluid Reasoning Index score but does not contribute to the measurement of the Visual Processing Index on the KABC-II NU at ages 7-18.

[6] At Ages 4-6, Pattern Reasoning is assigned to Visual Processing.

Table 1

Confirmatory Factor Analysis Fit Statistics for KABC-II NU Ten Subtest CHC Configuration for Normative Sample Participants Ages 7-18 (N = 500)

Model	χ^2	*df*	*p*[a]	CFI	SRMR	RMSEA	90% CI RMSEA	AIC
Ages 7-12 (*n* = 300)								
Higher-Order Models								
1. One factor (*g*)	173.21	35	.00	.873	.060	.115	[.098, .132]	14247
2. Four oblique factors	49.32*	29	.01	.981	.028	.048	[.023, .071]	14135
3. Five Oblique factors	46.27	25	.01	.981	.027	.053	[.028, .077]	14140
4. Five Oblique factors	42.86	24	.01	.983	.026	.051	[.025, .076]	14138
Hierarchical and Bifactor Models								
5. Hierarchical Model 2	**50.66**	**31**	**.01**	**.982**	**.029**	**.046**	**[.021, .068]**	**14132**
6. Hierarchical (SC on *g*)	55.44	31	.00	.978	.031	.051	[.028, .073]	14137
7. Hierarchical Model 3	63.59	30	.00	.969	.035	.061	[.040, .082]	14147
8. Bifactor Model 3	63.65	30	.00	.969	.035	.061	[.040, .082]	14147
9. Hierarchical Model 4	49.72	29	.01	.981	.029	.049	[.024, .071]	14135
10 Bifactor Model 4	50.18	29	.01	.981	.029	.049	[.025, .072]	14136
11. Bifactor Model 6[b]	54.18	29	.00	.977	.031	.054	[.031, .076]	14140
12. Bifactor Model 2[b]	45.97	28	.01	.984	.026	.046	[.020, .070]	14134
Ages 13-18 (*n* = 200)								
Higher-Order Models								
1. One factor (*g*)	189.86	35	.00	.809	.073	.149	[.128, .170]	9438
2. Four oblique factors	42.58*	29	.04	.983	.033	.048	[.002, .078]	9303
3. Five Oblique factors	40.89	25	.02	.980	.032	.056	[.021, .087]	9309
4. Five Oblique factors			Model Failed to Converge, Number of Iterations Exceeded					
Hierarchical and Bifactor Models								
5. Hierarchical Model 2	**44.75**	**31**	**.05**	**.983**	**.035**	**.047**	**[.000, .076]**	**9301**
6. Hierarchical (SC on *g*)	52.25	31	.01	.974	.041	.059	[.029, .085]	9308

7. Hierarchical Model 3	57.90	30	.00	.966	.044	.068	[.041, .094]	9316
8. Bifactor Model 3			Model Specification Error, Negative Residual Variance (WO, RL and RI)					
9. Hierarchical Model 4	46.64	29	.02	.978	.038	.055	[.022, .083]	9307
10. Bifactor Model 4			Model Specification Error, Negative Residual Variance (WO and RL)					
11. Bifactor Model 6[b]			Model Specification Error, Negative Residual Variance (WO and RL)					
12. Bifactor Model 2[b]			Model Specification Error, Negative Residual Variance (WO)					

Note. KABC-II = Kaufman Assessment Battery for Children-Second Edition. CFI = comparative fit index; SRMR = standardized root mean square residual; RMSEA = root mean square error of approximation; AIC = Akaike information criterion. *g* = general intelligence, Gc = Crystallized Ability, PRI = Perceptual Reasoning Index, Gv = Visual Processing, Gsm = Short-Term Memory, Glr = Long-Term Storage and Retrieval, PR = Pattern reasoning, SC = Story Completion. RI = Riddles, WO = Word Order, RL = Rebus Learning. In Model 2 for ages 7-12, correlations ranged from .50 to .77. In Model 2 for ages 13-18, correlations ranged from .59 to .72. In Model 4, PR was permitted to load on both Gv and Gf.

[a] Values rounded to nearest hundredth. Estimates of zero are not considered point estimates.

[b] Group-specific factors with less than three indicators were constrained in order to ensure identification.

[*] Statistically different ($p < .05$) from previous models.

Table X (Online Supplement).

R^2 values for KABC-II NU CHC Core Battery Configuration Indicators According to a Hierarchical Measurement Model with Four Latent Factors.

