BMC Public Health

RESEARCH

Open Access

Validation of the Global Scales of Early Development (GSED) tool in rural Western Kenya

Joshua Jeong[1*], Juliet K. McCann[1], Silas Onyango[2] and Michael Ochieng[3]

Abstract

Background Early childhood development (ECD) is a key determinant of long-term health, education, and wellbeing. However, one major global challenge is the lack of ECD assessment tools validated for use in low- and middle-income countries. To address this gap, the World Health Organization (WHO) launched in 2023 the Global Scales for Early Development, an open-access tool designed to generate reliable, valid, and internationally-comparable data on ECD for children aged 0–3 years globally. In this study, we examined the concurrent and convergent validity of the Global Scales for Early Development-long form (GSED-LF) for use with children aged 0–24 months in Kenya.

Methods We analyzed baseline data collected in October-November 2023 as part of a cluster-randomized controlled trial evaluating a parenting program for improving ECD in rural Western Kenya. Primary caregivers (91% mothers) with a child under 24 months were enrolled across 64 villages in Busia and Homabay counties. The GSED-LF was administered to all children ($N=647$). In a randomly selected sub-sample of children ($N=116$), the Bayley Scales of Infant and Toddler Development (Bayley-III) and the Caregiver Reported Early Development Instruments (CREDI) were also administered to compare their scores with those from the GSED-LF. Concurrent validity of GSED-LF was assessed in terms of its correlations with Bayley-III and CREDI. Convergent validity of GSED-LF was examined with respect to parenting outcomes, including parental stimulation, home caregiving environment, and maternal mental health.

Results GSED-LF scores had moderate associations with those on the Bayley and CREDI across the domains of cognitive, language, and motor development. GSED-LF had small associations with socioemotional development and relatively weaker concurrent validity for younger children under 12 months. GSED-LF also demonstrated good convergent validity in terms of showing moderate associations with maternal and paternal stimulation and the home caregiving environment.

Conclusions Overall, this study demonstrated the feasibility and initial validity of the GSED-LF as a direct assessment tool for use in rural Western Kenya. Additional psychometric analyses across diverse settings are needed to strengthen the reliability and validity evidence of the GSED-LF and establish it as a robust, globallyapplicable tool for assessing ECD in resource-limited settings.

Keywords Early child development, Global scales of Early Development, Measurement, Validation, Kenya

*Correspondence:
Joshua Jeong
joshua.jeong@emory.edu

[1]Hubert Department of Global Health, Rollins School of Public Health, Emory University, Atlanta, GA, USA
[2]African Population and Health Research Center, Nairobi, Kenya
[3]B&M Consult, Nairobi, Kenya

Introduction

Early child development (ECD), especially the period from birth to age 3 years, is a critical period of brain development during which children begin to acquire key skills that lay the foundations for later life development and wellbeing [1]. However, due to poverty and other co-occurring early life adversities, it is estimated that approximately 43% of young children in low- and middle-income countries (LMICs) are failing to meet their developmental potential [2]. In response, there has been an increasing investment in programs and policies aimed at promoting ECD globally [3, 4].

One key challenge amidst this bourgeoning field of global ECD is the lack of measurement tools that are feasible, reliable, and valid for use specifically in LMICs [5]. Most existing tools– such as the Bayley Scales of Infant and Toddler Development 3rd edition (Bayley-III) and Ages and Stages Questionnaire– were originally developed in Western contexts, are lengthy and costly to administer, and require significant training by qualified personnel like clinicians or psychologists. These features often make them impractical for use in large-scale community-based implementation research studies in resource-limited contexts across LMICs. Over the past decade, new open-access tools have been created to overcome these limitations and specifically for global use in diverse cultural contexts across LMICs settings. One notable example is the Caregiver Reported Early Developmental Index (CREDI) that was developed in 2017 as a caregiver-reported measure for assessing children's cognitive, language, motor, and socioemotional development among children 0–35 months of age and has been validated to date in over 20 LMICs [6, 7].

Building upon the CREDI, a new open-access tool has been developed recently by the World Health Organization (WHO) in collaboration with various global technical experts called the Global Scales for Early Development (GSED) [8]. Launched publicly in 2023, the GSED aims to become a global standardized tool that can provide international comparable data on ECD measured in terms of a single holistic score (D-score) among children 0–3 years of age [9]. There are two versions of the GSED with each serving a different purpose: the GSED-short form version that is a caregiver-reported measure intended for population-level monitoring and the GSED-long form a directly administered tool designed more for program evaluation purposes. The WHO team has been conducting a multi-country validation study of the GSED tools in seven countries and implemented in two rounds [10]. Round 1 included Bangladesh, Pakistan, Tanzania, and study results supporting GSED validity in these contexts have been published in the GSED technical report [11]. Round 2 included Brazil, China, Côte d'Ivoire, and the Netherlands, and a report of these results are forthcoming [10]. While the aspiration is for the GSED to become a globally validated tool for assessing ECD across cultures and contexts globally, current psychometric evidence is limited and especially in low-income countries settings. Specifically in sub-Saharan Africa, GSED validation studies have only been conducted so far to our knowledge in Tanzania and Côte d'Ivoire.

