



Published in final edited form as:

*Res Dev Disabil.* 2021 March ; 110: 103837. doi:10.1016/j.ridd.2020.103837.

## Cultural Adaptation of the Bayley Scales of Infant and Toddler Development, 3rd Edition for use in Kenyan Children Aged 18–36 Months: A Psychometric Study

Megan S. McHenry<sup>1,2</sup>, Eren Oyungu<sup>3</sup>, Ziyi Yang<sup>4</sup>, Abbey C. Hines<sup>1</sup>, Ananda R. Ombitsa<sup>2</sup>, Rachel C. Vreeman<sup>2,5</sup>, Amina Abubakar<sup>6</sup>, Patrick O. Monahan<sup>4</sup>

<sup>1</sup>Department of Pediatrics, Indiana University School of Medicine

<sup>2</sup>Academic Model Providing Access to Healthcare (AMPATH), Eldoret, Kenya

<sup>3</sup>Department of Child Health and Paediatrics, School of Medicine, College of Health Sciences, Moi University

<sup>4</sup>Department of Biostatistics, Indiana University School of Medicine

<sup>5</sup>Arnhold Institute for Global Health, Icahn School of Medicine, Mt Sinai Hospital

<sup>6</sup>Centre for Geographic Medicine Research (Coast), KEMRI, Kilifi, Kenya

### Abstract

**Corresponding Author:** Megan S. McHenry, MD, MS, FAAP, Assistant Professor of Pediatrics Division of Pediatric Infectious Diseases and Global Health Indiana University School of Medicine 705 Riley Hospital Drive Room 5853 Indianapolis, IN 46202, 317.274.4143 tel, 317.948.0860 fax, msuhl@iu.edu.

CRedit Author Statement:

**Megan McHenry, MD, MS** : conceptualization, conceptualization, methodology, supervision, project administration, writing-original draft, funding acquisition; **Eren Oyungu, MBChB, MMED, MSc, MPH**: conceptualization, methodology, supervision, project administration, writing- review and editing; **Ziyi Yang, MS**: formal analysis, writing-review and editing; **Abbey C. Hines, PsyD**: supervision; project administration; writing- review and editing; **Ananda R. Ombitsa BSc**: project administration; writing-review and editing; **Rachel C. Vreeman, MD, MS**: supervision, writing- review and editing; **Amina Abubakar, MED, PhD**: supervision, writing- review and editing, validation; **Patrick O. Monahan, PhD**: supervision, methodology, conceptualization, formal analysis, writing-review and editing.

Contributions

MSM conceptualized and sought funding for this study. Study design also informed by RCV and EO. ACH performed training on the Bayley Scales of Infant and Toddler Development, and AOR and ACH were responsible for monitoring quality of its administration. AOR oversaw data collection activities and data management processes. ZY completed all statistical analyses with significant input from MSM, POM, and AA. MSM completed the first draft of this manuscript. All authors contributed substantially to revising this article critically for important intellectual content. All authors contributed to the final manuscript and approved submission for publication.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of Interests

The authors of this manuscript declare no competing interests.

Data sharing

The de-identified data set and a data dictionary will be made available with publication of the trial after obtaining relevant Institutional Research Ethics Committee approval of a proposal and signed data access agreement. Inquiries can be made to author, Megan McHenry, MD, MS (msuhl@iu.edu).

**Background:** The Bayley Scales of Infant and Toddler Development, 3<sup>rd</sup> Edition (Bayley-III) is frequently used in international child development research. No studies examine its psychometric properties when culturally adapted within the Kenyan context.

**Aims:** To culturally adapt the Bayley-III for use in Kenya and evaluate its validity and reliability.

**Methods and Procedures:** Forward and backward translation, cognitive interviews, and a brief pilot of culturally adapted items were performed. This psychometric study was part of another study on children born to mothers with HIV in Eldoret, Kenya. One hundred seventy-two children aged 18–36 months were assessed for cognition, receptive/expressive communication, and fine/gross motor domains using the Bayley-III. Confirmatory factor analysis (CFA), inter-scale Pearson correlations, internal consistency, t-tests, and test-retest reliability were performed.

**Outcomes and Results:** The mean age of children was 22.8 (SD 4.5) months old; 52.7% (n = 89) were male. CFA revealed that both two- and three-factor indices had good and comparable fit. Pearson correlations were high between fine motor and receptive communication ( $r > 0.70$ ). Internal consistency was very strong for all of the subtests, with Cronbach coefficient alpha scores ranging from 0.88 to 0.96. Known groups/convergent validity was confirmed with stunting and parental concern for delays. Test-retest reliability was good and did not differ substantially across groups.

**Conclusions and Implications:** The Kenyan adapted Bayley-III is a psychometrically acceptable tool to assess child development. The scaled and composite scores should not be used to define Kenyan developmental norms, but it can be useful for comparing groups within research settings.

## Keywords

Psychometrics; validity; reliability; Bayley Scales of Infant and Toddler Development; child development; cultural adaptation

## 1 Introduction

Approximately 250 million children worldwide are at risk for not reaching their full developmental potential, with the vast majority residing in low- and middle-income countries (Black et al., 2017). International public health systems work to improve infrastructure to support child growth and development, and research is necessary to understand the factors impacting child development to provide guidance on policy and healthcare decisions. However, researchers face major challenges in identifying the appropriate tools for measuring development in cross-cultural settings. While there have been efforts to develop and validate measures for use within the local contexts of low- and middle-income settings, including the Malawi Developmental Assessment Tool (Gladstone et al., 2010), the Kilifi Developmental Inventory (Abubakar, Holding, Van Baar, Newton, & Van De Vijver, 2008), and the Rapid Neurodevelopmental Assessment (Khan et al., 2013), there is still a need to evaluate measures which are more widely used globally. Use of global measures provides the potential to compare the scores of these locally adapted measures across settings.

The Bayley Scales of Infant and Toddler Development, 3<sup>rd</sup> Edition (Bayley-III) is a well-known developmental assessment that is validated and normed for the United States (U.S.) population (Bayley, 2006). Bayley-III is often used internationally and adapted for local contexts; however, many studies do not outline the adaptation process or evaluate its psychometric properties (McHenry et al., 2018), raising concerns regarding the validity and reproducibility of the resulting data.

Psychometric studies of the Bayley-III have occurred in multiple countries across the world (Hua et al., 2019; Ranjitkar et al., 2018; Yu et al., 2013). Within sub-Saharan Africa, studies have been performed in South Africa (Ballot et al., 2017; Pendergast et al., 2018), Ethiopia (Hanlon et al., 2016) and Malawi (Cromwell et al., 2014). Most of these studies have found the Bayley-III to be a valid assessment tool for development; while some caution on the limited applicability of the U.S. norms within local populations (Cromwell et al., 2014). No studies on the validity of the Bayley-III have been performed in Kenya, and thus it is unclear how it will perform psychometrically within this setting. The objectives of this study were to culturally adapt the Bayley-III for use in Kenya and to evaluate the content, construct, convergent validity, and reliability measures for this adapted scale in Kenyan children aged 18–36 months.

## 2 Materials and Methods

This study was carried out as part of a larger project focused on development in children born to mothers infected with HIV (NEURODEV study- PI:MSM).

