Published in final edited form as:

Child Neuropsychol. 2019 May ; 25(4): 466–481. doi:10.1080/09297049.2018.1497588.

Building capacity in neurodevelopment assessment of children in Sub-Saharan Africa: A Quality Assurance Model to implement standardized neurodevelopment testing

Ruiseñor-Escudero Horacio¹, Familiar Itziar¹, Nyakato Mary², Kuteesa Agatha², Namukooli Jackie³, Ssesanga K. Titus³, Joyce Celeste⁴, Loughton Barbara⁵, Grab Janet⁶, Chernoff Miriam⁷, Vhembo Tichaona⁸, Fairlie Lee⁶, Kamthunzi Portia⁹, Boivin Michael¹, and the IMPAACT P1104s team

¹Psychiatry Department, Michigan State University, East Lansing, USA ²Makerere University-Johns Hopkins University Research Collaboration (MU-JHU), Kampala, Uganda ³Global Health Uganda, Kampala, Uganda ⁴Chris Hani HIV Unit, University of Witwatersrand, Soweto, South Africa ⁵Department of Pediatrics and Child Health, Tygerberg University, Tygerberg, South Africa ⁶Wits Reproductive Health & HIV Institute, Shandukani Clinic, Johannesburg, South Africa ⁷Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Harvard University, Boston, USA ⁸Harare Family Care, University of Zimbabwe, Harare, Zimbabwe ⁹Lilongwe Clinical Research Institute, Kamuzu Central Hospital, Lilongwe, Malawi

Abstract

Compromised neurodevelopment (ND) among infants and children is prevalent in Sub-Saharan Africa. Standardized testing of ND is frequently prohibitive in these contexts, as tests require skilled staff for their application. In this paper, we present a Quality Assurance Model (QualiND) for standardized ND testing, discussing findings and implications from our experience applying the Kauffman Assessment Battery for Children second edition (KABC-II). The QualiND model was implemented within IMPAACT P1104s study; a multi-site, prospective study including 615 children affected by HIV. From 2014–2016, the QualiND managed 18 testers across 6 sites located in 4 African countries applying the KABC-II in 9 local languages. The QualiND is a multilevel, video-assisted iterative model incorporating remote evaluation, feedback and supervision roles. Using an ad hoc rubric, videos of test application were evaluated by experienced staff in a centralized Quality Assurance (QA) center. At each study site, testers and supervisors reviewed feedback from videos received via email from the QA center and devised an action plan to address testing errors and deficiencies. There were few instances of invalid tests and few barriers to test completion. Over 97% of KABC-II tests across sites were considered to be valid by the QA center. Overall, the QualiND model was a useful platform for remote supervision to non-specialist and

Horacio Ruiseñor-Escudero, PhD [Corresponding author], Michigan State University, Department of Psychiatry Research Division, 909 Fee Rd., Rm A321 West Fee Hall, East Lansing, Michigan, 48824, USA. Phone: +(1) 517-884-2288. orcid.org/0000-0002-0367-6976. Twitter: [@horaciorues](https://twitter.com/horaciorues). ruiseore@msu.edu.

Geo-location Information: Kampala, Uganda, Harare, Zimbabwe, Johannesburg, South Africa, Lilongwe, Malawi. Qualiki ND

Conflict of interest: We declare we have no conflicts of interest

minimally trained research staff. The QualiND model may be useful to researchers and organizations involved in measuring early child development using standardized tests in LMIC.

Keywords

Neurodevelopment assessment; quality assurance; KABC-II; HIV; LMIC

Introduction

In low and middle-income countries (LMIC) more than 250 million children, 43% of which are under five years of age, are considered at risk of not developing “the intellectual skills, creativity, and well-being required to become healthy and productive adults” (Maureen M Black et al., 2017; Hack & Costello, 2007). More than half of these at-risk children (66%) live in sub-Saharan Africa (Maureen M Black et al., 2017). Globally, neurodevelopmental disorders in children are highly prevalent as a result of multiple factors such as nutritional deficiencies, environmental exposures, socioeconomic factors, and infectious diseases (Walker et al., 2007). All these factors have a cumulative and negative impact on brain development and early learning, resulting in neurodevelopmental delays in the first years of life that can have permanent impacts later in life (Maureen M Black et al., 2017). In sub-Saharan Africa an estimated 2.9 million (2.6–3.2 million) children are living with HIV and are at risk for a broad range of neurological (Donald et al., 2015; United Nations Program on HIV and AIDS, 2014) and generalized cognitive deficiencies (Blanchette, Lou, Fernandes-Penney, King, & Read, 2001), including motor (Boivin et al., 1995), visual, language and learning disorders (Nozyce et al., 1994). Consequently, assessment of neuropsychological functioning is rapidly becoming an important component of pediatric care and research. Early childhood is considered a point of entry on a pathway of increasing risk for later cognitive difficulties, especially among HIV-affected children in this region (Laughton, Cornell, Boivin, & Van Rie, 2013).

