Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately?

Saraswathy Sabanathan,1,2 Bridget Wills,1,2 Melissa Gladstone3

ABSTRACT
Global emphasis has shifted beyond reducing child survival rates to improving health and developmental trajectories in childhood. Optimum early childhood experience is believed to allow children to benefit fully from educational opportunities resulting in improved human capital. Investment in early childhood initiatives in low-income and middle-income countries (LMICs) is increasing. These initiatives use early childhood developmental assessment tools (CDATs) as outcome measures. CDATs are also key measures in the evaluation of programmatic health initiatives in LMICs, influencing public health policy. Interpretation of CDAT outcomes requires understanding of their structure and psychometric properties. This article reviews the structure and main methods of CDAT development with specific considerations when applied in LMICs.

INTRODUCTION
It is estimated that up to 200 million children fail to reach their developmental potential in low-income and middle-income countries (LMICs). This statistic underpins recent calls for urgent global action to improve early child development in the first 5 years of life. Poverty, malnutrition, recurrent and/or chronic infectious diseases plus inadequate cognitive stimulation, can all have lifelong consequences on development. Economic modelling studies demonstrate how early interventions and investment in disadvantaged children are effective in improving child development and cost-effective in improving the ‘human capital’ potential of individuals. Investment in early child development initiatives in LMICs is burgeoning and with this, prioritisation of global health research into interventions that have maximum impact on child development. Specific child development assessment tools (CDATs) are recommended by major research funding bodies as appropriate tools to evaluate child development interventions, despite limited evidence of the nature of the CDATs’ relationship with quality of life measures and later potential human capital. Since no CDAT is suitable for all populations, a variety of tools have been developed worldwide. Typically, knowledge of the structure, psychometric properties, applicability and limitations of CDAT is restricted to child development specialists. In practice, health workers from various backgrounds often administer and interpret these tests. This article describes the structure, main methods of CDAT development and their application, focusing on children less than 5 years of age in LMICs.

WHAT IS A CHILD DEVELOPMENTAL ASSESSMENT TOOL, WHAT DO THEY MEASURE AND HOW ARE THEY ADMINISTERED?
Early childhood is a period of rapid physical growth during which an individual acquires a complex set of skills and functional competencies that should facilitate achievement of their potential in life. Weight and height charts were developed on the premise that children from similar ethnic backgrounds follow similar growth trajectories. In a comparable fashion, acquisition of skills during early childhood is also expected to follow a set trajectory. For growth and development we know that it is best to intervene promptly in order to maintain a healthy trajectory.

Each domain of development is assessed by the child’s ability to carry out a series of increasingly complex activities that are thought to reflect the expected level of development at a particular age. The developmental domains evaluated by CDAT do not function as discrete entities but influence each other—for example, to successfully complete a jigsaw puzzle requires a combination of cognitive and fine motor skills, and sufficient understanding of language to process instructions.

Methods to assess child development may include: (A) direct assessment of standardised activities by a trained assessor in a clinical environment, (B) verbal reporting/completion of a questionnaire about the child’s abilities by parents or teachers and (C) unstructured observation by a trained assessor in an environment familiar to the child (eg, home/school). Direct testing in an unfamiliar environment by an unfamiliar adult may restrict a child’s engagement and participation in the assessment. Parental reporting, on the other hand, may be affected by recall bias. Unstructured observation can be difficult to reproduce and interpret. An ideal CDAT would include aspects of all three methods, but with limited financial resources, often one method is chosen.

TYPES OF ASSESSMENT TOOLS—SCREENING VERSUS FORMAL ASSESSMENT?
Screening tools are administered quickly, using a limited sample of items representing a domain and rely on predetermined cut-off points. Screening tools are designed to identify children who may...
have impairment and require a comprehensive assessment. However, screening tools have poor utility in assessing subtle delays that may have a significant impact on subsequent development. Screening tools are beneficial when used within the context of a development surveillance programme, where there are appropriate norms and known applicability in specific subgroups.

Most formal tools require specialised specific training of assessors. They allow administration of items above or below the age bracket of the child thus allowing the child’s strengths and weaknesses to be categorised more accurately. Formal assessments can be administered repeatedly to evaluate changes over time and/or following an intervention.

**IS THE TOOL MEASURING WHAT IT IS SUPPOSED TO—VALIDITY AND RELIABILITY?**

Developmental domains are a theoretical concept, typically referred to as ‘constructs’ in the psychometric literature. These cannot be directly measured but are inferred through the child’s performance on a number of observed variables (test items). Reliability is the variability of scores obtained by an individual if repeatedly given the same test. There are widely cited levels of acceptable reliability for testing. However, the result should be interpreted in context. For example, retest reliability can be influenced by developmental maturation over the time interval, or training by the caregiver following mistakes seen in the first assessment. Validity is the accuracy of the score representing the construct of interest. In addition, screening tools are evaluated for sensitivity, specificity, positive predictive value and negative predictive value. Reliability and validity of CDAT can be evaluated through several methods, summarised in table 2.