### 2.1 Setting

The cross-sectional study took place in Eldoret, Kenya, within the Academic Model Providing Access to Healthcare (AMPATH) consortium. The AMPATH HIV care program, born from a 20-year partnership between Indiana University School of Medicine (IUSM), Moi University School of Medicine (MUSM), and the Moi Teaching and Referral Hospital (MTRH) in Eldoret, Kenya, has enrolled over 160,000 patients and currently provides care for approximately 15,000 children who are HIV-positive and HIV-exposed in 65 clinics in western Kenya (Einterz et al., 2007; Inui et al., 2007). The children and their caregivers were recruited for the NEURODEV study from a large urban hospital and associated pediatric HIV clinic within AMPATH's catchment area between 12/2017 and 9/2019. Inclusion criteria were as follows: age 18–36 months and speaking primarily Swahili within the home. The study aimed to recruit up to 75 children per group, among the following groups: HIV-infected, HIV-exposed, and HIV-unexposed. All participants from the NEURODEV study with Bayley-III data were included in this study. No other exclusion criteria were applied.

### 2.2 Bayley Scales of Infant and Toddler Development, 3<sup>rd</sup> Edition (Bayley-III)

The Bayley-III is an internationally known assessment of development published in 2006 and is commonly used in research settings (Bayley, 2006; McHenry et al., 2018). The Bayley-III measures five broad developmental domains: adaptive behavior, cognitive, communication, motor, and social-emotional. Only the cognitive, communication, and motor domains were adapted due to the complexities of cross-cultural expectations for behavior

and social-emotional development. This is consistent with other studies that have used the Bayley-III in international settings (le Roux et al., 2018; Wedderburn et al., 2019). The communication domain consists of expressive and receptive communication subtests. The motor domain consists of fine and gross motor subtests. The scaled scores were reported and standardized against a normed population from the U.S. with a mean score of 10 and a standard deviation (SD) of 3 (Aylward GP & J, 2019).

## 2.3 Methods of Adaptation of the Bayley-III

The following steps were taken to ensure the appropriate interpretation of the test items and to support content validity:

**2.3.1 Forward and Backward Translation**—All spoken instructions and anticipated participant answers were pulled from the Bayley-III administration manual. The English text was translated into Swahili by one trained translator. The Swahili translations were then back-translated into English by a separate translator. The two English versions were compared for equivalence by the research team. When inconsistencies arose, they were discussed by a Kenyan pediatric neurologist (EO) and clinical officer (ARO), as well as a U.S.-based pediatrician (MSM), for consensus on harmonizing the language for functional equivalence. All changes were reviewed and approved by an academic clinical psychologist (ACH).

**2.3.2 Cognitive Interview**—Cognitive interviewing involves interviewers asking respondents to think out loud as they process candidate questions in detail, using the think-aloud strategy to better understand recall and thought processes related to a testing item and identify potential sources of response error (Willis & Miller, 2011). Cognitive interviewing was performed within this study to ensure that wording of the instructions and tasks, as part of the Bayley-III administration, were clear and unambiguous, which is essential for cross-cultural research. Cognitive interviews of Bayley-III test items were performed with 10 caregivers of young children. The average age of the caregiver was 33.5 years (Standard Deviation (SD)= 6.2), and all were female. Half of the participants identified as a housewife (5/10), three were casual laborers, one was a farmer, and one was a business owner. The average age of the youngest child within each household was 3.1 years (SD=0.9).

**2.3.3 Cultural Adaptation of Test Materials**—During the translation process, it was learned that gender-specific pronouns are absent from the Swahili language; therefore, the expert panel removed assessment questions requiring gender-specific pronouns from scoring. Cognitive interviews revealed unfamiliar images used within the Bayley-III; specifically, the washing machine and cooking on a stove. These images were exchanged for more familiar, culturally appropriate images. Alternative names were allowed for images of an apple and cookie, which are more commonly referred to within the culture as fruit and biscuit respectively. Children had difficulty with color-matching and naming within the cognitive, receptive communication and expressive communication subtests. Color names and concepts exist within the Swahili language and the Kenyan culture; however, these skills are not taught to young children with the same frequency that they are taught in the U.S.

After a review of expert consultation and cognitive interview data, the color items remained as an area appropriate for testing.

**2.3.4 Brief Initial Pilot**—The adapted and translated language and potential testing items were used in trial administrations of the adapted Bayley-III. Debriefing occurred after each administration, under the guidance of an expert panel consisting of a pediatric clinical psychologist (ACH), a neuropsychologist, a Kenyan pediatric neurologist (EO), and other Kenyan staff. Modifications were made by an iterative process to ensure appropriate understanding while keeping test administration as close to the original version as possible. A total of 17 trial administrations of the adapted Bayley-III were performed to ensure stable, consistent administrations without new cultural issues arising. The mean age of the children was 2.5 years ( $SD=0.6$ ), and 47% (8/17) were female.

## 2.4 Training and Quality Control of Bayley-III

One Kenyan clinical officer, the equivalent of an advanced practice provider within the U.S., was trained to administer the Bayley-III. This included in-person training on the theoretical background of the scale and the clinical administration of the Bayley-III by a highly trained clinical psychologist. The clinical officer performed video-taped administrations which were reviewed by the clinical psychologist, who provided feedback. This was repeated until the clinical psychologist was confident in the abilities of the clinical officer to perform the Bayley-III independently. Periodic review of administrations were performed to ensure quality control.

## 2.5 Administration of Bayley-III Within the NEURODEV Study

A brief questionnaire related to the NEURODEV study was performed with the caregiver before the administration of the Bayley-III. This questionnaire included the following information: maternal education level, anthropometrics, pre-term birth status, concern for and family history of developmental delay, time left alone during the week, and HIV-status. During the completion of the informed consent and questionnaire, the child participant was given a drink and snack prior to the administration of the Bayley-III. Five subtests of the culturally adapted Bayley-III, gross motor, fine motor, cognition, receptive communication, and expressive communication, were administered to children ages 18–36 months old. The test took between 45–90 minutes to administer to each child, and each session was video-recorded for quality control. Occasionally, a break was needed for a nap or additional snack during between subtests, and infrequently, children would be unable to complete the Bayley-III due to temperament. Caregivers were compensated 500 Kenyan shillings (\$5USD) for study-related time and transport. The children were given a small toy, such as a ball.

## 2.6 Study Sample

The characteristics of the study sample are shown in Table 1. The 184 children enrolled in the study were on average 22.8 ( $SD=4.5$ ) months of age, and 154 (91.1%) were born after 37 weeks. Approximately 65% of mothers had at least some secondary school education, which is aligned with national surveys indicating that 66% of Kenyan females ages 15–17 years are enrolled in secondary school (UNICEF, 2018). There were 31 (18.0%) mothers who reported concerns about developmental delay in their child and 23 (13.5%) reported having a

family history of children with a developmental delay. Twelve of the enrolled children were unable to complete the Bayley-III due to behavioral or temperament issues.

## 2.7 Ethical Approval

This study was approved by the Indiana University Institutional Review committee as an expedited study (#1601531540) and by the Moi University Institutional Research and Ethics Committee (IREC/2016/09) before initiation of study activities. The changes made within this adaptation were made under consultation with Pearson Licensing.

## 2.8 Statistical Analysis

**2.8.1 Baseline Characteristics**—Baseline variables were summarized using mean and SD for continuous variables along with frequency and percent for categorical variables. Differences were compared between males and females by using the *t*-test and Chi-square test.

