Assessing the Incremental Value of KABC-II Luria Model Scores in Predicting Achievement: What do they tell us beyond the MPI?

Ryan J. McGill

College of William & Mary

Angelia R. Spurgin

Texas Woman's University

Author note

Standardization data from the *Kaufman Assessment Battery for Children, Second Edition (KABC-II)*. Copyright © 2004 NCS Pearson, Inc. Used with permission. All rights reserved. Standardization data from the *Kaufman Test of Educational Achievement, Second Edition (KTEA-II)*. Copyright © 2004 NCS Pearson, Inc. Used with permission. All rights reserved.

Correspondence concerning this article should be addressed to Ryan J. McGill, School of Education, College of William & Mary, P. O. Box 8795, Williamsburg, VA 23187.

Abstract

The current study examined the incremental validity of the Luria interpretive scheme for the Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004a) for predicting scores on the Kaufman Test of Educational Achievement-Second Edition (KTEA-II; Kaufman & Kaufman, 2004c). All participants were children and adolescents ($N =$ 2,025) drawn from the nationally representative KABC-II/KTEA-II linked standardization sample. Consistent with previous studies, the full scale Mental Processing Index (MPI) score accounted for clinically significant portions of KTEA-II score variance in all of the regression models that were assessed. In contrast, the Luria factor scores collectively failed to provide meaningful incremental predictive variance after controlling for the effects of the MPI. Individually, the factor scores consistently accounted for trivial portions of achievement variance. Potential implications of these results for the correct interpretation of the KABC-II within clinical practice are discussed.

Keywords: incremental validity, KABC-II, clinical utility

Assessing the Incremental Value of KABC-II Luria Model Scores in Predicting Achievement: What do they tell us beyond the MPI?

The Kaufman Assessment Battery for Children-Second Edition (KABC-II; Kaufman & Kaufman, 2004a) measures the processing and cognitive abilities of children and adolescents between the ages of 3 years and 18 years. The KABC-II is a major structural and conceptual revision of the K-ABC (Kaufman & Kaufman, 1983): 10 measures were created and added to the current battery, item discrimination and scale ranges were increased, and the theoretical foundation for the battery was updated.

The KABC-II is grounded in a dual theoretical foundation featuring elements of Luria's neuropsychological model (1973) as well as the Cattell-Horn-Carroll (CHC) theory of cognitive abilities (Carroll, 1993; Schneider & McGrew, 2012). Examiners may select either the Luria or CHC interpretive models, though they must decide which model to use prior to testing. The interpretive models differ both in terms of factor structure (e.g., four factors versus five) as well as in content specifically as it relates to the inclusion of measures of acquired knowledge. Whereas, the CHC model includes measures of crystalized ability, the Luria model omits these measures. With an emphasis on the role of cognitive processing, the Luria model features eight core subtests, four first-order factor scores (Sequential Processing, Simultaneous Processing, Planning, and Learning), and a higher-order full scale Mental Processing Index (MPI) that is interpreted as reflecting an individual's overall level of cognitive processing.

Although the *Manual* (Kaufman & Kaufman, 2004b) advises users to interpret the KABC-II primarily from the CHC perspective, the Luria model is preferred in a variety of situations, including but not limited to, examining individuals from culturally and linguistically diverse backgrounds and assessing individuals known or suspected of having autism spectrum disorder (Kaufman, Lichtenberger, Fletcher-Janzen, & Kaufman, 2005). Accordingly, in a recent survey of 323 school psychologists, Sotelo-Dynega and Dixon (2014) found that the KABC-II was the preferred cognitive test battery of 20.4% of the practitioners surveyed when examining individuals from culturally and linguistically diverse backgrounds, making it the second most popular cognitive assessment used in those circumstances.

In terms of clinical interpretation, the Manual suggests that users should interpret the scores obtained from the KABC-II in a stepwise fashion beginning with the MPI and then proceeding to more specific measures (e.g., indexes and subtests). Despite this recommendation, it is suggested that the profile of strengths and weaknesses generated at the factor-level *is of more value* than the information provided by the MPI, and are valuable for diagnostic and educational purposes (p. 50). Inexplicably, Kaufman et al. (2005) suggest that the global score [MPI] provided by the KABC-II is not "considered particularly important for KABC-II interpretation" (p. 79). However, such prescriptive statements are rarely justified in applied practice and require adherence to high standards of empirical evidence (Marley & Levin, 2011). Although profile analysis and the primary interpretation of part scores on intelligence tests such as the KABC-II are popular in clinical practice, empirical support for the validity of these practices has repeatedly been found wanting (e.g., Macmann & Barnett, 1997; McDermott, Fantuzzo, & Glutting, 1990; McDermott, et al., 1992; Glutting, Watkins, & Youngstrom, 2003; Miciak et al., 2014; Watkins, 2000a; Watkins, Glutting, & Lei, 2007).

Additionally, simultaneous interpretation of KABC-II scores is potentially confounded by the hierarchical nature of the instrument. Interpreting scores at multiple levels (i.e., full scale, index scores, and subtest scores) ignores the fact that reliable subtest variance in all cognitive measures is multidimensional. That is, some common variance is apportioned to the general factor, some common variance is allocated to the broad cognitive abilities estimated by the first-order factors, and the remaining reliable variance (referred to as specificity) is unique to the subtest itself. Unfortunately, clinicians do not have a mechanism for disentangling these variances when interpreting measures at the level of the individual thus, interpretive redundancy may occur (Carroll, 1995; Schneider, 2013).