**HOW TO DETERMINE A CHILD’S PERFORMANCE ON THE CDAT—THE REFERENCE POPULATION**

All assessment tools have a threshold that a child must achieve for the assessor to be confident that there are no current developmental concerns. There are two main ways of setting this threshold, norm referencing and criterion referencing. A norm-referenced test is usually administered in a standardised manner and the individual child’s scores in each domain are compared with scores from a large representative sample of children of the same age and sex (normative data). Standard scores allow comparison between scales evaluating the same domain and monitoring of individuals at different ages. By contrast, criterion-referenced tests assess whether a child has acquired a particular skill by a certain age, according to a specific curriculum. Administration may not be standardised, since individual children sometimes require additional instructions or physical aids to complete a particular task but are then able to complete the specified tests. Criterion-referenced tests are often used in screening procedures, for example assessment of reading skills at school entry, or evaluation of an intervention such as a programme of physical therapy. In some cases criterion-based tests have been standardised on large representative populations.

There are concerns that the rapidly changing nature of society means that normative data become outdated very quickly. To counter this, inclusion of a control group is now considered crucial for research studies, even when the normative population is contemporaneous and comes from the same linguistic and cultural background as the study population.

**WHEN TO ASSESS DEVELOPMENT, WHAT DOES IT MEAN FOR THE FUTURE—PREDICTIVE VALIDITY?**

CDAT will only give a snapshot of the child at that time point. Traits that have not yet evolved in early childhood clearly cannot be assessed until such time as they might reasonably be expected to be present, and yet impairment within these characteristics may impact subsequent development significantly.

After an assessment, it is important to know the predictive validity of the CDAT score, whether normal or abnormal. Correlations between early childhood performance and school abilities in normal children are variable. Children with moderate developmental delay, despite the same initial difficulties, have great variability in developmental trajectories, with some children catching up while others exhibit evolving difficulties with age. Aspects of early development may be predictive of later academic ability. Preschool English language acquisition of vocabulary, grammar and descriptive phrasing predicted school reading ability in a prospective longitudinal US study.

Research in normal infants suggests that acquisition of efficient
inhibitory control and planning aspects of cognitive executive function at 4 years of age are associated with improved acquisition of school mathematical and literacy abilities at 6–7 years of age.\(^{21,22}\)

There is limited research examining the predictive nature between CDAT and longer-term outcomes in LMICs. Three longitudinal nutritional intervention studies collected CDAT scores in infants from Indonesia and Guatemala.\(^{23}\) Scores on 2-monthly assessments correlated poorly in infants less than 24 months, but score correlations increased on assessments carried out between 24 months and 30 months of age. The scores at age 20 months or younger had no predictive value for verbal reasoning and arithmetic scores at school age. Overall, predictive power in comparative US research was similar.\(^{24}\)

There is still a lack of evidence as to whether the interpretation of group differences in scores due to an intervention or exposure in infancy should be interpreted as predicting anything in terms of later intelligence.\(^{25}\)

### APPROACHES AND CHALLENGES OF CDAT DEVELOPMENT IN LMICs

Currently, only a few CDATs are available in LMICs and most are used in research settings. These CDATs typically follow one of the following formats: (A) a standard western CDAT with no adaptations; (B) a western CDAT translated (linguistic equivalence) and/or adapted for the local cultural environment (cultural equivalence); (C) an amalgamation of a number of translated and/or adapted items from several different western CDATs; or (D) a locally developed, culturally specific CDAT consisting of original items designed to be relevant to the population of interest.\(^{25}\) The format of CDAT adaptation depends on the aim of the application. Rie et al evaluated the impact of HIV on neurodevelopment of children in Kinshasa, Congo. The researchers deliberately used a direct translation of a western CDAT without cultural adaptation to compare between groups in the same setting and then compare between past and future studies in other settings.\(^{26}\) This method does not ensure that the tool is measuring the same underlying ability in all settings. Holding et al\(^{27}\) aimed to facilitate comparisons between different cultural groups in differing countries and needed to measure the same underlying ability in all settings. This was done by cultural modification of a western tool. The Kilifi Developmental Checklist (KDC) was also based on items from a range of CDAT.\(^{28}\) The items were chosen based on ease of observing item success, how well the item could differentiate within the population of interest and if the item could be readily described in the local languages. Additional considerations pertaining to LMICs include the prohibitive cost of the license fee to translate or adapt a western CDAT, the level of skill of the assessor and location of assessment. Fernandes et al\(^{29}\) developed a screening tool for use in high-income countries and LMICs. The team faced the challenge of developing a culture-free tool that was low in cost, could be administered after limited level of assessor training and completed in 30 min. To achieve this, an advisory panel listed project-specific criteria that the CDAT needed to fulfill.