**2.8.2 Validity**—Construct validity was evaluated using confirmatory factor analysis (CFA). CFA was conducted to test the hypothesized factor structure for Bayley-III. Three models with one to three factors were evaluated and compared. The one-factor model was organized with five subtests as indicators of a general factor. The two-factor model was arranged with two motor subtests and two communication subtests on one factor and the cognitive scale on another factor. The three-factor model was specified to have two motor subtests on the first factor, two communication subtests on the second factor, and the cognitive scale on the third factor. All models allowed factors to inter-correlate (based on *a priori* theory) and assumed that residual errors were uncorrelated. The raw scores of five subtests were used. The CFA models were evaluated based on six methods: (1) chi-square goodness-of-fit test comparing the sample and fitted covariance matrices; p-value greater than 0.05 indicates good fit; (2) adjusted goodness-of-fit index (AGFI) with good fit  $\geq 0.90$ ; (3) the root mean square error of approximation (RMSEA) with acceptable fit  $< 0.08$  and good fit  $< 0.05$ ; (4) Schwarz's Bayesian criterion (SBC) for which smaller is better; (5) Comparative Fit Index (CFI) for which  $\geq 0.90$  indicates good fit; and (6) standardized root mean square residual (SRMR), with good fit  $< 0.08$  (Kline, 2015). The CFA models were assessed for all ages combined, as well as separately by two age groups (18–22 months, and 23–35 months), gender (female and male), and HIV status (uninfected and exposed cohorts).

Pearson's correlation was used to examine the pairwise correlations among five raw subtests of Bayley-III in the three age groups. The following guideline was adopted to interpret the correlation magnitude, which was consistent with this area of research: very high (0.90 to 1.00), high (0.70 to 0.90), moderate (0.50 to 0.70), low (0.30 to 0.50), and negligible (0.00 to 0.30) (Mukaka, 2012). Additionally, to evaluate the sensitivity to developmental changes, the crude linear regression model and scatter plot was employed to assess the association between each raw subtest and age. The slope and  $R^2$  were estimated based on these models.

The Bayley-III cognitive, communication, and motor scale scores were compared across four indicators of potential developmental impairment to evaluate convergent validity. Malnutrition is commonly cited as a risk factor for poor development, with stunting most

closely associated. In these analyses, “stunting” refers to moderate-to-severe stunting (height-for-age (HFA) z-scores  $\leq -2$ ). “Underweight” refers to moderate-to-severe underweight status (weight-for-age (WFA) z-scores  $\leq -2$ ). “Wasting” refers to moderate-to-severe wasting (weight-for-height (WFH) z-scores  $\leq -2$ ). The WHO defines “moderate-severe malnutrition” based on these three variables (World Health Organization). Two groups were generated for each indicator, dichotomizing each respondent’s z-score at  $-2.0$ , representing a score of 2 SDs below the mean. The Bayley-III scores were also compared by groups defined by the presence of parental concern for developmental delay in the participating children. All comparisons were conducted using the two-sided *t*-test with parametric assumptions reasonably satisfied.

**2.8.3 Reliability**—Cronbach’s alpha was calculated for five raw subtests; an acceptable value is 0.70 to over 0.90 (Tavakol & Dennick, 2011), with 0.70–0.79 indicating fair internal consistency, 0.8–0.89 indicating good internal consistency, and 0.90 and above indicating excellent internal consistency. Clinically-informed imputation was performed for missing items using a method that has precedent for this tool (Bayley, 2006) which entailed the following: All items preceding the final starting point of test administration were assumed to be answered correctly; all items following the final testing end point were assumed to be incorrect. We removed testing items that had no variability amongst all participants. We also removed two items from the receptive communication domain because they required the differentiation of gender pronouns, which are not present in Swahili. Raw Cronbach Coefficient Alpha values were reported.

The parents or guardians of eight children were willing to return for retest within two weeks for the test-retest reliability evaluation. Mean and SD were computed for five raw subtests at first and second tests. The means of the two tests were compared using the paired *t*-test. We describe these results descriptively, due to the sample size. We calculated test-retest reliability using the absolute-agreement version of the intra-class correlation (ICC), and specifying occasions as a random effect. The Pearson correlation was calculated to evaluate its results in the context of prior studies that reported it. We used the following guideline to interpret ICC values: less than 0.50, between 0.50 and 0.70, between 0.70 and 0.90, and greater than 0.90 are indicative of poor, moderate, good, and excellent reliability, respectively, which is similar to the guideline by Koo et al. (Koo & Li, 2016), except when, being consistent with Nunnally and Bernstein (Nunnally & Bernstein, 1994), we used 0.70 instead of 0.75 as the cutoff between moderate and good reliability.

All analyses were conducted in SAS version 9.4.

### 3 Results

#### 3.1 Pilot Testing and Evaluation

A summary of the scaled scores of the Bayley-III subtests, by gender, is noted in Table 2. Females showed significantly higher fine motor scaled scores compared to males; however, all other domains showed no statistically significant difference between the groups.

### 3.2 Validity

Within the CFA, goodness of fit was calculated and results from three alternative CFA models can be seen in Table 3. The one-factor model was unacceptable, with  $p$ -value  $< 0.05$  and all other fit indices indicating poor fit across all groups. For the all-ages group, the two- and three-factor models fit fairly comparably, and their values were minimally different, according to the SRMR and CFI. For children aged 18–22 months, the three-factor model demonstrated superior fitted indices compared with the two-factor model; for ages 23–35, the three-factor model was very well-fitted model, and the two-factor was only reasonably fitted. The two- and three-factor models had a similar, reasonably good fit for males and females, as well as HIV-exposed, uninfected and HIV-unexposed groups. The raw factor loadings (their standard errors) and the standardized coefficients of subtests are presented for the three-factor models in Figure 1. The standardized coefficients, indicating the relative strength of the loadings on each factor, were  $>0.66$ , indicating strong associations between subtests and their factors. No concerns for estimation values existed within the models, as the unstandardized coefficients and standard errors were all in acceptable ranges without signs of inflation.

Table 4 displays the Pearson correlation between the raw scores for the five subtests. Receptive communication was most highly correlated with fine motor skill through nearly all sub-group analyses. Other domains were generally moderately correlated.

Scatterplots were used to reveal the relationships between subtest raw scores and age for the whole sample, along with the fitted line from linear regression models (Figure 2). All five subtests showed significantly positive linear associations with child age (all  $p$  values for slope  $< 0.001$ ). Moreover, 27% of the variance for cognitive score, 41% for receptive communication, 18% for expressive communication, 37% for fine motor, and 32% for gross motor was explained by age in the linear regression models.

Communication and motor scores were significantly lower in children with moderate-to-severe stunting status compared to those with no or little stunting (Table 5). No significant difference was found between healthy children and those with moderate-to-severe underweight status or wasting, although fewer than 15 children total met that criteria for being underweight or wasted. Additionally, children whose parents were concerned about their development had lower communication and motor scores compared to children whose parents were not concerned. While cognition scores followed patterns of difference similar to communication and motor scores in each of the known-groups/convergent validity analyses, none differed significantly.

### 3.4 Reliability

All of the five subtest domains demonstrated strong internal consistency, with scores ranging from 0.88 to 0.96. Four of the domains demonstrated excellent internal consistency, including the cognition domain with a score of 0.92; receptive communication domain with 0.90; expressive communication domain with 0.94; and gross motor domain with 0.96. The fine motor domain had the lowest score of 0.88, which is still considered good internal consistency. Test-retest reliability was considered preliminary since only eight participants



had two-week retest data (Table 6). The scores were increased slightly from the first to second tests across all subtests. All subtests demonstrated good test-retest reliability, according to the ICC (0.75 – 0.92), with both motor subtests showing excellent reliability (0.92). The ICC values should be considered preliminary given the small test-retest sample size.