Therefore, to follow the momentum around the launch of the GSED and build the psychometric evidence of the tool in additional country contexts, the objective of this study was to validate the GSED-long form version for use with children 0–24 months of age in Kenya. To our knowledge, the GSED has not been used before in Kenya. In this study, we assessed concurrent validity of the GSED by estimating its correlations with the Bayley-III and CREDI. We also assessed convergent validity in terms of its associations with maternal and paternal parenting outcomes, which are well-established correlates of ECD outcomes. These findings will establish the appropriateness and validity of the new GSED tool for use in Western Kenya and can guide future efforts to strengthen the psychometric evidence on the GSED.

Methods

Study setting and design

This measurement validation study was embedded within the baseline data collection for a cluster-randomized controlled trial evaluating the effectiveness of a parenting program in Western Kenya [12]. Six wards were selected across Ndhiwa subcounty in Homabay County and Bunyala subcounty in Busia County based on a collaborative process between ChildFund Kenya, which was the lead implementing partner, and community stakeholders. Sixty-four villages were randomly selected with stratification by ward (approximately 10 villages per ward) and randomly allocated to the intervention or control group. In each village, 10 primary caregivers with a child aged 0–24 months were randomly selected and enrolled into the trial. Specifically, 10 caregivers were selected per village because this was the number of participants per parenting group that was determined as the implementation strategy by NGO partners based on their prior experience. Only one caregiver was sampled per household. See trial protocol for more details about the study setting and design [12].

Data collection

Data were collected between October and November 2023 by a team of 14 research assistants and two supervisors. All research assistants were Kenyan, had a bachelor's degree, had prior experience conducting field research relating to child and family health and wellbeing, and were from Homabay and Busia counties. Research assistants conducted direct assessments

of ECD using the GSED-LF on all children enrolled into the trial and interviewed their primary caregivers using a structured survey that included modules on caregiver and household demographics, parenting practices, home learning environment, and mental health ($N = 647$). The purpose for this original data collection and the measurement of particular outcomes were because they aligned with the parenting intervention theory of change and were intended to serve as the baseline assessment for the RCT. Research assistants received a two-week training led by the first author that involved one week of classroom-based instruction and practice on the GSED (e.g., detailed instructions on the GSED, demonstrations by the trainer, practical exercises in pairs, mock assessments of GSED administration and personalized feedback by trainer) followed by a second week of field-based piloting of the tool in non-study communities.

For this GSED-LF validation substudy, a randomly selected subset of enrolled households was revisited within two days of the initial GSED assessment for additional assessments of ECD using the Bayley-III and the CREDI ($N = 116$). All caregivers reconsented for their children to be reassessed. These follow-up assessment visits were conducted by a separate team of two Kenyan research assistants who had extensive experience with conducting ECD assessments and received a similar and separate two-week training (one-week of classroom-based learning and one-week of piloting) specifically focused on the Bayley-III and the CREDI tools.

For the piloting phase, research assistants administered GSED or Bayley-III + CREDI (depending on which team they were assigned to) with two to three children per day. The trainers rotated across the research assistants to observe and provide real-time coaching and feedback to each research assistant individually and ensure their adherence to the administration manual and assessment protocols for the respective ECD assessment tool. Although formal reliability assessments comparing research assistants' scores to the trainers' scores were not conducted for the direct administration tools (GSED and Bayley-III), the trainers closely monitored each RA's performance during practice sessions and field piloting throughout the week to confirm that all RAs on both the GSED and Bayley-III teams demonstrated proficiency by the end of this training period to proceed with data collection. During data collection, trainers and supervisors conducted quality assurance checks of selected GSED, Bayley-III, and CREDI assessments done by each RA and provided real-time feedback as needed to ensure adherence to item administration manual and protocols. The ECD assessments and surveys with primary caregivers combined were approximately 1.5–2 h and were conducted in a private setting at the participant home and in the family's preferred language (i.e., Kiswahili, Luo, Luhya, or English). All data were collected on Android devices using the Kobotoolbox software application.

Measures

Early child development

The main outcome measure was the Global Scales for Early Development-Long Form (GSED-LF), which includes a maximum of 155 directly assessed items that collectively aims to provide one holistic score for ECD among children aged 0–3 years [9]. We selected the GSED-LF over the GSED-SF because it is a direct assessment tool, making it less susceptible to social desirability bias often associated with caregiver-reported measures [11]. Rather than being organized into developmental domains, GSED-LF items are categorized into three sections (A, B, and C) that are structured based on materials used and tasks that are expected to be completed in succession and ordered by level of difficulty. Items are either observed incidentally or elicited by the assessor or both and scored as yes (skill observed) or no. "Start" and "stop" rules are based on the age of child and their performance. A kit of locally available toys and materials was used for administration, which were easy to source in Kenya and culturally appropriate. Instructions for assessors were clear and easy to follow, facilitating smooth implementation of the assessment. A few items required translations for administering with children in local languages. We had these items translated by a Kenyan subject-matter expert in early childhood development assessments who was fluent in these languages. The translations were then reviewed by other Kenyan experts and pre-tested during the training and piloting phases to ensure linguistic and cultural appropriateness for the cultural context while maintaining fidelity to the original tool. In terms of scoring, GSED D-scores were then computed using the *dscore* package in R (version 1.9). This package implements Rasch model keys to calculate one single D-Score that represents a child's overall development. The GSED by design does not have developmental subscale scores [11]. More details about the open-access GSED tool and its latest materials can be found online at: https://www.who.int/teams/mental-health-and-substance-use/data-research/global-scale-for-early-development.