Multiple neuropsychological and behavioral tests are available and have been successfully used to screen or assess cognitive development in children in LMIC, including sub-Saharan Africa (Boivin et al., 2013) and Asia (M. M. Black et al., 2007), although many rely on questionnaires answered by parents or teachers. This limitation can be overcome by standardized neurocognitive tests, where a trained tester directly observes children perform specific tasks. In high-income countries, standardized ND assessment is typically carried out by highly trained professionals, including child psychologists, psychiatrists, and pediatricians. However, this implementation practice would be prohibitive in LMIC where the number of specialized health professionals is typically insufficient (Semrud-Clikeman et al., 2016; World Health Organization, 2017). Task-shifting, defined by the WHO as “the rational redistribution of tasks among health workforce teams” (Organization, 2007), has been increasingly promoted in the global health arena as means to address numerous barriers to the successful implementation of services in LMIC. Following a task-shifting approach, we trained and supervised minimally qualified staff in the implementation of a standardized neurocognitive test for children: the Kauffman Assessment Battery for Children second edition (KABC-II) within the IMPAACT P1104s study. The P1104s study was an

observational study carried out at six clinical sites (South Africa-3 sites, Malawi, Uganda and Zimbabwe) aimed at comparing neuropsychological outcomes among school age children of varying HIV status.

Models are useful as visual representations depicting complex interactions that lead to specific outcomes, such as human behavior, social, environmental events, among others, and can be applied to guide interventions in health related issues (Krieger, 1994; Parker & Aggleton, 2003). In this paper, we describe and discuss how a quality assurance model for ND testing can be used across settings and with personnel of varying experience and backgrounds.

Methods

The IMPAACT P1104S was a two-year multi-site observational study. It involved 615 children at six clinical sites in four sub-Saharan African countries (South Africa (3 sites), Malawi, Uganda, and Zimbabwe). The aim was to assess and compare the neurodevelopmental outcomes in the three cohorts of children ages 5–11 years; 1) HIV-infected, 2) HIV exposed (perinatally) but uninfected (HEU), and 3) HIV unexposed and uninfected infants (HUU). Full description of the study sample can be found elsewhere (Boivin et al., 2018). Demographic characteristics of study children are described in Table 1. In order to administer the KABC-II consistently across sites by personnel with only college-level education and basic neurodevelopmental (ND) assessment skills, we developed a quality assurance model for monitoring and supervision of assessments.

Performance-based measures of cognitive development

Performance-based measures of cognitive development, such as the KABC-II, involve direct observation of the child performing a particular task. In these tests, motor, language, and visual reception skills are evaluated by a trained evaluator using a standardized battery of tests that can include props and toys.

The P1104s study used the Kauffman Assessment Battery for Children second edition (KABC-II) to measure cognitive functioning in school-aged children. The KABC-II was chosen as the principal test for cognitive ability outcomes because it had been previously adapted for use in pediatric HIV research in Uganda (P. Bangirana, B. Giordani, et al., 2009; Boivin, Nakasujja, Sikorskii, Opoka, & Giordani, 2016) and validated in the sub-Saharan African context (Bangirana et al., 2009; Boivin et al., 1995). The KABC-II assesses cognitive abilities through a series of task-oriented instructions described to the child in the preferred language (Sesotho, Setswana, Zulu, Xhosa, Afrikaans, Chichewa, Luganda, Shona, or English). The KABC-II includes five cognitive domains: 1) Sequential Processing (short-term memory), 2) Simultaneous Processing (visual-spatial processing and problem-solving), 3) Learning (immediate and delayed memory), 4) Planning (executive reasoning), and 5) Knowledge (crystallized ability including verbal knowledge). The KABC-II also yields a composite global scale index, the Mental Processing Index (MPI), which provides an overall level of executive functioning, and a global Non-Verbal Index (NVI).

The Quality Assurance Model (QualiND)

The QualiND model was constructed with two supervision levels: local and remote supervision (Figure 1). The local level was composed of two units linked through an iterative procedure: a testing center, where KABC-II evaluations took place, and the quality assurance (QA) center at the Makerere University-Johns Hopkins University Research Collaboration Center in Uganda, where experienced KABC-II evaluators reviewed performed ND assessments and provided feedback. The remote supervision (RS) level represented a technical layer of on-demand supervision by outside experts (e.g. the PI and co-investigators). The remote supervision was set-up to address test issues on a case-by-case basis that the QA center supervisors were unable to resolve. Figure 1 shows the step-wise communication flow between the testing center and the QA center. Individual components of the QualiND model are described below.

The descriptive characteristics of each component of the QualiND model were summarized by site, as well as test completion rates. We also summarized the number and percent of participants for which site staff deemed the KABC-II to be valid. Reliability was assessed through looking at the consistency of test scores throughout the study using the intra-class correlation coefficient (ICC). The ICC was computed using linear mixed models regression approaches after adjusting for site, participant age at entry and sex. Models used restricted maximum likelihood (REML) estimation. Assuming no progression in illness, repeat testing on the same child with the same instrument should yield consistent results

Training and recruiting of ND evaluators

The first step in establishing the QualiND model was to identify individuals at the local study sites who could be trained as ND evaluators. For the P1104s study, potential evaluators were recruited from hospitals (e.g., nurses) and local partner organizations. Criteria for selecting test evaluators included core indicators of aptitude, being a high school graduate, and interest in working with children. Characteristics and information on evaluators are reported in Table 2. Supervisors at each testing center were selected based on previous experience with ND evaluations (e.g. at least 1 year experience administering KABC-II or Mullen Scales of Early Learning). Each of the study site evaluators and their supervisors were trained on the neuropsychological test battery for a one-week period with follow-up training on site by the principal investigator (MJB) during the initial week of enrolment and the initial assessments of study children. All evaluators were proficient in English and the KABC-II was not translated. Rather, evaluators were trained on how to provide test instructions for children in the local language. As part of the training, evaluators worked with the PI in arriving at consistent wordings in the local languages for that site, for each of the KABC-II spoken instructions scripted on the KABC-II administration easels. Training covered ND assessment and theory, introduction to KABC-II materials and manuals, and practice assessing volunteer children under close supervision. Active learning strategies used during training included small group work, where frequently used terms and testing instructions were rehearsed in the local languages with help from supervisors. For the most part, time devoted to practicing ND assessment skills was greater than the time spent in didactic training. By the end of the training, evaluators were expected to obtain at least 90% of maximum score on KABC evaluation rubric as scored by supervisors.