Whereas a wealth of psychometric information is provided in the Manual (Kaufman & Kaufman, 2004b) for the CHC model, equivalent information for the Luria model is absent. This is unfortunate as subsequent independent validity studies (e.g., Bangirana et al., 2009; Morgan, Rothlisberg, McIntosh, & Hunt, 2009; Reynolds, Keith, Flanagan, & Alfonso, 2013; Reynolds, Keith, Goldenring Fine, Fisher, & Low, 2007) of the KABC-II have been limited to examining the structural fidelity of the CHC model. Although results from these studies generally supported the five-factor model reported in the technical manual (Kaufman & Kaufman, 2004b), they cannot be taken as *de facto* evidence for the validity of the Luria model due to the fact that the CHC and Luria measurement models are not structurally equivalent (Cattell, 1978). Thus, until additional empirical evidence is provided, it cannot be assumed that the same latent constructs are measured in the Luria model (Nelson, Canivez, & Watkins, 2013).

Also missing from the KABC-II Manual were proportions of variance accounted for by the higher-order *g* factor and the proposed first-order factors, subtest *g* loadings, subtest specificity estimates, and incremental predictive validity estimates for the factors and subtest scores. Thus, clinicians do not have the necessary information for determining the relative importance of index and subtest scores relative to the MPI score. If the factor or subtest scores fail to capture meaningful portions of true score variance they will likely be of limited clinical utility. The omission of incremental predictive validity results is especially troubling given that users are encouraged to interpret the KABC-II beyond the MPI despite results from numerous incremental validity investigations of other cognitive measures that have not supported interpretation beyond similar full scale IQ (FSIQ) scores (see Canivez, 2013b for a review). Thus, in order to evaluate the utility of the interpretive practices that are recommended for practitioners in the KABC-II Manual (Kaufman & Kaufman, 2004b), it is necessary to examine the predictive validity provided by lower-order measures after controlling for the effects of variance already accounted for by the MPI (Glutting et al., 2006, Haynes, Smith, & Hunsley, 2011).

Incremental Predictive Validity

Since predicting achievement is a primary use of intelligence tests (Brown, Reynolds, & Whitaker, 1999; Gottfredson, 1997), examining relationships between cognitive variables (e.g., index scores) and external measures of achievement are an important component of establishing the external validity of a cognitive measure. According to Hunsley (2003), incremental validity is the “extent to which a measure adds to the prediction of a criterion beyond what can be predicted with other data” (p. 443). Incremental validity is rooted in the scientific law of parsimony which states “what can be explained by fewer principles is needlessly explained by more” (Jones, 1952, p. 620). These analyses directly address the question of whether data from one or more assessment sources increase the utility of clinical judgment. When applied to intelligence tests, interpretation of the FSIQ score is more parsimonious than interpretation at the first-order level of ability via an examinee’s profile of factor scores. Thus, to interpret primarily at the factor score level, practitioners should have a compelling reason for doing so (Glutting, Watkins, & Youngstrom, 2003).

Hierarchical multiple regression analysis is a well-established statistical procedure for assessing incremental validity in the social sciences and has been successfully applied in the technical literature in studies utilizing cognitive assessment data. In this procedure, the FSIQ is entered first into a regression equation followed by the lower-order factor scores to predict a criterion achievement variable. This entry technique allows for the predictive effects of the factor scores to be assessed while controlling for the effects of the FSIQ score and operates conceptually in very much the same way as the Schmid and Leiman technique (1957) for residualizing variance in exploratory factor analysis (EFA).

Incremental validity studies using hierarchical multiple regression analysis have been conducted on various iterations of the Wechsler scales (Canivez, 2013a; Glutting et al., 2006; Glutting, Youngstrom, Ward, Ward, & Hale, 1997), the Cognitive Assessment System (Canivez, 2011b), the Differential Ability Scales (Youngstrom, Kogos, & Glutting, 1999), the Reynolds Intellectual Assessment Scales (Nelson & Canivez, 2012), and the Woodcock-Johnson Tests of Cognitive Abilities (McGill, 2015a; McGill & Busse, 2015). Across these studies, it was consistently demonstrated that the omnibus FSIQ score on intelligence tests accounted for most of the reliable achievement variance in the regression models and that little additional incremental variance was accounted for by subordinate factor scores after controlling for the predictive effects of the FSIQ.

Although, it should be noted that Hale, Fiorello, Kavanaugh, Holdnak, and Aloe (2007) suggest these results are an artifact of method variance due to the threat posed by multicollinearity. According to Pedhazur (1997), multicollinearity is a potential threat to validity in multiple regression research that is introduced when a prediction model utilizes independent variables that are significantly correlated. As a result, it has long been suggested (e.g., Hale et al., 2007; Keith, 2015) that the predictive effects of lower-order scores are suppressed because they lack the freedom to vary from the FSIQ. Nevertheless, hierarchical multiple regression remains the "most common analytic strategy for quantifying the incremental contributions of specific methods, items, or measures to existing assessments" (McFall, 2005, p. 320).

With respect to the KABC-II, McGill (2015b) utilized similar procedures to assess the incremental predictive validity of the CHC model scores and found that whereas the Fluid-Crystallized Index (FCI) accounted for clinically significant portions of criterion achievement variance, the CHC factor scores accounted for trivial portions of achievement after controlling for the effects of the FCI. Again, these results are not useful for making inferences about the clinical utility of Luria model scores as the measurement model for that interpretive scheme differs from the CHC model. More germane to the present discussion, a recent higher-order EFA of the Luria model (McGill & Spurgin, 2015) found that the general factor produced from that model, was noticeably weaker and accounted for less common variance in KABC-II subtests when compared to equivalent estimates provided for the CHC model (e.g., Reynolds et al., 2007). Thus, it is possible that the first-order Luria scores may account for greater portions of incremental variance beyond the MPI as they contain less *g* variance as a result of the omission of measures of crystallized ability (DeThorne & Schaeffer, 2004; Jensen, 1984).

