Measuring early childhood development at a global scale: Evidence from the Caregiver-Reported Early Development Instruments

Dana Charles McCoy a,⁎, Marcus Waldman a, CREDI Field Team 1, Günther Fink b

a Harvard Graduate School of Education, 14 Appian Way, Cambridge, MA 02138, USA
b University of Basel and Swiss Tropical and Public Health Institute, Socinstrasse 57, CH-4051 Basel, Switzerland

A R T I C L E   I N F O

Article history:
Received 25 August 2017
Received in revised form 24 April 2018
Accepted 17 May 2018

Keywords:
Early childhood development
Measurement
Low- and middle-income countries

A B S T R A C T

Despite global interest in supporting and monitoring early childhood development (ECD), few valid and reliable tools exist for capturing ECD at scale across cultural contexts. This study describes the development and validation of the Caregiver Reported Early Development Instruments (CREDI) short form, a new tool for measuring the motor, cognitive, language, social–emotional, and mental health skills of children under age three in culturally diverse settings. Results from 8022 children living in 17 low-, middle-, and high-income countries suggest that the CREDI short form is valid, reliable, and acceptable for measuring population-level ECD. Data highlight differences in CREDI scores within and across countries based on maternal education, child nutritional status, and household stimulation practices. Implications for ECD policy and practice are described.

© 2018 Elsevier Inc. All rights reserved.

Over the past several decades, a growing body of research has highlighted the important role of early childhood development (ECD) for later-life success (Heckman, 2006; Moffitt et al., 2011; Nores & Barnett, 2010; Peet et al., 2015). The birth to three-year period is considered a time of particular biological and environmental sensitivity; it is during these early years that children begin to acquire some of the most basic yet transformative developmental skills (Black et al., 2017; Shonkoff et al., 2012). As a result, early childhood has increasingly been recognized by governments and non-governmental organizations (NGOs) as a “window of opportunity” for improving not only the developmental outcomes of individual children, but also the social and economic wellbeing of society as a whole (Black et al., 2017). This increased focus is also reflected in the recently ratified Sustainable Development Goals (SDGs), which directly incorporate early development under Target 4.2 (United Nations, 2015).

Despite this rapidly expanding interest in supporting ECD globally, no internationally validated tools of infants’ and toddlers’ skills currently exist that are sufficiently easy to administer, interpret, and compare across cultures. Such tools are necessary for monitoring global progress toward increasing developmental wellbeing and equity, as well as for generating better evidence on children’s developmental strengths and needs across diverse populations. To address this gap, we developed a new tool called the Caregiver-Reported Early Development Instruments (CREDI). The CREDI was designed to assist in monitoring progress toward meeting SDG Target 4.2, as well as for generating new data on children’s early wellbeing worldwide. The aim of this paper is to describe the development and validation of this new tool, as well as to provide initial data on the ECD status of children living in diverse parts of the world.

1. Existing approaches to estimating ECD status globally

Historically, several different approaches have been used to measure children’s ECD. At the individual level, developmental screeners and clinical assessments such as the Denver Developmental Screening Test, the Bayley Scales of Infant and Toddler Development, and the Ages & Stages Questionnaires have been used for decades to screen and diagnose children with developmental disabilities or delays (Bayley, 1969; Bricker & Squires, 1999; Frankenburg & Dodds, 1967). The strength of these individual assessments is their ability to provide detailed, normed information on children’s developmental skills and behaviors across multiple domains, including motor, language, and cognitive development.
Although all of these instruments have been used by researchers in international settings, the application of these tools in non-Western settings is somewhat contentious, given that both their items and standards were developed for very specific (typically U.S.-based) populations (Peña, 2007). Furthermore, the costs associated with purchasing, training, adapting, and implementing these measures often preclude their implementation at scale (Fernald, Kariger, Engle, & Raikes, 2009).

On a global level, several tools have recently emerged to assess ECD at the population level. Unlike individual assessments, these population-level tools are typically designed to provide information on the average skill levels of children at the community, national, or regional level. By necessity, population-level instruments must be simple and inexpensive to implement, and they must be cross-culturally comparable. Several population-level tools exist for measuring the development of preschool and early school-aged children. The Early Development Index (EDI), for example, is a teacher-reported questionnaire targeting multiple dimensions of school readiness for children between 3.5 and 6.5 years of age. It has been used nationally since 2004 to capture information on more than one million Canadian kindergarteners (Janus & Offord, 2007). The EDI has also been adapted and applied for population-level use in a variety of other countries, including Australia, the United States, Indonesia, China, Peru, Brazil, and Jamaica (Brinkman et al., 2007; Brinkman et al., 2016; Ip et al., 2013; Janus, Brinkman, & Duku, 2011; Janus et al., 2014). Other population-level initiatives have been developed specifically with low- and middle-income country (LMIC) contexts in mind. The Regional Project on Child Development Indicators (PRIDI) led by the Inter-American Development Bank provides population-level data across Latin America on two- to four-year-old children's cognitive, language, social–emotional, and motor development based on a brief set of performance-based indicators (Verdisco et al., 2014). Save the Children's International Development and Early Learning Assessment (IDEA) uses a similar format to assess 3.5- to 6.5-year-old children and has been applied in more than 40 culturally and linguistically diverse LMICs (Pisani, Borisova, & Dowd, 2015).

The population-level ECD measure with largest reach and coverage to date is the Early Childhood Development Index (ECDI), launched by UNICEF as part of the fourth round of the Multiple Indicator Cluster Surveys (MICS; UNICEF, 2014). The ECDI uses a parent-reported format to capture 10 basic items covering three- and four-year-olds' literacy–numeracy, learning, social–emotional, and physical development. The ECDI has been administered to more than 160,000 children in over 60 LMICs, providing the world's first population-level information on children's ECD status. In particular, data from the ECDI have been used to estimate that approximately one-third of three- and four-year-old children living in LMICs – or approximately 80 million children in total – are experiencing setbacks in either their cognitive or social–emotional development, and that these developmental setbacks are negatively correlated with a number of within- and across-country characteristics, including Human Development Index scores, nutritional status, urbanicity, and wealth (McCoy et al., 2016). Despite its reach, the ECDI remains limited in its current form due to its focus on a relatively small number of items and limited age range.