The Bayley Scales of Infant and Child Development–Third Edition (Bayley-III) is a widely used direct assessment of ECD that has been validated for use in Kenya [13]. Child development is directly assessed with respect to five subscales: cognitive, receptive language, expressive language, fine motor, and gross motor. The picture books and materials used in this assessment were adapted to be familiar to children in the local context. In addition to the five directly assessed domains, child socioemotional development was measured using the Bayley-III socioemotional subscale, which is collected

based on caregiver-reported. This subscale included 28 items in which the caregiver reports how their young child responds to their environment and interacts with others using a five-point Likert scale (1 = none of the time to 5 = all the time). Raw total scores were computed for each subscale. The internal consistency was strong for all domains of the Bayley-III within our sample (α = 0.92–0.97).

The Caregiver Reported Early Development Instruments (CREDI) Long Form is a caregiver-reported measure of ECD that was developed for global use with children aged 0–3 years and has previously been validated in East Africa [6, 14]. Caregivers reported on whether their child exhibited various milestones, behaviors, and skills across motor, cognitive, language, socioemotional, and mental health domains. Up to 108 items were asked of the caregiver depending on the child's age and responses to previous items, with caregivers responding yes or no for each item. The online CREDI Scoring Tool was used to produce subscale scores for the motor, cognitive, language, socioemotional domains. The internal consistency was strong for all domains of the CREDI (α = 0.92–0.95). A total raw score was also calculated separately for the child mental health subscale.

Parenting and sociodemographic characteristics

Data on parenting and sociodemographic characteristics were additionally collected through caregiver report. For maternal and paternal education, the primary caregiver reported on the highest level of education completed by the child's mother and father. Caregiver stimulation was assessed using an adapted version of the Family Care Indicators (FCI) [15], whereby the primary caregiver reported on the number of stimulation activities (e.g., play, naming things, telling stories; 7 items) completed by the child's mother and father with the index child over the past week. The caregiving environment was assessed using the Home Observation for Measurement of the Environment for infant and toddlers aged 0–3 years (HOME-IT) [16], which comprises both caregiver-reported and direct observation items used to assess quality of caregiving and the home learning environment. We calculated a total HOME score as well as subscale scores for particularly caregiver responsivity (11 items) and home learning materials (8 items). Both the FCI and HOME are widely used measures of parenting behaviors in LMICs with robust evidence supporting their validity and positive associations with ECD outcomes in these settings, including in Kenya [17–19].

Maternal parenting stress was self-reported using the parental distress subscale of the Parenting Stress Index Short Form (PSI-SF) [20]. This subscale included 12 items about the extent to which mothers felt restricted, unhappy, and stressed in their role as a parent, scored on a five-point Likert Scale (1 = strongly disagree to 5 = strongly agree). Items were summed to create a total score whereby higher scores indicated greater levels of parental distress. We classified high parenting stress as having a score above the 81st percentile (i.e., greater than or equal to a score of 32 in our sample), which is the cut-off scoring guideline established by the developers of the tool. The PSI-SF has been validated across LMIC settings and previously used in Kenya [21, 22]. Finally, maternal depressive symptoms were measured using the Centre for Epidemiological Studies Depression Scale-10 items (CESD-10) in which caregivers report on their experience of depressive symptoms in the past week (1 = rarely to 5 = most or almost all days) [23]. A sum score was calculated whereby higher scores corresponded to greater depressive symptoms. Risk of maternal depression was defined using a cut-off score of CESD-10 ≥ 10 per developer guidelines [23], which has been validated in prior studies including in Kenya [24, 25].

Statistical analysis

We calculated internally age-standardized scores for all measures. We tested concurrent validity by estimating Pearson correlations between age-standardized GSED D-scores and age-standardized domain-specific scores on the Bayley-III and CREDI. Analyses for concurrent validity were conducted in the subsample of children who were assessed using the GSED-LF as well as the Bayley-III and CREDI ($N = 116$). Analyses included dummy variables to control for child development assessors and reduce bias, considering that the 14 RAs were not randomly assigned to participants but were instead divided into two teams based on the counties where they resided. Analyses also accounted for clustering at the village level. We also stratified these analyses by child age group (0–12 months versus 13–24 months) and child gender (boys versus girls) to explore the robustness of concurrent validity results by child demographic characteristics.

Additionally, we assessed convergent validity by calculating Pearson correlations between GSED scores and a set of sociodemographic and parenting variables that are theoretically related to ECD. These factors included maternal and paternal education, maternal and paternal stimulation, HOME learning environment, and maternal mental health. Analyses for convergent validity were conducted in the full sample of children who were assessed using the GSED-LF and whose caregivers completed surveys regarding the parenting variables ($N = 647$). We also examined the associations between these same factors and Bayley and CREDI domain-specific scores to facilitate comparisons in the relative associations across tests. We interpreted correlation coefficients as small (0.10), moderate (0.30), and large (0.50) [26]. Analyses were performed in Stata version 16.1.