Therefore, direct examination of the predictive validity of the Luria interpretive scheme would benefit users of the measurement instrument. No incremental validity analyses were reported in the KABC-II Manual (Kaufman & Kaufman, 2004b), however, zero-order correlation coefficients between the MPI and the Luria factors ranged from .67 to .80 across the school-age span indicating moderate to strong relationships between these measures. To date, independent examination of the validity of the Luria interpretive scheme within the technical literature has been limited to the aforementioned EFA study (i.e., McGill & Spurgin, 2015), though it should be noted that construct validity studies alone are not sufficient for determining the relative importance of higher-order factors with respect to lower-order factors nor are they sufficient for answering questions of diagnostic utility or efficiency (Canivez, 2013b; Canivez, Konold, Collins, & Wilson, 2009).

Purpose of Current Study

To address this gap in the literature, the incremental validity of the KABC-II Luria scores in accounting for Kaufman Test of Educational Achievement-Second Edition (KTEA-II; Kaufman & Kaufman, 2004c) test scores beyond that already accounted for by the MPI was examined. Given the results of previous incremental validity studies, it was anticipated that there would be limited incremental prediction of achievement beyond the MPI. It is believed that results from the present study will be beneficial in establishing evidence-based interpretive procedures for the KABC-II in clinical practice.

Method

Participants

Participants were children and adolescents ages 7-0 to 18-11 ($N = 2{,}025$) drawn from the KABC-II/KTEA-II standardization sample. Demographic characteristics are provided in detail in the KABC-II Manual (Kaufman & Kaufman, 2004b). The standardization sample was obtained using stratified proportional sampling across demographic variables of age, sex, race/ethnicity, parent educational level, and geographic region. Examination of the tables in the Manual revealed a close correspondence to the 2001 U. S. census estimates across the stratification variables. The present sample was selected on the basis that it corresponded to the age ranges at which the Luria interpretive model could be fully specified as well as the fact that it permitted analyses of relationships between cognitive variables across a clinically relevant age span.

Measurement Instruments

Kaufman Assessment Battery for Children-Second Edition. The KABC-II is a multidimensional test of cognitive abilities for ages 3 to 18 years. The measure is comprised of 16 subtests, 9 of which contribute to the measurement of four Luria-based factor scores: Sequential Processing, Simultaneous Processing, Planning, and Learning. The eight core subtests combine linearly to form the full scale MPI composite. All factor and composite variables on the KABC-II are expressed as standard scores with a mean of 100 and a standard deviation of 15. The total norming sample (*N* = 3,025) is nationally representative based upon 2001 U.S. census estimates. Extensive normative and psychometric data can be found in the KABC-II Manual (Kaufman & Kaufman, 2004b). Mean internal consistency estimates for the included ages in this study ranged from .88 to .93 for the factor scores. The mean internal consistency estimate for the MPI was .95. Validity evidence is provided in several forms in the KABC-II manual and independent reviews are available (e.g., Bain & Gray, 2008; Thorndike, 2005).

Kaufman Test of Educational Achievement-Second Edition. The KTEA-II is a comprehensive academic assessment battery designed to measure four academic domains: Reading, Mathematics, Written Language, and Oral Language. The KTEA-II is comprised of 14 subtests that combine to yield 4 domain composites and a total achievement composite score. All scores are expressed as standard scores with a mean of 100 and a standard deviation of 15. Mean internal consistency estimates for the included ages in this study ranged from .90 to .97 for the subtest and composite scores that were assessed. Additional technical information for the KTEA-II can be found in the KTEA-II manual (Kaufman & Kaufman, 2004d).

Procedure

According to the KABC-II manual, all normative participants in the KABC-II dataset were administered measures from the KABC-II and the KTEA-II by trained examiners under the direct supervision of a standardization project member. Additionally, each examinee was given the KABC-II/KTEA-II in counterbalanced order to protect against potential threats to validity introduced by the sequencing of instruments.

Data Analyses

Hierarchical multiple regression analyses were conducted to assess the proportions of KTEA-II score variance accounted for by the observed KABC-II MPI and Luria factor scores. The MPI score was entered into the first block, and the Luria factor scores were entered jointly into the second block of the SPSS version 21 linear regression analysis. Luria factor effects also were individually assessed by entering each factor score alone into the second block of the regression equation. KTEA-II analyses included the Reading Composite, Reading Comprehension, Math Composite, Math Concepts and Applications, Math Calculation, Written Language Composite, Written Expression, Oral Language Composite, and Comprehensive Achievement Composite scores as criterion variables. The change in the KTEA-II achievement variance predicted by the Luria factor scores in the second block of the regression model provided an estimate of the incremental prediction beyond the MPI in the first block of the model. According to Dane and Dawes (2007), these variance partitioning procedures are appropriate for determining the degree to which including additional information beyond the MPI, significantly increments the prediction of a relevant criterion.

The results were interpreted using the resulting R^2 statistic as an effect size. Guidelines for interpreting R^2 as an effect size are found in Cohen (1988); they are “small,” .01; “medium,”

.09; and "large," .25. The critical coefficient in hierarchical multiple regression analysis is the incremental squared multiple correlation coefficient (ΔR^2). The ΔR^2 represents the amount of variance that is explained by an independent variable (IV) after controlling for the effects of IVs previously entered in the regression equation. At present there are no conventional guidelines for interpreting the ΔR^2 coefficient, thus Cohen's interpretive framework for R^2 was applied.

Due to the fact that the KABC-II utilizes different subtest combinations to form the Visual Processing factor across the 7 to 18 age span, separate analyses were conducted for ages 7 through 12 and 13 through 18 to account for these effects.

