Neurocognitive testing in West African children 3–6 years of age: Challenges and implications for data analyses

Florence Bodeau-Livinec, Leslie Davidson, Romeo Zoumenou, Achille Massougbodji, Michel Cot, Michael Boivin

Neurocognitive testing in West African children 3 to 6 years of age: challenges and implications for data analyses

Florence Bodeau-Livinec[a,b], Leslie L. Davidson[c], Roméo Zoumenou[d], Achille Masosugbodji[e], Michel Cot[f], Michael J. Boivin[g,h]

[a] Ecole des Hautes Etudes en Santé Publique, Département Épidémiologie et Biostatistiques, Rennes, France;

[b] Inserm UMR 1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team (Epopé), Center for Epidemiology and Statistics Sorbonne Paris Cité, DHU Risks in Pregnancy, Paris Descartes University, Paris, France; Email address : florence.bodeau-livinec@ehesp.fr

[c] Columbia University, Mailman School of Public Health and the College of Physicians and Surgeons, NY, USA; Email address: lld1@cumc.columbia.edu

[d] Institut de Recherche pour le Développement, Mère et enfant face aux infections tropicales, Faculté de Pharmacie de Paris, Université Paris Descartes, Paris, France; the Centre Biomédical des Cordeliers, Université Pierre et Marie Curie, Paris, France; and PRES Paris Sorbonne Cité, Université Paris Descartes, Paris, France; Email address: zoumenour@yahoo.fr

[e] Université d'Abomey-Calavi, Faculté des Sciences de la Santé, Cotonou, Bénin ; Email address: massougbodjiachille@yahoo.fr

[f] Institut de Recherche pour le Développement, Mère et enfant face aux infections tropicales, Faculté de Pharmacie de Paris, Université Paris Descartes, Paris, France; the Centre Biomédical des Cordeliers, Université Pierre et Marie Curie, Paris, France; and PRES Paris Sorbonne Cité, Université Paris Descartes, Paris, France; Email address: Michel.cot@ird.fr

[g] Departments of Psychiatry and Neurology & Ophthalmology, Michigan State University, 909 fee Road, Rm 321, West Fee Hall, East Lansing, MI, 48824 USA;

[h] Department of Psychiatry, University of Michigan, Ann Arbor, MI USA; Email address: boivin@msu.edu

Corresponding author: Florence Bodeau-Livinec, Quantitative Methods in Public Health Department, Ecole des Hautes Etudes en Santé Publique, Avenue Georges Sand, 93 La Plaine Saint-Denis, France

Email: florence.bodeau-livinec@ehesp.fr

Tel.: +33 (0)2 99 02 28 06

Declarations of interest: none.

ABSTRACT

Objective. When testing African children with developmental and cognitive standardized tests from high-income countries (HIC), investigators are uncertain as to whether to apply the HIC norms for these tests when standardizing a child's raw-score performance on the basis of age. The present study compared the construct validity of both raw and HIC-based standardized scores for the Mullen Scales of Early Learning (MSEL) and the Kaufman Assessment Battery in Children – 2nd edition (KABC-II) for Beninese children in a rural setting from three to six years of age.

Methods. Seventy-four children 3 to 4 yrs of age were assessed with the MSEL, and 61 eligible older children (5 to 6 yrs of age) were assessed with the KABC-II. Assessors spoke the instructions to the children and caregivers for the assessment items in the local language. The developmental quality of the home environment was evaluated with the Caldwell Home Observation for Measurement of the Environment (HOME) inventory, and a material possessions and housing quality checklist was used as a measure of socio-economic status (SES). Children's mothers were given the Raven's Progressive Matrices test (nonverbal cognitive ability), and the Edinburgh Postpartum Depression Scale (EPDS) (emotional wellbeing).

Results. For the MSEL, the 4-yr old group performed significantly better than the 3-yr old group on both the raw and standardized score comparisons for all scales. These differences were attenuated when using standardized scores, although the MSEL standardized cognitive composite score was still highly significant between years of age. When comparing 5- to 6-yr olds on KABC-II subtest and global scale performance, comparisons between the raw and standardized mean score performances were much less consistent. Generally, 6-yr olds performed significantly better than 5-yr olds on the raw score comparisons on the KABC-II subtests, but not so for standardized scores. Parent-child interactions assessed through the HOME measure was associated with both raw and standardized MSEL cognitive composite score outcomes on a multiple regression analysis. SES was the only significant predictor for KABC-II raw and standardized outcomes.

Conclusion. Standardization using HIC norms was not optimal, resulting in minimal impact to account for age when using the MSEL, and lower scores for oldest children compared with youngest children when using the KABC2. This is likely due to children in Benin drifting away from HIC-based norms with each passing year of age, systematically lowering standardized performance measures. These findings support the importance of having a local comparison group of reference or control children to allow for adjusted (for age, HOME, and SES) raw score comparisons when using western-based tests for developmental and neuropsychological evaluation.

KEY WORDS: Child development, neurocognition, Mullen Scales of Early Learning, Kaufman Assessment Battery for Children, HOME scale, Africa

HIGHLIGHTS

- 4-year old kids did better than 3-year olds on the Mullen Scales of Early Learning
- 6-year old kids did better than 5-yr olds on the Kaufman Assessment Battery – 2nd edition (KABC-II)
- These age differences were less consistent when standardizing with HIC norms
- Mullen and KABC-II performance predicted by maternal Ravens Matrices, HOME, & SES
- Mullen and KABC-II performance were not correlated with maternal depression

Neurocognitive testing in West African children 3 to 6 years of age: challenges and implications for data analyses

INTRODUCTION

Because of the paucity of neurocognitive testing developed and normed in sub-Saharan African countries (Semrud-Clikeman et al., 2016), several tests assessing child neurodevelopment developed in Western countries have been used in this region (Bangirana, John, et al., 2009b). Local or country-based normative data is usually not available for such measures, resulting in a variety of statistical strategies for standardizing test scores on the basis of age (Bergemann et al., 2012). With the Kaufman Assessment Battery for Children (KABC) for example, some studies standardize test scores using age-matched community controls ((Boivin et al., 2007; John et al., 2008), while others use standardized scores with foreign norms (Boivin et al., 2018).