Test	Factor	Ages 7-12	Ages 13-18
Verbal Knowledge	Gc	.659	.762
Riddles	Gc	.787	.798
Pattern Reasoning	PR	.534	.640
Story Completion	PR	.488	.281
Rover	PR	.358	.455
Triangles	PR	.454	*
Block Counting	PR	*	.412
Word Order	Gsm	.475	.719
Number Recall	Gsm	.505	.570
Atlantis	Glr	.455	.337
Rebus Learning	Glr	.642	.865

Note. Gc = Crystallized Ability, PR = Pattern Reasoning, Gsm = Short-Term Memory, Glr = Long-Term Storage and Retrieval.

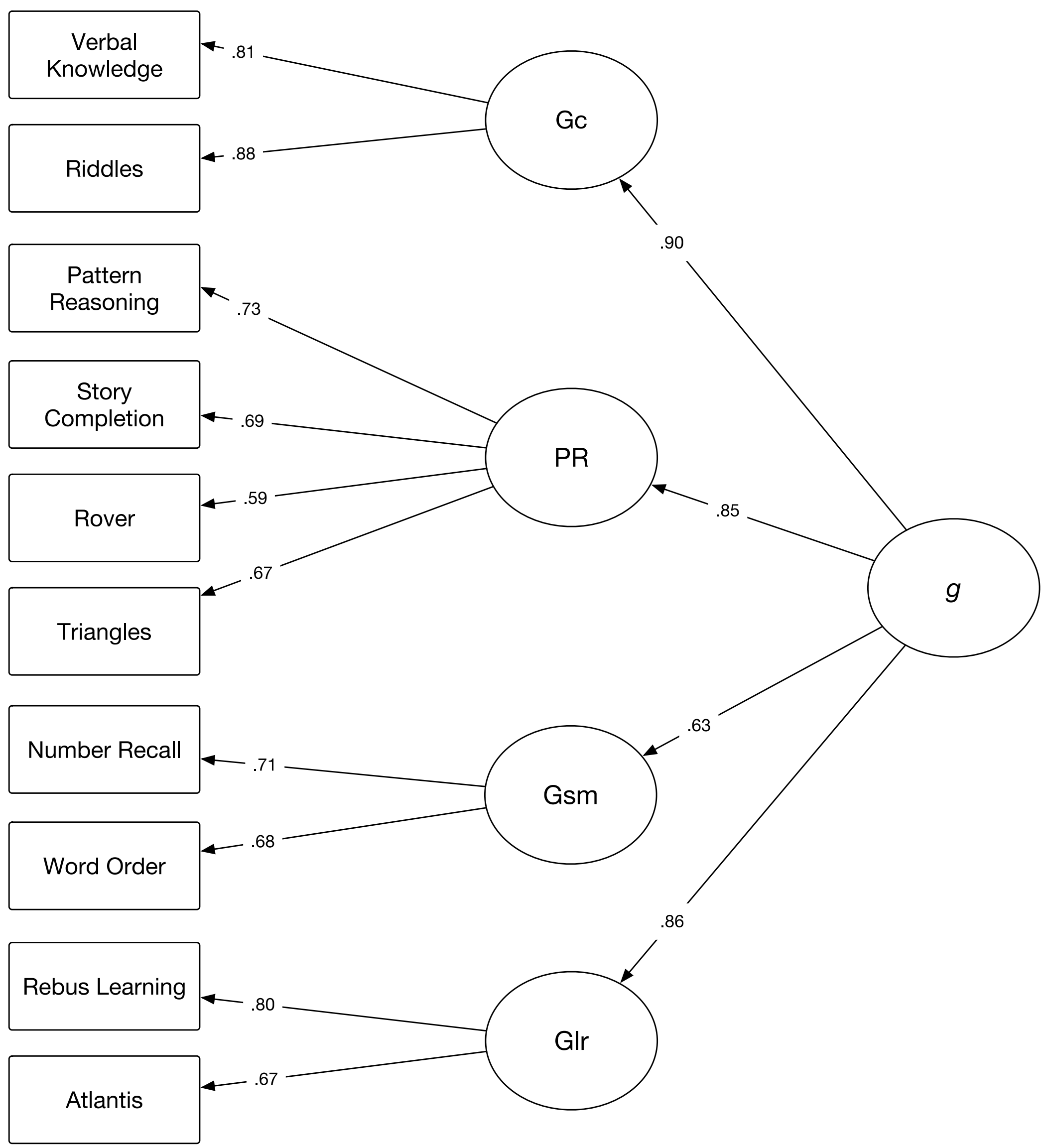

Figure 1. Four factor hierarchical measurement model (Model 5) with standardized loading coefficients for the KABC-II NU 10 subtest CHC interpretive configuration for ages 7-12. *g* = general intelligence, Gc = Crystallized Ability, PR = Perceptual Reasoning, Gsm = Short-Term Memory, Glr = Long-Term Storage and Retrieval. For the sake of clarity, residual terms are omitted.