### 2. The need for a global tool for children under age three

Despite the progress made in population-level assessment of ECD, to our knowledge there is no internationally validated tool currently available for measuring the early skills and behaviors of children under age three. Such a tool is needed for several reasons. First, population-level assessments of young children's develop-mental skills are necessary for monitoring the impact of national, regional, and global policies designed to improve ECD outcomes and reduce developmental inequities. The importance of the early childhood period has been recognized by several recent, ambitious policy initiatives. SDG Target 4.2, in particular, states that by 2030, all children must have access to "quality early childhood development, care and pre-primary education so that they are ready for primary education" (United Nations, 2015). To appropriately track progress toward meeting this milestone, a tool (or set of tools) is needed that can be implemented quickly and easily across many diverse yet often low-resourced settings. If appropriately planned, data generated from such a tool are also likely to have a number of additional practical benefits, including the ability to identify sub-populations of children who might be in need of additional supports, to track the effectiveness of large-scale intervention efforts, and to draw attention to developmental inequities (i.e., for advocacy purposes).

Current approaches to monitoring population-level outcomes for children under age three are insufficient. Most existing global estimates of young children's developmental needs have relied on proxy measures such as child stunting (a height-for-age z-score of <2 SDs below the median) and exposure to poverty, as these are often easier to quantify than complex developmental behaviors (Grantham-McGregor et al., 2007; Lu, Black, & Richter, 2016). And yet, even as rates of stunting and extreme poverty decrease worldwide (UNICEF, 2016; UNICEF, WHO, & World Bank, 2017), early developmental difficulties remain high (McCoy et al., 2016). Research has demonstrated that ECD is a product of a multidimensional set of environmental and biological inputs including not only malnutrition and poverty, but also the quality of caregiver–child interactions, cognitive stimulation, access to resources, and protection from violence and stress (Black et al., 2017; Nores & Barnett, 2010; Walker et al., 2011). Population-level measures that directly represent children's motor, language, cognitive, and social–emotional development are better positioned to capture the cumulative impacts of initiatives designed to target these multiple inputs, providing a more comprehensive “outcome-focused” perspective of true ECD status.

Second, population-level ECD instruments can generate improved evidence on children's early developmental status across diverse and often under-represented areas of the world. Historically, human development research and theory have largely focused on Western, educated, high-income samples (Henrich, Heine, & Norenzayan, 2010). At the same time, a large body of literature has emphasized the importance of children's environment for shaping their development (Bronfenbrenner & Morris, 2006; Sameroff, 2010), highlighting the need for additional research across diverse populations. Several studies from cross-cultural psychology have shown, for example, variation in the timing of basic motor milestone attainment (e.g., sitting, standing, walking) that is thought to be linked to cultural expectations and practices related to these skills (e.g., Werner, 1972). As noted above, more modern evidence has also shown country-level differences in preschoolers' cognitive and social–emotional skills (as measured by the ECDI) that are linked to countries' socioeconomic and nutritional status (McCoy et al., 2016). Generating further evidence on children's ECD in diverse parts of the world can inform a clearer understanding of developmental commonalities and differences, as well as what characteristics might be associated with positive population-level ECD outcomes. Together, this information can be useful for informing the design of more effective intervention strategies.

Descriptive data on ECD may be particularly important for the youngest children. Mounting evidence has confirmed the birth to three period as a time in which individuals are developing most rapidly and are most sensitive to environmental input (Farah et al.,
It is during this period that children begin to achieve basic developmental milestones that allow them to further engage with – and benefit from – their environments (Phillips & Shonkoff, 2000). At the same time, measuring development within this age period can be particularly challenging, as children often lack the verbal and social skills to effectively communicate and demonstrate their skills to an observer. These issues are compounded in low-resourced areas, where access to trained staff who can reliably assess children is limited. As such, alternatives to direct assessment may be more appropriate – and scalable – for capturing population-level ECD in this age period.

3. The present study

The aim of the present study is to describe the development, validation, and initial evidence from the Caregiver-Reported Early Development Instruments (CREDI). The CREDI was designed as a caregiver-reported, cross-culturally comparable, population-level measure of ECD for children under three years. The main objective of the CREDI is to provide an accurate and easy-to-administer assessment of ECD for children between 0 and 35 months that functions across a wide variety of cultural, linguistic, and socioeconomic contexts. The CREDI is similar to the existing population-level assessments of ECD status discussed above in that it focuses on a core set of motor, cognitive, language, social–emotional, and mental health skills of young children that are believed to be relevant across cultural and linguistic settings. It differs in that it focuses exclusively on children under age three and relies on caregivers’ reports of these skills, rather than requiring direct interaction with children.

As part of the larger project, both a short and a long form of the CREDI were developed from the same broad item set. The goal of the long form is to provide detailed information to researchers interested in measuring specific developmental domains. Information on the long form will be presented in a separate paper. In the present study, we focus on the CREDI’s short form, which is intended as a measure that can be integrated into household survey and monitoring efforts in order to provide a “snapshot” of overall ECD status at the population level. In doing so, we aim to contribute to efforts to monitor national, regional, and global changes in ECD over time, as well as to inform the evidence base on developmental similarities and differences around the world. In particular, rather than providing specific domain scores (e.g., for motor or language skills), the CREDI’s short form provides a single, overall score capturing children’s developmental progress across domains. To simplify implementation, the CREDI’s short form (which we will identify simply as the “CREDI” for the remainder of this paper) contains 20 items selected to characterize children’s development within predefined six-month age bands (i.e., 0–5 mo, 6–11 mo, etc.). All 20 items on each age-specific form use a yes/no response scale by each child’s primary caregiver.

The first aim of this paper is to describe the conceptual and empirical steps taken to develop the CREDI, as well as its strengths and limitations as a tool for global ECD monitoring and evaluation. The second aim is to describe the psychometric properties of the CREDI using data collected from 0- to 35-month-old children living in 17 high-, middle-, and low-income countries. In particular, we complement quantitative data on the CREDI’s criterion and discriminant validity, test–retest reliability, and item characteristics with qualitative data on caregivers’ comprehension and interpretation of each CREDI item. The final aim of this study is to use the CREDI to present new information on children’s ECD skills and behaviors across age and country context. In doing so, we aim to build the evidence base on global ECD in the important, but often neglected, 0–3 age period.

4. Methods

4.1. CREDI development process

The CREDI was developed using a multi-phase process that included (1) the development of a conceptual framework, (2) the development of an initial item set, and (3) four rounds of qualitative and quantitative pilot testing. Each of these phases was conducted with feedback from an advisory panel of international ECD and measurement experts representing multiple disciplines and backgrounds.