ND assessment training is ideally conducted at the testing center, where evaluators can familiarize with testing conditions specific to the context and local language. When new evaluators replaced staff in the study, a final practice video was sent to the QA center for approval before the evaluator started formal testing of study children. Initial training is necessary but not sufficient to build the skill-level of evaluators required for a complex test like the KABC-II, which involves many variables depending on the child's age. This is why the training phase is considered the foundation of the QualiND model upon which improvements can be made through iterative cycles of supervision.

Neurodevelopmental assessment: the KABC-II

ND assessments were initiated at each study site shortly after training was completed and took place in private, clean, and comfortable rooms, within a quiet environment. Before assessment started, evaluators made every effort to lower the risk of the child becoming distracted, such as instructing the caregiver to sit the child facing the door, making sure the child had been fed before the test, and to turn cell phones off. Throughout the assessment period, at least one on site supervisor was available, in addition to the regional QA center staff that was available via email or online conference call to address any testing issues evaluators might encounter. After each test was completed, evaluators reviewed the scoring sheet before the child left to ensure completeness. Data from the scoring sheets were entered with double data-entry system into a database, where subscale and standardized scores were obtained (by age and gender using American norms as per standard application procedures).

Standardized operating procedures (Kerr et al.) were developed for the KABC-II at each study site detailing procedures for test administration, scoring, and quality assurance. All evaluators reviewed the SOPs monthly in a session led by the supervisor at the testing center and reviewed the KABC training video, reporting the important aspects discussed and lessons learned to the QA center.

As part of training, evaluators were instructed on how to videotape themselves while performing the KABC-II with a study child. Videos were composed of 3 minutes of each of the 4 KABC-II scales in the Luria model (Kaufman & Kaufman, 1983): Sequential Processing Scale, Simultaneous Processing Scale, Learning Ability and Planning Ability, for a total of 12 minutes of recording. The total 12-minute recording was deemed an appropriate sample of KABC-II testing abilities; typically evaluators require 25–50 minutes to complete the core battery of tests in the Luria model.

The caregiver was required to provide a separate informed consent for the video recordings. To protect the child's identity, the video camera was positioned behind the child. The frame of the video had to show the child's hand and the evaluator's face and actions. After completing the recording, evaluators uploaded videos into a previously setup file hosting service (e.g., Dropbox) where the QA center staff could remotely access files for internal review. The file hosting service used was secure and IRB approved. In addition, every 6 months evaluators were required to record a full KABC-II, which was viewed by other evaluators at the testing center as a training exercise. The number of videos sent during the study period varied by site (Table 2), depending on the number of study evaluators and their access to reliable Internet.

Children identified as having significant developmental problems (defined as a Mental Processing Index score <60) were referred to on-site rehabilitative services. Type and intensity of service differed by site, from simple pediatric consultations (Malawi, Zimbabwe), and psychological evaluations (South Africa), to computerized cognitive rehabilitation programs (Health).

Evaluation

The QA center carried out monthly evaluations based on a standardized review method of KABC-II videos using a tool developed *ad hoc*. This scoring rubric had a 3-point scale (ranging from 0 for “evaluator does not show the skill” to 2 for “evaluator shows the skill all the time”) in four proficiency areas for each of the KABC-II subtests (5) based on correctness of administration. The proficiency areas were defined as: 1) Explaining (how the test was explained to the child), 2) Easel and materials (evaluator’s ability with assessment materials included with the test), 3) Familiarity (evaluator’s knowledge of commonly used test rules such as basal rules, drop back points, and stop points), and 4) Efficiency (time required to complete and fluency of test). A total score for each video was derived adding all points obtained across the rubric areas (possible range 0–240). Videos were reviewed on a first come, first served basis with QA staff reviewing approximately 2–4 videos per day.

In addition, the video rubric included a comment section that explained to the evaluator the score obtained in each KABC-II sub-scale. Finally, the rubric provided an overall comment rating the performance of the evaluator as well as instructions on how the he/she could improve future performance.

Staff at the QA center reviewed videos within seven days of being submitted by the evaluator, and entered the rubric scores into a calculating sheet for recording purposes and to obtain the total video score. This database was used to keep track of evaluator’s performance and to generate feedback. Staff at the QA center kept records of feedback from monthly onsite training on SOPs and video reviews by testing center, and assisted with queries that were raised by evaluator groups during these sessions.

Feedback

After evaluating and scoring evaluators KABC-II performance videos, QA center staff generated a written report. In addition to conveying the summary scores, this report detailed errors/omissions during the test and suggested appropriate action for improvement. To provide positive reinforcement, the report also included any positive actions displayed by evaluators during their interactions with the child.