## 4 Discussion

Within this study, we described the process by which we culturally adapted the Bayley-III for use in young Kenyan children and evaluated the validity and reliability of the test. To our knowledge, this is the first study to explore the validity and reliability of Bayley-III within this population; thus, we shed light on the suitability of the Bayley-III for use with young Kenyan children. Content validity was evaluated through forward-and-backward translations, cognitive interviews, and a small pilot administration to ensure that testing items would be interpreted as intended. Both the two- and three-factor models were acceptable within the confirmatory factor analysis, indicating construct validity. Children had lower communication and motor composite Bayley-III scores when stunted or when a parental concern for developmental delay was expressed, which are factors that tend to be associated with delays, demonstrating convergent validity. Furthermore, raw scores increased as expected as children aged. This validation and our limited, yet encouraging, reliability results indicate that our culturally adapted Bayley-III is a reasonably valid and reliable tool for use with Kenyan children aged 18–36 months.

The CFA was a major component of our construct validity analysis. The CFA demonstrated that the culturally adapted Bayley-III performed reasonably well. Both the two- and three-factor models fit the data well overall, and a minimal change in optimal factor structure was found between age groups. Change across age groups is not unexpected as their structure is unstable due to new differing exposures in their home experiences at different stages of their developmental trajectory (Martins, Alves, & Almeida, 2016). Similar findings were noted in a psychometric study of Bayley-III in Vietnam, although a single factor (general development) was specified for their model (Sun et al., 2019). Despite the CFA's robust ability to aid in validity testing, few studies have used this analysis. One study examined factor structure and invariance the Bayley-III across seven international sites and while the tool was found to be valid at each site, challenges with invariances between international site suggest that mean comparisons between groups within a given location would be useful. (Pendergast et al., 2018). Another study used principal components analysis, instead of CFA, for a psychometric study of the Bayley-III in Persian children, which is more exploratory in nature than CFA (Soleimani et al., 2016). Conceptually, the three-factor model is most consistent overall, as both expressive and receptive communication loaded onto one factor, fine and gross motor loaded on another, and cognition loaded upon the third factor. These CFA results support the validity of the adapted Bayley-III.

Within this study, the convergent validity of the cognition subtest findings differed from the communication and motor subtests, and the cognition subtest was the only test that did not have significantly lower scores in children who were moderately or severely stunted, a known risk factor for poor development, or those whose caregivers believed they had a delay

(Walker et al., 2011). It is unclear why this discrepancy may have occurred. One possible hypothesis may be that the tasks within the cognition section more likely required the children to interact with items for which they had varying degrees of familiarity. Many households in sub-Saharan Africa have few stimulating items specifically for children. In Uganda, approximately 2% of households have a book for children and 50% of households have toys (UNICEF, 2016). While many other aspects of the cognitive domains performed well for the validity and reliability analysis, a small degree of caution might be considered when interpreting cognitive scores. The key component to our reliability analysis was evaluating internal consistency. Internal consistency is the degree of interrelations among items on a particular test, or subtest within our analysis (American Psychological Association, 2020). Our study demonstrated that the adapted version of the Bayley-III had strong internal consistency, with most of the subtests having excellent internal consistency. The combination of both a well-performing CFA model and strong internal consistency for each of the five subtest make a strong argument in support of the reliability and strong psychometric properties of this culturally adapted version of the Bayley -III for young Kenyan children.

The test-retest Pearson correlation within our results corresponded approximately with the ICC and are comparable with prior studies measuring reliability of the Bayley-III internationally (Hanlon et al., 2016; Yu et al., 2013). However, the ICC that we used may be theoretically preferred as its formula is based on absolute agreement, whereas the Pearson coefficient assesses relative consistency. The Pearson coefficient allows mean systematic shifts in scores over time to count toward reliability, which represents an inappropriate “inflation” of the reliability index if a researcher believes, as we do, that reliability should be conceptualized as agreement and not solely consistency in these domains. Because of the limited number of participants returning for retesting, the paired t-test p-value was not emphasized (it found no significant mean change), although an improvement of scores was descriptively noted between test administrations. These results are aligned with the reliability data from the Bayley -III within the normative population (Bayley, 2006). Interrater reliability was not performed as only one individual administered the Bayley-III. The individual self-reflected on test videos and submitted these at random for review by the clinical psychologist for quality control. Future studies using this tool within this population should focus on a more robust reliability analysis with a larger test-retest sample size.

While the focus of our manuscript was to evaluate various forms of validity for the subtests of the Bayley-III, the low mean scaled scores of our study population were also notable. Within our study, our scaled scores were much lower compared with the U.S. norms, with most of our sample’s mean scaled subtest scores falling over one SD below the normed scaled mean. The Bayley-III has shown that reliance on its U.S. population reference curves would misclassify children in other cultures (Cromwell et al., 2014). In a population of young Malawian children, raw scores of the Bayley-III normative data and the Malawian cohort tracked similarly until children reached between 9–12 months of age, after which the slope of the Malawian children’s raw scores increased at a slower pace compared to the normative data (Cromwell et al., 2014). Some may state that these results suggest that the environments in which these children are developing may not be optimally stimulating. While not atypical for the population, over a 35% of the children in our study had a mother

that had only completed up to a primary school education. Lower levels of maternal education are known to be a strong predictor of lower scores around two years of age (Lennon, Gardner, Karmel, & Flory, 2010).

Alternatively, it is also very likely that certain aspects of development are more adaptive and highly prioritized in settings like Kenya compared to the U.S, but are not measured with assessments like the Bayley-III. If measured, U.S. children may score sub-optimally, while Kenyan children may excel. As with all assessments, the child is only being evaluated on the constructs deemed important by the test developers. Similarly, some of the alterations we made to the test administration, such as eliminating the pronoun-specific items, negatively impacted scoring when compared to the standard administration. The score was then not an accurate reflection of the child's ability or intellectual capacity, as it was being compared to a context within which pronoun-specific items existed. As such, alternations in the scale should not be interpreted to reflect any ethnic, cultural, or racial differences in a given population's abilities overall.

Factors like these influence the scores of the adapted Bayley-III, limit its use as a diagnostic tool for developmental delay, and complicate the interpretation of the Bayley-III scores across cultures. The Bayley-III has had challenges with invariance across international sites (Pendergast et al., 2018), and this challenge is also likely to be faced when comparing the use of the Bayley-III between the U.S. and Kenya. More research is needed in this area. Nevertheless, we believe the tool is useful in most research settings. With the results of our psychometric analysis, we feel confident that the adapted Bayley-III validly and reliably evaluates constructs within cognitive, communication, and motor domains. Thus, the numeric scores of the adapted Bayley-III may be used for comparisons between groups within a research setting, but not for cross-comparison with studies in different cultures. Raw scores are often used for these comparisons; however, due to the narrow age windows required for interpreting raw scores, this may not always be a feasible option. A large validation study would be useful to compare normative data from a representative Kenyan cohort to the U.S. or develop a Kenyan-specific standardized norm. However, this would require significant resources and may not capture the developmental constructs that are most meaningful to the local Kenyan context. For those interested in comparing developmental data across cultures and settings, focusing efforts on the development of tools that can be used cross-culturally with minimal adaptation, such as the Malawi Developmental Assessment Tool, is an important future direction (Gladstone et al., 2010).