4.1.1. Phase I: conceptual framework

First, based on a review of the literature, we defined five core domains of ECD as the central foci of the CREDI: motor development (including fine and gross motor skills), language development (including expressive and receptive language skills), cognitive development (including executive function, reasoning, problem solving, and pre-academic knowledge), social–emotional development (including emotional and behavioral self-regulation, emotion knowledge, and social competence), and early mental health difficulties (including internalizing and externalizing symptoms). These domains and constructs were selected based on their cross-cultural applicability, their centrality in the clinical pediatric and ECD literatures, and their ability to predict school readiness and later-life outcomes (e.g., Bornstein & Haynes, 1998; Duncan et al., 2007; Grissmer, Grimm, Ayier, Murrah, & Steele, 2010; Lewis & Brooks-Gunn, 1981; Lyons-Ruth, Easterbrooks, & Cibelli, 1997; Piek, Dawson, Smith, & Gasson, 2008). Importantly, we acknowledge that many early skills do not fall neatly into one domain or another. Reaching for an object, for example, may involve motor skills (to move the arm), cognitive skills (to establish attention toward the object), and language skills (to non-verbally communicate interest in the object to someone else in the room). Nevertheless, we chose these five domains to ensure that CREDI items cover a diverse range of ECD skills and to create links with the existing literature, which relies heavily on these domains.

4.1.2. Phase II: item development

Second, we developed an initial set of items for pilot testing. To begin this process, we reviewed the literature to identify existing ECD direct assessments and caregiver reports from high-, middle-, and low-income country settings (see Appendix A in Supplementary material). Following this review, and based on a pre-specified set of criteria (see McCoy et al., 2017), we developed a set of items to represent each of our five core domains, with some items deemed to represent multiple domains (e.g., children’s ability to express their emotions using words was categorized as representing both language and social–emotional functioning). Example items include: “Can the child maintain a standing position on his/her own, without holding on or receiving support?” (motor); “Can the child follow simple directions (e.g., ‘Stand up’ or ‘Come here’)?” (language and cognitive); and “Does the child involve others in play (i.e., play interactive games with other children)?” (social–emotional). The full list of items is shown in Appendix C in Supplementary material.

Items were restricted to questions that could plausibly be answered as “yes” or “no” by caregivers from all socioeconomic groups. To facilitate administration and ensure comparability within and across countries, no direct observation items were included. All items also included a “don’t know/do not wish to disclose” response option for caregivers who felt they were unable to determine whether the response should be “yes” or “no” or were uncomfortable sharing for any reason. We developed a detailed description of each item’s meaning and target construct(s) to facilitate translation and implementation across a wide range of cultural and linguistic contexts. For motor items, we also created a stylized
drawing of the intended action or behavior. These illustrations were meant to be shown to caregivers during CREDI administration in order to improve their understanding. Drawings were developed to be culturally neutral and were tested as part of the pilot testing (see Fig. 1 for sample drawing). All CREDI items and accompanying materials are freely available on a publicly accessible website: https://sites.sph.harvard.edu/cried.

4.1.3. Phase III: pilot testing

The third phase of the CREDI development process was the pilot testing phase. As noted above, we conducted four iterative rounds of qualitative and quantitative pilot testing. Qualitative pilot testing was conducted in seven sites (Ghana, Guatemala, Hong Kong, Laos, Lebanon, Pakistan, and Tanzania) and included one-on-one semi-structured cognitive interviews with caregivers. During these interviews, caregivers were asked to respond to each CREDI item as well as one or more of a set of pre-determined follow-up questions probing caregivers’ understanding of the item (e.g., “In your own words, what do you think this question is asking?” or “What do you picture when you think of a child doing this type of behavior?”). An additional set of follow-up questions probed caregivers’ general perceptions of the CREDI items, including whether they found the items to be difficult to understand, inappropriate, or incomplete.

Quantitative pilot testing was conducted with 16,029 children through 25 surveys (data collection rounds) across 21 sites in 17 countries (n = 2 low-income; n = 9 lower-middle-income; n = 4 upper-middle-income; n = 2 high-income). Pilot data collection was done in partnership with local researchers and NGOs familiar with the needs, cultures, and languages of each setting. Typically, these local teams integrated the CREDI into their own ongoing projects (see Appendix A in Supplementary material for details). Relative to an independent, multi-site project, this collaborative approach was both more resource-efficient and representative of the CREDI’s intended use. Nevertheless, the majority of pilot testing was conducted using convenience samples, meaning that results from each site cannot be generalized to the full population of children living in that country. Given that the items tested in each country varied based on the pilot round and, in some cases, the objectives of the local partner teams (see Appendix B in Supplementary material), item-level coverage varied substantially across sites. As such, although data were collected on 16,029 children, we restrict our analyses to different subsamples. First, our IRT sample focused on the 14,113 who were under age 36 months at the time of assessment. Second, our final analytic sample for whom CREDI scores were computed was limited to the 8022 children under age 36 months for whom there were responses to at least 75 percent of the age-band specific items selected by the IRT analysis (see Analytic Plan section, below, for details).

All surveys included administration of CREDI items as well as a set of socio-demographic items, which varied somewhat across survey rounds. To reduce administration burden, each caregiver received only a subset of the CREDI items that was most relevant to his/her child’s age (i.e., 0–11 mo, 12–23 mo, 24–35 mo). CREDI administration time varied across setting and age group for the pilot testing, but did not exceed 25 min.

As part of the quantitative testing, a subsample of 509 children (6.3% of the final analytic sample) in Brazil, Ghana, Guatemala, Jordan, and Lebanon included work targeting test–retest reliability in which CREDI items were re-administered to caregivers approximately one week later. A separate subsample of 1814 children (22.6% of the analytic sample) in Brazil, Chile, Pakistan, Tanzania, and Zambia also included concurrent collection of alternative measures of ECD used to test criterion validity. These criterion measures were selected by local field teams based on their cultural appropriateness in the given setting, and included the Ages & Stages Socioemotional Questionnaire (ASQ:SE; Chile; Squires, Bricker, & Twombly, 2002), the Bayley Scales of Infant and Toddler Development III (BSID–III; Pakistan and Tanzania; Bayley, 1993), the Intergrowth Neurodevelopmental Assessment (INTER–NDA; Zambia; Fernandes et al., 2014), the MacArthur–Bates Communicative Development Inventory (Chile; Fenson et al., 2007), and the PRIDI (Brazil; Verdisco et al., 2014). Finally, children’s length/height was measured in a subsample of 3741 children (46.6% of the analytic sample) in Brazil, Colombia, Ghana, Guatemala, Pakistan, Tanzania, and Zambia. Height measures were converted into normed height–for–age z-scores (HAZ) using the WHO Anthro software package (Blossner et al., 2007).