Feedback was sent via email to the evaluator and to the supervisor at the testing center within a week of the video submission. Providing rapid review and input was considered fundamental in the QualiND model, allowing the evaluator the ability to recall mistakes, weaknesses, strengths, and identify areas where improvement was needed. The rapid review and feedback processes promoted corrections in performance if needed and these were implemented as soon as possible.

Plan and Action

Evaluators and supervisors at each testing center reviewed the feedback reports sent by the QA center and jointly agreed on the test areas that required improvement or practice as well as the specific actions needed to achieve these. In this phase, supervisors at testing centers played a central role in helping evaluators correct testing practices. Supervisors were encouraged to work with evaluators by observing their testing sessions and providing on-site feedback and direct oversight. Supervisors were responsible of working with the evaluator in applying recommendation and implementation practices made by the QA center with the evaluator. Because supervisors received the feedback reports for all evaluators, they were in the unique position of identifying frequent challenges met by all evaluators, and they organized small sessions where issues were reviewed as a group. In-site implementation of the QualiND model was measured by compliance with timely sending of videos, compliance with sending monthly summaries, review and acknowledgment of feedback reports.

Results

A total of 15 evaluators and 10 supervisors attended the pre-study workshop training delivered by the PI with support of the QA center staff. The QA center was based in Kampala, Uganda and composed of 3 experienced (eg. > 2 years of experience in ND tetsing) ND assessors with Bachelors degrees in Nursing and Psychology. Evaluators and supervisors received a refresher training every year to review testing practices and study procedures. Each testing center reviewed the SOP every 3–4 months.

Staff at the QA center supervised a total of 15 evaluators from study sites located in 4 different countries (Zimbabwe, South Africa (3), Malawi, and Uganda) and 9 languages. From 2014 to 2016, the QA center supervised evaluator's performance on a monthly basis by evaluating 5 min videos with the specific rubric and generating the feedback report.

Overall video compliance was high; across sites 85% of expected monthly videos were sent to the QA center. The Lilongwe and Kampala sites had the lowest compliance rate (64%). The staff at the QA center reviewed an average of 10 videos per evaluator (range 1–26). The mean total score at the beginning of the evaluation (e.g., evaluator's first video reviewed by the QA center) was 161 (range 118–180), while the mean total score over the 10 months duration of QA center supervision was 178 (range 175–180). At the last video scoring conducted, evaluators on average had a mean total score of 165 (range 140–180). Rubric scores were considered sufficient (eg. Mean score >70% of maximum score) and none of the evaluators were terminated. Evaluators and supervisors at each site were required to acknowledged feedback reports.

Test completion rates were high across all sites; 99% of participants (n=611) evaluated at baseline completing the KABC-II. There were few instances of invalid tests and few barriers to test completion (e.g. disruptive behaviors in the child such as crying or refusing to follow instructions); at each assessment time point, 97% or more of the KABC-II across sites was considered to be valid by the supervisor. Intra-class correlation measures (ICC) indicating the consistency of KABC-II scores within participants during 12 months of study evaluation were high, ranging from 0.5 to 0.71.

Discussion

This paper describes the QualiND model; an approach to standardized ND testing using the same assessment protocol in six Sub-Sahara Africa study sites in four (Malawi, South Africa, Uganda, and Zimbabwe) countries with children and their caregivers speaking 9 different languages. Given the high rate of valid and completed tests as well as video submissions across sites supporting this, the QualiND model was a useful platform for remote supervision to non-specialist and minimally trained research staff.

The validity of adapting “Western-based” tests to the African context has been questioned, recommending instead the development of new neurodevelopmental assessment tools that are more culturally-appropriate and use normative data specific to the target population. (Holding et al., 2016) Boivin and Giordani (2009, 2013) have challenged this view, proposing that foundational brain/behaviour functions are universal to human neurodevelopment, and therefore, can be assessed by neuropsychological assessments (even “Western” ones) that are well designed, carefully adapted, and appropriately applied. The QualiND model begins to address this merging the efforts of researchers and staff to develop a feasible, practical, and flexible scheme of supervision that was easy to implement across different contexts and settings for clinic-based research purposes. The QualiND model describes, to the best of our knowledge, the first such model for quality assurance of a neurocognitive test while addressing the current neurodevelopment evaluation gap in low-resource settings.

In the present study, the KABC-II was adapted to six different study sites in four different sub-Saharan African countries (South Africa, Zimbabwe, Malawi, & Uganda) in nine different local languages. Such an ambitious multi-site study presumably makes for research findings that are more rigorous in terms of greater external validity and more generalizable across different regions of sub-Sahara Africa, as compared to the typical published one-site pediatric HIV neurodevelopmental studies. However, such multi-site studies with more complex cognitive assessment tests might also give rise to problems in standardizing the testing procedures for these tests when adapting them to the local cultural and language contexts. Despite the manner in which the standardized administration of the tests could have been undermined by speaking the instructions in the local languages, the KABC-II maintained good factor structure, good construct validity, good reliability, and proved sensitive to the proximal neurocognitive effects of HIV disease in a consistent manner across all study sites. The psychometric evidence using KABC-II standardized scores from American norms for this is presented in detail in a recent publication pertaining to the present study by (Chernoff, 2018).