Table 1 Sample characteristics (*N* = 647)

Sociodemographic characteristics	Mean (sd) or *n* (%)
Child characteristics	
Age (months), *mean (sd)*	11.7 (6.1)
Age (categorical), *n (%)*	
0–12 months	375 (58.0%)
13–24 months	272 (42.0%)
Sex, *n (%)*	
Female	315 (48.7%)
Male	332 (51.3%)
Primary caregiver relation to index child	
Mother	590 (91.2%)
Grandmother	52 (8.0%)
Other	5 (0.8%)
Mother characteristics	
Age (years), *mean (sd)*	27.0 (6.4)
Education, *n (%)*	
No education	3 (0.5%)
Some primary school (incomplete)	203 (31.5%)
Completed primary education	283 (43.9%)
Completed secondary education	156 (24.2%)
Stimulation practices, *mean (sd)*	3.4 (1.5)
Parenting stress (total score), *mean (sd)*	27.4 (8.2)
High parenting stress (PSI-SF > 32)	202 (34.2%)
Depressive symptoms, *mean (sd)*	9.9 (5.0)
Risk of depression (CESD-10 ≥ 10)	283 (48.0%)
Father characteristics	
Age (years), *mean (sd)*	34.2 (8.8)
Education, *n (%)*	
No education	6 (1.0%)
Some primary school (incomplete)	121 (20.5%)
Completed primary education	223 (37.9%)
Completed secondary education	239 (40.6%)
Stimulation practices, *mean (sd)*	2.5 (2.0)
Household characteristics	
Number of children in household, *mean (sd)*	3.5 (2.1)
Improved water source, *n (%)*	317 (49.0%)
Improved toilet facilities, *n (%)*	183 (28.3%)
Improved roof, *n (%)*	613 (94.7%)
Improved floor, *n (%)*	176 (27.2%)
HOME total score, *mean (sd)*	24.4 (5.0)
HOME responsivity subscale score, *mean (sd)*	8.8 (2.4)
HOME learning materials subscale score, *mean (sd)*	1.6 (1.6)

Note: CESD-10 = Center of Epidemiologic Studies Depression Scale, 10-item version. PSI-SF = Parenting stress index-short form

Results

Sample characteristics

In total, 647 children were assessed on the GSED and had surveys completed by their primary caregiver (Table 1). The mean child age was 11.7 months (SD = 6.1, range = 0–24 months). Less than half of children were female (48.7%). Nearly all primary caregivers were the index child's mother (91.2%). Mothers were on average 27.0 years of age (SD = 6.4), with about one quarter (24.2%) having completed secondary education. Maternal mental health problems were prevalent with 34.2% of mothers having high parenting stress (PSI-SF > 32) and 48% at risk of depression (CESD-10 ≥ 10). Fathers were older at 34.2 years (SD = 8.8) and more having completed secondary education (40.6%). On average, mothers engaged in 3.4 out of seven stimulation activities with their children in the past week, compared to fathers who engaged in 2.5 stimulation activities. On average, households had 3.5 children under 18 years of age (SD = 2.1). Most households (94.7%) had an improved roof (metal or concrete), 49.0% had an improved water source (piped, public tap, covered well, rainwater, or bottled), 28.3% had an improved toilet facility (flush, ventilated pit latrine, or pit latrine with slab), and 27.2% had an improved floor (wood, tiles or concrete). Raw scores on the ECD assessments are summarized in Supplementary Table 1.

Table 2 Correlations between GSED D-score and other ECD measures (*N* = 116)

	Correlation with GSED	*p*-value
Bayley-III		
Cognitive	**0.24**	**0.007**
Receptive language	**0.24**	**0.004**
Expressive language	**0.22**	**0.002**
Fine motor	**0.27**	**0.003**
Gross motor	**0.30**	**< 0.001**
Socioemotional	0.10	0.260
CREDI		
Cognitive	**0.20**	**0.018**
Language	0.11	0.219
Motor	**0.29**	**< 0.001**
Socioemotional	**0.19**	**0.034**
Mental health	-0.03	0.743

Note: Estimates represent Pearson correlation coefficients (*r*). All ECD scores are age-standardized and analyses adjust for child development assessors and are clustered at the village level

Concurrent validity between GSED and other ECD tools

Among the subsample of children who were assessed using all three measures of ECD (*N* = 116), GSED D-scores were positively associated with cognitive, language, and motor domains of the Bayley-III (*r* = 0.22–0.30) (Table 2). GSED D-scores were most strongly associated with Bayley-III fine motor and gross motor development (*r* = 0.27 and *r* = 0.30, respectively), relative to the other Bayley-III domains. There was no significant association observed between the GSED D-score and the Bayley-III socioemotional subscale. Positive correlations were also observed between the GSED and the cognitive, motor, and socioemotional subscales of the CREDI. Again, the motor subscale of the CREDI has the largest associations with the GSED (*r* = 0.29). CREDI language and mental health subscale scores were not significantly associated with the GSED.

Upon stratifying by child age, GSED D-scores were more consistently and strongly associated with both the Bayley-III and CREDI scores among older children aged 13–24 months compared to younger children aged 0–12 months (Table 3). For older children, GSED D-scores had moderate-sized associations with most domains of the Bayley-III and the CREDI ($r = 0.32$–0.49), except for the Bayley-III's expressive language, fine motor, and socioemotional domains and the CREDI child mental health domain. On the other hand, for younger children, GSED D-scores had smaller associations with Bayley-III and CREDI domain scores, most of which were not statistically significant ($r = 0.02$–0.31).

In terms of child gender differences, the magnitude of associations between GSED and Bayley-III domains were generally larger among girls than boys (Table 3). On the other hand, magnitude of associations between GSED and CREDI were generally similar between boys and girls, apart from CREDI socioemotional motor development for which the associations were larger among boys than girls.