Results

The means, standard deviations, skewness, and kurtosis statistics for all of the study variables are listed in Table 1. The mean (99.60 to 100.38) and standard deviation ranges (14.08 to 15.14) for the cognitive and achievement variables generally reflect values that would be expected for normally distributed standard score variables. Skewness values provided evidence of normally distributed symmetry for all the variables, ranging from -0.06 to 0.09. Additionally, inspection of the residual plots of the data indicated that the regression models utilized in this study met the assumptions for homoscedasticity of the residuals.

Ages 7 to 12 (n = 1,141)

Table 2 presents the results from hierarchical multiple regression analyses for the KTEA-II scores for participants aged 7 to 12 years. The MPI accounted for statistically significant ($p <$ .05) portions of each of the KTEA-II achievement scores. Across the 9 regression models utilized to predict Reading, Reading Comprehension, Math, Math Concepts and Applications, Math Calculation, Written Language, Written Expression, Oral Language, and Comprehensive Achievement skills on the KTEA-II, the MPI accounted for 28% (Math Calculation) to 55%

(Comprehensive Achievement; *Mdn* = 43%) of the criterion variable variance. The R^2 values that corresponded to those variance increments all indicate large effects using Cohen's (1988) interpretive guidelines.

Luria factor scores entered jointly into the second block of the regression equations accounted for 0% (Comprehensive Achievement) to 2% (Math Applications; *Mdn* = 1%) of the incremental variance. The ΔR^2 values that corresponded to those variance increments were indicative of small to negligible effects. The incremental variance coefficients attributed to individual KABC-II Luria factor scores ranged from 0% to 1%. Although tests of significance indicated that the Luria factors on the KABC-II contributed statistically significant portions of incremental achievement variance beyond the effects of the MPI, effect size estimates were consistently small across all of the prediction models for the 7 to 12 age group. A post hoc power analysis revealed that for each of the IV blocks, $R^2/\Delta R^2$ effect sizes of less than .02 could be reliably detected with α set at .05, with power at .97 or greater in all of the regression models.

Ages 13 to 18 (*n* = 883)

Table 3 presents the results from hierarchical multiple regression analyses for the KTEA-II scores for participants aged 13 to 18 years. The MPI accounted for statistically significant ($p <$ .05) portions of each of the KTEA-II achievement scores. Across the 9 regression models utilized to predict Reading, Reading Comprehension, Math, Math Concepts and Applications, Math Calculation, Written Language, Written Expression, Oral Language, and Comprehensive Achievement skills on the KTEA-II, the MPI accounted for 37% (Written Expression) to 56% (Comprehensive Achievement; *Mdn* = 39%) of the criterion variable variance. The R^2 values that corresponded to those variance increments all indicate large effects using Cohen's (1988) interpretive guidelines.

Luria factor scores entered jointly into the second block of the regression equations accounted for 2% (Comprehensive Achievement, Reading Composite, Reading Comprehension, Math Composite, Math Calculation, and Math Applications) to 4% (Written Expression; *Mdn* = 2%) of the incremental variance. The ΔR^2 values that corresponded to those variance increments were indicative of small to negligible effects. The incremental variance coefficients attributed to individual KABC-II Luria factor scores ranged from 0% to 2%. Although tests of significance indicated that the Luria factors on the KABC-II contributed statistically significant portions of incremental achievement variance beyond the effects of the MPI, effect size estimates were consistently small across all of the prediction models for the 13 to 18 age group. A post hoc power analysis revealed that for each of the IV blocks, $R^2/\Delta R^2$ effect sizes of less than .02 could be reliably detected with α set at .05, with power at .92 or greater in all of the regression models.

Discussion

The present study assessed the incremental validity of the KABC-II Luria model scores in predicting academic achievement beyond that provided by the MPI. Hierarchical multiple regression analyses were used to determine the extent to which KABC-II Luria model scores provided meaningful improvements in prediction of KTEA-II scores beyond the MPI. For the KTEA-II, the MPI provided statistically significant prediction with large effect sizes for all of the KTEA-II achievement scores that were assessed across the 7-12 and 13-18 age ranges. These findings are consistent with those obtained from many intelligence and academic achievement batteries (Canivez, 2013b). While statistically significant improvements in prediction of KTEA-II scores were observed for the combined KABC-II Luria factor scores, effect size estimates (R^2 < .05) were consistently small. Additionally, none of the factor scores individually accounted for meaningful effects in any of the regression models that were assessed.

These results are similar to previous findings (McGill, 2015b) with the alternative CHC interpretive model. However, direct comparison with the incremental validity results from McGill (2015b) showed that in the present study, the combined KABC-II Luria factor scores provided less incremental prediction of the KTEA-II Reading Comprehension score at ages 13-18 (2%) and the Oral Language Composite across the school-age span (1%-3%). In contrast, increased prediction was observed for Written Expression at ages 13-18. Additionally, the MPI accounted for smaller amounts of achievement when compared to predictive validity estimates for related FSIQ scores on comparable intelligence tests (e.g., Canivez, 2013a, Glutting et al., 2006). According to DeThorne and Schaeffer (2004), this predictive attenuation at the full scale level is likely the result of omitting measures of crystallized ability in the Luria interpretive scheme. However, despite the fact that the Luria model seems to estimate a relatively *weaker* general ability dimension, the incremental predictive effects of the first-order factor scores were consistently trivial after controlling for the effects of the MPI.