Once items are selected, the first step is pilot studies on a small number of children. These aim to discover the range of responses elicited from the tested population and identify problems with unfamiliar stimulus materials. Prado et al\(^{37}\) adapted a cognitive test for Indonesian children, who had difficulty identifying a picture of a bunny, so replaced it with a picture of a

---

**Table 2: Basic psychometric properties used to evaluate child development assessment tools (CDATs)**

<table>
<thead>
<tr>
<th>Relevance/Importance</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internal consistency</td>
<td>Evaluates the similarity of test items assessed in one domain. One measure is split-half reliability, which compares the scores on two halves of a test in a single domain.</td>
</tr>
<tr>
<td>Interobserver</td>
<td>Evaluates variability between different assessors on the same subject.</td>
</tr>
<tr>
<td>Intraobserver</td>
<td>Evaluates variability within a single assessor on a single subject.</td>
</tr>
<tr>
<td>Validity</td>
<td></td>
</tr>
<tr>
<td>Test-retest</td>
<td>Evaluates variability within the subject (influenced by random factors such as familiarity with items and mood).</td>
</tr>
<tr>
<td>Content</td>
<td>Experts in the field make consensus agreement on whether the individual item and the range of items adequately sample and represent the domain of interest.</td>
</tr>
<tr>
<td>Criterion</td>
<td>Ideally assessed by comparison to an established ‘gold standard’ test assessing the same construct.</td>
</tr>
<tr>
<td>Discriminant/convergent</td>
<td>Evaluates expected positive and negative correlations between scores in different domains or between different tests of the same or differing underlying construct.</td>
</tr>
<tr>
<td>Construct</td>
<td>Statistical evaluation to see whether values of observed data fit a theoretical model of the constructs (confirmatory) or to explore a possible model of the ‘underlying traits’ being measured.</td>
</tr>
</tbody>
</table>

Difficult to interpret in early childhood when changes in development occur over a short time. Usually the repeat assessment should be carried out within 2 weeks of the first test.

Subjective measure that cannot be used in isolation to evaluate validity.

Scores from two independent tests (eg, one using report method the other a direct test) of one domain should correlate where neither test is considered a ‘gold-standard’. To ensure the test is not overlapping with constructs not of interest, the scores evaluating different constructs should poorly correlate, for example, test scores on ‘fine motor’ should correlate poorly with ‘social emotional’.

Large numbers of assessments are required to evaluate this.
Table 3  Examples of CDATs (children <5 years) used in LMIC countries sourced from published literature