Convergent validity between GSED and parenting variables

In the full sample of children ($N = 647$), GSED D-scores were also positively correlated with various parenting-related outcomes (Table 4). More specifically, GSED scores had moderate-sized correlations with maternal stimulation, paternal stimulation, and multiple indicators of the HOME-Inventory (i.e., total score, responsivity subscale score, and learning materials subscale score), which ranged from $r = 0.34$ to $r = 0.57$. No significant associations were observed between GSED D-scores and maternal or paternal education or maternal mental health in our sample. These correlations between the GSED D-scores and parenting outcomes were consistent with the corresponding associations with the Bayley-III and CREDI.

Discussion

In this study, we examined the psychometric properties of the GSED-LF tool for assessing ECD among young children 0–2 years in rural Western Kenya. As one of the first known studies to use GSED in Kenya, we found evidence supporting its concurrent and convergent validity for use in our study context. Our results indicate that the GSED holds promise as new tool for measuring ECD in resource-limited settings.

We established concurrent validity of the GSED in terms of its moderate associations with other previously validated ECD assessment tools in Kenya and similar LMIC settings [6, 13, 27]. More specifically, we found that the holistic D-score on the GSED was consistently associated with cognitive, language, and motor development subscale scores on both the Bayley and CREDI. These associations were slightly stronger in magnitude with Bayley than CREDI, which may be because the Bayley is similarly a direct administration measure compared to CREDI that is a caregiver-reported measure. Compared to the other ECD subscales, GSED had small and inconsistent associations with socioemotional development as measured by the Bayley and CREDI as well as the CREDI mental health subscale. This finding, along with the relatively modest effect sizes observed between the GSED and other outcome domains, may reflect the fact that the GSED generates a single overall development score that captures more variability rather than focusing on specific sub-domains that represent more cohesive

Table 3 Correlations between GSED D-score and other ECD measures stratified by child's age and gender ($N = 116$)

	Children aged 0–12 months		Children aged 13–24 months		Girls		Boys	
	Correlation with GSED	*p*-value	Correlation with GSED	*p*-value	Correlation with GSED	*p*-value	Correlation with GSED	*p*-value
Bayley-III								
Cognitive	0.11	0.343	**0.38**	**0.004**	**0.42**	**0.001**	0.11	0.324
Receptive language	0.18	0.080	**0.38**	**0.012**	**0.39**	**0.003**	0.16	0.125
Expressive language	**0.22**	**0.007**	0.23	0.135	**0.33**	**0.018**	0.16	0.070
Fine motor	**0.24**	**0.036**	0.28	0.060	**0.32**	**0.029**	0.20	0.077
Gross motor	**0.31**	**0.004**	**0.32**	**0.040**	**0.31**	**0.013**	**0.35**	**0.003**
Socioemotional	0.03	0.788	0.18	0.256	0.17	0.280	0.07	0.546
CREDI								
Cognitive	0.15	0.152	**0.34**	**0.015**	0.14	0.339	0.21	0.047
Language	-0.03	0.804	**0.37**	**0.007**	0.07	0.613	0.09	0.427
Motor	**0.23**	**0.018**	**0.49**	**0.001**	0.26	0.065	**0.29**	**0.003**
Socioemotional	0.11	0.284	**0.33**	**0.019**	0.07	0.604	**0.23**	**0.033**
Mental health	-0.02	0.867	-0.20	0.199	0.03	0.834	-0.09	0.489

Note: Estimates represent Pearson correlation coefficients (*r*). All ECD scores are age-standardized and analyses adjust for child development assessors and are clustered at the village level

Table 4 Correlations between parenting variables and GSED D-score as well as with other ECD measures

Parenting variables	GSED (N=647)	Bayley-III (N=116) Cognitive	Receptive Language	Expressive Language	Fine motor	Gross motor	Socio-emotional	CREDI (N=116) Cognitive	Language	Motor	Socio-emotional	Mental health
Maternal education	0.02	-0.02	-0.07	-0.07	0.01	0.02	-0.03	0.01	-0.03	0.04	0.00	-0.04
Paternal education	0.01	-0.05	-0.05	-0.08	-0.08	-0.04	-0.06	-0.10	-0.11	-0.08	-0.09	0.02
Maternal stimulation	**0.45*** **	**0.28** **	**0.26** **	**0.38*** **	**0.29** **	**0.30** **	**-0.26** **	**0.38*** **	**0.38*** **	**0.33** **	**0.40*** **	0.13
Paternal stimulation	**0.38*** **	**0.27** **	**0.25* **	**0.30** **	**0.25* **	**0.26* **	**0.27** **	**0.27* **	**0.27** **	**0.24* **	**0.29** **	0.12
HOME (total score)	**0.57*** **	**0.42*** **	**0.44*** **	**0.33*** **	**0.48*** **	**0.48*** **	**0.36*** **	**0.42*** **	**0.35*** **	**0.47*** **	**0.43*** **	-0.06
HOME (responsivity subscale)	**0.44*** **	**0.32*** **	**0.24** **	0.17	**0.34*** **	**0.35*** **	**0.28** **	**0.30** **	**0.22* **	**0.33*** **	**0.29** **	-0.06
HOME (learning materials subscale)	**0.34*** **	**0.26** **	**0.27** **	**0.20* **	**0.33*** **	**0.32*** **	0.15	**0.26** **	**0.21* **	**0.31*** **	**0.29** **	-0.05
Maternal parenting stress	-0.03	0.10	0.06	0.06	0.08	0.12	0.06	0.13	0.12	0.12	0.11	0.11
Maternal depressive symptoms	-0.05	0.02	0.04	0.04	-0.02	0.03	-0.06	0.03	0.01	0.01	0.03	0.17