One important advantage of the KABC-II compared to other comprehensive tests of child cognitive ability is that it provides for a global index (Nonverbal Index or NVI) using only subtests where an understanding of spoken English on the part of the child is not required. The mental processing index (MPI) on the other hand, uses all age-appropriate core subtests in estimating the global performance of the child on the KABC-II, across all global cognitive ability domains. In the Chernoff et al. validation study, important predictors for children’s neuropsychological performance were statistically correlated with both the NVI

and the MPI global scores. The multiple regression statistical prediction for NVI and MPI was comparable for such factors as study site/language, HIV status, WHO-standardized stunting and wasting indicators, child disability screening score, and quality of the home environment – in some cases – almost identical in terms of regression coefficients. The only differences obtained between the NVI and MPI were with the predictors of WHO body mass index (BMI) standardized score, and whether the principal caregiver at the time of assessment was the biological mother; and in both cases MPI was the strong statistical outcome. Therefore, in the present analyses, we were confident in using the MPI as our principal global index of performance for the children in our P1104s study, despite the fact that by using more verbal subtests in its composite score, it might be more vulnerable to the polyglot features of our present study.

Another issue related to quality assurance for such complex tests as the KABC-II, is whether American norms should be used to standardize the subtest performance for age of the child at assessment. Some form of standardization is necessary for the individual subtest in order to compute performance scores for the global cognitive ability domains, as well as arrive at a composite performance score for the tests as a whole such as the NVI or MPI. Such norms for a test as the KABC-II are usually not available for a given cross-cultural study site in Africa, necessitating comparison or reference groups to compare to the clinical exposure or intervention groups in that study context. This was done for the present P1104s study whereby comparison groups of HIV-exposed/uninfected and unexposed/uninfected children were enrolled at each of our study sites, age-matched to the HIV-infected cohort at that site. However, such reference groups are usually statistically powered to be large enough for between-group comparisons, but not sufficiently large to provide for a normative reference for standardizing test performance by year of age, especially if there is a large range in ages in the clinical group sample (6 to 11 years of age in the present sample).

Despite the fact that we used American norms to standardize our scores for age, Chernoff et al. (2018) obtained consistent results for the factor structure of our KABC-II subtests, compared to a recent study of rural, HIV-unexposed isiZulu-speaking South African children from 7 to 11 years of age (Mitchell et al., 2017), and consistent to the global performance domains as validated for the KABC-II test validated the factor structure of the KABC-II in American children. Our present factor structure for the KABC-II was also consistent with that of study in Uganda (Bangirana, Musisi Seggane, et al., 2009), testing children with cerebral malaria, used the same subtests as in this study, but with raw scores adjusted statistically for age for the entire Ugandan sample, rather than using American norms. We believe that the robust nature of the factor structure for the KABC-II in our validation studies, even when administered within a multiplicity of sites in various local languages and when using American-based norms to standardize subtest performance for age, is due to the characteristics of the test itself. Each subtest of the KABC-II begins with sample items followed by at least a few scripted teaching items if the child does not provide the correct. This better ensures that the child understands the cognitive task, so that the score on the subtest better reflects the child's actual abilities, rather than a lack of understanding of what was expected for the task at hand.

The QualiND model had several strengths. First, its use enabled the application of a complex standardized test for ND assessment (i.e. KABC-II) by non-specialist research staff (Semrud-Clikeman et al., 2016) relying on video-based remote supervision. Second, the model not only delivered tiered supervision to staff but also allowed for continuous theoretical and practical training throughout the course of the study. In a separate report (Chernoff, 2018), we evaluated the feasibility, reliability and validity of cross-cultural assessment of ND in this sample of HIV affected children. Results demonstrate that KABC-II scores obtained in the P1104s study were consistent across all sites. Validity and reliability measures, as well as expected correlations with other variables (e.g., congruent validity), confirmed that the non-specialist staff involved in the study were capable of conducting ND assessments to the highest standards and the same degree of proficiency as specialist staff in other studies. Although there is considerable debate among the scientific community about the validity of ND assessments carried out by non-specialist research staff with minimal training (Holding et al., 2016), results obtained in the P1104s study across six research sites in Sub-Saharan Africa (Chernoff, 2018) challenge that notion and suggest that ND assessment is feasible when used in combination with a supervision template such as the QualiND model. To this end, we believe that the QualiND model could be used with other standardized tests of ND such as the Mullen Scales of Early Learning or the Bayley Scales of Infant and Toddler Development and extend the use of these measures in low-resource settings. Future studies should assess evaluate the transportability of the QualiND model in the ND assessment of younger (eg. less than 5 years old) children.

Third, during QualiND implementation, the continued interaction between testing centers, QA center, and remote supervision levels played a major role in improving evaluator's performance. Because the number of supervisors and their availability differed across sites (Table 1), the QA center was paramount in providing technical support and supervision through direct communication with evaluators and using standardized and regular evaluations. Centralizing the QA process for the KABC-II was critical in systematizing or QA method in a consistent and efficient manner through the duration of this multi-site study. In sum, our elaborate QA process, perhaps unique to the present study, resulted in study findings for the KABC-II, that were consistent, valid and reliable; even when adapting those tests to the local language context and even when using American-based norms to provide for standardized scores (Boivin et al., 2018; Chernoff, 2018).