Quantitative pilot testing was conducted primarily through one-on-one verbal interviews with caregivers by trained research staff who had the equivalent of a high school education or above and spoke the relevant local language(s) fluently. This approach was used as an alternative to written surveys to minimize issues associated with low levels of caregiver literacy and to best match with standard approaches to household survey research in LMICs. A subsample of 1651 children (20.6% of the analytic sample) in the United States, India, and Brazil participated in online surveys covering the same CREDI items. Data from 68 U.S. caregivers that received both the in-person and online formats suggested that responses did not significantly or systematically differ based on format of administration (Kane, 2016).

In consultation with the CREDI authors, local teams were responsible for the translation and back-translation of all CREDI items and materials into the local language(s). Teams were encouraged to reference the item descriptions in the CREDI Item Guide to facilitate this process of meaningful rather than direct translation. Teams were also encouraged to use colloquial rather than formal language in all translations. In a small number of cases, local teams also made minor adaptations to the examples provided in the items. For example, some CREDI items provide examples of simple words that, when translated to different languages, become more linguistically complex than originally intended (e.g., the single-syllable “ball” becomes the three-syllable “pelota” in Spanish). When this...
occurred, local teams were instructed to identify examples that adhere to the intention of the original item.

4.2. Validation sample

Pilot Round 1 took place in Tanzania (Ifakara) and included a total of 70 items selected from an initial set of 92 items targeting children ages 18–35 months. Results of Pilot Round 1 are described in McCoy et al. (2017). Based on these results, as well as feedback from our local team and advisory panel, two main revisions were made to the CREDI prior to Pilot Round 2. First, we expanded the age range to cover all children under age three. Second, we included additional, more difficult items for 18- to 35-month-olds to avoid ceiling effects. A revised set of 117 items was tested in a sample of 4472 0- to 35-month-old children living in Bangladesh, Brazil, Laos, Tanzania (Dar es Salaam), the United States, and Zambia as part of Pilot Round 2. After completion of Pilot Round 2, additional item revisions were made, and further items were added to ensure appropriate coverage in all domains and age ranges. A total of 147 items were then administered to a sample of 993 children in Lebanon, Jordan, and Pakistan as part of Pilot Round 3. After a final round of revisions/additions, we then administered a set of 149 items to a sample of 7807 children in Brazil, Cambodia, Chile, Colombia, Ghana, Guatemala, India, Nepal, Philippines, and the United States as part of Pilot Round 4. In addition to the quantitative work conducted in these larger samples, additional qualitative pilot testing was conducted with approximately 60 caregiver–child pairs in each of the following sites: Ghana, Guatemala, Hong Kong, Laos, and Lebanon. Details regarding each site’s pilot sample and methods can be found in Appendix B in Supplementary material. Further details of the changes made to each item between pilot rounds and the rationale for these changes are available upon request from the first author.

Table 1 provides an overview of the full and analytic samples within each of the 17 countries where piloting activities were conducted. It also includes estimates of average daily income per capita and malnutrition for children under age five in these countries as proxy measures for local living conditions (UNICEF, 2017). Country-level average income per person and day ranged between US$6 in Nepal and US$144 in the United States. Similarly, large differences were observed for the physical development of children, with only 2.1 percent of children experiencing stunting (HAZ < −2) in the U.S. and Chile, and over 40 percent of children experiencing stunting in Cambodia, Guatemala, Laos, and Zambia.

4.3. Analytic plan

Several steps were used to construct a scale from the CREDI and assess the scale’s psychometric properties. First, a set of analyses was conducted using the 14,113 children under age 36 months to restrict items to a subset meeting four, pre-specified empirical criteria. In particular, retained items must:

1) have fewer than 10% “don’t know” responses;
2) be understood by at least 80% of caregivers in qualitative interviews;
3) have a test-retest reliability (i.e., a kappa value) of greater than .40; and
4) be invariant across low-, middle-, and high-income countries (i.e., show the same relation with the latent development scale across country income groups).

Item understanding on qualitative interviews was determined based on coding of English interview transcripts. Specifically, two coders – the local interviewer and one trained, master’s level CREDI team member – reviewed caregiver responses to the qualitative interview questions to determine whether the caregiver interpreted the item as originally intended. When the two coders disagreed, another CREDI team member served as a blind tiebreaker. Item invariance across country income categories was established based on fitting separate logistic curves for each item in high-, upper middle-, lower middle-, and low-income countries. Using the curves fitted in low-income countries as reference, we then tested whether intercepts and slopes in any of the three other income groups were statistically significantly different from those in the reference group using a cutoff of $p < .05$ with Bonferroni’s family wise error rate correction for multiple testing.

4.3.1. Item response theory (IRT)

Using the responses from the 14,113 children under 36 months, a two-parameter (2PL) IRT model was fit to the items that met the above-mentioned criteria in order to select the 20 items that mini-