Note: Estimates represent Pearson correlation coefficients (r). *$p<0.05$, **$p<0.01$, ***$p<0.001$

sets of skills or behaviors. Thus, studies particularly interested in assessing socioemotional development may benefit from using an additional tool besides GSED to more accurately capture these specific skills. Since the GSED-LF was not designed to include a wide range of socio-emotional development items, WHO and the GSED team are currently developing a separate GSED-Psychosocial Form (GSED = PF) subscale to assess children's psychosocial development and behavioral wellbeing [8]. However, to our knowledge, this measure has not yet been publicly released at the time of this study.

We found notable differences in concurrent validity results when we stratified by child age. The associations between the GSED and both the Bayley and CREDI were smaller and mostly not statistically significant for children under one year, whereas they were larger and more robust for children older than one year. Several prior studies in LMICs have compared concurrent validity of ECD tools by child age, such as between the CREDI and ASQ-3 with children 5–36 months in China [28] and between the Bayley, Battelle Developmental Inventory, and Denver with children 6–42 months in Colombia [29]. Similar to our results, both studies found stronger evidence of concurrent measurement validity with older aged children. This could suggest potential content validity issues for some assessment items of the younger children or be influenced by child age-related sources of measurement bias (e.g., greater variability in child temperament or stronger tester effects due to challenges in managing younger children's behaviors during assessments) [30, 31]. At the same time, the stronger concurrent validity observed among older children is understandable as older children tend to exhibit a broader range of developmental skills and greater variability in scores, enabling ECD measures to potentially differentiate more effectively and thus enhancing validity.

We also found that concurrent validity results between GSED and Bayley-III were stronger among girls than boys. Although few prior studies have specifically compared measurement validity by child gender, several observational studies have shown girls having higher ECD scores including for socioemotional development [32, 33] and described gender differences in temperament and attentional regulation during structured assessments, which could potentially contribute to gender differences in their performance on ECD tests and measurement validity results. These gender differences in concurrent validity may also reflect cultural influences in terms of styles of play for boys versus girls, gendered expectations around play, and relatedly perhaps children's prior exposure to certain toys based on their gender [34]. Future analyses such as testing for differential item functioning by gender could help identify potential sources underlying these observed differences. Taken

together, future validation studies should pay attention to its use with younger aged children particularly in the first year of life and potential gender differences to ultimately strengthen the validity of ECD assessment tools for use with boys and girls across its full intended age range.

In addition to concurrent validity, GSED demonstrated good convergent validity in terms of its associations with parenting outcomes. GSED had moderate positive associations with maternal and paternal stimulation and the home caregiving environment. These associations are consistent with a strong body of evidence highlighting these parenting-related factors as key inputs of ECD globally [4, 35, 36]. On the other hand, GSED was not associated with maternal or paternal education or maternal mental health in our study context. Convergent validity results for the GSED were consistent with those observed for the Bayley and CREDI, further reinforcing the robustness of the GSED measure.

There are several limitations worth noting. First, we specifically tested concurrent and convergent validity but did not assess other additional types of psychometric evidence, such as predictive validity, test-retest reliability, and measurement invariance. Future validation studies of the GSED should comprehensively assess these various psychometric properties. Second, this validation study was conducted in a relatively small subsample of participants that were enrolled into a larger trial. We did not conduct a formal sample size calculation as this measurement validation study was exploratory in nature. At the time of our study design and data collection, we were unaware of prior studies that presented expected effect sizes for the different types of measurement validity, so our focus was on gathering preliminary evidence. The sample size was also fixed under the predetermined sample for the cluster-RCT and then largely driven by resource and logistical constrains. Despite this limitation, our use of robust measurement tools for the caregiving-related outcomes assessed in our validation analysis provide valuable preliminary insights to inform future research. Appropriately powered studies are needed in the future to confirm the robustness of results and ensure findings are not subject to small sample bias. Finally, it is worth noting that our analytic sample focused on children aged 0–24 months, and our validation results do not generalize to the full age band of children on which the GSED can be used (0–35 months). Therefore, future studies should assess the validity of GSED particularly with children aged 25–35 months in Kenya.

Conclusion

In summary, the GSED demonstrated adequate overall concurrent and convergent validity as a new ECD assessment tool for use in the context of rural Western Kenya. We discovered two important insights namely that the GSED tool exhibited weaker validity when used with children under 12 months of age and had small associations with the specific domain of children's socioemotional development and mental wellbeing. Future validation studies should confirm these findings and strengthen the tool as needed to ensure psychometrically robust, cross-culturally equivalent measurement of ECD across LMICs [37]. As the GSED tool begins to be used for various purposes such as for program evaluation, a critical research question moving forward will be its appropriateness as a tool for impact evaluation purposes, specifically whether the GSED can detect group-level differences and its sensitivity to capture domain-specific effects.