Using norms from foreign countries is of worry when comparing data from different countries. However, comparing standardized scores using American norms amongst exposure or intervention groups within the same setting may be reasonable, assuming that the age distributions are comparable. In fact, this approach may be necessary if a test requires standardized subtest or scale scores in order to arrive at a global or composite measure. This may often be the case, given that local or even in-country norms for a given standardized test are usually not available in the African context when assessing children with developmental or neuropsychological test developed in high-income countries (HIC).

Irrespective, unless investigators have a representative control or reference group available, they may not be able to standardize their performance measures for age unless they use norms for that test originating in HICs. When doing so, construct validity is important but not often evaluated when deciding whether to use “western-based” norms in standardizing a child’s test performance on the basis of age. The principal objective of the present study was to compare the construct validity of raw and standardized scores from the Mullen Scales of Early Learning and the Kaufman Assessment Battery in Children – 2nd edition (KABC-II) in Beninese children of 3 to 6 years of age.

In general, when attempting to establish the construct validity of a given construct, it is necessary to consider a range of factors, including not only convergent or predictive validity (i.e., to demonstrate that the construct correlates with things that it should correlate with), but also divergent/discriminant validity (i.e., to demonstrate that the construct does not correlate with things that it should not correlate with). A multi-trait multi-method matrix is often utilized to assess these factors. In this case, we will attempt to do the same, comparing convergent validity over divergent/discriminant validity, which should enhance our ability to make reliable judgments regarding construct validity.

MATERIAL AND METHODS

Study Site and Population. The study population included children born of mothers enrolled in 2011 in a trial comparing 2 intermittent preventive treatments for malaria, the MiPPAD (Malaria in Pregnancy Preventive Alternative Drugs) study in Benin, West Africa. Pregnant women were followed from the second trimester of pregnancy through delivery, and offspring were followed from birth to 12 months of life. About 747 offspring were assessed within the TOVI project at one year of age. The aim of the overall project was to assess the impact of hemoglobin concentration during pregnancy on offspring development (Mireku et al., 2015). Our study population included 135 siblings of children enrolled in the TOVI project (Kobto G. Koura et al., 2013). The study was conducted in the district of Allada, a semi-rural area located 50 km north of Cotonou, the capital of Benin. The study participants were recruited in two health centers: Attogon, and Sékou.

Data Collection Procedure. On the occasion of a home visit for the sibling of one year of age (the TOVI infant) between June 2011 and March 2013, a questionnaire on socioeconomic status, the Raven matrices (Raven, 2000b) and the Edinburgh Postnatal Depression Scale (EPDS) for the mother (Cox, Holden, & Sagovsky, 1987a), and the Home Observation for the Measurement of the

Environment (HOME) Inventory (Caldwell BM, 2003) were administered. Between January 2013 and April 2013, child development was assessed by research nurses trained specifically in the use of the Mullen Scales of Early Learning (MSEL) and the Kaufman Assessment Battery in Children – 2nd edition (KABC-II) at local health centers in siblings of the TOVI child. These siblings were between three and six years of age. The MSEL and KABC-II instructions for parents were spoken in Fon, the local language. A psychology graduate from the university of Benin along with several study nurses worked together to translate the spoken instructions for these tests in to the local dialect of Fon spoken by mothers in the study site region. These process of adapting these measures to the local context is described in a previous study where the maternal questionnaires and MSEL were used with the same cohort of children at 1 year of age (K. G. Koura et al., 2013). Study nurses and the coordinator psychologist were specifically trained by a psychologist (MJB) and the principal investigator to administer the MSEL, the KABC-II, and the HOME.

Measures

Mullen Scales of Early Learning (MSEL). The MSEL covers various domains to assess childhood development (Mullen, 1995). The five Mullen Scales are Gross Motor, Fine Motor, Visual Reception, Receptive Language, and Expressive Language. After scoring all items and computing raw scores, these raw scores are converted into a standardized score (*T* score) for each of the five Mullen Scales according to American norms. *T* scores from the Fine Motor, Visual Reception, Receptive Language, and Expressive Language scales are converted into the Early Learning Composite score, which provides the general cognitive factor underlying all cognitive performance (Mullen, 1995). Because the MSEL was only available in English, spoken instructions were given in the local language of Fon, as described previously (Kobto G. Koura et al., 2013).

Kaufman Assessment Battery for Children, second edition (KABC-II). The 2nd edition of this test was published in 2004. In addition to simultaneous processing and sequential processing, and compared with the original KABC, it includes the global domain measures of learning and planning (reasoning). This test was already available in French (by Les Editions du Centre des Psychologie Apliquee (ECPA); French translation published and distributed exclusively by ECPA with the permission of AGS Publishing, Pearson Products Inc.) normed on a population of children in France, and used as such in Benin for the purpose of the study.

Socioeconomic status. We used two variables to assess family socioeconomic status family wealth and maternal education. The family wealth scale has been described elsewhere (Kobto G. Koura et al., 2013). Briefly, it was assessed using a scoring instrument incorporating a checklist of material possessions (radio, television, bike, motorbike, and car), possession of cows and access to electricity. Maternal education included schooled and unschooled.

The Home Observation for the Measurement of the Environment (HOME) (Caldwell & Bradley, 2003). This evaluation was done in the home of the children and their mother. It was adapted and piloted for this setting (Kobto G. Koura et al., 2013).

Raven's Progressive Matrices (Raven, 2000a). This is a nonverbal test of cognitive ability for adults, often used in cross-cultural settings and used in this study to assess the mother's global nonverbal cognitive ability. The matrices are made up of a series of diagrams or designs with one part missing.

Edinburgh Postpartum Depression Scale (EPDS). We used the EPDS to assess maternal depressive symptoms (Cox, Holden, & Sagovsky, 1987b; Hanlon et al., 2008; Kakyo, Muliira, Mbalinda, Kizza, & Muliira, 2012). Scores derived from the EPDS were analyzed as a continuous variable. The EPDS, already available in French, was translated into Fon by consensus by our study psychologist and several nurses as a team, in that they knew the local dialect spoken my mothers in at our two study sites (K. G. Koura et al., 2013).