Some limitations exist in the interpretation of our results. Our study population overrepresented children who were HIV-exposed but uninfected due to the primary study. This may have impacted the number of children who were stunted and underweight, as HIV-exposed children have higher rates of malnutrition (McHenry et al., 2019), which is known to be associated with worse developmental scores in general, as well as within this population (McDonald et al., 2013). Similarly, our population was recruited from an academic center at a public referral hospital, illustrating another way in which the patient population might not be representative of other populations in Kenya. Another limitation is our small sample size, further limited by the age range that was included. Additionally, 12 children enrolled in the study were unable to complete the Bayley-3 assessment. We believe

this was due to the prolonged time spent at the clinic prior to recruitment. While families were offered the opportunity to return for study activities on a different day, many preferred to engage in activities on the day of recruitment to avoid transportation issues. This likely led to more children failing the Bayley-III than would have otherwise. We had initially hoped to perform test-retest with 10% of our study population, but only 5% returned for testing. Failure to return may have been due to additional travel required for the follow-up visit, as their initial visit often coincided with a previously scheduled appointment. This limited our ability to perform a robust test-retest analysis. Finally, while we used paternal concern for delay as a measure of convergent validity, we acknowledge that it is not a solid measure for validation. In Kenya, no formal counseling occurs with families on expected developmental milestones; thus, each parent may have a different individual threshold for what constitutes as a development delay.

Within this study, we conducted factor analytic model fit directly for a-priori selected subgroups, as well the overall sample, due to the hypothesized importance of these subgroups in assessments like the Bayley-III. Our goal was to assess whether the fit of the overall model was reasonably adequate for subgroups, not just for the overall sample, when comparing models with a different number of factors. However, we thank a reviewer for pointing out that other standard procedures for testing measurement invariance (van de Schoot, Lugtig, & Hox, 2012) could also be applied for statistically testing whether subgroups statistically differ on factor measurement features such as loadings, intercepts, and residuals.

## 5 Conclusions

This psychometric analysis found that the culturally adapted Bayley-III performed reasonably well and can be used in Kenya. When scores are scaled, children within this study performed worse than the normative population. The reasons for this finding are multifactorial, but likely primarily due to cultural differences in testing items and scoring and possibly due to lower levels of maternal education. While minimal caution may be considered when interpreting results for cognition testing, overall, the culturally adapted Bayley-III is a valid and psychometrically acceptable test for use in young Kenyan children.

### Funding:

This study was funded by Indiana University's Morris Green Physician Scientist Development Program and Indiana CTSI at Indiana University School of Medicine. MSM's time during the implementation of this study was funded by a training grant entitled "Training in STIs and Other Infections of Global Health Significance" (T32AI007637, PI: Dr. Wools Kaloustian) from the National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA. The validity analysis was performed while MSM was supported by an NIH K23 Mentored Patient-Oriented Research Career Development Award (K23MH116808).

### References

Abubakar A, Holding P, Van Baar A, Newton CRJC, & Van De Vijver FJR (2008). Monitoring psychomotor development in a resource-limited setting: an evaluation of the Kilifi Developmental Inventory. *Annals of tropical paediatrics*, 28(3), 217–226. doi:10.1179/146532808X335679 [PubMed: 18727851]

- American Psychological Association. (2020). APA Dictionary of Psychology: Internal Consistency. Retrieved from <https://dictionary.apa.org/internal-consistency>
- Aylward GP, & J, Z. (2019). The Bayley Scales: Clarification for Clinicians and Researchers. Retrieved from <https://www.pearsonassessments.com/content/dam/school/global/clinical/us/assets/bayley-4/bayley-4-technical-report.pdf>
- Ballot DE, Ramdin T, Rakotsoane D, Agaba F, Davies VA, Chirwa T, & Cooper PA (2017). Use of the Bayley Scales of Infant and Toddler Development, Third Edition, to Assess Developmental Outcome in Infants and Young Children in an Urban Setting in South Africa. *Int Sch Res Notices*, 2017, 1631760. doi:10.1155/2017/1631760 [PubMed: 28835912]
- Bayley N (2006). *Bayley Scales of Infant and Toddler Development*, 3rd edition. San Antonio, Texas: Pearson.
- Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, ... Grantham-McGregor S (2017). Early childhood development coming of age: science through the life course. *The Lancet*, 389(10064), 77–90. doi:10.1016/S0140-6736(16)31389-7
- Cromwell EA, Dube Q, Cole SR, Chirambo C, Dow AE, Heyderman RS, & Van Rie A (2014). Validity of US norms for the Bayley Scales of Infant Development-III in Malawian children. *Eur J Paediatr Neurol*, 18(2), 223–230. doi:10.1016/j.ejpn.2013.11.011 [PubMed: 24423629]
- Einterz RM, Kimaiyo S, Menegech HN, Khwa-Otsyula BO, Esamai F, Quigley F, & Mamlin JJ (2007). Responding to the HIV pandemic: the power of an academic medical partnership. *Acad Med*, 82(8), 812–818. doi:10.1097/ACM.0b013e3180cc29f1 [PubMed: 17762264]
- Gladstone M, Lancaster GA, Umar E, Nyirenda M, Kayira E, van den Broek NR, & Smyth RL (2010). The Malawi Developmental Assessment Tool (MDAT): The Creation, Validation, and Reliability of a Tool to Assess Child Development in Rural African Settings. *PLOS Medicine*, 7(5), e1000273. doi:10.1371/journal.pmed.1000273 [PubMed: 20520849]
- Hanlon C, Medhin G, Worku B, Tomlinson M, Alem A, Dewey M, & Prince M (2016). Adapting the Bayley Scales of infant and toddler development in Ethiopia: evaluation of reliability and validity. *Child Care Health Dev*, 42(5), 699–708. doi:10.1111/cch.12371 [PubMed: 27381579]
- Hua J, Li Y, Ye K, Ma Y, Lin S, Gu G, & Du W (2019). The reliability and validity of Bayley-III cognitive scale in China's male and female children. *Early Hum Dev*, 129, 71–78. doi:10.1016/j.earlhumdev.2019.01.017 [PubMed: 30711715]
- Inui TS, Nyandiko WM, Kimaiyo SN, Frankel RM, Muriuki T, Mamlin JJ, ... Sidle JE (2007). AMPATH: living proof that no one has to die from HIV. *J Gen Intern Med*, 22(12), 1745–1750. doi:10.1007/s11606-007-0437-4 [PubMed: 17972138]
- Khan NZ, Muslima H, Shilpi AB, Begum D, Parveen M, Akter N, ... Darmstadt GL (2013). Validation of rapid neurodevelopmental assessment for 2- to 5-year-old children in Bangladesh. *Pediatrics*, 131(2), e486–494. doi:10.1542/peds.2011-2421 [PubMed: 23359579]
- Kline R (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford publications.
- Koo TK, & Li MY (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, 15(2), 155–163. doi:10.1016/j.jcm.2016.02.012 [PubMed: 27330520]
- le Roux SM, Donald KA, Brittain K, Phillips TK, Zerbe A, Nguyen KK, ... Myer L (2018). Neurodevelopment of breastfed HIV-exposed uninfected and HIV-unexposed children in South Africa. *AIDS (London, England)*, 32(13), 1781–1791. doi:10.1097/QAD.0000000000001872
- Lennon E, Gardner J, Karmel B, & Flory M (2010). Bayley Scales of Infant Development. In Benson JB; & Haith MM (Eds.), *Language, Memory, and Cognition in Infancy and Early Childhood* (pp. 37–48): Elsevier.
- Martins AA, Alves AF, & Almeida LS (2016). The factorial structure of cognitive abilities in childhood. *European Journal of Education and Psychology*, 9(1), 38–45. doi:10.1016/j.ejeps.2015.11.003
- McDonald CM, Manji KP, Kupka R, Bellinger DC, Spiegelman D, Kisenge R, ... Duggan CP (2013). Stunting and wasting are associated with poorer psychomotor and mental development in HIV-exposed Tanzanian infants. *The Journal of nutrition*, 143(2), 204–214. doi:10.3945/jn.112.168682 [PubMed: 23256148]