<table>
<thead>
<tr>
<th>Tool and country</th>
<th>Type of tool and item selection (translation±adaptation)</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kilifi Developmental Checklist (KDC)</td>
<td>Mixture of items from government local screening test, Western tools and LMIC tools developed elsewhere</td>
<td>Excellent*</td>
<td>Items range from poor to good correlation with parental report†</td>
<td>Aimed to be used by low-skilled workers. Identified that secondary school level education was minimal for an assessor.</td>
</tr>
<tr>
<td>Kenya</td>
<td>Direct and Report</td>
<td>Excellent†</td>
<td></td>
<td>Done</td>
</tr>
<tr>
<td></td>
<td>11–109 months</td>
<td>Excellent†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kilifi Developmental Inventory (KDI)</td>
<td>Assessment based on KDC with additional items from western tools</td>
<td>Excellent*</td>
<td>Good correlation with maternal reports†</td>
<td>Designed to monitor changes over time. Involved parental focus groups so procedures acceptable to community.</td>
</tr>
<tr>
<td>Kenya</td>
<td>Direct observation</td>
<td>Excellent†</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>6–35 months</td>
<td>Excellent†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Developmental Milestone Checklist</td>
<td>Screening items from Western tools Report</td>
<td>Good*</td>
<td>Very good correlation with KDI†</td>
<td>Structured interview. Future plans to develop cut-offs using clinical samples. Can be administered by those with little training.</td>
</tr>
<tr>
<td>Kenya</td>
<td>3–24 months</td>
<td>Good†</td>
<td></td>
<td>Stunted children had lower scores. Done</td>
</tr>
<tr>
<td>Guide to monitoring child development</td>
<td>Screening Core ideas were adapted from three seminal western models of child development. Report</td>
<td>Excellent*</td>
<td>Excellent agreement with comprehensive evaluation§ Sensitivity 0.88; 95% CI 0.69 to 0.96 Specificity 0.93; 95% CI 0.83 to 0.97</td>
<td>Brief, open-ended precoded interview with parent. No specific cognitive domain questions. May not be appropriate for monitoring ‘at-risk’ populations.</td>
</tr>
<tr>
<td>Turkey</td>
<td>0–24 months</td>
<td>Good§</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapid neurodevelopmental assessment</td>
<td>Screening Mixture of neurological examination and neurodevelopment from western tools Direct</td>
<td>Excellent§</td>
<td>Significance differences in scores in children with neurodevelopment impairment Detected difference in scores between urban rural children in older age groups</td>
<td>Used in a small high-risk population and identified infants requiring intervention for retinopathy of prematurity. RNA identified additional problems such as impairments in vision and hearing.</td>
</tr>
<tr>
<td>Bangladesh 0–2 years</td>
<td>0–24 months</td>
<td>Sensitivity 80–90% Specificity 60–78%</td>
<td>Comparison with ‘gold-standard’ tests already developed for Bangladesh.</td>
<td></td>
</tr>
<tr>
<td>Rapid neurodevelopmental assessment</td>
<td>Screening Mixture of neurological exam and development assessment from western tools Direct</td>
<td>Excellent§</td>
<td>Stunted children had lower scores.</td>
<td></td>
</tr>
<tr>
<td>Bangladesh 2–5 years</td>
<td>24–60 months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malawi Developmental Assessment Tool.</td>
<td>Screening Western developmental tool and new items locally devised Direct and report</td>
<td>Excellent49</td>
<td></td>
<td>Good strategy to devise socioemotional items.</td>
</tr>
<tr>
<td>Malawi</td>
<td>0–72 months</td>
<td>60–80% of items at least fair§</td>
<td>Sensitivity 97% Specificity 82%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>99–100% scoring at least fair§</td>
<td>Sensitive to malnutrition except social developmental domain</td>
<td>50 51</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Tool and country</th>
<th>Type of tool and item selection (translation±adaptation)</th>
<th>Reliability</th>
<th>Validity</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTERGROWTH-21st Neurodevelopment Assessment²⁹</td>
<td>Screening Items from selection of Western and LMIC tools</td>
<td></td>
<td></td>
<td>Project specific criteria to create a tool for use in children from middle-class and upper class families across low-income, middle-income and high-income settings, carried out by non-specialists. Also evaluates vision, hearing and sleep-wake cycle.</td>
</tr>
<tr>
<td></td>
<td>Direct and report</td>
<td>Good</td>
<td></td>
<td>Clear hypothesis driven tool creation. May not be suitable as a general developmental assessment.</td>
</tr>
<tr>
<td>Systematic approach to assess nutritional influences. Indonesia¹⁷</td>
<td>Assessment Items from selection of Western tools</td>
<td></td>
<td></td>
<td>Expected relationship with maternal depression and maternal education</td>
</tr>
<tr>
<td></td>
<td>Direct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>24 months</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Interobserver Discriminant Test-retest</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>22–55 months</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Interpretation of statistical tests are listed below.¹⁶ ³² These need to be interpreted in context of application and population.¹⁷ ³³

<table>
<thead>
<tr>
<th>Levels of ( \kappa ) Levels§, ( R_I ), intraclass† correlation and ( \alpha ) coefficients*.</th>
<th>Level of proposed agreement¶ (%)</th>
<th>Levels of clinical or practical significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;0.4</td>
<td>&lt;70%</td>
<td>Poor</td>
</tr>
<tr>
<td>0.4–0.59</td>
<td>70–79</td>
<td>Fair</td>
</tr>
<tr>
<td>0.60–0.74</td>
<td>80–89</td>
<td>Good</td>
</tr>
<tr>
<td>0.75–1.00</td>
<td>90–100</td>
<td>Excellent</td>
</tr>
</tbody>
</table>

CDAT, child development assessment tool; LMICs, low-income and middle-income countries.
Global child health