Statistical analysis. Subjects characteristics were first described in total and according to the test involved (MSEL or KABC-II). Then, for each test, mean raw and composite scores were compared between 3 and 4 years of age for the MSEL, as well as a comparison between children 5 and 6 years of age for the KABC-II. Finally, socioeconomic variables and other psychometric testing known to be associated with poor child development were included in the model without and then with age at assessment. The dependent variable was the Early Learning Composite raw and standardized scores for the MSEL, respectively, and the Mental Composite Index (MPI) raw and standardized scores for the KABC-II, respectively. Chi-square tests (or Fischer's exact test as appropriate) and *t*-tests were used to compare percentages and means, respectively. Pearson correlations were computed between the HOME subscales, EPDS, Raven Progressive Matrices, and marital status and scores of Mullen (ECL) and KABC-II (MPI). Multiple linear regressions were used for multivariate analyses, in order to evaluate the significance of various home environment and maternal predictive variables for both the raw and standardized score outcomes. A P-value < 0.05 was considered significant for these predictors in a multiple regression model, and for significant between age-group differences. All the analyses were performed with SAS version 9.2 software.

Human Subjects Protection. The study's protocol was approved by the University of Abomey-Calavi's institutional review board (IRB), New York University, Michigan State University IRBs, and the French Institut de Recherche pour le Développement's (IRD) Consultative Ethics Committee. All women who participated in this study signed informed consent before enrollment.

RESULTS

Table 1 describes the population characteristics. In total, 135 children were included in the study, 74 being assessed with the MSEL and 61 for the KABC-II. Only 38% of mothers reported completing any formal schooling. There was no difference regarding maternal education, HOME, EPDS, SES, infant's gender, and maternal marital status between children assessed by the MSEL and those assessed by the KABC-II.

Each scale from the MSEL showed higher raw scores at 4 years of age compared with 3 years of age (Table 2). As far as standardized scores were concerned, the four MSEL scales still showed higher scores at 4 years of age compared with 3 years of age, but the differences were not significant for two of the scales (receptive and expressive language).

For the KABC-II, raw scores at 6 years of age were all higher than scores at 5 years of age on all but one subtest. However, the between age-group difference was significant for only three of the ten subtests (Table 3). For the standardized scores using French norms, mean performances were lower at 6 years of age compared with 5 years of age for seven out of nine subtests (Table 3). The difference was statistically significant for two of these.

In establishing the construct validity present MSEL composite cognitive development and KABC-II global cognitive ability (MPI) assessment measures, HOME total score as a measure of caregiving quality was significantly correlated with the Raw score sums of both of these and with the KABC-II MPI standardized score (Table 4). HOME caregiving quality was also significantly correlated with Raven Progressive Matrices performance of the mother and the Family Wealth Index. Likewise, maternal Raven Progressive Matrices performance and Family Wealth Index was significantly related to KABC-II MPI raw and standardized global scores for the child, but not the MSEL cognitive composite. Edinburgh maternal depression and maternal marital status was not significantly correlated to HOME caregiving quality or to any of the raw or standardized MSEL cognitive composite or KABC-II MPI measures.

The multivariate analyses evaluated the predictive significance of SES, EPDS, Raven's Progressive Matrices, HOME, maternal marital status, and year of age. These were assessed within a multiple regression analysis model (with and without year of age), for both raw and standardized MSEL composite cognitive and KABC-II Mental Processing Index (MPI) as the outcomes (Table 5). Parent-child interactions assessed with the HOME inventory was associated with both raw and standardized MSEL cognitive composite score outcomes. SES was the only predictor significant for KABC-II raw and standardized outcomes. Furthermore, the total proportion of explained variance for the outcome (R squared) for the multiple regression model increased dramatically from .05 to .68 when adjusting for year of age in the model. For the KABC-II, SES was associated with the Mental Processing Index in both models, and R squared did not change after adjustment for age at assessment (Table 5).

DISCUSSION

In our Beninese sample, the MSEL and KABC-II showed reasonable construct validity in terms of the older age group performing better. This was especially the case for the raw score outcomes, and less so for the standardized scores using American (MSEL) or French (KABC-II) norms to adjust for age. Likewise, for both raw and standardized overall performance outcomes, overall performance measures were associated with known risk factors for poor child development, including the lower caregiving quality as assessed with the HOME inventory for the MSEL cognitive composite performance, and lower SES scores for the KABC-II MPI overall cognitive performance measures.

The MSEL was used within the TOVI study in one year old children (K. G. Koura et al., 2013). In this previous research, cognitive composite ability scores for the MSEL were associated significantly with the HOME, the Raven progressive matrices, SES, and maternal education. In this present study, the MSEL composite was only significantly associated with the HOME.

As noted previously, when attempting to establish the construct validity of a given construct, it is necessary to consider a range of factors, including not only convergent or predictive validity (i.e., to demonstrate that the construct correlates with things that it should correlate with), but also divergent/discriminant validity (i.e., to demonstrate that the construct does not correlate with things that it should not correlate with). HOME caregiving quality and Family Wealth Index (a measures of

SES) were significantly correlated with both the raw and standardized KABC-II MPI global cognitive performance measure of the children in the present study. This finding has been noted with the KABC-II in a previous study of Ugandan school-age children (Bangirana, John, et al., 2009a).