Abbreviations

Bayley-III	Bayley Scales of Infant and Toddler Development 3rd edition
CREDI	Caregiver Report of Early Development Index
ECD	Early child development
GSED	Global Scales of Early Development
GSED-LF	Global Scales for Early Development-Long Form
LMICs	Low- and middle-income countries
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12889-025-21801-9.

Supplementary Material 1

Acknowledgements

We thank the team at B&M Consult for data collection as well as the study participants for their time. We also acknowledge the World Health Organization, its technical experts, partners, and funders for their contributions to the development of the Global Scales for Early Development (GSED) tool.

Author contributions

JJ conceptualized the study, developed study methodology, conducted data analysis, drafted the article, and takes responsibility for all aspects of the work. JKM managed data cleaning, contributed to data analysis, and contributed to drafting the article. SO and MO contributed to the data collection. All authors critically reviewed and approved the final version submitted for publication.

Funding

This study received funding from the Conrad N. Hilton Foundation. JJ is supported in part by a grant from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R00HD105984). The funders had no role in study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Data availability

Data are available on request from the corresponding author.

Declarations

Ethics approval and consent to participate

This study received ethics approvals from Institutional Review Board at Emory University (STUDY00006385) and the Ethics and Review Committee at the Jaramogi Oginga Odinga Teaching and Referral Hospital in Kenya (Protocol #: ISERC/JOOTRH/736/23). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Caregivers provided written informed consent for their participation after the enumerator read aloud the consent form and addressed any questions that caregivers may have had.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 5 September 2024 / Accepted: 5 February 2025
Published online: 10 February 2025