<table>
<thead>
<tr>
<th>Country</th>
<th>Full sample size (total number of children assessed)</th>
<th>Analytic sample size (children with sufficient data to calculate CREDI score)</th>
<th>Country estimated stunting prevalence</th>
<th>Country estimated average daily income per capita in USD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bangladesh</td>
<td>280</td>
<td>12</td>
<td>39%</td>
<td>$8.6</td>
</tr>
<tr>
<td>Brazil</td>
<td>2359</td>
<td>1281</td>
<td>7%</td>
<td>$39.8</td>
</tr>
<tr>
<td>Cambodia</td>
<td>493</td>
<td>409</td>
<td>40%</td>
<td>$9.0</td>
</tr>
<tr>
<td>Chile</td>
<td>244</td>
<td>236</td>
<td>2%</td>
<td>$60.8</td>
</tr>
<tr>
<td>Colombia</td>
<td>378</td>
<td>277</td>
<td>13%</td>
<td>$35.6</td>
</tr>
<tr>
<td>Ghana</td>
<td>3000</td>
<td>1468</td>
<td>19%</td>
<td>$10.8</td>
</tr>
<tr>
<td>Guatemala</td>
<td>205</td>
<td>195</td>
<td>47%</td>
<td>$19.9</td>
</tr>
<tr>
<td>India</td>
<td>200</td>
<td>200</td>
<td>38%</td>
<td>$15.7</td>
</tr>
<tr>
<td>Jordan</td>
<td>317</td>
<td>248</td>
<td>8%</td>
<td>$28.1</td>
</tr>
<tr>
<td>Laos</td>
<td>46</td>
<td>3</td>
<td>44%</td>
<td>$14.6</td>
</tr>
<tr>
<td>Lebanon</td>
<td>426</td>
<td>331</td>
<td>17%</td>
<td>$35.9</td>
</tr>
<tr>
<td>Nepal</td>
<td>363</td>
<td>363</td>
<td>37%</td>
<td>$6.3</td>
</tr>
<tr>
<td>Pakistan</td>
<td>250</td>
<td>222</td>
<td>45%</td>
<td>$12.9</td>
</tr>
<tr>
<td>Philippines</td>
<td>720</td>
<td>719</td>
<td>30%</td>
<td>$19.0</td>
</tr>
<tr>
<td>Tanzania</td>
<td>3715</td>
<td>759</td>
<td>34%</td>
<td>$6.9</td>
</tr>
<tr>
<td>USA</td>
<td>1021</td>
<td>791</td>
<td>2%</td>
<td>$144.4</td>
</tr>
<tr>
<td>Zambia</td>
<td>2012</td>
<td>508</td>
<td>40%</td>
<td>$9.9</td>
</tr>
<tr>
<td>Total/average</td>
<td>16,029</td>
<td>8022</td>
<td>27%</td>
<td>$28.1</td>
</tr>
<tr>
<td>Min</td>
<td>46</td>
<td>3</td>
<td>2%</td>
<td>$6.3</td>
</tr>
<tr>
<td>Max</td>
<td>3715</td>
<td>1468</td>
<td>47%</td>
<td>$144.4</td>
</tr>
</tbody>
</table>

Notes: Income per person and day computed by dividing purchasing-power-parity adjusted per capita income in each country by 365 days. Stunting prevalence is defined as the percentage of children under the age of five with a height-for age z-score < −2. Stunting data were retrieved from [http://data.unicef.org/topic/nutrition/malnutrition/](http://data.unicef.org/topic/nutrition/malnutrition/)
mized the standard error of measurement (i.e., maximized the total information) within each six-month age band (e.g., 0–5 mo, 6–11 mo, etc.). The number of items within each age band was limited to 20 to reduce administration burden while minimizing loss of reliability and precision. IRT traditionally assumes that scaled scores are normally distributed, but preliminary analysis revealed that estimated scale scores exhibited substantial left-skew. Consequently, latent regression was used to relax the normality assumption so that scaled scores were assumed to instead be centered around a mean determined by a cubic function of age. We further allowed for heteroscedasticity in the dispersion of scaled scores around the age-conditional mean by allowing separate variances to be estimated for each two-month age interval. Formally, the model equation for the 2PL IRT model fit to the response data is given by

$$\Pr \left( Y_{ij} = 1 \right) = \frac{e^{\theta_i - \delta_j}}{1 + e^{\theta_i - \delta_j}} \sigma_i \sim N \left( \beta_i \text{AGE}_i + \beta_2 \text{AGE}_i^2 + \beta_3 \text{AGE}_i^3, \sigma^2 \left( \text{AGE}_i \right) \right)$$ (1)

where $Y_{ij}$ is the caregiver’s response to the $j$th item, $\theta_i$ is the underlying latent development scale score for the $i$th child, and $\sigma^2 \left( \text{AGE}_i \right)$ is given by the function defined in Eq. (2).

$$\sigma^2 \left( \text{AGE}_i \right) = \begin{cases} \sigma_1^2 & \text{if 0mo < AGEl < 2mo} \\ \vdots & \text{if 3mo < AGEl < 36mo} \\ \sigma_{18}^2 & \text{if 34mo < AGEl < 36mo} \end{cases}$$ (2)

A maximum-marginal likelihood estimator (MMLE) was used to estimate the discrimination (e.g., $a_i$), difficulty (e.g., $\delta_i$), and structural parameters in Eqs. (1) and (2). MMLE is a full information estimator that does not assume responses are missing completely at random (MCAR). Instead, MMLE makes the less restrictive assumption that missing response data are missing at random (MAR) so that the available item responses effectively inform the missing responses, resulting in improved robustness to potential problems associated with missing data. When estimating our IRT models in the full sample of children under age 36 months, we observed an average percent missing across items of 23.6% (range = 0–82.3%), which can be attributed to the considerations listed above (i.e., updates made to the item sets across pilot rounds and sites’ idiosyncratic decisions on which domains to administer). Mplus version 8 was used to fit all IRT models (Muthén & Muthén, 2017).

Within each age band, individuals’ total scores for the CREDI were determined using the test characteristic curve (TCC). The TCC relates the number of items passed to the underlying latent development scale by summing the 20 item response functions (denoted IRF$_j$($\theta$)). Given that optimal scoring procedures for the 2PL require computational programs likely unavailable to users, the TCC is simply a convenient approach to approximate a child’s scale score based on the number of items passed. Formally, the TCC was calculated for each age band using the formula

$$\text{TCC} \left( \theta \right) = \sum_j \text{IRF}_j \left( \theta \right) = \sum_j \frac{e^{\theta_i - \delta_j}}{1 + e^{\theta_i - \delta_j}}$$ (3)

where the item response function is summed over the $j = 1, \ldots, 20$ selected highest-information items using the discrimination and difficulty estimates obtained after fitting the IRT model defined in Eqs. (1) and (2) to the data. Since the precision of the scaled score estimates depend critically on the number of items with non-missing responses, scores were only calculated for children if caregivers responded to at least 75 percent of the 20 items selected for the child’s age band. Missing item-level responses were imputed using multiple imputation before calculating scores (average% missing across items = 12.3%; range = 0–25%).

A final set of analyses was then conducted to describe the overall properties of the newly generated CREDI short form in the final analytic sample of children with available scores. Test-level invariance (or test bias) by country income and age form was evaluated using the percent unsigned differential test functioning (SuDTF) statistic introduced by Chalmers, Counsell, and Flora (2016) and explained in detail in Appendix G in Supplementary material. Internal consistency of the CREDI was assessed using Cronbach’s alpha within each age band. Discriminant validity of the CREDI was tested through multi-level linear regression models assessing score differentials with respect to child gender, household wealth quintile (as calculated based on the first principal component of a set of site-specific asset indicators), caregiver education, child HAZ, and household provision of stimulation using six items (e.g., whether an adult read, counted, sang, or played with a child over the prior three days) taken from UNICEF’s Multiple Indicator Cluster Survey. These models included random intercepts to account for individuals’ nesting within countries. We examined the CREDI’s criterion validity based on its raw and age-standardized correlations with alternative measures of ECD in the subsamples for which these additional data were available, including one low-income country (Tanzania), two lower-middle-income countries (Pakistan and Zambia), one upper-middle-income country (Brazil), and one high-income country (Chile). Finally, we also report several general findings from qualitative interviews and feedback from field teams.