Likewise, the MSEL composite cognitive development score for our present sample was significantly related to HOME caregiving quality, also a finding previously noted in Ugandan preschool-age children affected by HIV (Bass et al., 2016). However, we did not see a significant relationship between Edinburgh postpartum depression of mothers and MSEL cognitive composite development; nor between Raven Progressive Matrices test performance for the mothers and MSEL cognitive composite (raw or standardized). This was the case as well in a larger sample of our Benin children at one-year of age with the MSEL cognitive composite (K. G. Koura et al., 2013). However, in that study, Raven Progressive Matrices was significant related to MSEL standardized composite cognitive, although this was with 1-year olds whereas our children were assessed with the MSEL at 3 to 4 years of age in the present study. Unexpectedly, maternal depression and the Ravens Progressive Matrices were not associated with the HOME subscales as a measure of quality of caregiving. Maternal depression of Ugandan mothers with HIV can be associated with quality of caregiving as measured by the HOME (Bass et al., 2016; Boivin, Bangirana, Nakasuja, et al., 2013; Boivin, Bangirana, Nakasujja, et al., 2013).Maternal depression was not associated with marital status and family wealth in the present samples, although socio-economic factors can contribute to maternal depression and anxiety in impoverished communities in Uganda for mothers with HIV (Familiar, Murray, et al., 2016).

Finally, the fact that maternal depression was not predictive of cognitive skill either at 1 or at 3 to 4 yrs of age in our cohort of children could be seen as an instance of divergent validity. Maternal depression typically does not directly impact on a child's cognitive ability level, but is more likely to impact on children's cognitive skill perhaps indirectly through compromised quality of caregiving in mothers clinically depressed. Boivin and colleagues have observed this relationship with Ugandan mothers with HIV and their exposed but non-infected children (Bass et al., 2016; Familiar, Nakasujja, et al., 2016; Murray et al., 2016).

Likewise, to see maternal performance on the Ravens Progressive Matrices test as not predictive of their infants MSEL cognitive performance at 1-year of age, but predictive at 3 to 4 years of age in the same cohort, may also be taken as evidence of divergent validity within the present cohort analysis. This is because the MSEL composite cognitive ability scores in these preschool-age children (3 to 4 yrs) would be more evident as it could relate to maternal nonverbal cognitive ability (Ravens Progressive Matrices), than would likely be the case for the infant at 1 year of age.

One limitation of our study is that the HOME was scored in relation to the TOVI child, not the older sibling included in the present KABC-II assessment and analyses. The HOME may vary from a child to another one in the same family. However, maternal education, the family wealth and the Raven matrices should not vary. One explanation is that the neurocognition, especially assessed by MSEL scores, may be more sensitive to socioeconomic factors in infancy and very early childhood, as compared to school-age cognitive ability as assessed with the KABC-II. However, the quality of the home environment was significantly related to KABC performance at school age for Ugandan children (Bangirana, John, et al., 2009a; Bangirana, Menk, John, Boivin, & Hodges, 2013; Bangirana, Seggane, et al., 2009).

Standardized scores are getting lower when child's age increases. Both tests were normed on a mean of 100 with a standard deviation of 15. Our results show low standardized scores: about 70 for the MSEL composite standardized score, with a SD of 10 at 3 to 4 years of age. Average standardized performance for the Benin children was below 60 for the KABC-II, with a SD of 12 for 5 to 6 years of age. Furthermore, as children get older, standardized scores tend to decrease in longitudinal assessment using the MSEL for Ugandan children (Bass et al., 2017; Boivin et al., 2017). This may reflect the differential impact of the environment and culture on such standardized developmental assessments as children age in HICs versus LMICs. Furthermore, in the Beninese context, these low scores are probably affected by both the culture fairness of the test in such setting (test not developed for this population) and the impact of the risk factors potentially affecting potentially contributing to developmental delays (asymptomatic or symptomatic chronic infection such as malaria, helminthes, schistosomiasis; anemia and micronutrient deficiencies, toxic environmental exposure factors) (Boivin, Kakooza, Warf, Davidson, & Grigorenko, 2015).

In the present study, scores varied according to the age at assessment in raw and standardized scores. Ideally, children should be tested at a same age. If this is not possible, stratifying for age should be performed in evaluating exposure/outcome relationships. In case of an age difference among exposure groups, the statistical model should control for age, even when using standardized scores that supposedly have already been age adjusted.

In addition, the use of different test batteries (MSEL vs. KABC-II) – and different normative samples (US and French, respectively) – across age groups introduces a number of potential confounds that complicate interpretation of results. This could well have been the case because differences were found between the younger and older groups on some of the comparisons between years of age for those tests. Therefore, it is not clear that those differences were necessarily driven by older children drifting away from HIC-based norms with each passing year of age. The findings may be driven, to some degree, by fundamental differences in the tests and their respective normative sets across age groups.

Also, it should not be taken as a given that a particular test measures the same constructs across cultural settings, as has been considered by other child development assessment specialists working in the sub-Saharan African setting (P. Holding et al., 2016a). Attempts should first be made to establish the invariance of the constructs across settings, and only after invariance has been established should the test be interpreted as measuring the same cognitive ability construct across those settings. Conducting this kind of analysis often requires larger samples than were collected in this study (Kitsao-Wekulo, Holding, Taylor, Abubakar, & Connolly, 2013). Nonetheless, it would be important to see in future studies if the MSEL taps into the same domains in both 3 and 4 year olds, or if the KABC-II taps into the same domains in both 5 and 6 year olds.

Other studies found good construct and factor structure validity for the KABC tests in other countries in sub-Sahara Africa (Bangirana, Musisi, et al., 2009; Giordani et al., 1996; van Wyhe, van de Water, Boivin, Cotton, & Thomas, 2017). Some tests have been developed in SSA allowing for more relevant tools and dimensions assessed (Gladstone et al., 2010; P. Holding et al., 2016b; P. A. Holding et al., 2004; P. a. A. Holding, A. , 2005). However, even tests that have been well validated and highly contextualized for a given African setting can still be systematically biased in a different cultural, rural/urban, religious, or linguistic African context (Semrud-Clikeman et al., 2016). Even if the capacity and resources were available to validate a given test for a given context, Western tests and their corresponding norms from HICs still tend to be more readily available and have already undergone a great deal of development and validation for those country settings. The purpose of this study is not to emphasize on the utility of one test other another, but to empirically evaluate an important issue in the adaptation of western-based assessments for children – that is, how best to adjust performance measures for a child's age.