WHAT HAVE WE LEARNT FROM CDATs IN LMICS: FACTORS THAT INFLUENCE EARLY CHILD DEVELOPMENT IN LMICS

Debate on the influence of nurture versus nature continues, although an increasing body of opinion now considers the early perinatal environment to be as important in determining cognitive ability as an individual’s genetic background. Intrauterine environment, nutritional deficiency in infancy, infectious diseases and poverty, all shape the developmental trajectory of a child in a LMIC. What we know about these threats to development come from research using CDAT. Walker et al13 reviewed the risk and protective factors of early child development that can be modified by interventions in children under the age of 5 years in LMICs, including iron deficiency, malaria and inadequate cognitive stimulation. These three factors illustrate challenges of CDAT interpretation. Three prospective iron supplementation trials unexpectedly found no change in cognitive development measured by an adapted western CDAT.35 It was plausible that the CDAT was not sensitive enough to the effects of supplementation of a nutritional deficiency, resulting in a shift to a more hypothesis-driven, focused testing of specific areas of development.35 This shift to ensure the CDAT can evaluate hypothesised outcomes based on biological mechanisms was incorporated into developing an appropriate CDAT for severe malaria.36 A team in Kilifi adapted a western tool for school-age children and then successfully evaluated its construct validity and sensitivity for discriminating cognitive deficits following severe malaria (high-risk), mild malaria (medium risk) and no previous admission with malaria (low-risk).27 There were significant differences in scores between unschooled high-risk children and unschooled low-risk children. There was no significant difference in scores between high-risk and low-risk malaria children who were at school, highlighting the plausible benefit of formalised education in protecting from neurocognitive sequelae in LMIC settings. This finding was put into context by the authors, who reported that the group attending school was not necessarily representative of severe children, since school attendance is limited by economic restraints and parental perception of whether their child would benefit from formal schooling. A child perceived to have cognitive deficits by their parents might not be sent to school. In a setting with a multitude of biological threats to early child development, the addition of poverty will limit the beneficial opportunities of a child’s environmental influence on their development. However, there is variability in parental attitudes and behaviours within one socioeconomic class and also variability within the same socioeconomic class between cultures. A South African Home Screening Questionnaire, found mental development was primarily accounted for by the mother’s ability to structure the child’s environment for learning.16 This was through the provision of play materials and maternal involvement in the child’s activities, which were believed to promote development. The authors concluded that the socioeconomic variables evaluated did not solely determine maternal behaviour. Supporting this premise are studies suggesting maternal stimulation programmes within early childhood community-based stimulation interventions that can have long-lasting benefits on cognitive, socioemotional wellbeing despite adverse socioeconomic environments.77 CDAT research will continue to inform the agenda for early intervention, before lifelong trajectories leading to large health and economic inequalities become fixed.

CONCLUSIONS AND FUTURE DIRECTIONS

Childhood developmental outcomes are firmly on the current global health agenda particularly now that it is time to set the new millennium developmental goals. However the relationship between future employability, quality of life measures and performance on a CDAT before the age of 5 years, is not known. Longitudinal research is needed to evaluate these relationships. In LMICs where early intervention support is limited, robust CDAT construction within culturally acceptable early child development programmes will help explore the relationship with later outcomes and promote the factors which protect against the neurocognitive detrimental effects of adversities faced by children in LMICs.38

QUICK CHECKLIST FOR APPRAISING A CDAT

1. Does the CDAT adequately measure all aspects of the domain(s) theoretically affected by a risk factor or intervention?
2. Has the CDAT been shown to be reliable and valid in the population of interest?
3. Is the CDAT sensitive enough in the setting required to identify the changes expected for the risk factor or intervention?
4. Are the number of evaluations and duration of follow-up suitable for evaluating the outcome of interest?
5. In research studies, is there a suitable control group and have potential influences been considered?

Contributors SS conceived the article and wrote the final manuscript and agree to be accountable for all aspects of the work.

Funding The Wellcome Trust.

Competing interests None.

Provenance and peer review Commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See: http://creativecommons.org/licenses/by/4.0/
REFERENCES


Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately?
Saraswathy Sabanathan, Bridget Wills and Melissa Gladstone

Arch Dis Child 2015 100: 482-488 originally published online March 30, 2015
doi: 10.1136/archdischild-2014-308114

Updated information and services can be found at:
http://adc.bmj.com/content/100/5/482

These include:

References
This article cites 42 articles, 12 of which you can access for free at:
http://adc.bmj.com/content/100/5/482#BIBL

Open Access
This is an Open Access article distributed in accordance with the terms of the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited. See:
http://creativecommons.org/licenses/by/4.0/

Email alerting service
Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Topic Collections
Articles on similar topics can be found in the following collections
ADC Global child health (48)
Open access (206)
Developmental paediatrics (128)

Notes

To request permissions go to:
http://group.bmj.com/group/rights-licensing/permissions

To order reprints go to:
http://journals.bmj.com/cgi/reprintform

To subscribe to BMJ go to:
http://group.bmj.com/subscribe/