5. Results

5.1. Item selection

Details on each of the 149 CREDI items tested in the final pilot (Pilot Round 4) can be found in Appendix C in Supplementary material. Of these 149 items, a total of 70 items (47.0%) were dropped during the item selection process (see Table 2 for details). All items had fewer than 5% missing values due to a “don’t know” response, and therefore none were dropped based on this criterion. Four items were dropped as a result of low levels of understanding identified during qualitative interviews. A total of 28 items were dropped due to low test-retest reliability scores. The average for the remaining items’ kappa value was 0.60 (SD = 0.13, range = 0.40–0.88). Test–retest reliability was lowest in the social–emotional and motor health domains and highest in the social, cognitive, and language domains. Finally, 38 additional items were dropped because their slopes and/or intercepts were not invariant across country income classification, indicating that they were functioning differently in different sites. These dropped items came from all domains, with mental health showing the least amount of variation.

5.2. Item response theory and scale formation

2PL IRT was used to select CREDI items for the short form. These results are shown in Appendix D in Supplementary material. Overall, 62 unique items representing multiple domains of ECD were selected across the six age groups. Cognitive skills were represented by 40 items, language skills were represented by 28 items, motor skills were represented by 24 items, social–emotional items were represented by 21 items, and mental health was represented by 4 items. Selected items for the younger age groups were more heavily weighted toward the motor domain, whereas items for the older age groups were more heavily weighted toward the cognitive, language, social–emotional, and mental health domains. The average test–retest reliability for the final set of selected items was kappa = .62 (SD = .13, range = .41–.86; see Appendix C in Supplementary material for item-specific values). The internal consistency (Cronbach’s alpha) of the CREDI was .89 for 0–5 mo, .86 for 6–11 mo, .84 for 12–17 mo, .86 for 18–23 mo, .84 for 24–29 mo, and .80 for 30–35 mo.
We find that the selected items are optimally selected to maximize precision. Appendix E in Supplementary material displays the item difficulty across all 62 items relative to the distribution of children's positions on the latent development scale. We find that item locations pair nicely with person locations. Moreover, Appendix F in Supplementary material displays the test characteristic curves and conditional standard error of measurement by age group. Together, these figures provide further evidence of sufficient and approximately equal precision of scaled score estimates across all six age groups.

Finally, we find little evidence of test bias. In studying invariance by country income category and age group, we find that the substantive size of the bias averaged less than approximately $\text{SD}_{\text{diff}} = 5\%$, or the equivalent of approximately one additional item endorsed or not endorsed. Although we are unaware of any formal guidelines that stipulate sufficient test-level invariance, it is our judgment that the level of test bias we found is insufficient to invalidate inferences about between-group differences in development.

CREDI scores across site can be found in Fig. 2. These results show variability in CREDI scores both within and across countries. In keeping with the levels of adversity faced and general living conditions across data collection sites (see Table 1 and Appendix B in Supplementary material), average CREDI scores were highest in Brazil, Chile, Lebanon, and the U.S., and lowest in Ghana, Guatemala, Jordan, Pakistan, Philippines, and Zambia. At the same time, heterogeneity within sites was consistently larger than heterogeneity across sites, with site differences (site-specific random intercepts) accounting only about 7% of total variation in z-scores observed.

5.3. Discriminant validity

Differences in CREDI scores across several sociodemographic subgroups are shown in Table 3. These results suggest that, on average across sites, girls scored 0.08 SDs higher than boys. Household wealth – a site-specific relative measure of households’ socio-economic status ranging from 1 to 5 – was associated with significant but small increments in CREDI scores, with children from the top wealth quintile scoring an average of 0.20 SDs higher than children from the bottom wealth quintile within each site. Caregiver education also showed robust associations with CREDI scores, with children from households in which the caregiver had more than 12 years of schooling scoring 0.20 SDs higher than children from households where the caregiver reported no formal education. Children’s nutritional status was also associated with CREDI scores, where each SD increment in height-for-age was associated with a 0.12 SD increment in CREDI scores. Finally, household stimulation was also significantly associated with CREDI outcomes, even when controlling for other characteristics. Specifically, children from households that reported enacting six stimulating activities (e.g., reading, counting, playing, telling stories) in the prior three days showed CREDI scores that were 0.95 SDs higher than children from households reporting no stimulation.

5.4. Criterion validity

Table 4 summarizes the raw and age-normalized correlations between the CREDI and alternative “gold standard” metrics of ECD by site. Overall, the CREDI showed adequate concurrent criterion validity, with raw correlations ranging between $r = .34$ and $r = .92$, and age-normalized correlations (which minimize age as a confound in developmental scores) ranging between $r = .23$ and $r = .47$. Importantly, the specific domain of focus varied across these criterion metrics, with only the ASQ-SE including a specific focus on social–emotional wellbeing.

5.5. Qualitative findings

In addition to examining the CREDI’s psychometric properties within this pilot sample, results of qualitative interviews and feedback from field teams revealed that the CREDI was acceptable to both caregivers and data collectors. In terms of the general feasibility of the items, caregivers did not report any consistent concerns regarding the social desirability or acceptability of the CREDI items in the qualitative interviews. Feedback from field teams also suggested that administering 20 CREDI items takes less than five minutes per caregiver, with some variation across site.

6. Discussion

The overall aim of the present paper was to describe the development and validation of the Caregiver-Reported Early Development Instruments (CREDI) short form, a new tool for capturing the overall early development of children under three living in global contexts. The objective of the CREDI is to provide as a scalable source of population-level data on ECD that can be compared within and across diverse cultural and linguistic settings. In particular, the CREDI short form is intended to provide as a tool for measuring the developmental status of populations of children from diverse contexts around the world, for monitoring progress toward achieving global development goals (e.g., SDG Target 4.2), and for generat-
Table 3
Results of linear regression models estimating CREDI scores by gender, wealth, caregiver education, physical growth status, and home stimulation scores.