When considering why the Luria factor scores generally failed to account for meaningful achievement variance beyond the MPI, it is important to remember that all factor-level scores on intelligence tests are comprised of various mixtures of common, unique, and error variances (Carroll, 1993; 1995). In an examination of the structural validity of the Luria model, McGill and Spurgin (2015), found that all of the Luria subtest measures contained non-trivial portions of common variance associated with the higher-order dimension (*g*). The results from the present study suggest that when that *g* variance is accounted for in in the first-order Luria factors there is little reliable unique variance left that is useful for predicting norm-referenced achievement on the KTEA-II. If assessment and interpretation of first-order Luria scores is of critical importance, the test authors will likely need to increase the number of subtests estimating those dimensions in the forthcoming revision of the measurement instrument in order to expand the amount of specificity at that level (Canivez, 2011a). The current findings are especially relevant for practitioners who utilize the KABC-II to assess individuals suspected of having a learning disability (LD). Emerging models of LD identification (e.g., Flanagan, Alfonso, & Mascolo, 2011) encourage practitioners to evaluate relationships between an individual's profile of cognitive strengths and weaknesses and their relationship to achievement indicators. Whereas it may be possible for practitioners to account for general factor effects when interpreting primarily at the first-order ability level, contemporary LD identification models have yet to provide a mechanism for doing so.

As in other investigations (e.g., Canivez, 2013a; Glutting et al., 2006; McGill & Busse, 2015), multicollinearity of the MPI and the factor scores in the hierarchical multiple regression analyses was observed in the present study due to the linear combination of subtests to produce factor scores and the MPI. However, this redundancy is precisely the problem that practitioners must confront when simultaneously interpreting full scale and factor-level scores on intelligence tests such as the KABC-II. Additionally, it should be noted that multicollinearity is not a threat to validity in regression studies that are limited to interpreting the R^2 statistic (Cohen, Cohen, West, & Aikem, 2003), nor does it invalidate the use of hierarchical multiple regression analysis to detect improvements in R^2 such as those provided by the Luria factor scores beyond the MPI (Schneider, 2008). As previously discussed, all Luria factor scores on the KABC-II contain common variance associated with the MPI. As a result, practitioners who interpret Luria factor scores on the KABC-II, without accounting for the effects of the MPI risk overestimating the predictive effects of those lower-order scores (Glutting, et al., 2006).

Proponents of cognitive profile analysis (e.g., Hale et al., 2008; Hale et al., 2007) have long argued that the results produced from incremental validity studies are an artifact of entering factor scores into the second block of a regression equation, thereby constraining their predictive effects at the expense of the FSIQ. However, this argument fails to take into account the fact that whereas correlations between factor-level scores and FSIQ composites are strong, they are certainly not unitary. As a result, first-order scores have sufficient freedom to vary at differential levels of prediction and entering them subordinate to the FSIQ does not automatically obviate their potential predictive utility (Dawes, 1979; Wiggins, 1988). These results demonstrate well that while the additional specificity provided by these measures resulted in increased predictive efficacy, the clinical utility of these effects is likely to be minimal (e.g., Glutting et al., 1997).

Limitations

This study is not without limitations that should be considered when interpreting the results. The most important limitation of the present study is the use of an archived standardization sample. Although the sample was relatively large and nationally representative, additional research is needed to determine if these results generalize to more focal clinical populations (e.g., individuals suspected of having a learning disability).

Additionally, it is important to remember that this study was designed to be predictive in nature, which limits the explanatory inferences that can be drawn from the data. According to Pedhazur (1997), predictive studies are concerned primarily with determining the most optimal variables or set of variables for predicting an external criterion with a specific sample, whereas in explanatory investigations the goal of the research is to shed light on a relationship with results that will generalize to the population. Despite philosophical distinctions, the line between explanation and prediction often is blurred in behavioral research (Hempel, 1965). That is, the results of predictive research have implications for explanation however; they cannot be used alone for such purposes (Schneider, 2008).Whereas this necessarily limits the generalizability of the findings, it is worth noting that the present study utilized the same variables and reference sample upon which users of the KABC-II base their interpretive judgments in clinical practice. Nevertheless, alternative analyses using latent variable modeling would be beneficial for examining explanatory relationships between latent KABC-II\KTEA-II (as well as their revisions) cognitive-achievement dimensions.

Finally, although we have employed conventional guidelines (e.g., Cohen, 1988) for interpreting the R^2 statistic as an effect size, ridged application of these criteria has been criticized (e.g., Keith, 2015). For example, Dawes (1999) noted the incorporation of additional measures or scores with low incremental validity may be of value when predicting important outcomes such as risk for suicidality or other diagnostic conditions (e.g., LD). In these circumstances, clinicians may seek to account for as much of the variance in the criterion as possible therefore, an additional 2%-5% of predictive power may be important. However, clinicians have been advised to consider the cost-benefit ratio of additional time and resources required to obtain such information (Haynes & Lench, 2003; Yates & Taub, 2003). Nevertheless, users would benefit from additional research examining the sensitivity-specificity of the KABC-II in predicting KTEA-II outcomes to determine if the factor scores account for meaningful criterion variance at specific cut-points.

Conclusion

The results of this study do not support recommendations in the KABC-II Manual (Kaufman & Kaufman, 2004b), or other interpretive resources (e.g., Kaufman et al., 2005; Singer et al., 2012) that the first-order factor scores should be the primary point of interpretation with this instrument. In contrast, the results indicate that examiners should focus most of their interpretive weight on the MPI when using the Luria interpretive scheme because it consistently accounted for the largest amounts of variance across achievement indicators on the KTEA-II. The results from the current study add to the growing literature on the validity of the KABC-II, and raise additional questions (e.g., Braden & Ouzts, 2005; McGill & Spurgin, 2015) about the clinical and diagnostic utility of the factor scores produced by the Luria model. As a consequence, users are encouraged to utilize these measures cautiously, if at all, for important diagnostic decisions in educational settings.