Our study was limited by its study sample. Only about thirty children were included in each age category. Therefore, although a trend was seen for the associations between socioeconomic variables and outcomes, some of them were not significant. One test was normed on the American population (MSEL) whereas the other one was normed on the French population from a more recent sample (KABC-II). This may have led to some differences in scores. Another limitation is the constraint of having to use norms of HICs for a given western-based test in order to arrive at the global or composite performance measure. For example, the cognitive ability composite score for the MSEL must be derived from the standardized scale scores (Visual Reception, Fine Motor, Receptive Language, Expressive Language). It is usually not compiled directly from the scale raw score performance, as was done in the present analysis for comparison sake. The same is true for the present comparison of raw to standardized global performance comparison for the KABC-II MPI score.

CONCLUSIONS

When using western-based tests developed in high-income countries (HIC), investigators working in Africa are sometimes uncertain as to whether to use the norms originally developed for those tests in order to standardize their performance measures on the basis of age. The Mullen Scales of Early Learning (MSEL) developmental assessment (3, 4 yr olds) and the Kaufman Assessment Battery for Children (2nd edition) neuropsychological assessment (KABC-II) (5, 6 yr olds) were used to assess children as a means of evaluating the comparative construct validity of raw and standardized score performances for these tests. Standardization using HIC norms was not optimal, resulting in minimal impact to account for age when using the MSEL, and lower scores for oldest children compared with youngest children when using the Kabc2. This is likely due to the rural African Benin children drifting away from HIC-based norms with each passing year of age, towards weaker standardized performance measures. These findings support the importance of having a local comparison group of reference children when using western-based tests for neurodevelopmental and neuropsychological evaluation, statistically adjusting with important predictors of performance outcomes such as age, HOME, and SES.

ACKNOWLEDGEMENTS

We would like to thank the entire staff of the three health centers (Allada, Attogon, Sékou), and the study's participants. We gratefully thank our field team who collected the data and provided the medical care. We thank Micheline Garel and Nour Tabbara for translation of Mullen Scales of Early Learning. We acknowledge the contributions of an anonymous reviewer in the interpretation and explanation of the principal study findings.

Funding: This research investigation was supported by the National Institutes of Health (NIH), grant R21 HD060524

REFERENCES

Bangirana, P., John, C. C., Idro, R., Opoka, R. O., Byarugaba, J., Jurek, A. M., & Boivin, M. J. (2009a). Socioeconomic predictors of cognition in Ugandan children: Implications for community based interventions. *PLOS ONE, 4*(11), e7898. doi:7810.1371/journal.pone.0007898.

Bangirana, P., John, C. C., Idro, R., Opoka, R. O., Byarugaba, J., Jurek, A. M., & Boivin, M. J. (2009b). Socioeconomic Predictors of Cognition in Ugandan Children: Implications for Community Interventions. *Plos One, 4*(11). doi:e7898

10.1371/journal.pone.0007898

Bangirana, P., Menk, J., John, C. C., Boivin, M. J., & Hodges, J. S. (2013). The association between cognition and academic performance in Ugandan children surviving malaria with neurological involvement. *PLOS ONE, 8*(2), e55653. doi:10.1371/journal.pone.0055653

PONE-D-12-26328 [pii]

Bangirana, P., Musisi, S., Allebeck, P., Giordani, B., John, C. C., Opoka, R. O., . . . Boivin, M. J. (2009). A preliminary investigation of the construct validity of the KABC-II in Ugandan children with prior cerebral insult. *African Health Sciences, 9*(3), 186-192.

Bangirana, P., Seggane, M., Allebeck, P., Giordani, B., John, C. C., Opoka, O. R., . . . Boivin, M. J. (2009). A preliminary examination of the construct validity of the KABC-II in Ugandan children with a history of cerebral malaria. *Afr Health Sci, 9*(3), 186-192.

Bass, J. K., Nakasujja, N., Familiar-Lopez, I., Sikorskii, A., Murray, S. M., Opoka, R., . . . Boivin, M. J. (2016). Association of caregiver quality of care with neurocognitive outcomes in HIV-affected children aged 2-5 years in Uganda. *AIDS Care, 28 Suppl 1*, 76-83. doi:10.1080/09540121.2016.1146215

Bass, J. K., Opoka, R., Familiar, I., Nakasujja, N., Sikorskii, A., Awadu, J., . . . Boivin, M. (2017). Randomized controlled trial of caregiver training for HIV-infected child neurodevelopment and caregiver well being. *AIDS, 31*(13), 1877-1883. doi:10.1097/QAD.0000000000001563

Bergemann, T. L., Bangirana, P., Boivin, M. J., Connett, J. E., Giordani, B. J., & John, C. C. (2012). Statistical approaches to assess the effects of disease on neurocognitive function over time *J Biomet Biostat, S7:016. doi:10.4172/2155-6180.S7-016*

Boivin, M. J., Bangirana, P., Byarugaba, J., Opoka, R. O., Idro, R., Jurek, A. M., & John, C. C. (2007). Cognitive impairment after cerebral malaria in children: a prospective study. *Pediatrics, 119*(2), e360-366.

Boivin, M. J., Bangirana, P., Nakasuja, N., Page, C. F., Shohet, C., Givon, D., . . . Klein, P. S. (2013). A year-long caregiver training program to improve neurocognition in preschool Ugandan HIV-exposed children. *J Dev Behav Pediatr, 34*(2), 269-278.

Boivin, M. J., Bangirana, P., Nakasujja, N., Page, C. F., Shohet, C., Givon, D., . . . Klein, P. S. (2013). A year-long caregiver training program improves cognition in preschool Ugandan children with human immunodeficiency virus. *J Pediatr, 163*, 1409-1416.