References

1. Black MM, Walker SP, Fernald LCH, Andersen CT, DiGirolamo AM, Lu C, McCoy DC, Fink G, Shawar YR, Shiffman J, et al. Early childhood development coming of age: science through the life course. Lancet. 2017;389(10064):77–90.
2. Lu C, Black MM, Richter LM. Risk of poor development in young children in low-income and middle-income countries: an estimation and analysis at the global, regional, and country level. Lancet Glob Health. 2016;4(12):e916–22.
3. Britto PR, Lye SJ, Proulx K, Yousafzai AK, Matthews SG, Vaivada T, Perez-Escamilla R, Rao N, Ip P, Fernald LCH, et al. Nurturing care: promoting early childhood development. Lancet. 2017;389(10064):91–102.
4. Jeong J, Franchett EE, Ramos de Oliveira CV, Rehmani K, Yousafzai AK. Parenting interventions to promote early child development in the first three years of life: a global systematic review and meta-analysis. PLoS Med. 2021;18(5):e1003602.
5. Boggs D, Milner KM, Chandna J, Black M, Cavallera V, Dua T, Fink G, Kc A, Grantham-McGregor S, Hamadani J, et al. Rating early child development outcome measurement tools for routine health programme use. Arch Dis Child. 2019;104(Suppl 1):S22–33.
6. McCoy DC, Waldman M, Fink G. Measuring early childhood development at a global scale: evidence from the caregiver-reported Early Development instruments. Early Child Res Q. 2018;45:58–68.
7. McCoy DC, Sudfeld CR, Bellinger DC, Muhihi A, Ashery G, Weary TE, Fawzi W, Fink G. Development and validation of an early childhood development scale for use in low-resourced settings. Popul Health Metr. 2017;15(1):3.
8. Global Scales for Early Development (GSED). [https://www.who.int/teams/mental-health-and-substance-use/data-research/global-scale-for-early-development]
9. McCray G, McCoy D, Kariger P, Janus M, Black MM, Chang SM, Tofail F, Eekhout I, Waldman M, van Buuren S et al. The creation of the global scales for Early Development (GSED) for children aged 0–3 years: combining subject matter expert judgements with big data. BMJ Glob Health 2023, 8(1).
10. Cavallera V, Lancaster G, Gladstone M, Black MM, McCray G, Nizar A, Ahmed S, Dutta A, Anago RKE, Brentani A, et al. Protocol for validation of the global scales for Early Development (GSED) for children under 3 years of age in seven countries. BMJ Open. 2023;13(1):e062562.
11. Global Scales for Early Development v1.0. Technical report [https://www.who.int/publications/i/item/WHO-MSD-GSED-package-v1.0-2023.1]
12. Jeong J, McCann JK, Onyango S, Ochieng M. A parenting program delivered through existing community-based peer groups to improve early child development in Homabay and Busia Counties, Kenya: study protocol for a cluster-randomized controlled trial. BMC Pediatr. 2024;24(1):592.
13. McHenry MS, Oyungu E, Yang Z, Hines AC, Ombitsa AR, Vreeman RC, Abubakar A, Monahan PO. Cultural adaptation of the Bayley scales of Infant and Toddler Development, 3rd Edition for use in Kenyan children aged 18–36 months: a psychometric study. Res Dev Disabil. 2021;110:103837.
14. Waldman M, McCoy DC, Seiden J, Cuartas J, Fink G. Validation of motor, cognitive, language, and socio-emotional subscales using the Caregiver reported Early Development instruments: an application of multidimensional item factor analysis. Int J Behav Dev. 2021;45(4):368–77.
15. Kariger P, Frongillo EA, Engle P, Britto PM, Sywulka SM, Menon P. Indicators of family care for development for use in multicountry surveys. J Health Popul Nutr. 2012;30(4):472–86.
16. Bradley RH, Caldwell BM. The HOME Inventory and family demographics. Dev Psychol. 1984;20(2):315–20.
17. Jeong J, Bartoli B, McCann JK. Development and validation of a measure for father involvement during early childhood in a resource-limited context. BMC Public Health. 2024;24(1):2970.
18. Onyango S, Kitsao-Wekulo P, Langat N, Okelo K, Murdock DE, Utzinger J, Fink G. Maternal stimulation and early child development in sub-saharan Africa: evidence from Kenya and Zambia. BMC Public Health. 2023;23(1):2418.
19. Jeong J, Bliznashka L, Sullivan E, Hentschel E, Jeon Y, Strong KL, Daelmans B. Measurement tools and indicators for assessing nurturing care for early childhood development: a scoping review. PLOS Glob Public Health. 2022;2(4):e0000373.
20. Abidin R, Flens JR, Austin WG. The parenting stress index. Forensic uses of clinical assessment instruments. edn. Mahwah, NJ, US: Lawrence Erlbaum Associates; 2006. pp. 297–328.
21. Hunt X, Laurenzi C, Skeen S, Swartz L, Sundin P, Weiss RE, Tomlinson M. Family disability, poverty and parenting stress: analysis of a cross-sectional study in Kenya. Afr J Disabil. 2021;10:744.
22. González-López KT, Vásquez-Chingay SN, Rodrigo-Tintaya RA, Leiva-Colos FV, Morales-García WC, Adriano-Rengifo CE. Psychometric properties of the parenting stress index-short form in a Peruvian sample. Psicol Reflex Crit. 2024;37(1):42.
23. Andresen EM, Malmgren JA, Carter WB, Patrick DL. Screening for depression in well older adults: evaluation of a short form of the CES-D (center for epidemiologic studies Depression Scale). Am J Prev Med. 1994;10(2):77–84.
24. Larsen A, Pintye J, Odhiambo B, Mwongeli N, Marwa MM, Watoyi S, Kinuthia J, Abuna F, Gomez L, Dettinger J, et al. Comparing depression screening tools (CESD-10, EPDS, PHQ-9, and PHQ-2) for diagnostic performance and epidemiologic associations among postpartum Kenyan women: implications for research and practice. J Affect Disord. 2023;324:637–44.
25. Kilburn K, Prencipe L, Hjelm L, Peterman A, Handa S, Palermo T. Examination of performance of the Center for epidemiologic studies Depression Scale Short Form 10 among African youth in poor, rural households. BMC Psychiatry. 2018;18(1):201.
26. Cohen J. A power primer. Psychol Bull. 1992;112(1):155–9.
27. Pendergast LL, Schaefer BA, Murray-Kolb LE, Svensen E, Shrestha R, Rasheed MA, Scharf RJ, Kosek M, Vasquez AO, Maphula A, et al. Assessing development across cultures: Invariance of the Bayley-III scales Across Seven International MAL-ED sites. School Psychol Q. 2018;33(4):604–14.
28. Li Y, Tang L, Bai Y, Zhao S, Shi Y. Reliability and validity of the Caregiver reported Early Development Instruments (CREDI) in impoverished regions of China. BMC Pediatr. 2020;20(1):475.
29. Rubio-Codina M, Araujo MC, Attanasio O, Munoz P, Grantham-McGregor S. Concurrent validity and feasibility of short tests currently used to measure early Childhood Development in large Scale studies. PLoS ONE. 2016;11(8):e0160962.
30. Feldman HM, Dollaghan CA, Campbell TF, Kurs-Lasky M, Janosky JE, Paradise JL. Measurement Properties of the MacArthur Communicative Development inventories at Ages one and two years. Child Dev. 2000;71(2):310–22.
31. Miller LE, Perkins KA, Dai YG, Fein DA. Comparison of parent report and direct assessment of child skills in toddlers. Res Autism Spectr Disorders. 2017;41–42:57–65.
32. Weber A, Darmstadt GL, Rao N. Gender disparities in child development in the east Asia-Pacific region: a cross-sectional, population-based, multicountry observational study. Lancet Child Adolesc Health. 2017;1(3):213–24.
33. McCoy DC, Peet ED, Ezzati M, Danaei G, Black MM, Sudfeld CR, Fawzi W, Fink G. Early Childhood Developmental Status in Low- and Middle-Income countries: National, Regional, and global prevalence estimates using Predictive modeling. PLoS Med. 2016;13(6):e1002034.
34. Ertem IO, Krishnamurthy V, Mulaudzi MC, Sguassero Y, Balta H, Gulumser O, Bilik B, Srinivasan R, Johnson B, Gan G, et al. Similarities and differences in child development from birth to age 3 years by sex and across four countries: a cross-sectional, observational study. Lancet Glob Health. 2018;6(3):e279–91.
35. Scherer E, Hagaman A, Chung E, Rahman A, O'Donnell K, Maselko J. The relationship between responsive caregiving and child outcomes: evidence from direct observations of mother-child dyads in Pakistan. BMC Public Health. 2019;19(1):252.
36. Jeong J, Mapendo F, Hentschel E, McCann JK, Yousafzai AK. Validation of an observational tool for assessing mother-child and father-child interactions in Mara, Tanzania. Dev Psychol 2025.
37. Munoz-Chereau B, Ang L, Dockrell J, Outhwaite L, Heffernan C. Measuring early child development across low and middle-income countries: a systematic review. J Early Child Res. 2021;19(4):443–70.

Publisher's note