References

Bain, K., & Gray, R. (2008). Test reviews: Kaufman, A. S., & Kaufman, N. L. (2004). Kaufman Assessment Battery for Children, Second edition. Circle Pines, MN: AGS. *Journal of Psychoeducational Assessment, 26,* 92-101. doi: 10.1177/0734282907300461

Bangirana, P., Seggane, M., Allebeck, P., Giordani, B., Chandy, C. J., Opoka, O. R., Byarugaba, J., Ehnvall, A., & Boivin, M. J. (2009). A preliminary examination of the construct validity of the KABC-II in Ugandan children with a history of cerebral malaria. *African Health Sciences, 9,* 186-192. Retrieved from http://www.ajol.info/index.php/ahs/article/view/49010/35360

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, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since "Bias in Mental Testing." *School Psychology Quarterly, 14,* 208-238. doi: 10.1037/h0089007

Canivez, G. L. (2011a). Hierarchical factor structure of the Cognitive Assessment System: Variance partitions from the Schmid–Leiman (1957) procedure. *School Psychology Quarterly, 26,* 305–317. doi:10.1037/a0025973

Canivez, G. L. (2011b, February). *Incremental predictive validity of Cognitive Assessment System PASS scores*. Poster presented at the 2011 Annual Convention of the National Association of School Psychologists, San Francisco, CA.

Canivez, G. L. (2013a). Incremental criterion validity of WAIS-IV factor index scores: Relationships with WIAT-II and WIAT-III subtest and composite scores. *Psychological Assessment, 25,* 484-495. doi: 10.1037/a0032092

Canivez, G. L. (2013b). 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., Konold, T. R., Collins, J. M., & Wilson, G. (2009). Construct validity of the Wechsler Abbreviated Scale of Intelligence and Wide Range Intelligence Test: Convergent and structural validity. *School Psychology Quarterly, 24,* 252-265. doi: 10.1037/a0018030

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.

Carroll, J. B. (1995). On methodology in the study of cognitive abilities. *Multivariate Behavioral Research, 30,* 429–452. doi:10.1207/s15327906mbr3003_6

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum Press.

Cohen, J. (1988). *Statistical power analysis for the behavior sciences* (2nd ed.). New York: Psychology Press.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Flanagan, D. P., Alfonso, V. C., & Mascolo, J. T. (2011). A CHC-based operational definition of SLD: Integrating multiple data sources and multiple data-gathering methods. In D. P. Flanagan & V. C. Alfonso (Eds.), *Essentials of specific learning disability identification* (pp. 233-298). Hoboken, NJ: John Wiley.

Dana, J., & Dawes, R. M. (2007). Comment of Fiorello et al. “Interpreting intelligence test results for children with disabilities: Is global intelligence relevant?” *Applied Neuropsychology, 14,* 21-25. doi: 10.1080/09084280701280379

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571-582. Doi: 10.1037/0003-066x.34.7.571

Dawes, R. M. (1999). Two methods for studying the incremental validity of a Rorschach variable. *Psychological Assessment, 11,* 297-302. doi: 10.1037/1040-3590.11.3.297

DeThorne, L. S., & Schaefer, B. A. (2004). A guide to child nonverbal IQ measures. *American Journal of Speech-Language Pathology, 13,* 275-290. doi: 10.1044/1058-0360(2004/029)

Glutting, J. J., Watkins, M. W., Konold, T. R., & McDermott, P. A. (2006). Distinctions without a difference: The utility of observed versus latent factors from the WISC-IV in estimating reading and math achievement on the WIAT-II. *Journal of Special Education, 40,* 103-114. doi: 10.1177/00224669060400020101

Glutting, J. J., Watkins, M. W., & Youngstrom, E. A. (2003). Multi-factored and cross-battery assessments: Are they worth the effort? In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological & educational assessment of children: Intelligence, aptitude, and achievement*. (2nd ed., pp. 343-376). New York: Guilford Press.

Glutting, J. J., Youngstrom, E. A., Ward, T., Ward, S., & Hale, R. L. (1997). Incremental efficacy of the WISC-III factor scores in predicting achievement: What do they tell us? *Psychological Assessment, 9,* 295-301. doi: 10.1037/1040-3590.9.3.295

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79-132. doi: 10.1016/S0160-2896(97)90014-3

Hale, J. B., Fiorello, C. A., Domont, R., Willis, J. O., Rackley, C., & Elliott, C. (2008). Differential Ability Scales-Second Edition (neuro)psychological predictors of math performance for typical children and children with math disabilities. *Psychology in the Schools, 45,* 838-858. Doi: 10.1002/pits.20330

Hale, J. B., Fiorello, C. A., Kavanuagh, J. A., Holdnack, J. A., & Aloe, A. M. (2007). Is the demise of IQ interpretation justified? A response to special issue authors. *Applied Neuropsychology, 14,* 37-51. doi:10.1080/09084280701280445.

Haynes, S. N., & Lench, H. C. (2003). Incremental validity of new clinical assessment measures. *Psychological Assessment, 15,* 456-466. doi: 10.1037/1040-3590.15.4.456

Haynes, S. N., Smith, G. T., & Hunsley, J. D. (2011). *Scientific foundations of clinical assessment*. New York: Routledge

Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.

Hunsley, J. (2003). Introduction to the special section on incremental validity and utility in clinical assessment. *Psychological Assessment, 15,* 443– 445. doi:10.1037/1040-3590.15.4.443

Jensen, A. R. (1984). The black-white difference on the K-ABC: Implications for future tests. *Journal of Special Education, 18,* 377-408. doi: 10.1177/002246698401800316

Jones, W. T. (1952). A history of western philosophy. New York: Harcourt-Brace.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.

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

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

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

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

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

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

Luria, A. R. (1966). *Human brain and psychological processes*. New York: Harper Row.