The testing rubric specifically developed for this study and implemented as part of the evaluation in the QualiND model was useful at several levels. Staff at the QA center had a standardized metric to evaluate videos and send feedback to evaluators, who in turn received specific recommendations on test areas that needed correction or improvements as well as instructions on how to achieve this. Supervisors at each site used the scored rubric to appraise evaluator's performance individually and as a group. Finally, the rubric yields a global score that allowed tracking evaluators' progress on test proficiency. The rubric score increased progressively from the beginning of QA center evaluations up to the end of the study, suggesting an improvement in the identified testing skills and can be interpreted as an average learning curve. Interestingly, rubric scores decreased by the end of the study, probably due to evaluator's sense of mastery of the test and less use of the manual and easel materials. It must be noted the evaluation rubric and scores were developed *ad hoc* to follow

evaluators' performance and should be interpreted as such. Also of note is that 5/6 sites incorporated a new evaluator at some point during the study, possibly impacting rubric scores. Reasons for evaluator attrition varied across sites and included staff changes in projects and personal reasons. High staff turnover in Malawi and Uganda was one the major reason behind the low video compliance. These results reinforce the need for continuous training and supervision of standardized ND assessment, where evaluators are constantly reviewing good test practices and procedures.

KABC-II's item instructions were spoken to the children as needed in their local language. However, the KABC-II was sensitive enough in its design, and proved highly adaptable to the various cultural and language contexts as demonstrated by Chernoff and colleagues' (2018) report on the validity of scores obtained. This was probably because the initial ND assessment training was followed by periodic SOP revisions and continuous test-specific corrections made to evaluators. We believe this type of supervised iterative learning coupled with remote supervision resulted in valid and reliable assessment of ND across sites.

The most challenging piece of the QualiND model was building the QA center as it required a cadre of staff with background in the administration of the specific ND test. Although experienced staff in ND assessment is hard to come across in LMIC, the QualiND model as proposed here starts addressing this issue by building local capacity through continued training and supervision of non-specialist research staff.

Although the QualiND model relied on a cloud-based, file-hosting service (e.g., Dropbox) to provide remote supervision, data and experience from all sites indicated that this was feasible. Videos were in most part uploaded according to schedule and communication between testing centers and QA staff was reasonably maintained. A frequent challenge encountered at all study sites was internet availability. The QualiND model depends on reliable internet access to upload testing videos to a file-sharing platform and for teams to maintain communication. Slow and interrupted internet connections frequently resulted in frustration for staff, and in a few incomplete video submissions. In 2015, Blantyre, Malawi was severely affected by flooding that partially disrupted communication between the Malawian site and the QA center. However, despite this challenge the Malawi testing center team was able to upload video files to the file-sharing platform in a timely manner, with little to no disruption of the assessments, underscoring the minimal Internet requirements of the QualiND model.

With better medical treatment and supportive care for HIV/AIDS now available in low- and middle-income countries, more children now survive into adulthood. More attention is now being paid to the quality of life of these children who live with persisting cognitive deficits. In order for these children to achieve their full potential, specific ND assessments are required (Boivin et al., 2018) to implement the non-pharmacological interventions that have been called for, such as cognitive rehabilitation, speech and physical therapy and caregiver training (Boivin, Ruisenor-Escudero, & Familiar-Lopez, 2016) that might mitigate deficits.

There are a number of technical and conceptual limitations to consider. First, the QualiND model relies on Internet connection for communication between the testing centers and the

QA center that might not be equally available across LMIC. However, access to broad-band internet is growing rapidly in many LMIC, especially in the Sub-Saharan African region. Increased network availability and coverage in the coming years will most likely make this limitation purely anecdotal. Second, videotaping meant additional consent from the caregiver of the videotaped child and the need to safeguard videos sent to the assessment center, securing them both on site and at the assessment center to protect the child's identity. Nevertheless, protection of privacy was met through password-enhanced protocols that were strictly followed at each site, complying with ethical standards. Third, the QualiND model was designed for use among research staff sharing a common language (e.g., English) at the testing and QA centers, and its applicability needs to be tested in studies where different languages of the site staff are involved.

Despite limitations, the assessment center resulted in excellent reliability across sites over successive years (Chernoff, 2018), enhanced time to train replacement evaluators in the event of turnover at distant sites, and provided for a much more efficient means of quality assurance than the usual model of periodic site visits and testing observation by a "supervisor", allowing for evaluators themselves to assess their own performance as a group, in the absence of a visiting site supervisor.

Conclusion

To the best of our knowledge, the QualiND model represents the first attempt at a concise, yet comprehensive, multi-tiered supervision model to provide quality assurance in the implementation of a standardized ND test by non-specialist research staff. Use of the QualiND model across six study sites in Sub-Saharan Africa proved to be an extremely useful framework for the implementation of a standardized test of neurodevelopment as part of a large-scale research study. We believe that given that ND test results were valid, reliable and consistent (Chernoff, 2018), the QualiND model may be of use to researchers and organizations involved in measuring early child development using standardized tests in, multi-site, low and middle-income settings.

Acknowledgments:

The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Drs. E. Pim Brouwers (NIH/NIMH) and Sonia Lee (NIH/NICHD) served as protocol advisors the research leadership team for P1104s for their respective NIH institutes. We gratefully acknowledge their expertise and counsel during the study. The authors also acknowledge the protocol administrative support provided by J.L. Ariansen from FHI360, as well as the support provided by Dr. Elizabeth Petzold throughout the protocol approval, finalization and study initiation process.