Boivin, M. J., Barlow-Mosha, L., Chernoff, M. C., Laughton, B., Zimmer, B., Joyce, C., . . . Team, I. P. S. (2018). Neuropsychological performance in African children with HIV enrolled in a multisite antiretroviral clinical trial. *AIDS, 32*(2), 189-204. doi:10.1097/QAD.0000000000001683

Boivin, M. J., Kakooza, A. M., Warf, B. C., Davidson, L. L., & Grigorenko, E. L. (2015). Reducing neurodevelopmental disorders and disability through research and interventions. *Nature, 527*(7578), S155-160. doi:10.1038/nature16029

Boivin, M. J., Nakasujja, N., Familiar-Lopez, I., Murray, S. M., Sikorskii, A., Awadu, J., . . . Bass, J. K. (2017). Effect of Caregiver Training on the Neurodevelopment of HIV-Exposed Uninfected Children and Caregiver Mental Health: A Ugandan Cluster-Randomized Controlled Trial. *J Dev Behav Pediatr*. doi:10.1097/DBP.0000000000000510

Caldwell BM, B. R. (2003). *HOME Inventory Administration Manual. . Arkansas: Print Design Inc.*

Caldwell, B. M., & Bradley, R. H. (2003). *Home inventory administration manual.* Arkansas: Print Design Inc.

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987a). DETECTION OF POSTNATAL DEPRESSION - DEVELOPMENT OF THE 10-ITEM EDINBURGH POSTNATAL DEPRESSION SCALE. *British Journal of Psychiatry, 150*, 782-786. doi:10.1192/bjp.150.6.782

Cox, J. L., Holden, J. M., & Sagovsky, R. (1987b). Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *Br J Psychiatry, 150*, 782-786.

Familiar, I., Murray, S., Ruisenor-Escudero, H., Sikorskii, A., Nakasujja, N., Boivin, M. J., . . . Bass, J. K. (2016). Socio-demographic correlates of depression and anxiety among female caregivers living with HIV in rural Uganda. *AIDS Care, 28*(12), 1541-1545. doi:10.1080/09540121.2016.1191609

Familiar, I., Nakasujja, N., Bass, J., Sikorskii, A., Murray, S., Ruisenor-Escudero, H., . . . Boivin, M. J. (2016). Caregivers' depressive symptoms and parent-report of child executive function among young children in Uganda. *Learn Individ Differ, 46*, 17-24. doi:10.1016/j.lindif.2015.01.012

Giordani, B., Boivin, M. J., Opel, B., Dia Nseyila, D., Diawaku, N., & Lauer, R. E. (1996). Use of the K-ABC with children in Zaire, Africa: An evaluation of the sequential-simultaneous processing distinction within an intercultural context. *International Journal of Disability, Development and Education, 43*(1), 5-24.

Gladstone, M., Lancaster, G. A., Umar, E., Nyirenda, M., Kayira, E., van den Broek, N. R., & Smyth, R. L. (2010). The Malawi Developmental Assessment Tool (MDAT): the creation, validation, and reliability of a tool to assess child development in rural African settings. *PLoS Med, 7*(5), e1000273. doi:10.1371/journal.pmed.1000273

Hanlon, C., Medhin, G., Alem, A., Araya, M., Abdulahi, A., Hughes, M., . . . Prince, M. (2008). Detecting perinatal common mental disorders in Ethiopia: validation of the self-reporting questionnaire and Edinburgh Postnatal Depression Scale. *J Affect Disord, 108*(3), 251-262. doi:10.1016/j.jad.2007.10.023

Holding, P., Anum, A., van de Vijver, F. J., Vokhiwa, M., Bugase, N., Hossen, T., . . . Gomes, M. (2016a). Can we measure cognitive constructs consistently within and across cultures? Evidence from a test battery in Bangladesh, Ghana, and Tanzania. *Appl Neuropsychol Child*, 1-13. doi:10.1080/21622965.2016.1206823

Holding, P., Anum, A., van de Vijver, F. J. R., Vokhiwa, M., Bugase, N., Hossen, T., . . . Gomes, M. (2016b). Can we measure cognitive constructs consistently within and across cultures? Evidence from a test battery in Bangladesh, Ghana, and Tanzania. *Appl Neuropsychol Child*, 1-13. doi:10.1080/21622965.2016.1206823

Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., . . . Stevenson, J. (2004). Assessing cognitive outcomes in a rural African population: development of a neuropsychological battery in Kilifi District, Kenya. *J Int Neuropsychol Soc, 10*(2), 246-260.

Holding, P. a. A., A. . (2005). Developing a Measure of the Consequences of Prenatal and Peri-Natal Insults in Early Childhood in Kenya: Preliminary Report of the Results for R21(TWOO6805-01/TWO3 007) *FIC/NIH RFA, Brain Disorders in the Developing World: Research Across the Lifespan*.

John, C. C., Panoskaltsis-Mortari, A., Opoka, R. O., Park, G. S., Orchard, P. J., Jurek, A. M., . . . Boivin, M. J. (2008). Cerebrospinal fluid cytokine levels and cognitive impairment in cerebral malaria. *Am J Trop Med Hyg, 78*(2), 198-205. doi:78/2/198 [pii]

Kakyo, T. A., Muliira, J. K., Mbalinda, S. N., Kizza, I. B., & Muliira, R. S. (2012). Factors associated with depressive symptoms among postpartum mothers in a rural district in Uganda. *Midwifery, 28*(3), 374-379. doi:10.1016/j.midw.2011.05.001

Kitsao-Wekulo, P. K., Holding, P. A., Taylor, H. G., Abubakar, A., & Connolly, K. (2013). Neuropsychological testing in a rural African school-age population: evaluating contributions to variability in test performance. *Assessment, 20*(6), 776-784. doi:10.1177/1073191112457408

Koura, K. G., Boivin, M. J., Davidson, L. L., Ouedraogo, S., Zoumenou, R., Alao, M. J., . . . Bodeau-Livinec, F. (2013). Usefulness of child development assessments for low-resource settings in francophone Africa. *J Dev Behav Pediatr, 34*(7), 486-493. doi:10.1097/DBP.0b013e31829d211c

Koura, K. G., Boivin, M. J., Davidson, L. L., Ouedraogo, S., Zoumenou, R., Alao, M. J., . . . Bodeau-Livinec, F. (2013). Usefulness of Child Development Assessments for Low-Resource Settings in Francophone Africa. *Journal of developmental and behavioral pediatrics, 34*(7), 486-493.