Macmann, G. M., & Barnett, D. W. (1997). Myth of the master detective: Reliability of interpretations for Kaufman's "intelligent testing" approach to the WISC-III. *School Psychology Quarterly, 12,* 197-234. doi: 10.1037/h0088959

Marley, S. C., & Levin, J. R. (2011). When are prescriptive statements in educational research justified? *Educational Psychology Review, 23*, 197-206. doi: 10.1007/s10648-011-9154-y

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

McDermott, P. A., Fantuzzo, J. W., Glutting, J. J., Watkins, M. W., & Baggaley, A. R. (1992). Illusions of meaning in the ipsative assessment of children's ability. *The Journal of Special Education, 25,* 504-526. doi: 10.1177/002246699202500407

McFall, R. M. (2005). Theory and utility: Key themes in evidence-based assessment. Comment on the special section. *Psychological Assessment, 17,* 312-323. doi: 10.1037/1040-3590.17.3.312

McGill, R. J. (2015a). Incremental criterion validity of the WJ-III COG clinical clusters: Marginal predictive effects beyond the general factor. *Canadian Journal of School Psychology, 30,* 51-63. doi: 10.1177/0829573514560529

McGill, R. J. (2015b). Interpretation of KABC-II scores: An evaluation of the incremental validity of Cattell-Horn-Carroll (CHC) factor scores in predicting achievement. *Psychological Assessment*. Advance online publication. doi: 10.1037/pas0000127

McGill, R. J., & Busse, R. T. (2015). Incremental validity of the WJ III COG: Limited predictive effects beyond the GIA-E. *School Psychology Quarterly, 30,* 353-365. doi: 10.1037/spq0000094

McGill, R. J., & Spurgin, A. R. (2015). Exploratory higher order analysis of the Luria interpretive model on the Kaufman Assessment Battery for Children-Second Edition (KABC-II) school-age battery. *Assessment*. Advanced online publication. doi: 10.1177/1073191115614081

Miciak, J., Fletcher, J. M., Stuebing, K. K., Vaughn, S., & Tolar, T. D. (2014). Patterns of cognitive strengths and weaknesses: Identification rates, agreement, and validity for learning disabilities identification. *School Psychology Quarterly, 29,* 21-37. Doi: 10.1037/spq0000037

Morgan, K. E., Rothlisberg, B. A., McIntosh, D. E., & Hunt, M. S. (2009). Confirmatory factor analysis of the KABC-II in preschool children. *Psychology in the Schools, 46,* 515-525. doi: 10.1002/pits.20394

Nelson, J. M., & Canivez, G. L. (2012). Examination of the structural, convergent, and incremental validity of the Reynolds Intellectual Scales (RIAS) with a clinical sample. *Psychological Assessment, 24,* 129-140. doi: 10.1037/a0024878

Nelson, J. M., Canivez, G. L., & Watkins, M. W. (2013). Structural and incremental validity of the Wechsler Adult Intelligence Scale-Fourth Edition with a clinical sample. *Psychological Assessment, 25,* 618-630. doi: 10.1037/a0032086

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

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

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

Schneider, W. J. (2008). Playing statistical Ouija board with communality analysis: Good questions, wrong assumptions. *Applied Neuropsychology, 15,* 44-53. doi: 10.1080/09084280801917566

Schneider, W. J. (2013). What if we took our models seriously? Estimating latent scores in individuals. *Journal of Psychoeducational Assessment, 31,* 186-201. doi: 10.1177/0734282913478046

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

Singer, J. K., Lichtenberger, E. O., Kaufman, J. C., Kaufman, A. S., & Kaufman, N. L. (2012). The Kaufman Assessment Battery for Children-Second Edition and the Kaufman Test of Educational Achievement-Second Edition. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 269-296). New York: Guilford Press.

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

Thorndike, R. 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. 520–522). Lincoln: University of Nebraska.

Watkins, M. W. (2000). Cognitive profile analysis: A shared professional myth. *School Psychology Quarterly, 15,* 465-479. doi: 10.1037/h0088802

Watkins, M. W., Glutting, J. J., & Lei, P. (2007). Validity of the full-scale IQ when there is significant variability among WISC-III and WISC-IV factor scores. *Applied Neuropsychology, 14,* 13-20. doi: 10.1080/09084280701280353

Wiggins, J. S. (1988). *Personality and prediction: Principles of personality assessment*. Malabar, FL: Krieger Publishing.

Yates, B. T., & Taub, J. (2003). Assessing the costs, benefits, cost-effectiveness, and cost-benefit of psychological assessment: We should, we can, and here's how. *Psychological Assessment, 15,* 478-495. doi: 10.1037/1040-3590.15.4.478

Youngstom, E. A., Kogos, J. L., & Glutting, J. J. (1999). Incremental efficacy of Differential Ability Scales factor scores in predicting individual achievement criteria. *School Psychology Quarterly, 14,* 26-39. doi: 10.1037/h0088996

Table 1

Univariate Descriptive Statistics for KABC-II/KTEA-II Cognitive-Achievement Variables

Variables	*N*	*M*	*SD*	Skewness	Kurtosis
MPI	2025	100.11	14.90	0.01	0.06
Sequential	2025	100.11	14.99	0.01	-0.06
Simultaneous	2025	100.13	14.94	0.01	0.01
Learning	2025	100.38	15.05	0.00	0.05
Planning	2024	100.09	14.97	0.02	-0.01
Reading Composite	2147	99.60	15.03	0.03	-0.08
Reading Comprehension	2147	99.75	14.75	-0.03	0.37
Mathematics Composite	2328	99.95	14.72	0.01	-0.05
Math Concepts & Applications	2520	99.84	14.86	0.09	0.18
Math Calculation	2328	100.02	14.08	-0.04	0.28
Written Language Composite	2145	99.90	14.95	0.01	-0.03
Written Expression	2520	99.69	15.14	-0.06	0.07
Oral Language Composite	2520	100.19	14.84	0.02	-0.02
Comprehensive Achievement	2145	99.86	15.00	0.01	-0.08

Note. MPI = Mental Processing Index. Obtained values rounded to the nearest hundredth.