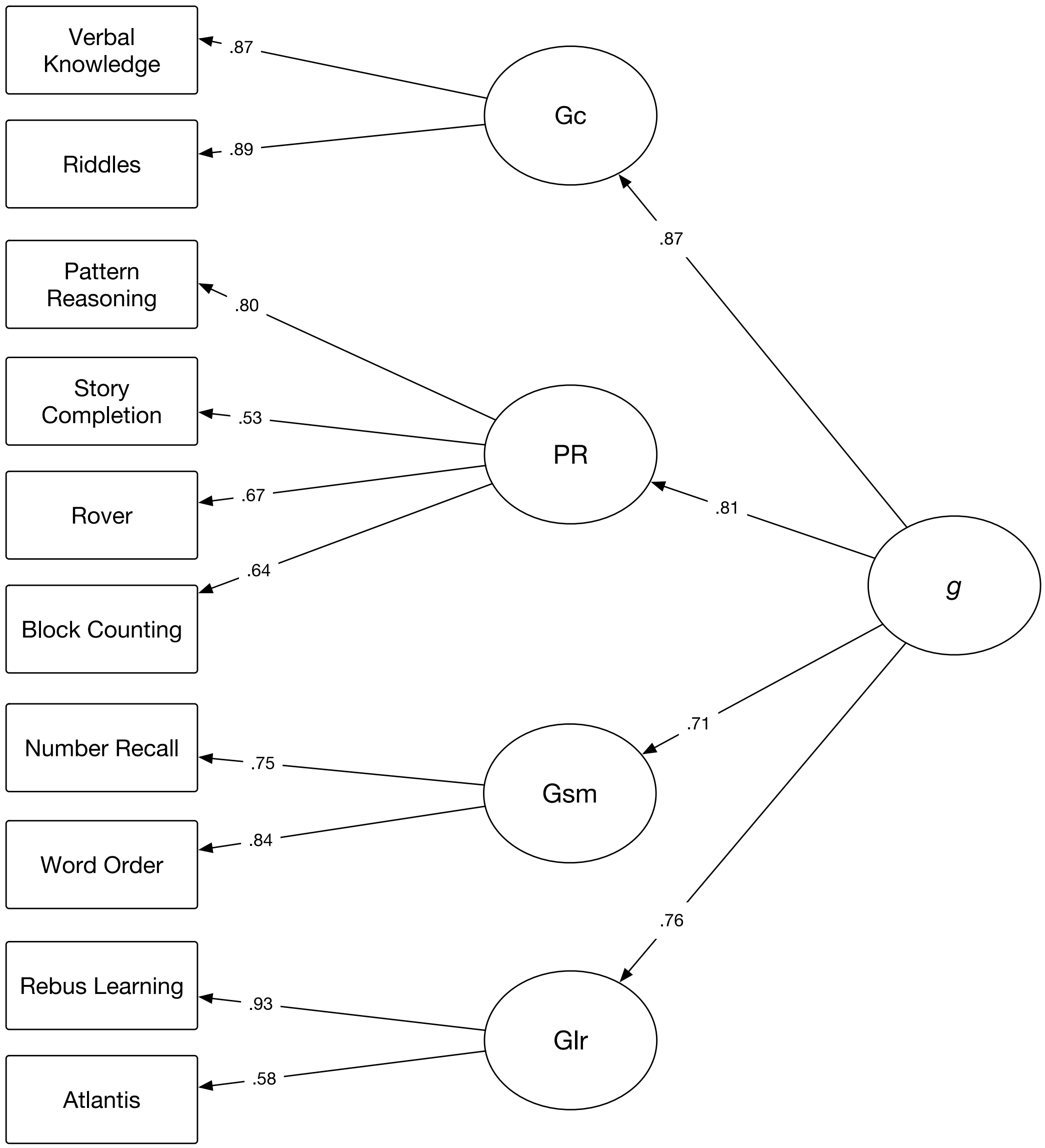

Figure 2. Four factor hierarchical measurement model (Model 5) with standardized loading coefficients for the KABC-II NU 10 subtest CHC interpretive configuration for ages 13-18. *g* = general intelligence, Gc = Crystallized Ability, PR = Perceptual Reasoning, Gsm = Short-Term Memory, Glr = Long-Term Storage and Retrieval. For the sake of clarity, residual terms are omitted.