<table>
<thead>
<tr>
<th>Dependent variable</th>
<th>CREDI short form Z-score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Female (male=0, female=1)</td>
<td>0.081***</td>
</tr>
<tr>
<td>Wealth quintile (1 poorest, 5 richest)</td>
<td>0.049***</td>
</tr>
<tr>
<td>Caregiver education (reference: none)</td>
<td></td>
</tr>
<tr>
<td>Primary</td>
<td>0.056*</td>
</tr>
<tr>
<td>Secondary</td>
<td>0.096***</td>
</tr>
<tr>
<td>Tertiary</td>
<td>0.196***</td>
</tr>
<tr>
<td>HAZ</td>
<td></td>
</tr>
<tr>
<td>MICS stimulation (range = 0–6)</td>
<td></td>
</tr>
<tr>
<td>N countries</td>
<td>16</td>
</tr>
<tr>
<td>N individuals</td>
<td>7541</td>
</tr>
</tbody>
</table>

Note: * p < 0.05, ** p < 0.01, *** p < 0.001; Standard errors shown in parentheses; HAZ = height-for-age z-score. Sample sizes vary due to differential availability of predictors. All models include random intercepts for countries.

Table 4
Raw and age-normalized correlations between the CREDI and other ECD metrics.

<table>
<thead>
<tr>
<th>Measure</th>
<th>Site</th>
<th>Raw correlation</th>
<th>Age-normalized correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>r</td>
<td>p-value</td>
<td>r</td>
</tr>
<tr>
<td>ASQ:SE</td>
<td>Chile</td>
<td>0.764</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BSID cognition</td>
<td>Pakistan</td>
<td>0.922</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BSID cognition</td>
<td>Tanzania (Ilakara)</td>
<td>0.337</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>INTERNDA</td>
<td>Zambia (Chipata)</td>
<td>0.634</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MacArthur Bates CDI</td>
<td>Chile</td>
<td>0.742</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>PRIDI</td>
<td>Brazil (Sao Paulo)</td>
<td>0.480</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Notes: ASQ:SE = Ages & Stages Questionnaire; Social–Emotional; INTERNDA = INTERGROWTH Neurodevelopmental Assessment; PRIDI = Regional Project on Child Development Indicators; BSID = Bayley Scales of Infant and Toddler Development; MacArthur Bates CDI = MacArthur Bates Communicative Development Inventory.

ing new data on developmental similarities and differences across contexts. Results of the present analyses suggest that the CREDI can provide valid, reliable, and practical information regarding children's ECD across a diverse range of cultural and linguistic contexts. Several specific findings warrant further discussion.

First, although the final CREDI short form includes items from all identified developmental domains, fewer items were selected by our statistical (information-maximizing) model from the social–emotional and mental health domains compared to the motor, cognitive, and language domains (see Appendix C in Supplementary material). This was particularly true for younger children, whose CREDI scores are strongly weighted toward the motor domain. These results are not particularly surprising. Previous research and clinical standards have emphasized the importance of motor development in infancy, as these foundational skills enable children to explore their environments and acquire progressively more advanced skills in problem solving, communication, and social interaction (Bushnell & Boudreau, 1993; Iverson, 2010; Pick, Dawson, Smith, & Gasson, 2008). Previous research has also shown stronger correlations amongst motor, cognitive, and language skills compared with social–emotional and mental health behaviors. Indeed, although there is evidence for heterogeneity in how and when children acquire motor, cognitive, and language skills across contexts (Kelly, Sacker, Schoon, & Nazroo, 2006), most of the core milestones in these areas tend to emerge discretely in somewhat consistent patterns (Committee on Children with Disabilities, 2001; Onis, 2006). For example, barring any severe developmental delay or disability, the vast majority of children eventually walk, talk, and follow basic instructions reliably. Social–emotional and mental health characteristics, on the other hand, are much more strongly linked with a number of environmental and biological inputs, creating patterns of non-linear change and cross-cultural variation (Carter, Garrity-Rokous, Chazan-Cohen, Little, & Briggs-Gowan, 2001; Keller et al., 2004). For example, children may increase or decrease their levels of irritability and reactivity depending not only on their age or temperament, but also on the amount of stress and support that they are experiencing day-to-day, the mental health of their caregivers, or the overall acceptance of these behaviors in a given culture or setting.

The CREDI was intentionally designed to focus on social–emotional and mental health items that are most likely to develop in predictable ways across diverse settings and remain stable over short periods of time, rather than those that might be subject to short-term or cross-cultural variability. Indeed, our exclusion of items with poor test–retest reliability and invariance, as well as the use of IRT analyses for item selection helps to ensure that the selected items are reliable and collectively represent a singular underlying construct of development. Given this approach, the CREDI must be interpreted as a general measure of development that is dependent on the specific domains and items included within each age group, rather than as a complete representation of all possible skills, behaviors, and competencies that might matter for young children in diverse contexts.

Second, results of pilot analyses revealed mostly adequate evidence for the reliability and validity of the CREDI in a diverse group of settings. In terms of criterion validity, the CREDI showed significant and positive correlations with “gold standard” direct assessment and caregiver-reported metrics covering a wide range of developmental domains and cultural settings, including when controlling for age. Internal consistency was also adequate, sup-
porting the IRT results and suggesting that the CREDI items collectively represent a single underlying construct of “development.” We also identified evidence for discriminant validity of CREDI scores based on several demographic and environmental characteristics. Consistent with the prior literature, small differences in CREDI scores emerged based on child gender, with girls scoring on average 0.08 SDs higher than boys (Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006; Fenson et al., 1994; Lung, Shu, Chiang, Chen, & Lin, 2009). Children’s height-for-age z-scores were also correlated with their CREDI scores, supporting a long-standing – but incomplete – link between nutritional status and ECD (Sudfeld et al., 2015). In keeping with prior research, within-country wealth – as measured by household assets – and maternal education were positively associated with CREDI scores, yet their effect sizes were somewhat weak compared with those of household stimulation (McCoy, Zulkowski, & Fink, 2015; Wolf & McCoy, 2017). Indeed, although the association between stimulation and CREDI scores may be inflated due to same-reporter bias, its strength highlights the relevance of ongoing efforts to support positive parenting practices around the world, regardless of family SES (e.g., Baker-Henningham & López Bóo, 2010; Black et al., 2017).

Test–retest reliability in the present analysis was more variable, with some items showing high levels of stability across one to two weeks, and others showing greater heterogeneity in responses. Although we have limited the final set of CREDI items to those meeting baseline standards for test–retest reliability, future work should consider additional testing in this area to maximize reliability. Similarly, although we have conducted preliminary analyses to ensure a basic level of item invariance, additional research is needed to ensure that these psychometric properties are stable across and within diverse settings using population-representative samples.