- McHenry MS, Apondi E, Ayaya SO, Yang Z, Li W, Tu W, ... Vreeman RC (2019). Growth of young HIV-infected and HIV-exposed children in western Kenya: A retrospective chart review. *PLoS One*, 14(12), e0224295. doi:10.1371/journal.pone.0224295 [PubMed: 31800588]
- McHenry MS, McAteer CI, Oyungu E, McDonald BC, Bosma CB, Mpofu PB, ... Vreeman RC (2018). Neurodevelopment in Young Children Born to HIV-Infected Mothers: A Meta-analysis. *Pediatrics*, 141(2). doi:10.1542/peds.2017-2888
- Mukaka MM (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal : the journal of Medical Association of Malawi*, 24(3), 69–71. [PubMed: 23638278]
- Nunnally J, & Bernstein I (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill, Inc.
- Pendergast LL, Schaefer BA, Murray-Kolb LE, Svensen E, Shrestha R, Rasheed MA, ... Seidman JC (2018). Assessing development across cultures: Invariance of the Bayley-III Scales Across Seven International MAL-ED sites. *Sch Psychol Q*, 33(4), 604–614. doi:10.1037/spq0000264 [PubMed: 30507236]
- Ranjitkar S, Kvestad I, Strand TA, Ulak M, Shrestha M, Chandyo RK, ... Hysing M (2018). Acceptability and Reliability of the Bayley Scales of Infant and Toddler Development-III Among Children in Bhaktapur, Nepal. *Front Psychol*, 9, 1265. doi:10.3389/fpsyg.2018.01265 [PubMed: 30087639]
- Soleimani F, Azari N, Vameghi R, Sajedi F, Shahshahani S, Karimi H, ... Gharib M (2016). Is the Bayley Scales of Infant and Toddler Developmental Screening Test, Valid and Reliable for Persian Speaking Children? *Iran J Pediatr*, 26(5), e5540. doi:10.5812/ijp.5540 [PubMed: 28203335]
- Sun L, Sabanathan S, Thanh PN, Kim A, Doa TTM, Thwaites CL, ... Wills B (2019). Bayley III in Vietnamese children: lessons for cross-cultural comparisons. *Wellcome Open Res*, 4, 98. doi:10.12688/wellcomeopenres.15282.1 [PubMed: 31448341]
- Tavakol M, & Dennick R (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53–55. doi:10.5116/ijme.4dfb.8dfd [PubMed: 28029643]
- UNICEF. (2016, 6 2020). Percentage of children aged 0–59 months who have learning materials at home. Retrieved from <https://data.unicef.org/topic/early-childhood-development/home-environment/>
- UNICEF. (2018). *Situation Analysis of Children and Women in Kenya, 2017*. In UNICEF (Ed.). Nairobi, Kenya: UNICEF.
- van de Schoot R, Lugtig P, & Hox J (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology*, 9(4), 486–492. doi:10.1080/17405629.2012.686740
- Walker SP, Wachs TD, Grantham-McGregor S, Black MM, Nelson CA, Huffman SL, ... Richter L (2011). Inequality in early childhood: risk and protective factors for early child development. *The Lancet*, 378(9799), 1325–1338. doi:10.1016/S0140-6736(11)60555-2
- Wedderburn CJ, Yeung S, Rehman AM, Stadler JAM, Nhapi RT, Barnett W, ... Donald KA (2019). Neurodevelopment of HIV-exposed uninfected children in South Africa: outcomes from an observational birth cohort study. *Lancet Child Adolesc Health*, 3(11), 803–813. doi:10.1016/s2352-4642(19)30250-0 [PubMed: 31515160]
- Willis GB, & Miller K (2011). Cross-Cultural Cognitive Interviewing: Seeking Comparability and Enhancing Understanding. *Field Methods*, 23(4), 331–341. doi:10.1177/1525822X11416092
- World Health Organization. (1 7 2020). WHO Anthro Survey Analyser and other tools.
- Yu YT, Hsieh WS, Hsu CH, Chen LC, Lee WT, Chiu NC, ... Jeng SF (2013). A psychometric study of the Bayley Scales of Infant and Toddler Development - 3rd Edition for term and preterm Taiwanese infants. *Res Dev Disabil*, 34(11), 3875–3883. doi:10.1016/j.ridd.2013.07.006 [PubMed: 24029804]

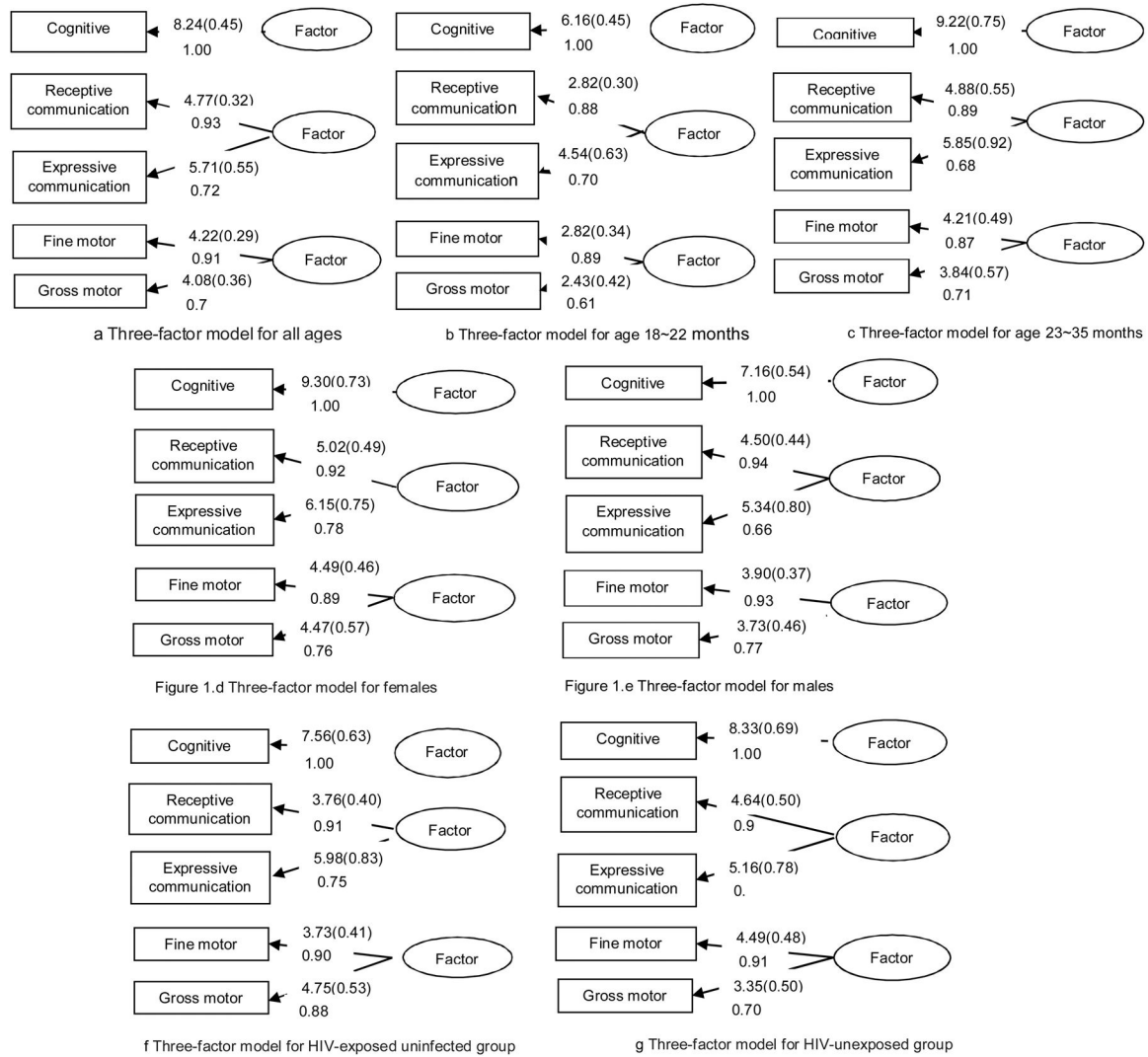
**Highlights:**

- This study described the development of the Kenyan adapted Bayley Scales (BSID-III)
- Scale found to be psychometrically acceptable tool to measure child development
- The scaled/composite scores should not be used to define Kenyan developmental norms
- Adapted BSID-III can be a useful tool to compare groups within research settings