Table 2

Incremental Contribution of Observed Kaufman Assessment Battery for Children-Second Edition Luria Model Scores in Predicting Achievement Scores from the Kaufman Test of Educational Achievement Test-Second Edition Beyond the MPI for Ages 7-12 (n = 1,141)

	Comprehensive Achievement			Reading Composite			Reading Comprehension		
Predictor	R^2	Δ R^2	Increment (%)[b]	R^2	Δ R^2	Increment (%)[b]	R^2	Δ R^2	Increment (%)[b]
MPI	.55*	-	55%	.47*	-	47%	.43*	-	43%
Luria Factors (*df* = 4)[a]	.55	.00	0%	.47	.01*	1%	.45	.01*	1%
Sequential	.55	.00	0%	.47	.00	0%	.43	.00	0%
Simultaneous	.55	.00	0%	.47	.01*	1%	.44	.00*	0%
Planning	.55	.00*	0%	.47	.00	0%	.44	.01*	1%
Learning	.55	.00	0%	.47	.00*	0%	.44	.00	0%

	Math Composite			Math Calculation			Math Applications		
Predictor	R^2	Δ R^2	Increment (%)[b]	R^2	Δ R^2	Increment (%)[b]	R^2	Δ R^2	Increment (%)[b]
MPI	.44*	-	44%	.28*	-	28%	.46*	-	46%
Luria Factors (*df* = 4)[a]	.46	.01	1%	.30	.01*	1%	.48	.02*	2%
Sequential	.44	.00	0%	.28	.00	0%	.46	.00	0%
Simultaneous	.45	.00	0%	.28	.00	0%	.47	.01*	1%
Planning	.45	.01*	1%	.29	.01*	1%	.46	.00*	0%
Learning	.45	.01*	1%	.28	.00	0%	.47	.01*	1%

	Written Language Composite			Written Expression			Oral Language Composite		
Predictor	R^2	Δ R^2	Increment (%)[b]	R^2	Δ R^2	Increment (%)[b]	R^2	Δ R^2	Increment (%)[b]
MPI	.39*	-	39%	.34*	-	34%	.36*	-	36%
Luria Factors (*df* = 4)[a]	.40	.01*	1%	.34	.01*	1%	.37	.01*	1%
Sequential	.39	.00	0%	.34	.00	0%	.36	.00	0%
Simultaneous	.40	.01*	1%	.34	.01*	1%	.36	.00	0%
Planning	.39	.00	0%	.34	.00	0%	.36	.00	0%
Learning	.39	.01*	1%	.34	.00	0%	.36	.00	0%

Note. MPI = Mental Processing Index composite score. All coefficients rounded to nearest hundredth, may not equate due to rounding.

[a] Degrees of freedom reflects controlling for the effects of the MPI.

[b] Represents proportion of variance accounted for by variables at their entry point into regression equation. $R^2/\Delta R^2$ values multiplied by 100.

*$p < .05$.

Table 3

Incremental Contribution of Observed Kaufman Assessment Battery for Children-Second Edition Luria Model Scores in Predicting Achievement Scores from the Kaufman Test of Educational Achievement Test-Second Edition Beyond the MPI for Ages 13-18 (n = 883).

	Comprehensive Achievement			Reading Composite			Reading Comprehension		
Predictor	R^2	ΔR^2	Increment (%)[b]	R^2	ΔR^2	Increment (%)[b]	R^2	ΔR^2	Increment (%)[b]
MPI	.56*	-	56%	.47*	-	47%	.38*	-	38%
Lurian Factors (*df* = 4)[a]	.57	.02*	2%	.49	.02*	2%	.40	.02*	2%
Sequential	.56	.00	0%	.47	.00	0%	.38	.00	0%
Simultaneous	.56	.01*	1%	.48	.01*	1%	.39	.01*	1%
Planning	.56	.01*	1%	.47	.00*	0%	.39	.01*	1%
Learning	.56	.00	0%	.47	.00	0%	.38	.00	0%

	Math Composite			Math Calculation			Math Applications		
Predictor	R^2	ΔR^2	Increment (%)[b]	R^2	ΔR^2	Increment (%)[b]	R^2	ΔR^2	Increment (%)[b]
MPI	.48*	-	48%	.39*	-	39%	.47*	-	47%
Lurian Factors (*df* = 4)[a]	.50	.02*	2%	.41	.02*	2%	.48	.02*	2%
Sequential	.48	.00*	0%	.40	.00	0%	.47	.01*	1%
Simultaneous	.48	.00	0%	.39	.00	0%	.47	.00	0%
Planning	.49	.01*	1%	.40	.01*	1%	.47	.01*	1%
Learning	.48	.00	0%	.39	.00	0%	.47	.00	0%

	Written Language Composite			Written Expression			Oral Language Composite		
Predictor	R^2	ΔR^2	Increment (%)[b]	R^2	ΔR^2	Increment (%)[b]	R^2	ΔR^2	Increment (%)[b]
MPI	.38*	-	38%	.37*	-	37%	.38*	-	38%
Lurian Factors (*df* = 4)[a]	.41	.03*	3%	.40	.04*	4%	.41	.03*	3%
Sequential	.38	.00	0%	.37	.00	0%	.39	.00	0%
Simultaneous	.40	.02*	2%	.38	.01*	1%	.39	.01*	1%
Planning	.39	.01*	1%	.38	.01*	1%	.39	.01*	1%
Learning	.39	.01*	1%	.37	.01*	1%	.38	.00	0%

Note. MPI = Mental Processing Index composite score. All coefficients rounded to nearest hundredth, may not equate due to rounding.

[a] Degrees of freedom reflects controlling for the effects of the MPI.

[b] Represents proportion of variance accounted for by variables at their entry point into regression equation. $R^2/\Delta R^2$ values multiplied by 100.

$*p < .05$.