In addition to providing details on the psychometric properties of the CREDI, a further goal of this study was to present descriptive information on children’s development across the 17 sites with CREDI pilot data. Results of descriptive analyses in these convenience-based samples revealed both similarities and differences in young children’s ECD skills within and across sites. As shown in Fig. 2, average CREDI scores varied substantially across sites, with the lower overall mean scores observed in particularly at-risk samples from Cambodia, Ghana, Guatemala, Pakistan, Philippines, and Zambia. Higher-income contexts such as those from Brazil, Chile, Colombia, and Lebanon, on the other hand, showed higher average CREDI scores. As highlighted in the literature, countries can use internationally comparable data in a way that is most relevant and useful to them, such as setting national standards or monitoring the effectiveness of programs and policies that support children to fulfill their potential (Gove & Black, 2016). When using the CREDI, we recommend that any comparisons across populations are made with a careful eye to sampling. In particular, nationally representative data will be needed to make generalizations about the ECD status of children living in countries, which has not happened to date. The CREDI will create a population-level measure of average development, but should not be used to estimate proportions of children not reaching specific milestones. Furthermore, in addition to focusing on population-average CREDI scores, we also recommend attending to within-site heterogeneity, which is great even within our non-representative samples.

6.1. Practical implications

Feedback from qualitative interviews with caregivers and field team members suggested that the CREDI was well understood by participants, while also being quick and simple to implement. Evidence from this study also suggests that a caregiver-report format is appropriate for capturing population-level ECD for children less than three. Relative to more cost-intensive direct assessment formats, caregiver reports provide a more generalizable view on children’s skills and behaviors across time and setting, are more appropriate for capturing the important but difficult to directly observe domains of social–emotional and mental health development, and are less likely to be biased against children who are unfamiliar with clinical assessments, are shy with strangers, and/or do not understand verbal instructions (Fernald et al., 2009). Asking caregivers about their children’s development may also serve as an intervention into and of itself, helping them to be more attentive toward their children’s skills and behaviors in the future. Although the translation process is somewhat arduous, training for the CREDI is relatively minimal, as the items are intended to be read exactly as written with no additional interpretation on the part of the data collector. Finally, the CREDI is an open-source, freely available tool, eliminating costs associated with copyrights.

Together, these considerations suggest that the CREDI is practically feasible to implement within low-resourced, low-literacy settings and in the context of large-scale household surveys. As such, the CREDI may serve as a useful tool for monitoring ECD status and generating new evidence on development in diverse parts of the world. At the same time, the CREDI focuses on a narrow range of ECD skills that are common across countries. As such, to ensure a more holistic population-level measurement strategy, we recommend that the CREDI be used alongside other metrics, including those that (1) capture culturally-specific ECD skills, (2) examine ECD in older age groups, and (3) measure the quality of the environments to which children are exposed.

More generally, the results presented here suggest that assessing population-level development across countries is possible despite large differences in local culture and context. Compared to clinical assessments at the local level, global population-level tools will always be more restricted in scope, but will nevertheless create urgently needed data to guide ECD policy globally.

6.2. Limitations and future directions

Although the present study has numerous strengths, there are several limitations of this work that should be considered before implementing the CREDI. First, although the present validation study included a large number of children from a wide range of cultural and linguistic contexts and although the data were representative for local populations in Brazil, Ghana, Tanzania, and Zambia, the overall sample is not representative of any country or a global population of children. Related to this, the country-level data provided in Table 1 and used for country-level correlations may not have been representative of the specific subgroups of children selected for the present study by convenience. Next steps for this work include collecting data using the CREDI within larger, nationally and/or globally representative samples. Doing so will allow for additional analyses of invariance across subpopulations (e.g., caregivers of different education levels), and will also allow for the development of age norms and better guidelines for whether children are developmentally “on” versus “off track.” Research is also needed targeting specific subpopulations of interest, including atypically developing children or those with disabilities.

Second, although we collected data on criterion validity and test–retest reliability of the CREDI, we lack information in this study on the CREDI’s long-term predictive validity. Future research is needed to identify the ECD constructs and items that are most strongly related to meaningful outcomes later in life. Such information will provide clearer guidelines for early intervention and policy.

Third, although the CREDI appears to have met its goal as a cross-culturally valid population-level tool of overall development status, as noted above, it cannot and should not take the place
of more in-depth, culturally-specific metrics. As a population-level assessment, the CREDI is not designed to screen or diagnose individual children with specific developmental problems or delays, to provide feedback to caregivers on their child’s development status, or to detect subtle individual-level changes that may be induced through intervention. Furthermore, to facilitate cross-cultural comparisons, the domains and concepts constructed by the CREDI are intentionally general and therefore fail to capture meaningful but culturally-specific developmental phenomena. Although the long form of the CREDI is intended to address some of these gaps, we strongly recommend that locally-derived and culturally-sensitive items and measures be used alongside the CREDI when attempting to capture the needs and status of specific individuals or populations, or to research the impacts of smaller-scale programs or policies. Many such culturally-specific tools currently exist or are under development, including the Kilifi Developmental Inventory, the Malawi Developmental Assessment, the Developmental Milestones Checklist, and the East Asia-Pacific Early Child Development Scales (Abubakar, Holding, Van Baar, Newton, & van de Vijver, 2008; Gladstone et al., 2010; Prado et al., 2014; Rao et al., 2014). Moving forward, additional work is needed to make these tools more accessible not only for population-level monitoring and research purposes, but also for providing feedback to clinicians, caregivers, and field staff aiming to understand individual children’s developmental wellbeing.

7. Conclusions

Results of the present study suggest that the CREDI has the potential to serve as a valid, reliable, and practical metric of population-level ECD for children under the age of three. In particular, the short form of the CREDI may provide utility within household survey and global monitoring efforts to quickly and easily capture a “snapshot” of children’s overall ECD status across domains. Such information may be useful for monitoring children’s developmental status, identifying subgroups of children facing particular challenges, allocating scarce resources, and tracking progress in meeting global developmental goals. In the future, additional research using the CREDI and other more culturally-specific ECD assessments in population-representative samples is needed to best serve the specific needs of young children globally.

Funding

This work was supported by the Saving Brains Program from Grand Challenges Canada (Grant Number 0073-03). We would also like to acknowledge the intellectual contributions of the CREDI Advisory Panel members and our data collection partners. Finally, we would like to thank the thousands of children and caregivers who participated in this research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ecresq.2018.05.002.

References