Funding: Overall support for the International Maternal Pediatric Adolescent AIDS Clinical Trials (IMPAACT) Network was provided by the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) under Award Numbers UM1AI068632 (IMPAACT LOC), UM1AI068616 (IMPAACT SDMC) and UM1AI106716 (IMPAACT LC), with co-funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and the National Institute of Mental Health (NIMH).

References

- Bangirana P, Giordani B, John CC, Page C, Opoka RO, & Boivin MJ (2009). Immediate neuropsychological and behavioral benefits of computerized cognitive rehabilitation in Ugandan

pediatric cerebral malaria survivors. *Journal of developmental and behavioral pediatrics: JDBP*, 30(4), 310. [PubMed: 19668094]

Bangirana P, Giordani B, John CC, Page C, Opoka RO, & Boivin MJ (2009). Immediate neuropsychological and behavioral benefits of computerized cognitive rehabilitation in Ugandan pediatric cerebral malaria survivors. *J Dev Behav Pediatr*, 30(4), 310–318. doi:10.1097/DBP.0b013e3181b0f01b [PubMed: 19668094]

Bangirana P, Seggane M, Allebeck P, Giordani B, John CC, Opoka OR, . . . Boivin MJ (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. [PubMed: 20589149]

Black MM, Baqui AH, Zaman K, McNary SW, Le K, Arifeen SE, . . . Black RE (2007). Depressive symptoms among rural Bangladeshi mothers: implications for infant development. *Journal of child psychology and psychiatry, and allied disciplines*, 48(8), 764–772. doi:JCPP1752 [pii]

Black MM, Walker SP, Fernald LC, Andersen CT, DiGirolamo AM, Lu C, . . . Shiffman J (2017). Early childhood development coming of age: science through the life course. *The Lancet*, 389(10064), 77–90.

Blanchette N, Lou M, Fernandes-Penney A, King S, & Read S (2001). Cognitive and motor development in children with vertically transmitted HIV infection. *Brain and cognition*, 46(1), 50–53. [PubMed: 11527362]

Boivin MJ, Bangirana P, Nakasujja N, Page CF, Shohet C, Givon D, . . . Klein PS (2013). A year-long caregiver training program improves cognition in preschool Ugandan children with human immunodeficiency virus. *J Pediatr*, 163(5), 1409–1416. e1405. [PubMed: 23958115]

Boivin MJ, Barlow-Mosha L, Chernoff MC, Laughton B, Zimmer B, Joyce C, . . . Palumbo PE (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 [PubMed: 29112069]

Boivin MJ, Green SD, Davies AG, Giordani B, Mokili JK, & Cutting WA (1995). A preliminary evaluation of the cognitive and motor effects on pediatric HIV infection in Zairian children. *Health Psychology*, 14(1), 13. [PubMed: 7737068]

Boivin MJ, Nakasujja N, Sikorskii A, Opoka RO, & Giordani B (2016). A Randomized Controlled Trial to Evaluate if Computerized Cognitive Rehabilitation Improves Neurocognition in Ugandan Children with HIV. *AIDS Res Hum Retroviruses*, 32(8), 743–755. doi:10.1089/AID.2016.0026 [PubMed: 27045714]

Boivin MJ, Ruisenor-Escudero H, & Familiar-Lopez I (2016). CNS Impact of Perinatal HIV Infection and Early Treatment: the Need for Behavioral Rehabilitative Interventions Along with Medical Treatment and Care. *Curr HIV/AIDS Rep*, 13(6), 318–327. doi:10.1007/s11904-016-0342-8 [PubMed: 27783207]

Chernoff ML,B; Ratswana M; Familiar I; Fairlie L; Vhembo T; Kamthunzi P; Kabugho E; Joyce C; Zimmer B; Ariansen JL; Jean-Philippe P; Boivin M (2018). Validity of Neuropsychological Testing in Young African Children Affected by HIV. *Journal of Pediatric Infectious Diseases*(01), 1–88.

Donald KA, Walker KG, Kilborn T, Carrara H, Langerak NG, Eley B, & Wilmschurst JM (2015). HIV Encephalopathy: pediatric case series description and insights from the clinic coalface. *AIDS research and therapy*, 12(1), 2. [PubMed: 25598835]

Hack M, & Costello DW (2007). Early childhood development: the global challenge.

Health, U. M. o. (2010). National Implementation guidelines for HIV Counselling and Testing in Uganda. Retrieved from Kampala:

Holding P, Anum A, van de Vijver FJ, Vokhiwa M, Bugase N, Hossen T, . . . Bangre O (2016). Can we measure cognitive constructs consistently within and across cultures? Evidence from a test battery in Bangladesh, Ghana, and Tanzania. *Applied Neuropsychology: Child*, 1–13.

Kaufman AS, & Kaufman NL (1983). *Kaufman assessment battery for children*: Wiley Online Library.

Kerr SJ, Puthanakit T, Vibol U, Aupibul L, Vonthanak S, Kosalaraksa P, . . . Luesomboon W (2014). Neurodevelopmental outcomes in HIV-exposed-uninfected children versus those not exposed to HIV. *AIDS Care*, 26(11), 1327–1335. [PubMed: 24878112]

Krieger N (1994). Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med*, 39(7), 887–903. [PubMed: 7992123]

- Laughton B, Cornell M, Boivin M, & Van Rie A (2013). Neurodevelopment in perinatally HIV-infected children: a concern for adolescence. *J Int AIDS Soc*, 16, 18603. doi:10.7448/ias.16.1.18603 [PubMed: 23782482]
- Mitchell JM, Tomlinson M, Bland RM, Houle B, Stein A, & RoCHAT TJ (2017). Confirmatory factor analysis of the Kaufman assessment battery in a sample of primary school-aged children in rural South Africa. *South African Journal of Psychology*, 0081246317741822.
- Nozyce M, Hittelman J, Muenz L, Durako SJ, Fischer ML, & Willoughby A (1994). Effect of perinatally acquired human immunodeficiency virus infection on neurodevelopment in children during the first two years of life. *Pediatrics*, 94(6 Pt 1), 883–891. [PubMed: 7971006]
- Organization, W. H. (2007). Task shifting: rational redistribution of tasks among health workforce teams: global recommendations and guidelines.
- Parker R, & Aggleton P (2003). HIV and AIDS-related stigma and discrimination: a conceptual framework and implications for action. *Soc Sci Med*, 57(1), 13–24. [PubMed: 12753813]
- Semrud-Clikeman M, Romero RAA, Prado EL, Shapiro EG, Bangirana P, & John CC (2016). Selecting measures for the neurodevelopmental assessment of children in low-and middle-income countries. *Child Neuropsychology*, 1–42.
- United Nations Program on HIV and AIDS. (2014). The GAP Program. Retrieved from Geneva:
- Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, Pollitt E, . . . Group, I. C. D. S. (2007). Child development: risk factors for adverse outcomes in developing countries. *The Lancet*, 369(9556), 145–157.
- World Health Organization. (2017). Global Health Observatory (GHO) data: Psychiatrist and nurses (100,000 population). Retrieved from http://www.who.int/gho/mental_health/human_resources/psychiatrists_nurses/en/

Figure 1.
The Quali ND model for standardized neurodevelopmental test implementation

Table 1.
Child demographics and growth characteristics at P1104s study entry by site.

Characteristic	Site							P-value
	Joburg (n=69)	Soweto (n=100)	Tygerberg (n=140)	Malawi (n=81)	Uganda (n=89)	Zimbabwe (n=132)	Total (n=611)	
HIV+	27 (39%)	42 (42%)	56 (40%)	33 (41%)	37 (42%)	51 (39%)	246 (50%)	1.000 ^(a)
HEU	21 (30%)	29 (29%)	42 (30%)	24 (30%)	26 (29%)	41 (31%)	183 (30%)	
HUU	21 (30%)	29 (29%)	42 (30%)	24 (30%)	26 (29%)	40 (30%)	182 (30%)	
Males	32 (46%)	50 (50%)	65 (46%)	43 (53%)	49 (55%)	51 (39%)	290 (47%)	0.182 ^(a)
Age (years) mean (S.D.)	8.0 (1.6)	7.9 (1.3)	7.4 (1.3)	6.6 (0.9)	6.7 (1.2)	6.8 (1.4)	7.2 (1.4)	<0.001 ^(b)
In school	58 (84%)	84 (86%)	94 (71%)	55 (77%)	41 (46%)	74 (57%)	406 (69%)	
WHO weight z-score, mean (S.D.)	-0.31 (1.11)	-0.19 (1.09)	-0.18 (1.06)	-1.14 (0.76)	-0.52 (0.83)	-0.48 (0.87)	-0.44 (1.01)	<0.001 ^(b)
WHO height z-score, mean (S.D.)	-0.64 (1.04)	-0.57 (1.05)	-0.38 (0.95)	-1.55 (0.98)	-0.83 (0.96)	-0.38 (1.00)	-0.66 (1.06)	<0.001 ^(b)
Residential zone	7 (10%)	3 (3%)	1 (1%)	34 (42%)	25 (28%)	39 (30%)	109 (18%)	<0.001 ^(a)
Rural	30 (43%)	10 (10%)	138 (99%)	27 (33%)	60 (67%)	3 (2%)	268 (44%)	
Peri-urban Urban	32 (46%)	87 (87%)	1 (1%)	20 (25%)	4 (4%)	90 (68%)	234 (38%)	

^(a) Chi-Square Test

^(b) Analysis of Variance

HEU, HIV exposed-uninfected

HUU, HIV unexposed-uninfected

Table 2.
Descriptive characteristics of P1104s study sites and evaluators

	South Africa			Malawi	Uganda	Zimbabwe
	Johannesburg	Soweto	Tygerberg	Lilongwe	Kampala	Harare
Site Characteristics						
Number of evaluators	01	01	05	03	03	02
Total number of tests performed	196	300	403	241	266	392
Total number of Videos sent	15	21	92	40	38	62
Number of supervisors on site	1	1	2	3	1	2
IT support on site	Yes	Yes	Yes	Yes	Yes	Yes
Evaluator's Characteristics						
Languages used on site	English, Zulu, Setswana, Sesotho	English, Zulu, Setswana, Sesotho	English, Afrikaans, Xhosa	English, Chichewa	English, Luganda	Shona, English
Tester's expertise (college, technical or educational degree of each tester)	Nursing Diploma	Bachelor's Psychology	Senior certificate (Grade 11) Senior Certificate (Grade 12) Bachelor's Psychology Senior certificate (Grade 12) Bachelors in Psychology	3 Diploma in nursing	Bachelor's Education Bachelors Community Psychology Bachelor's Community Psychology	2 Registered Nurse/Midwife