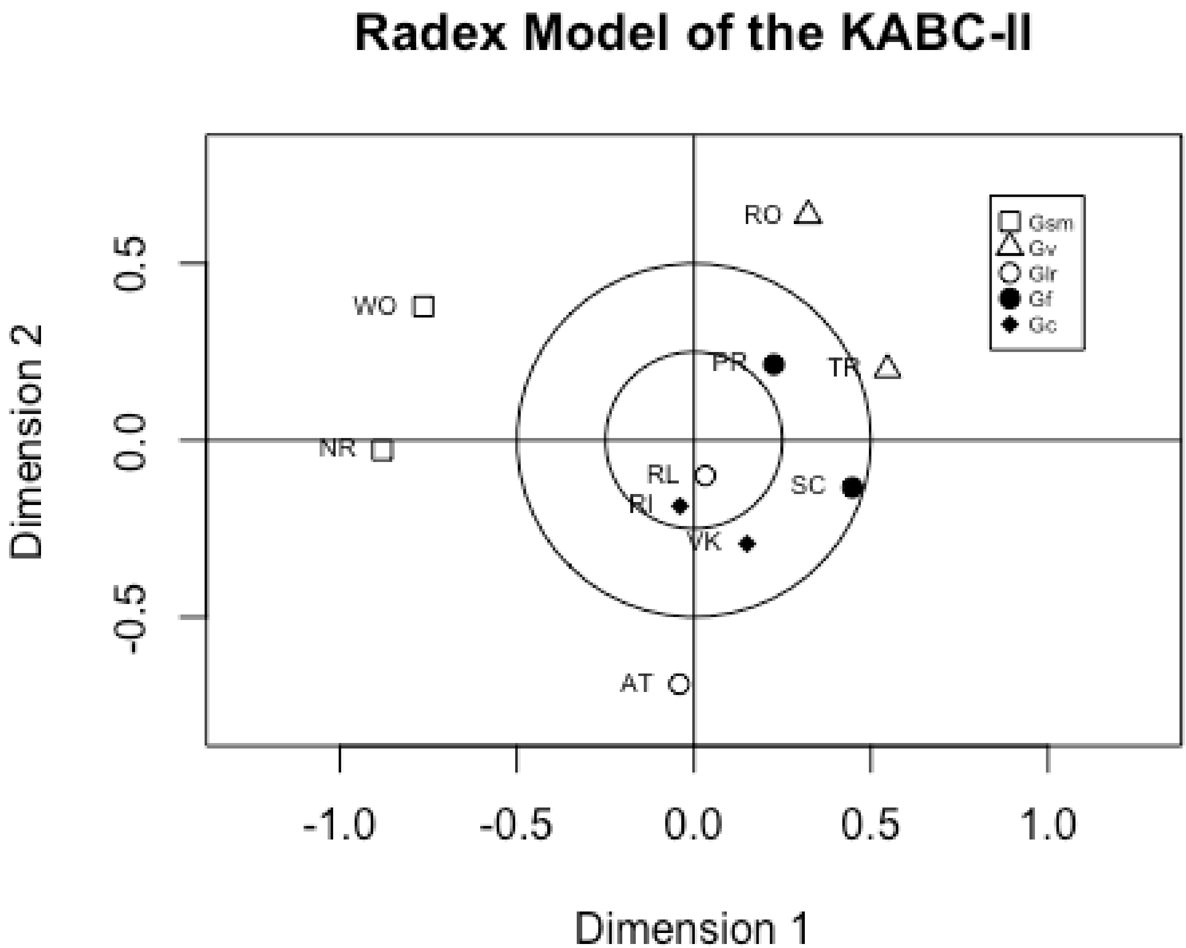

Figure 3. 2-D Nonmetric MDS Map for the KABC-II NU 10 Subtest CHC Core Battery Configuration for Ages 7-12. Gsm = Short-Term Memory, Gv = Visual Processing, Glr = Long-Term Storage and Retrieval, Gf = Fluid Reasoning, Gc = Crystallized Ability. WO = Word Order, NR = Number Recall, RL = Rebus Learning, AT = Atlantis, RI = Riddles, VK = Verbal Knowledge, RO = Rover, TR = Triangles, PR = Pattern Reasoning, SC = Story Completion.

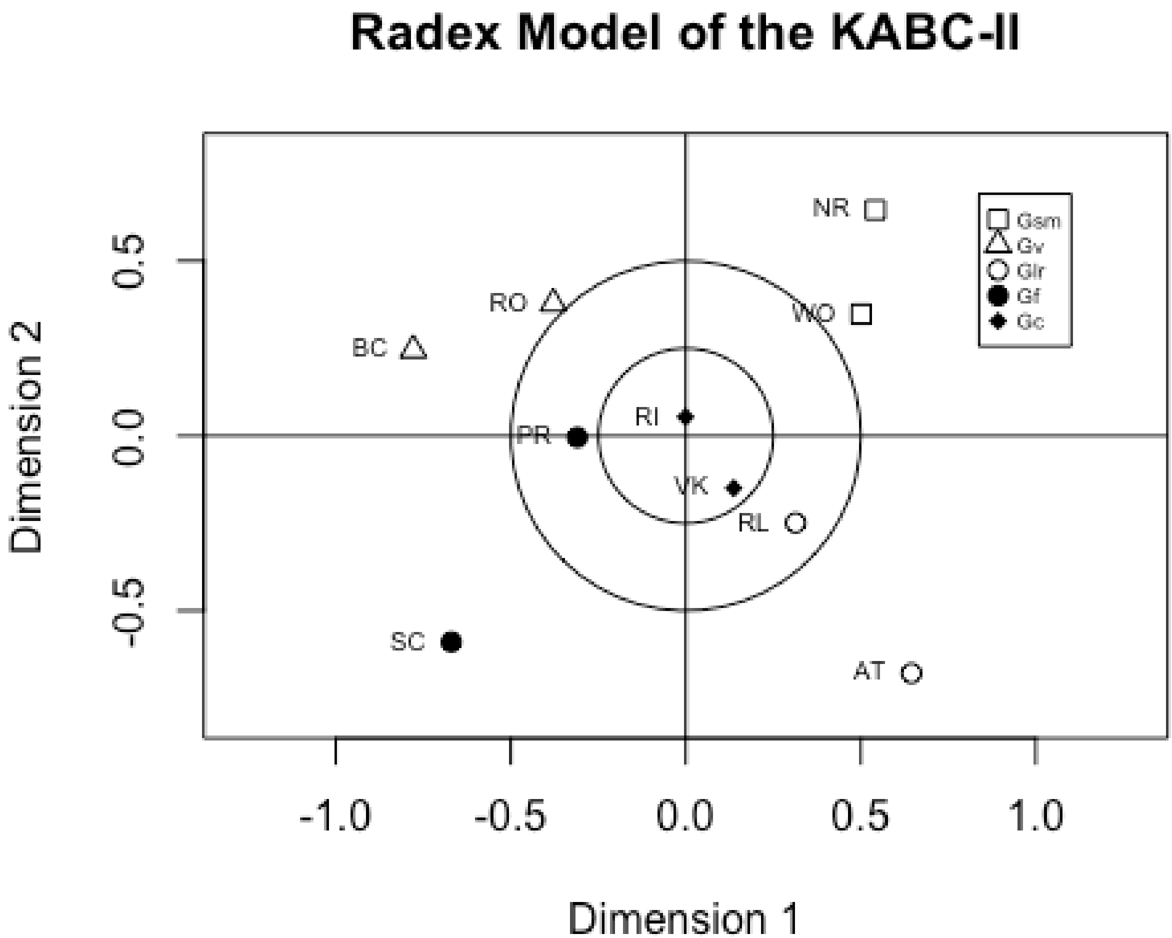

Figure 4. 2-D Nonmetric MDS Map for the KABC-II NU 10 Subtest CHC Core Battery Configuration for Ages 13-18. Gsm = Short-Term Memory, Gv = Visual Processing, Glr = Long-Term Storage and Retrieval, Gf = Fluid Reasoning, Gc = Crystallized Ability. WO = Word Order, NR = Number Recall, RL = Rebus Learning, AT = Atlantis, RI = Riddles, VK = Verbal Knowledge, RO = Rover, BC = Block Counting, PR = Pattern Reasoning, SC = Story Completion.