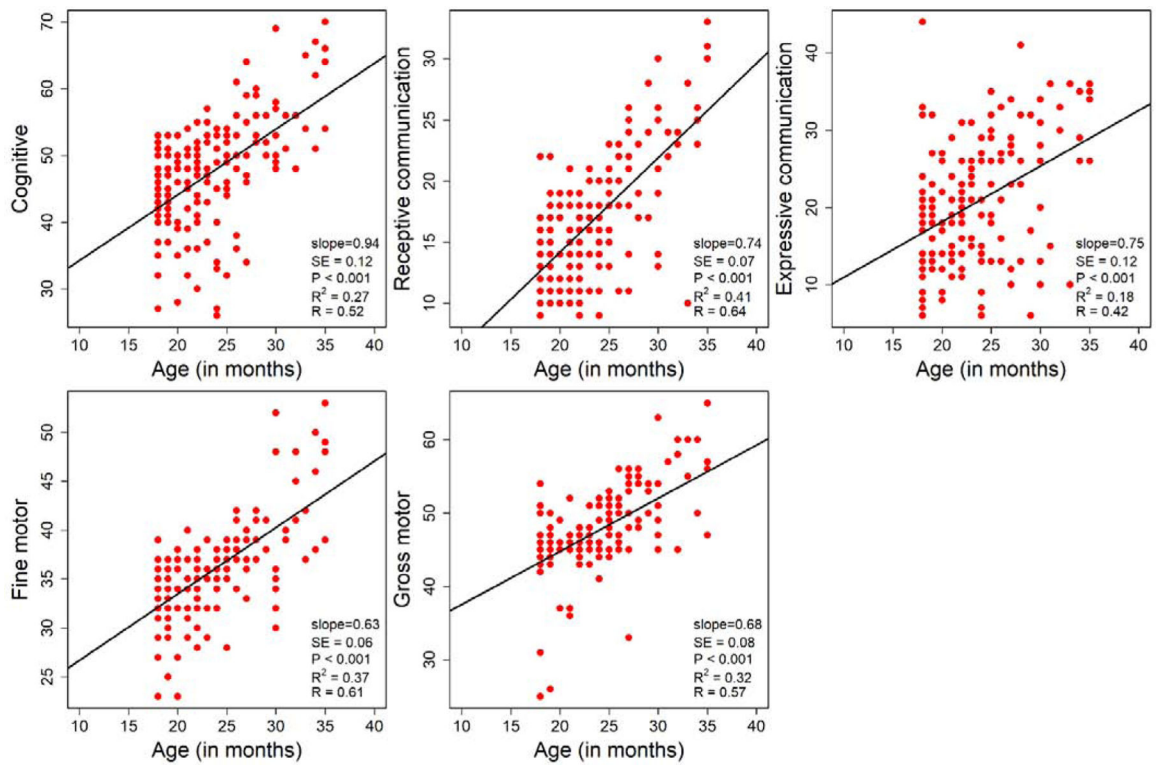
### What this paper adds?

This study describes the process of culturally adapting the Bayley-III for use in young Kenyan children and an evaluation of the validity and reliability of this test. To our knowledge, this is the first study to explore the validity and reliability of the Bayley-III within this population, and, thus, sheds light on the suitability of the Bayley-III for use with young Kenyan children. Content validity and semantic equivalence was evaluated through forward-and-backward translations, cognitive interviews, and a small pilot administration to ensure that testing items would be interpreted as intended. For construct validity, both the two- and three-factor models were acceptable within the confirmatory factor analysis. For convergent validity, children had lower Bayley-III for language and motor composite scores when factors known to be associated with delays were present, e.g., they were stunted, or parental concern for developmental delay was indicated. Furthermore, raw scores increased as expected as children aged. These validity results, in addition to the limited, yet encouraging, reliability results, suggest that the culturally adapted Bayley-III is a valid and reliable research tool for use in Kenyan children aged 18–36 months when comparing two or more groups within a population.





**Figure 1:** Confirmatory Factor Analysis- loading standards. The top row of each path shows the estimated unstandardized coefficient (standard error). The bottom row shows the **standardized coefficient**. And all p values are less than 0.001 for all paths in all two- and three-factors models.



**Figure 2.**  
Relationships between subtests and age, fit with linear regression models

**Table 1.**

## Characteristics of Study Sample

Characteristics	Overall (n= 72) N (%)	Males (n=89) n (%)	Females (n=83) n (%)	$\chi^2$ or t	P value
<b>Preterm (&lt;37 weeks)</b>					
Yes	13 (7.7)	9 (10.3)	4 (4.9)		
No	154 (91.1)	77 (88.5)	77 (93.9)		
Not sure	2 (1.2)	1 (1.2)	1 (1.2)	1.8	0.456
<b>Maternal education</b>					
No school or partial primary school	27 (15.7)	16 (18.0)	11 (13.3)		
Complete primary school	34 (19.8)	17 (19.1)	17 (20.5)		
Some or complete secondary school	57 (33.1)	34 (38.2)	23 (27.7)		
Post-secondary school	54 (31.4)	22 (24.7)	32 (38.6)	4.7	0.195
<b>Developmental Delay</b>					
Maternal concern for developmental delay	31 (18.0)	17 (19.1)	14 (16.9)	0.1	0.703
Family history of developmental delay	23 (13.5)	10 (11.4)	13 (15.9)	0.7	0.392
<b>Scores, Mean (SD)</b>					
Score on PHQ-9	3.5 (4.0)	3.4 (3.9)	3.5 (4.0)	0.2	0.880
Score on WAMI	0.6 (0.2)	0.7 (0.2)	0.6 (0.2)	0.3	0.804
Child height-for-age Z score	-1.2 (1.5)	-1.3 (1.6)	-1.1 (1.3)	1.0	0.299
Child weight-for-age Z score	-0.2 (1.1)	-0.3 (1.1)	-0.1 (1.1)	1.2	0.224
Child weight-for-height Z score	0.5 (1.0)	0.5 (1.1)	0.5 (0.9)	0.2	0.832

Notes: PHQ-9- Patient Health Questionnaire; WAMI- Water, Assets, Maternal education, Income; SD- Standard Deviation

**Table 2.**Scaled scores on sub-tests of the Bayley Scales of Infant and Toddler Development, 3<sup>rd</sup> edition, by gender