Mireku, M. O., Davidson, L. L., Koura, G. K., Ouedraogo, S., Boivin, M. J., Xiong, X., . . . Bodeau-Livinec, F. (2015). Prenatal Hemoglobin Levels and Early Cognitive and Motor Functions of One-Year-Old Children. *Pediatrics, 136*(1), E76-E83. doi:10.1542/peds.2015-0491

Mullen, E. M. (1995). *Mullen Scales of Early learning : Manual*. Circle Pines, MN: American Guideline Service, Inc.

Murray, S. M., Familiar, I., Nakasujja, N., Winch, P., Galloa, J., Opokae, R., . . . Bass, J. K. (2016). Caregiver mental health and HIV-infected child wellness: perspectives from Ugandan caregivers. *AIDS Care, 28*(in press). doi:http://dx.doi.org/10.1080/09540121.2016.1263722

Raven, J. (2000a). The Raven's progressive matrices: change and stability over culture and time. *Cogn Psychol, 41*(1), 1-48. doi:10.1006/cogp.1999.0735

S0010-0285(99)90735-1 [pii]

Raven, J. (2000b). The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology, 41*(1), 1-48. doi:10.1006/cogp.1999.0735

Semrud-Clikeman, M., Romero, R. A., Prado, E. L., Shapiro, E. G., Bangirana, P., & John, C. C. (2016). Selecting measures for the neurodevelopmental assessment of children in low- and middle-income countries. *Child Neuropsychol*, 1-42. doi:10.1080/09297049.2016.1216536

van Wyhe, K. S., van de Water, T., Boivin, M. J., Cotton, M. F., & Thomas, K. G. (2017). Cross-cultural assessment of HIV-associated cognitive impairment using the Kaufman assessment battery for children: a systematic review. *J Int AIDS Soc, 20*(1), 1-11. doi:10.7448/IAS.20.1.21412

Table 1. Subjects characteristics.

	Total	*MSEL*	KABC2	*p-value*
	N=135	*N=74*	*N=61*	
Home Observation for Measurement of the Environment (HOME) Inventory (median, standard deviation (SD)	27.0 (2.0)	27.2	27.4	.60
Raven Progressive Matrices scale	15.0 (3.2)	15.4	15.1	.66
Edinburg Postpartum depression scale (EPDS)	8.0 (3.7)	8.7	7.9	.23
Family wealth index	5.0 (2.6)	5.4	5.6	.61
Infant's age (years, n (%))				
3	39 (28.9)	39 (52.7)		
4	35 (25.9)	35 (47.3)		
5	30 (22.2)		30 (49.2)	
6	31 (23.0)		31 (50.8)	
Maternal education				.25
None	83 (62.4)	43 (58.1)	40 (67.8)	
Some schooling	50 (37.6)	31 (41.9)	19 (32.2)	
Infant's gender				.37
Male	74 (54.8)	38 (51.4)	36 (59.0)	
Female	61 (45.2)	36 (48.6)	25 (41.0)	
Maternal marital status				.64
Monogamous	82 (61.7)	47 (64.4)	35 (60.3)	
Polygamous	49 (36.8)	26 (35.6)	23 (39.7)	
Widow	2 (1.5)			

Table 2 Mullen Scales of Early Learning (MSEL) raw and standardized scale and composite scores. The USA normative mean for the standardized scales is 50 and the raw score totals for the scales tend to be in the same range. For the MSEL composite cognitive standardized score, the mean is 100 (SD = 15). The raw score totals for this composite tend to be higher.

	Raw scores		p-value	Standardized scores		p-value
	3 years of age	*4 years of age*		*3 years of age*	*4 years of age*	
	Mean (SD)	*Mean (SD)*		*Mean (SD)*	*Mean (SD)*	
Scales						
Gross Motor	29.8 (3.5)	32.4 (3.2)	.001	NA	NA	-
Visual Reception	31.6 (5.1)	34.9 (5.8)	.01	31.4 (7.1)	36.6 (8.2)	.004
Fine Motor	30.3 (4.1)	34.7 (5.6)	<.001	29.9 (7.1)	35.2 (7.5)	.003
Receptive Language	30.8 (4.7)	33.5 (4.1)	.01	32.0 (7.4)	33.8 (4.8)	.22
Expressive Language	31.7 (7.4)	35.8 (8.7)	.04	33.5 (9.4)	37.1 (7.5)	.07
MSEL Cognitive Composite Score						
Mean (SD)	126.9 (24.6)	142.8 (21.2)	.004	66.6 (10.3)	73.2 (9.4)	.005
Mean minus 1 SD (n, %)	6 (15.4)	6 (17.1)	.84	6 (15.4)	5 (14.3)	.89
Mean minus 2 SD (n, %)	0 (0)	1 (2.9)	.47*	0 (0)	1 (2.9)	.47

* Fisher's test. Raw scores for Cognitive Composite Score is the sum of the four standardized subscales visual reception, fine motor, receptive language, and expressive language.

Table 3 Kaufman Assessment Battery for Children, 2nd edition (KABC-II) raw and standardized scale and global scores. The USA normative mean for the subtest scales is 10 (SD = 2.5) and the raw score totals for the scales tend to be in the same range, except for the Leaning scales of Atlantis and Rebus. For the KABC-II nonverbal (NVI) and global standardized domains, the mean is standard score is 100 (SD = 15). The raw score totals for these values tend to be lower.

	Raw scores		p-value	Standardized scores		p-value
	5 years of age	*6 years of age*		*5 years of age*	*6 years of age*	
	Mean (SD)	*Mean (SD)*		*Mean (SD)*	*Mean (SD)*	
Scales						
Atlantis	22.0 (11.3)	26.1 (15.0)	.23	5.0 (2.4)	4.7 (2.2)	.55
Conceptual thinking	4.0 (2.5)	5.1 (3.5)	.18	2.5 (1.8)	1.7 (1.6)	.07
Number recall	4.9 (2.2)	6.7 (2.6)	.004	6.8 (3.7)	7.4 (3.0)	.54
Gestalt Closure	4.8 (3.2)	6.4 (3.3)	.07	2.9 (2.2)	3.4 (1.5)	.35
Rover	7.5 (5.4)	9.3 (4.5)	.16	NA	7.4 (2.8)	-
Rebus	16.1 (8.2)	17.1 (8.5)	.64	7.9 (2.9)	6.0 (2.4)	.007
Triangles	5.4 (3.2)	7.5 (3.2)	.02	2.4 (2.3)	2.3 (1.8)	.88
Word order	7.5 (2.4)	8.5 (2.4)	.10	5.1 (3.0)	4.1 (2.8)	.18
Pattern reasoning	2.7 (2.0)	2.6 (1.9)	.76	5.6 (1.8)	4.3 (1.4)	<.001
Hand movements	4.3 (1.6)	5.3 (1.5)	.01	6.4 (2.3)	5.7 (2.0)	.21
Nonverbal Index (NVI)	19.5 (7.6)	17.1 (5.4)	.15	52.9 (11.4)	53.6 (7.9)	.79
Global Domains						
Sequential Processing	11.9 (5.3)	11.5 (4.8)	.71	76.2	74.5	.65
Simultaneous Processing	10.9 (4.3)	15.2 (5.6)	.002	55.7	57.3	.50
Learning	12.3 (4.8)	11.0 (4.2)	.24	78.6	72.0	.05
Mental Processing Index (MPI)						
Mean (SD)	35.5 (11.7)	37.6 (12.0)	.48	59.3 (13.5)	58.6 (11.6)	.83
Mean minus 1 SD (n, %)	4 (13.3)	5 (16.1)	.76	5 (16.7)	5 (16.1)	.95
Mean minus 2 SD (n, %)	1 (3.3)	0 (0)	.49	0 (0)	0 (0)	-

Raw scores for global domains and MPI are the sum of the relevant scales.

Table 4. Pearson product-moment correlation coefficient matrix for principal assessments in the study

	Raw Mullen Scales Early Learning cognitive composite	Standardized Mullen Scales of Early Learning cognitive composite	Raw KABC-II Mental Processing Index	Standardized KABC-II Mental Processing Index
HOME caregiving quality	0.27*	0.16	0.33**	0.34**
Edinburgh Postpartum Depression Scale	0.16	-0.09	-0.13	-0.17
Raven Progressive Matrices test	0.15	0.07	0.22	0.24
Family wealth scale	0.18	0.11	0.38**	0.38**
Maternal marital status	0.08	0.03	-0.14	-0.15

	HOME caregiving quality	Edinburgh Postpartum Depression	Raven Progressive Matrices	Family Wealth Index	Maternal marital status

HOME caregiving quality	-0.07	0.26**	0.36**	-0.14
Edinburgh Postpartum Depression Scale		-0.05	0.08	0.08
Raven Progressive Matrices test			0.16	-0.12
Family Wealth Index				-0.03

*p<0.05, **p<0.01 statistical significance. Measures: Maternal marital status 1=monogamous, 2=polygamous; Mullen Scales of Early Learning cognitive composite score; Kaufman Assessment Battery for Children, 2nd edition (KABC-II) mental processing composite (MPI) score.

Table 5. Predictors for Mullen Scales of Early Learning (MSEL) cognitive composite standardized score and Kaufman Assessment Battery for Children, 2nd edition (KABC-II) Mental Processing Index global performance scores.

	MSEL Cognitive Composite Standardized Score				KABC-II Mental Processing Index Standardized Score			
	Model I beta (95% CI)	Model II beta (95% CI)	Model III beta (95% CI)	Model IV beta (95% CI)	Model I beta (95% CI)	Model II beta (95% CI)	Model III beta (95% CI)	Model IV beta (95% CI)
	R^2=.12	R^2=.25	R^2=.05	R^2=.68	R^2=.27	R^2=.27	R^2=.28	R^2=0.28
HOME score	3.2 (-0.3;6.6)	3.4 (0.1;6.6)*	1.6 (-1.2;3.4)	1.3 (0.2;2.3)*	0.9 (-0.5;2.3)	0.9 (-0.5;2.3)	0.9 (-0.6;2.4)	0.9 (-0.6;2.4)
Edinburgh Depression	0.9 (-0.8;2.5)	1.1 (-0.4;0.6)	-0.5 (-1.4;0.4)	-0.1 (-0.6;0.4)	-0.7 (-1.5;0.2)	-0.7 (-1.6;0.1)	-0.8 (-1.7;0.1)	-0.8 (-1.7;0.1)
Raven Matrices	0.5 (-1.2;2.1)	0.1 (-1.5;1.7)	0.0 (-1.0;1.1)	0.4 (-0.3;1.0)	-0.1 (-1.2;1.0)	0.0 (-1.1;1.1)	0.0 (-1.1;1.1)	0.0 (-1.2;1.2)
Family wealth	0.9 (-1.3;3.1)	1.2 (-0.8;3.3)	0.4 (-1.0;1.7)	0.4 (-0.4;1.1)	1.7 (0.5;3.0)*	1.7 (0.5;3.0)*	1.9 (0.6;3.2)*	1.9 (0.5;3.2)*
Maternal marital status	7.3 (-4.7;19.3)	5.5 (-5.7;16.8)	3.0 (-3.8;9.8)	1.2 (-2.7;5.2)	-0.4 (-6.3;5.5)	-0.1 (-6.1;5.9)	-0.4 (-6.6;5.8)	-0.5 (-6.9;5.8)
Age at assessment		17.6 (7.0;28.3)*		13.1 (11.5;14.8)*		2.1 (-3.6;7.7)		-0.7 (-6.7;5.2)

Models using raw scores:

Model I - Linear model adjusting for HOME score, EPDS, Raven, Family wealth, maternal education, and maternal marital status

Model II - Linear model adjusting for HOME score, EPDS, Raven, Family wealth, maternal education, and maternal marital status and age at assessment

Models using standardized scores:

Model III - Linear model adjusting for HOME score, EPDS, Raven, Family wealth, maternal education, and maternal marital status

Model IV - Linear model adjusting for HOME score, EPDS, Raven, Family wealth, maternal education, and maternal marital status and age at assessment.