	Overall (n= 72) N (%)	Males (n=89) N (%)	Females (n=83) N (%)	<i>t</i>	P value
Cognitive, Mean (SD)	4.9 (2.3)	4.8 (2.1)	5.0 (2.4)	0.4	0.657
Receptive communication, Mean (SD)	5.2 (2.1)	5.0 (2.0)	5.4 (2.1)	1.4	0.158
Expressive communication, Mean (SD)	6.4 (3.2)	6.4 (3.2)	6.5 (3.2)	0.2	0.821
Fine motor, Mean (SD)	8.4 (2.5)	8.0 (2.3)	8.9 (2.7)	2.3	0.022*
Gross motor, Mean (SD)	5.5 (1.9)	5.4 (1.7)	5.6 (2.2)	0.8	0.413

Notes. SD = standard deviation; t = t value

\* - denotes significance

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**

Goodness-of-Fit Indices from Confirmatory Factor Analysis Models Based on Raw Subtest Scores by Age Groups

Model	Chi-square test				AGFI	RMSEA	SBC	CFI	SRMR
	$\chi^2$	df	$\chi^2/df$	P value					
<b>All Ages (n=172)</b>									
One factor	180.58	6	30.10	<0.001	0.19	0.41	226.91	0.65	0.17
Two factors	14.27	5	2.85	0.014	0.90	0.10	65.75	0.98	0.03
Three factors	7.13	3	2.38	0.068	0.92	0.09	68.90	0.99	0.02
<b>Ages: 18–22 months (n=95)</b>									
One factor	78.97	6	13.16	<0.001	0.31	0.36	119.95	0.53	0.20
Two factors	9.86	5	1.97	0.079	0.88	0.10	55.40	0.97	0.04
Three factors	1.90	3	0.63	0.594	0.96	<0.001	56.55	1.00	0.02
<b>Ages: 23–35 months (n=77)</b>									
One factor	57.31	6	9.55	<0.001	0.37	0.34	96.41	0.72	0.15
Two factors	6.02	5	1.20	0.304	0.91	0.05	49.46	0.99	0.03
Three factors	4.34	3	1.45	0.227	0.89	0.08	56.46	0.99	0.03
<b>Females (n=83)</b>									
One factor	90.53	6	15.09	<0.001	0.14	0.41	130.30	0.68	0.17
Two factors	5.54	5	1.11	0.353	0.93	0.04	49.73	1.00	0.02
Three factors	4.73	3	1.58	0.193	0.89	0.08	57.76	0.99	0.02
<b>Males (n=89)</b>									
One factor	94.84	6	15.81	<0.001	0.22	0.41	135.24	0.61	0.18
Two factors	13.23	5	2.65	0.021	0.82	0.14	58.11	0.96	0.04
Three factors	4.07	3	1.36	0.254	0.91	0.06	57.93	1.00	0.03
<b>HIV-exposed uninfected (n=74)</b>									
One factor	99.62	6	16.60	<0.001	0.14	0.46	138.35	0.57	0.20
Two factors	36.60	5	7.32	<0.001	0.48	0.29	79.64	0.86	0.08
Three factors	14.20	3	4.73	0.003	0.63	0.23	65.85	0.95	0.05
<b>HIV-unexposed (n=74)</b>									
One factor	90.50	6	15.08	<0.001	0.08	0.44	129.24	0.59	0.20
Two factors	3.65	5	0.73	0.602	0.94	<0.01	46.69	1.00	0.02
Three factors	3.31	3	1.10	0.346	0.91	0.04	54.96	1.00	0.02

Notes. AGFI = adjusted goodness of fit index; RMSEA = root mean square error of approximation; SBC = Schwarz's Bayesian criterion; CFI = comparative fit index; SRMR = standardized root mean square residual

**Table 4.**

Pearson Correlation Matrix for Five Raw-Score Subtests

	Cognitive	Receptive Comm.	Expressive Comm.	Fine Motor	Gross Motor
<b>All Age Groups (n=172)</b>					
Cognitive	1.00	0.66	0.59	0.65	0.55
Receptive communication		1.00	0.67	0.78	0.64
Expressive communication			1.00	0.56	0.49
Fine motor				1.00	0.69
Gross motor					1.00
<b>Ages: 18–22 months (n=95)</b>					
Cognitive	1.00	0.51	0.47	0.47	0.32
Receptive communication		1.00	0.62	0.62	0.41
Expressive communication			1.00	0.45	0.34
Fine motor				1.00	0.54
Gross motor					1.00
<b>Ages: 23–35 months (n=77)</b>					
Cognitive	1.00	0.63	0.57	0.64	0.53
Receptive communication		1.00	0.61	0.73	0.58
Expressive communication			1.00	0.49	0.43
Fine motor				1.00	0.61
Gross motor					1.00
<b>Females (n=83)</b>					
Cognitive	1.00	0.69	0.66	0.68	0.54
Receptive communication		1.00	0.72	0.79	0.68
Expressive communication			1.00	0.63	0.58
Fine motor				1.00	0.67
Gross motor					1.00
<b>Males (n=89)</b>					
Cognitive	1.00	0.62	0.52	0.61	0.55
Receptive communication		1.00	0.62	0.75	0.60
Expressive communication			1.00	0.49	0.40
Fine motor				1.00	0.72
Gross motor					1.00
<b>HIV-exposed uninfected (n=74)</b>					
Cognitive	1.00	0.65	0.66	0.48	0.57
Receptive communication		1.00	0.68	0.70	0.65
Expressive communication			1.00	0.44	0.47
Fine motor				1.00	0.79
Gross motor					1.00

	Cognitive	Receptive Comm.	Expressive Comm.	Fine Motor	Gross Motor
<b>HIV-unexposed (n=74)</b>					
Cognitive	1.00	0.63	0.52	0.69	0.46
Receptive communication		1.00	0.62	0.80	0.67
Expressive communication			1.00	0.63	0.45
Fine motor				1.00	0.64
Gross motor					1.00

Notes. Orange depicts highly correlated values; Green depicts moderately correlated values

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 5.**

Known-Groups/Convergent Validity: Comparisons on Indicators of Potential Development Impairment

	<b>Bayley-III Composite Scores</b>		<i>t</i>	<b>P value</b>
	<b>Mean (SD)</b>			
	<b>HFA z-score -2 (n=44)</b>	<b>HFA z-score &gt;-2 (n=120)</b>		
Cognitive	72.50 (11.39)	74.71 (11.36)	1.1	0.272
Communication	70.84 (12.46)	77.19 (14.03)	2.6	0.009*
Motor	78.25 (10.98)	82.78 (11.43)	2.3	0.024*
	<b>Parental concern = Yes (n=31)</b>		<b>Parental concern = No (n=141)</b>	
Cognitive	72.58 (10.64)	74.75 (11.47)	1.0	0.335
Communication	67.84 (10.34)	77.20 (13.98)	3.5	0.001*
Motor	76.68 (12.24)	82.77 (11.01)	2.7	0.007*

Notes. Bayley-III = Bayley Scales of Infant and Toddler Development, 3<sup>rd</sup> edition; HFA = height-for-age; SD= standard deviation

\*-denotes significance

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 6.**

## Test-Retest Reliability for Raw-Score Subtests

Raw score	First test (N=8)		Second test (N=8)		Difference (Second-First)		ICC	Pearson correlation
	Mean	SD	Mean	SD	Mean	SD		
Cognitive	50.63	14.17	55.75	15.12	5.13	14.65	0.84	0.84
Receptive communication	20.00	8.94	23.75	8.91	3.75	8.93	0.75	0.75
Expressive communication	20.38	11.45	24.88	7.92	4.50	9.84	0.81	0.87
Fine motor	37.63	10.31	40.13	8.69	2.50	9.53	0.92	0.94
Gross motor	48.75	10.74	51.50	8.42	2.75	9.65	0.92	0.95

Notes: ICC= intraclass correlation; SD = standard deviation

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript