

RESEARCH

Open Access



Feasibility and acceptability of implementing the Global Scales for Early Development (GSED) package for children 0–3 years across three countries

Ambreen Nizar Merchant^{1*} , Raghbir Kaur², Gareth McCray³, Vanessa Cavallera², Ann Weber⁴, Melissa Gladstone⁵, Magdalena Janus⁶, Imran Nisar¹, Patricia Kariger⁷, Sunil Sazawal⁸, Arup Dutta⁸, Salahuddin Ahmed⁹, Yunting Zhang¹⁰, Mariana Pacifico Mercadante¹¹, Arsene Zongo¹², Yvonne Schönbeck¹³, Tarun Dua², Iris Eekhout¹³, Fahmida Tofail¹⁴, Maureen Black^{15,16}, Abdullah H. Baqui¹⁷, Dana McCoy¹⁸, Farzana Begum¹, Romuald Kouadio E. Anago¹², Alexandra Brentani¹¹, Fan Jiang¹⁰, Symone Detmar¹³, Michelle Perez Maillard², Marcus Waldman¹⁹, Stef van Buuren^{13,20}, Abbie Raikes¹⁹, Katelyn Hepworth¹⁹, Marta Rubio Codina²¹, Shirina Akhtar⁹, Fyezah Jehan¹, Rasheda Khanam⁶, Hassan Naqvi¹, Abrarul Haque Asif⁹, Junaid Mehmood¹, Sidra Afzal¹ and Gillian Lancaster³

Abstract

Background To assess the neurodevelopment of children under three years, a multinational team of subject matter experts (SMEs) led by the World Health Organization (WHO) developed the Global Scales for Early Development (GSED). The measures include (1) a caregiver-reported short form (SF), (2) a directly administered long form (LF), and (3) a caregiver-reported psychosocial form (PF). The feasibility objectives of this study in Bangladesh, Pakistan, and the United Republic of Tanzania were to assess (1) the study implementation processes, including translation, training, reliability testing, and scheduling of visits and (2) the comprehensibility, cultural relevance, and acceptability of the GSED measures and the related GSED tablet-based application (app) for data collection for caregivers, children, and assessors.

Methods In preparation for a large-scale validation study, we implemented several procedures to ensure that study processes were feasible during the main data collection and that the GSED was culturally appropriate, including translation and back translation of the GSED measures and country-specific training packages on study measures and procedures. Data were collected from at least 32 child-caregiver dyads, stratified by age and sex, in each country. Two methods of collecting inter-rater reliability data were tested: live in-person versus video-based assessment. Each country planned two participant visits: the first to gain consent, assess eligibility, and begin administration

Raghbir Kaur, Vanessa Cavallera, Tarun Dua, Michelle Perez Maillard are members of the World Health Organization. The authors alone are responsible for the views expressed in this publication, and they do not necessarily represent the decisions, policy or views of the World Health Organization.

*Correspondence:

Ambreen Nizar Merchant
ambreen.nizar@aku.edu

Full list of author information is available at the end of the article



© The Author(s) 2025, corrected publication 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

of the caregiver-reported GSED SF, PF, and other study measures and the second to administer the GSED LF directly to the child. Feedback on the implementation processes was evaluated by in-country assessors through focus group discussions (FGDs). Feedback on the comprehensibility, relevance, and acceptability of the GSED measures from caregivers was obtained through exit interviews in addition to the FGD of assessors. Additional cognitive interviews were conducted during administration to ensure comprehension and cultural relevance for several GSED PF items.

Results The translation-back translation process identified items with words and phrases that were either mis-translated or did not have a literal matching translation in the local languages, requiring rewording or rephrasing. Implementation challenges reiterated the need to develop a more comprehensive training module covering GSED administration and other topics, including the consent process, rapport building, techniques for maintaining privacy and preventing distraction, and using didactic and interactive learning modes. Additionally, it suggested some modifications in the order of administration of measures. Assessor/supervisor concurrent scoring of assessments proved to be the most cost-effective and straightforward method for evaluating inter-rater reliability. Administration of measures using the app was considered culturally acceptable and easy to understand by most caregivers and assessors. Some mothers felt anxious about a few GSED LF items assessing motor skills. Additionally, some objects from the GSED LF kit (a set of props to test specific skills and behaviors) were unfamiliar to the children, and hence, it took extra time for them to familiarize themselves with the materials and understand the task.

Conclusion This study generated invaluable information regarding the implementation of the GSED, including where improvements should be made and where the administered measures' comprehensibility, relevance, and acceptability needed revisions. These results have implications both for the main GSED validation study and the broader assessment of children's development in global settings, providing insights into the opportunities and challenges of assessing young children in diverse cultural settings.

Keywords Early childhood development, Global, Scales, Measurement, Feasibility, Monitoring

Key messages regarding feasibility

- What uncertainties existed regarding feasibility?

Before testing the psychometric properties of the GSED measures in the main validation study, it was crucial to ensure that the translated items retained their original meaning and were understandable to caregivers, that objects used in the GSED LF kit were familiar to children, and that the administration of the entire set of GSED measures, including the time it took to complete each tool and the overall duration, was feasible. Since the measures were presented on a customized GSED app, it was also essential to determine the app's acceptability to assessors and caregivers.

- What are the key feasibility findings?

The GSED measures and their customized application were well received by children, caregivers, and assessors overall. However, some items in the GSED LF and PF were found to be incomprehensible. Valuable feedback was received, and after meetings with SMEs to review back translation, sentence structuring was refined in the local language for better understandability. Since Urdu was not the only language spoken at the Pakistan study site, the measures also needed to be translated into Sindhi for the main validation phase. During focus group discussions (FGDs) of assessors, challenges faced during field implementation regarding tool administration and visit scheduling were reported. All these inputs helped

strengthen the existing training module for more effective preparation of teams for the main validation study.

- What are the implications of the feasibility findings for the design of the main study?

The need to translate measures in alternative languages in one national context was highlighted from the feasibility phase, along with some rephrasing and restructuring that was needed in a few items. Additionally, a revised in-depth training module with more detailed segments was added to address challenges faced during the feasibility phase in the form of specific scenarios. A more comprehensive standard operating procedures (SOP) document for data collection was written, focusing on visit schedules and timelines. Additionally, changes suggested in the app setup were made to allow real-time data collection and more robust ways to track data collection in preparation for the main validation study.

Background

Rapid brain development occurs in the first 1000 days of life; prenatal and early postnatal experiences significantly impact early childhood development (ECD), influencing lifelong learning and health [1, 2]. Healthy development in this period is associated with future educational achievement, well-being, and life success [3–7]. With the ratification of the United Nations (UN) Sustainable Development Goals (SDGs) and the inclusion of Target

4.2.1, which aims at monitoring the proportion of children under five years that are developmentally on track [8], accurate measurement of child development at the population level has become a priority. However, measuring child development is complex [9], and the tools available are often not culturally sensitive [10], or easy to administer, or globally applicable [11]. Furthermore, more than 150 instruments are presently available that capture child development [12], with different domain structures (e.g., gross-motor, fine-motor, cognitive, language, socioemotional) and varying scoring mechanisms—some of which are outdated given advances in measurement science [12]. A culturally comparable population and programmatic measurement package with a single score capturing multiple domains based on item response theory and the Rasch model did not previously exist for global use.

This study reports on assessing the feasibility of a large-scale study in Bangladesh, Pakistan, and the United Republic of Tanzania to validate a globally applicable, unidimensionally scored but multidomain child development assessment measure, for children up to 36 months of age, for population level and programmatic evaluation [13]. Under the leadership of the World Health Organization (WHO), three independent and experienced research teams convened to the Global Scales for Early Development (GSED) [14–16]. The GSED was constructed from large-scale datasets containing 66,075 children assessed on 2211 items from 18 measures of child development from 32 countries [17]. Subject matter experts made in-depth judgments to inform final item selection based on conceptual matches between items from different measures, developmental domains measured by each item, perceptions of the feasibility of administration of each item in diverse contexts, and a good fit on the Rasch model [18]. The final GSED prototypes of caregiver-reported short-form (SF), directly administered long-form (LF), and caregiver-reported psychosocial form (PF) were created for tablet-based and paper-based assessments.

The GSED SF includes 139 items indicative of skills and behaviors related to cognitive functioning, motor, language, and social-emotional development. For example, “*Can your child bang objects together?*” assesses a child’s fine motor hand/eye coordination. Sixty items include prompts in the form of culturally neutral images, short animations, and audio recordings that assist in understanding the question. All items are presented as questions to the caregiver, with a binary response option “Yes/No.” Start rules based on the child’s age and expected level of development and stop rules based on varying performance are used to ensure that all pertinent data have been collected. The GSED LF includes 155 items

assessing similar skills and behaviors to the GSED SF, administered directly by an assessor, following start-and-stop rules based on the child’s age and responses. A locally constructed, low-cost kit with props that the child interacts with to show their developmental skills, such as rattles or toy cars, was used in the assessments. Scores of GSED LF items are also binary (i.e., either the skill was observed or not observed). For example, “*Picks the longest stick of three*” or “*Finds toy hidden under the cloth*” are both considered to assess cognition. Both the GSED SF and LF provide a unidimensional “Developmental-score (D-score)” representing the child’s development level and a single Development for Age-adjusted Z score (DAZ) [12], which takes into account the child’s age—with developmental curves being developed analogous to those in the WHO Multicentre Growth Reference Study, allowing scores to be compared across the age range. The third measure, the GSED PF, aims to provide a population-level indication of the extent to which children (up to 3 years) exhibit early precursors of nonnormative behaviors and regulatory issues, which can occur at any age. For example, “*Does your child avoid looking you in the eye?*” is an item that is not necessarily related to age.

A large-scale study is planned to validate the measures in seven countries worldwide: Bangladesh, Pakistan, and the United Republic of Tanzania were initially recruited in phase 1 [19], and another four countries (Brazil, China, Ivory Coast, and the Netherlands) will be recruited to phase 2. The aim is to collect data on a planned sample size of 1248 children per country in a 1-year prospective design to evaluate the psychometric properties of the GSED measures, including concurrent validity, short-term predictive validity, convergent and discriminant validity, and test-retest and inter-rater reliability.

Previous studies have shown that cultural adequacy and cross-cultural comparability are two major challenges of ECD measurement [20, 21]. Before the GSED validation study could commence, each country team completed a feasibility study to ensure that the preparatory processes were complete (e.g., item adaptation and translation, kit preparation), the data collection procedures were clearly understood and could be well managed using a tablet-based application (app) in each country, and the assessments were acceptable to the caregivers and assessors alike in terms of content, administrative capability, and length. This feasibility phase was considered necessary to address and mitigate any anticipated or unanticipated challenges and to ensure optimal consistency across countries in data collection. In this paper, we focus on the three phase 1 countries, as the phase 2 countries feasibility phase is still ongoing.

The specific aim of the feasibility study was to assess the acceptability of the large-scale study setup,

implementation processes, and the GSED measures in three countries to determine whether any changes needed to be made in preparation for the main validation study. The specific objectives were as follows:

1. To evaluate the feasibility of the implementation processes by:
 - a) Assessing the adaptation processes and fidelity of the translation of the GSED measures
 - b) Critically evaluating and refining the training processes
 - c) Trialing visit scheduling and study measures administration processes
 - d) Assessing the robustness of the data management systems
 - e) Comparing “in-person” inter-rater reliability assessment with “video-based” assessments to determine the most appropriate method for the main study.
2. To evaluate the comprehensibility, cultural relevance, and acceptability of:
 - a) The GSED measures and the supplementary battery of other study measures to be administered.
 - b) Using a tablet-based GSED App for data collection.

Specific progression criteria [22, 23] were not created for any objectives, as the main validation study was already funded; instead, this feasibility study was used to gather evidence for and inform changes to the implementation processes in the main study. Such evidence will likely also prove useful for the broader field of early childhood development by providing insights regarding the opportunities and challenges of assessing young children in global settings.

Methods

This study complies with the International Ethical Guidelines for Biomedical Research Involving Human Subjects [24] and received ethical approval from the WHO Ethics Board (Ref 004583), followed by ethical approval from institutional ERCs of individual study sites. From Pakistan, approval was sought from the National Bioethics Committee NBC (Ref 4-87/NBC-/422/19/1170) and Aga Khan University AKU (Ref. 1567). For the Bangladesh site, approval was obtained from the Institutional Review Board (IRB) of the Projahnmo Research Foundation (PR-190002) and Johns Hopkins Bloomberg School of Public Health (IRB No.: 00009615). In the United

Republic of Tanzania-Pemba, the study was approved by the Zanzibar Health Research Ethics Committee (Ref: ZAHREC/03/PR/Sept/2019/02).

Study settings and participants

The feasibility study was conducted from January to March 2020 in Bangladesh, Pakistan, and the United Republic of Tanzania. In all three countries, most children were enrolled from existing cohorts of the Alliance for Maternal and Newborn Health Improvement (AMANHI) study group [25]. In sites where children had outgrown the needed age groups, newborn and younger children were recruited from the Antenatal CorTicosteroids for Improving Outcomes in preterm Newborns (ACTION) trial in Bangladesh (JHSPH IRB # 00007684) [26] and the Demographic Surveillance System in Pakistan [27].

In Bangladesh, the GSED study was implemented in Sylhet district, particularly the two subdistricts of Zakiganj and Kanaighat, where the AMANHI study group maintains a health and demographic surveillance of 500,000 people with an annual birth cohort of approximately 12,500, and the catchment areas include three tertiary care hospitals in Sylhet city. In Pakistan, the study site was a fishing village (Ibrahim Hyderi) located on the outskirts of the metropolitan city of Karachi. In 2022, the number of children under the age of 5 years was approximately 15,393, and the annual birth cohort was 3500 (unpublished data). The Department of Pediatrics and Child Health at Aga Khan University maintains a Primary Health Centre (PHC) at the site staffed by medical doctors, paramedical staff, and community health workers. In the United Republic of Tanzania, the study was undertaken on Pemba Island in Wete and Chake Chake districts, covering a population of ~ 450,000 with an annual birth rate of ~ 12,000 (data from the ongoing surveillance system of AMANHI-Pemba). The AMANHI-Pemba study group has digitized the whole island with each household numbered and geo-referenced, and therefore census of the whole island has been undertaken.

Recruitment and consent

Children and caregivers were approached at home during a first visit by GSED-trained community health workers. Eligibility criteria included the presence of a respondent who was the biological mother, legal guardian if the mother was deceased, or the primary caregiver who spent the most time with the child. In addition, the caregiver respondent was eligible if they were over 18 years, understood the local language used in the GSED forms (i.e., Bangla, Swahili, and Urdu), and spoke to the child in the same language as translated

for the forms. Last, children who were acutely ill in the previous 5 days were rescheduled for a later date. Standard formal consenting procedures were followed.

Sample size and sampling scheme

A minimum sample of 32 caregiver-child dyads from each country site was deemed sufficient based on the joint judgment of statistical and subject matter experts regarding the amount of data needed to be collected to achieve the feasibility objectives [23, 28]. A quota sampling scheme was drawn up to ensure comprehensive coverage of the target age range, stratified into eight age groups (0–2, 3–5, 6–8, 9–11, 12–17, 18–23, 24–29, 30–41 months) and balanced by sex (see Additional file 1). Although our study focused on children aged 0–3 years, we sampled children up to 41 months because older children were needed for the psychometric evaluation of the items in the main study.

Data collection

Study measures

The complete set of GSED measures and other contextual measures, listed in Table 1, were administered to all participants. The kit with props used in the GSED LF administration is shown in Additional file 2.

GSED app

The data were collected via a newly created tablet-based GSED application (app) developed by the Center for Public Health Kinetics Global (United Republic of Tanzania) in collaboration with the social enterprise company Universal Doctor (www.universaldoctor.com). The GSED app is built on a core Open Data Kit (ODK) platform (available at: <http://Getodk.org/>), a free and open-source software platform for off-grid electronic data collection and management in resource-constrained environments. The data collection version v.1.25 of the ODK Collect app was adapted and customized for the GSED project. In

Table 1 Summary of GSED and other contextual measures used in the feasibility study

Construct	What the measure captures	Measure	Administration mode	Average administration time (minutes)
Child neurological development	Cognitive, motor, language, and social-emotional development	GSED SF [18] GSED LF [18]	Caregiver report Child assessment	15–25 30–75
Child behavioral and regulatory function	Indication of early precursors of nonnormative behaviors and regulatory issues	GSED PF	Caregiver report	20
Child health and household socioeconomic status (SES)	<ul style="list-style-type: none"> • Eligibility (exclusion criteria) • Demographic information • Information about acute child health • Delivery and perinatal conditions • Child's health history • Maternal health/chronic illness 	Eligibility and household form (specifically developed for the study)	Caregiver report	35
Child anthropometry	<ul style="list-style-type: none"> • Weight • Infant length/child height • Child's mid-upper arm circumference • Child's head circumference 	Anthropometry form (according to WHO Child Growth Standards) [29]	Child assessment	15
Family environment	<ul style="list-style-type: none"> • Home environment (HOME only) • Play/stimulation/interactions between the child and other family members in the home (HOME) • Child neglect/abuse • Exposure to violence or conflict • Family resilience • Family social support 	Home observation for measurement of the environment inventory (HOME) [30]	Caregiver report and observation	45
Caregiver health and wellbeing	<ul style="list-style-type: none"> • Caregiver depressive symptoms 	<p>Childhood Psychosocial Adversity Scale (CPAS) [31]^a</p> <p>Brief Resilience Scale (BRS) [32]^a</p> <p>Family Support Scale (FSS) [33]^a</p> <p>The Patient Health Questionnaire-9 (PHQ-9) [34]</p>	<p>Caregiver report</p> <p>Caregiver report</p> <p>Caregiver report</p> <p>Caregiver report</p>	15 1 5 5

^aThese measures have been minimally adapted for the study



Fig. 1 a Home page of the GSED app on a tablet. b GSED long form grid

addition to the overall appearance, the app incorporated a grid-based interface for the GSED LF to aid administration. Additionally, the GSED app provided other utility tools, such as a timer and information button, which facilitated the long-form administration by displaying administrative guidelines and images for each item in the grid-based user interface. ODK aggregate with MySQL 5.7 community edition was used as the aggregator at the back end. The data were collected on Android-based tablets with a 10-inch screen for better visibility and user

interface. A screenshot of the app's home page is given in Fig. 1a, and the GSED LF grid is shown in Fig. 1b.

Feasibility outcomes

The methods for addressing each feasibility objective are detailed below. The feasibility of the implementation processes is addressed in section 1, and the acceptability of the processes and measures is explained in more detail in section 2. It should be noted that only one FGD was held with each country team at the end of the study to collect

feedback on the feasibility and acceptability of the processes described.

1. Assessing the feasibility of the implementation process:

- a) Fidelity of translation and adaptation processes of GSED and other measures

Translation was needed for all the GSED measures (LF, SF, and PF) and other contextual measures described in Table 1. The forms were translated from English to Bangla, Urdu, and Swahili for Bangladesh, Pakistan, and the United Republic of Tanzania, respectively. A standardized translation and back translation process was carried out in each country. First, the forms were translated from English to the local language by two independent local professional translators recruited by the study managers at each site [35]. Second, each translation was reviewed by the local study teams to reach a consensus on the wording. Third, the agreed-upon local language versions were back-translated by two separate independent translators into English, and back translations were then compared with the original English version. Finally, the back translations underwent an iterative review and revision process by the WHO team and SMEs, identifying and revising items where the meaning had altered from the original before being finalized and approved for data collection [36]. For the PHQ9 and HOME, local translations were already available, so they were only back-translated once and then reviewed and approved. Eligibility forms also went through a single round of translation and back translation, as they were brief questions with direct and easy meaning.

Further feedback from assessors regarding clarity and perceived comprehensibility for caregivers was obtained via the structured FGD at the end of the feasibility study.

- b) Refining the training processes

The feasibility study was used to test and refine the training processes and packages that had been developed for the validation study. An in-person Training of Trainers (ToT) event for supervisors of all three country teams was conducted for 1 week in the United Republic of Tanzania, led by a team from the WHO and SMEs from various international universities and institutions with sizable experience in developmental psychology, pediatrics, early childhood development, and psychometrics and measure creation. The training involved (i) theoretical sessions about child development principles and measurement, (ii) a detailed review of study procedures,

and (iii) an item-by-item review of the GSED measures and other measures used in the study. This was followed by live demonstrations of best-practice GSED implementation by SMEs and practice sessions that gave further explanations for the “difficult-to-administer” items. Training participants also played a role under supervision to ensure that they understood the administration of items correctly. Draft standard operating procedures (SOPs) for study implementation were developed during the ToT event. The SOPs outlined processes for approaching eligible households, seeking informed consent, administering the measures, and data collection and management, along with item guides and manuals for the GSED measures.

The site supervisors who were participants in this training then served as local “master trainers” who trained their respective country team assessors. To train the assessors at each study site, the site supervisors designed a 2-week training program in consultation with the WHO team. The training and certification process included the following:

1. Pre- and post-training quizzes helped participants focus on the set objectives. In addition, post training quizzes were part of the certification process.
2. Each assessor was needed to perform three administrations of the GSED SF, LF, PF, CPAS, HOME, PHQ9, BRS, and FSS on children aged (1) less than 6 months, (2) 7–18 months, and (3) 19–36 months. The supervisors simultaneously scored assessments. To be approved to collect data for the GSED study, field assessors were needed to undergo a certification process that involved achieving an agreement of 90% on the forms’ scoring between the assessor and the local supervisor.
3. For certification of anthropometric measurements of head circumference, mid-upper arm circumference, length, height, and weight, assessors were trained on standardized procedures [37]. Each country site already had master trainers trained by anthropometry specialists. They served as “gold standard” assessors during training. For inter-rater and intra-rater agreement, assessors and trainees were needed to take anthropometric measurements on ten children in two rounds. Their measurements were checked for intra-rater agreement (precision), and against the measurements, the gold standard assessor took for inter-rater agreement (accuracy). Differences in measurements falling within the defined margins of error (MOE) were considered acceptable. The MOE for length, height, and head circumference was ± 0.5 cm, and the mid-upper arm circumference

Table 2 Summary of visit schedules

Group 1	Group 2
1st visit at home	
Same for both groups	<ul style="list-style-type: none"> • Eligibility and consent form • Household information • GSED short form (SF) (audio-recorded) • GSED psychosocial form (PF) (audio-recorded) • HOME Inventory tool • Anthropometric assessment
Qualitative data collection only for group 1	<ul style="list-style-type: none"> • GSED psychosocial form (PF) cognitive testing (audio-recorded and notes on paper) • GSED psychosocial form (PF) exit interview (audio-recorded and notes on paper)
2nd visit at center/clinic (within 48 h of visit 1)	
Same for both groups	<ul style="list-style-type: none"> • GSED long form (LF) (video recorded) • CPAS • PHQ9 • Family support and resilience
No qualitative data collection during visit 2 for mothers in group 1	<ul style="list-style-type: none"> • GSED long form (LF) (video recorded) • CPAS • PHQ9 • Family support and resilience • GSED long form exit (LF) interview [immediately after GSED LF] (audio-recorded and notes on paper) • Comprehensive visit exit interview [at the end of all testing] (audio-recorded and notes on paper)

was ± 0.2 cm. Additional rounds of standardization were implemented for those who did not pass the initial round.

The FGDs held with assessors and supervisors at the end of the feasibility study elicited their feedback on the training sessions. They were asked (i) if they thought the training objectives were met, (ii) whether any modifications were needed, and (iii) what challenges they faced during data collection.

c) Trialing visit scheduling and administration processes

One of the essential objectives of the feasibility study was to trial and devise the most practical way of scheduling visits to administer all the study measures. Due to the large number of measures to be administered, the schedule was divided into two visits to minimize the burden on the families. In all three sites, the first visit was performed at home. In the United Republic of Tanzania and Pakistan, the second visit was performed in a mobile clinic or clinic setting. In Bangladesh, it was performed at home due to unavailability of clinic or center facilities.

The visit schedule is shown in Table 2. Within each visit, half of the children/caregivers (group 1) received the GSED PF cognitive testing (see section 2a for details) and GSED PF exit interview, and half (group 2) received the GSED LF exit interview and comprehensive exit interview. In addition, at the Bangladesh site, the feasibility sample was divided into two subgroups to assess the feasibility of having one or two study visits to see if conducting all the assessments in one day was feasible. The risk of conducting the assessments over 2 days was that caregivers might not return to the clinic the next day with their child. However, the risk of conducting the assessments in 1 day was that the caregivers and children would feel overburdened and become too restless or tired.

We conducted exit interviews to gather feedback from caregivers about their experience. We asked them about the length of the visits, whether they found it to be a major disruption to their routines, how well the study teams maintained confidentiality and privacy, and the order in which questionnaires were asked. Feedback from assessors was collected regarding the overall challenges they faced during the scheduling of visits and administration of the measures during the FGD administration.

Table 3 Summary of qualitative data collection

Qualitative measures	Tool assessed	Administered to	Intent
Exit interviews	GSED PF, GSED LF, and overall for all other study measures	Caregivers	To understand acceptability, the ease of administration, workflows and visit schedules, and respondent comprehension for the battery of measures used in the GSED validation
Focus group discussions (FGDs)	All	Site supervisors and assessors	Feedback on the experience of various aspects of the study: <ul style="list-style-type: none"> • Consenting process • Ease of administration of the forms • Feedback on visit schedules • Use of GSED app • Training needs • Comprehension of the items • Familiarity of objects in the GSED LF kit
Cognitive testing of GSED PF	GSED PF	Caregivers	Evaluation of how the caregiver understood the items to construct his or her answers

d) Assessing the robustness of the data management systems

Data were checked for completeness, accuracy, and quality by manually monitoring the data collection process at the end of each day. Data were collected on tablets and extracted to CSV format for each data collection form. These CSV files were then merged using pre-written software and shared with the WHO in a password-locked folder by each country's data manager for analysis purposes.

e) Comparing “in-person” inter-rater reliability assessment with “video-based”

A further objective of the feasibility study was to evaluate two methods to assess the inter-rater reliability for the GSED measures to be implemented in the main validation study. The first method consisted of an assessor administering the measure while recording a video (for the GSED LF using a camera fixed on a tripod) or audio (for the GSED SF and PF). The videos and audio were then independently assessed and scored by other assessors. The second method consisted of an independent supervisor (acting as master rater) in-person scoring live assessments simultaneously with the primary assessor.

2. To evaluate the acceptability of:

- The GSED measures and the supplementary battery of contextual measures to be administered

A further objective of the GSED feasibility study was to establish the overall acceptability of the GSED measures in terms of item appropriateness to context and

comprehensibility. Feedback was sought from (i) caregivers ($n=16$ per country) whose feedback regarding cultural acceptability and comprehensibility of GSED measures was critical via exit interviews, (ii) field-site supervisors and assessors from the three countries involved in operationalizing each step of the study process via FGDs conducted at the end of the feasibility study, and (iii) a subsample of caregivers reviewing 9 items in the newly created GSED PF via cognitive interviews. Table 3 summarizes the data collected.

The FGDs helped understand the viewpoints of both caregivers and assessors within each country, which were fed back by the supervisors and assessors. Table 4 lists the prompts given in the FGDs.

The caregiver exit interviews comprised semi-structured questions about (i) the GSED LF, (ii) the GSED PF, and (iii) the overall administration experience at the end of the second visit. As the GSED LF was directly administered to a child, it was important to know how easy or difficult this interaction was for the families. Hence, a question asked during the GSED LF exit interview was, “*Was there anything during the administration of the tests with your child that you did not feel comfortable with?*”. Another question asked during the comprehensive caregiver exit interview was, “*Did you feel uncomfortable with any of the questions or how any of the questions were asked?*”. The GSED SF was not included specifically in this part of the work as it was very similar both in content and methodology to the Infant and Young Child Development (IYCD) [28] and Caregiver Reported Early Developmental Instrument (CREDI) [15] where these exercises with caregivers have already been carried out, and thus it was deemed as conveying unnecessary burden on caregivers. An example of an exit interview is given in Additional file 3.

Table 4 Topics and examples of prompts during the FGD sessions held with assessors

Domains	Few examples
Consenting process	<ul style="list-style-type: none"> • Can you tell us about the consenting process? • Did parents have follow-up questions?
Overall experience of administration	<ul style="list-style-type: none"> • What is your overall experience with administering the forms to the parents? • Were there specific forms/questions they struggled/hesitated with? • For you, what was the most difficult form to administer? • How do you feel the flow of the form administration went- that is how the forms are sequenced • Did you feel the length of the interviews was a challenge for the respondents or the child?
GSED LF	<ul style="list-style-type: none"> • How did parents respond to the administration of the GSED LF? • Were there any activities that the parents did not understand? • Were there any activities that seemed to make the parents feel uncomfortable? • Were there any test-related equipment materials or pictures that were difficult to use?
Training	<ul style="list-style-type: none"> • Please share your training experience with us • Did you feel you had enough practice with the children/respondents prior to administration? • Did your training include videos of administrations—if yes, was this helpful? • Did the training include any reliability assessment? • What training activity did you enjoy the most?
App	<ul style="list-style-type: none"> • Please share your experience with the app use • Please share the challenges that you faced during the app use • Did you have any concerns with the app distracting the rapport with respondents and children? • Which forms were easiest to use with the app? • Which forms were challenging to use and what were the challenges? • What changes do you suggest to improve the app? • What was your experience with the video-recording? • Did you feel that it was disruptive to the process of form administration? • Please share your challenges and concerns with the video recording
Video recording	

Table 5 Subset of 9 items from GSED PF used in cognitive testing

- Does your child seem to look through or past people as if they were not there?
- When upset, does your child repeatedly rock back and forth?
- Does your child become extremely distressed/upset/disturbed (e.g., cries, screams) in response to loud sounds or bright lights?
- Does your child have extreme changes in emotions (e.g., quickly changing from very happy to very angry) for no apparent reason?
- Does your child act impulsively without thinking (e.g., running into the street without looking or doing other dangerous behaviors)?
- Does your child get worried or anxious?
- Does your child scream while still asleep and cannot be comforted?
- After you have been separated, does your child seem upset (e.g., angry or withdrawn) when you are reunited?
- Is your child unable to sit still?

The GSED PF was a newly created measure comprising 62 items. In preliminary field work, 9 items (see Table 5) had been identified with unusual response patterns, and we took the opportunity to refine and retest these items in this study. Caregiver feedback was gathered while administering the form through cognitive testing. “Think-aloud” techniques were used to improve the instrument’s reliability by ensuring that the meanings of the items were clear to respondents and matched

the conceptual framework of the instrument developers [38]. The method consisted of administering open-ended questions about the items on the measure to the caregiver and asking them to (1) rephrase or explain the items and (2) explain what the items would look like in their child. eliciting their interpretation and understanding of them [39]. The question asked for each item was “*Can you tell me in your own words what you think this question is asking OR describe what you picture when you*

think of this behavior?". These two questions aimed at eliciting an explanation of what the caregiver interpreted and whether any rephrasing, restructuring, or cultural adaptations were needed.

b) Using a tablet-based GSED app for administration

Following the development of the GSED app, web-based training sessions were held to train country supervisors and assessors on its usage which led to the setup of a system of data transfer to the server and cloud storage for each site. Challenges in developing the GSED app, web-based training, and setting up the data management system will be discussed in detail in a separate paper.

Data analysis Information about cultural acceptability and comprehensibility of GSED measures was gathered from the exit interviews and cognitive interviews in parallel as the administration of the GSED measures progressed. The country specific FGDs were conducted after data collection had been completed. The qualitative data were compiled and synthesized with Dedoose, an online tool for examining qualitative data [40]. It allowed researchers to identify themes and extract excerpts from the FGDs as well as compile quantitative data about how participants responded (e.g., number of comments made that included a certain response or theme, such as feeling that some materials were unfamiliar or suited for older children). The Yes and No responses received from exit interviews are summarized using counts and percentages.

After the data analysis had been completed, the feedback and lessons learned were shared at a virtual technical meeting between the WHO coordinators, SMEs, and country teams to discuss whether further revision of the measures and the overall administration processes was needed before the main validation study began.

Results

A total of 110 child-caregiver dyads (Bangladesh $n=32$; Pakistan $n=32$; the United Republic of Tanzania $n=46$) were enrolled in the study. Given that all three sites had a list of children from the AMANHI cohort or from ongoing pregnancy surveillance (updated every 2 months), the quota sampling scheme to cover all age ranges proved easily achievable. The results section describes important corrections made during the review process of back translations before the start of feasibility phase (see Table 6). During feasibility phase, the feedback obtained from assessors and caregivers guided the changes made in revised training module (see Table 8). The challenges experienced regarding the administration of GSED and other tools are listed in Table 10, and the items that needed revision or omission are discussed in Table 11.

1) Feasibility of the implementation process:

a) Fidelity of translation and adaptation processes of GSED and other measures

The rigorous translation and back translation process proved beneficial, as many translation errors across all three languages were identified by SMEs when back-translated items were compared with original English items. Additionally, site supervisors and SMEs had online meetings where site supervisors explained how some words used in the original English items did not have an exact translation in the country's local language or that sometimes adding a few more words would make more sense to the overall item translation than using just single translated words.

See Table 6 for examples of errors identified in back translations.

Another critical finding during FGDs from site assessors in Pakistan was that several eligible families could not participate because Urdu was not spoken in their families. Therefore, measures (using the back translation

Table 6 Examples of errors identified during the translation-back translation process

Original items from GSED SF	Back-translated items that needed correction
When lying on his/her back, does your child move his/her arms and legs?	Does your baby shake hands when lying on his back?
Can your child unscrew the lid from a bottle or jar?	Can your child open the lid of a bottle? As here, the word <i>unscrew</i> does not have any literal translation in Urdu
Does your child grasp your finger if you touch his/her hand?	If you hold the baby's hand, does he hold your finger?
Can your child walk several steps while holding on to a person or object (e.g., wall or furniture)?	Can your child walk several steps without touching (such as walls or furniture) to someone or object?
Can your child tell you or someone familiar his/her own name [nickname] when asked to?	Can your child tell your name or someone you know his or her name (nickname) when asked?
Does your child smile?	Can your child show happiness by smiling?
Does your child stop what he/she is doing when you say "Stop!" even if just for a second?	When your child is being cautioned to "Stop!" doing something does he/she stop even if for a second?

process detailed above) were translated into Sindhi as well, for the main validation phase. This extensive activity of translation and back translation proved beneficial.

b) Refining the training processes

FGDs, where assessors and supervisors participated, gave insight into the challenges faced during administration and interaction with the caregiver-child dyad. The suggested solution was to add an in-depth training module to prepare assessors for the anticipated challenges during data collection. Table 7 provides examples for each of the challenges/difficulties identified during the FGD and their solutions, which were integrated into the revised training module.

During the virtual technical meeting held together with all country teams, after consensus from site investigators, the WHO coordinators and SMEs compiled a structured training module based on thematic/didactic sessions in the classroom and practice sessions at study sites. Details of the revised training module are listed in Table 8.

c) Trialing visit scheduling and administration processes

The feasibility study also assessed the convenience of the overall visit schedules for caregivers and assessors at each site. During the exit interviews, a few mothers said that the duration of the visit could have been shorter, some questionnaires made them uncomfortable, they did not want to answer them, and others felt that the visits posed a disruption in their routine. Caregivers also responded that they found a few materials in the toolkit unfamiliar to their child. See Tables 9 and 10 for a summary of responses from caregivers during GSED PF, LF, and overall comprehensive exit interviews, respectively.

During the FGDs, assessors explained the challenges they faced. This led to further discussion and the decisions made during the virtual technical meeting in preparation for the next phase of the study. See Table 11 for examples.

The two visits' schedule was found to be feasible. During the first visit, families were approached for the first time at home, and consent for participation was obtained from the caregivers and other family decision-makers, which avoided later refusals. The second visit, which was performed at a center or clinic, allowed for a more controlled environment with minimal distraction for the directly observed GSED LF administration.

d) Robustness of the data management systems

The data management system was revised after the feasibility study for data collection, monitoring, and quality control purposes. Data (for example, child name, ID, sex, gestation age, and date of birth) from the eligible participant list for each site were linked to data collection forms on the app and prepopulated for verification at the time of data collection. This helped minimize data entry errors and saved time for data entry. A separate utility module was developed as a desktop-based application for overall study data management. The utility module allowed for scheduling of study visits, monitoring of study recruitment rates in age and sex bins, data completion status for each child, and data visualization and generation of anonymous data files for the analysis and data transfers. An app-based quality control module was developed as part of the data management system to ensure fidelity to the data collection process. The time-intensive procedures for monitoring laid out in the manual would be a key challenge when applied to the large sample size needed for the main validation study at each site. Therefore, an advanced data management system was planned for the main study for a standardized data monitoring and transfer approach for all sites.

e) Comparing "in-person" inter-rater reliability assessment with "video-based"

For the GSED SF and PF, the method of assessing inter-rater reliability by listening to audio recordings was deemed adequate but had several drawbacks that assessors pointed out during the FGD. The main drawback was that the gestures, body language, and nodding used by caregivers could not be recorded. Additionally, the quality of the voice recordings remained a challenge for scoring. Similarly, for the GSED LF, the video recordings used to assess inter-rater reliability had several limitations reported by the site assessors. First, the camera, once placed at a fixed location in a tripod stand, could not capture all actions, especially for the motor component where the child was required to move. Second, where sites were performing assessments at home or in mobile clinics, high levels of lighting were needed for the recording but were found to disturb both children and caregivers, which was a threat to the ecological validity of the data collection. Third, assessors found that the video recorder equipment was a distraction for the children. Finally, country site leads were concerned that some caregivers would not provide consent for making video and audio recordings of the administration given the intrusiveness of the process. The collection of reliability data through supervisors' simultaneous scoring with the

Table 7 Important findings from FGD yielding refinement in the training module

Challenges shared by assessors faced during data collection/administration	How did assessors resolve the issue?	How should it be resolved for the main validation phase
Mothers complained of the lengthy consenting process	Assessors counselled mothers on the importance of the consenting process	A separate topic on the "Consenting process" was introduced in the training module to train assessors on the consenting process and make caregivers understand its importance
Mothers felt anxious about audio and video recording	When the purpose was explained to them, most participants agreed	Audio and video recordings were removed from the next study phase
Refusals: Mother not available, withdraw consent because of the unwillingness of other family members	Accepted refusals	Culturally specific: Assessors were trained to explain the consent form in front of all decision-makers in the family for caregiver-child participation
Follow-up questions from caregivers on the child's development	Assessors followed instructions of staying neutral on the child's performance Filled in referral forms where the child needs a visit to pediatrician	The training presentation for assessors was updated to emphasize that GSED is not a diagnostic tool and to encourage assessors to convey this message to caregivers
Often other family members accompanied the caregiver-child for their second visit when PHQ9 and CPAS are administered. Caregivers were often found worried about the post-interview reaction of the family member(s)	It remained a challenging situation during the feasibility phase	After discussion, it was added as a challenging scenario to be discussed during training, where assessors can be encouraged to counsel other accompanying family members to wait in the waiting area because the child should only have the caregiver present when all the measures are administered

Table 8 Revised training module

Training format/component	Individual strategies	Details
Classroom sessions	Pre-recorded and live presentations on:	<ul style="list-style-type: none"> Introduction to early childhood development ECD, Sustainable development goals and emphasis on SDG 4.2.1, Introduction to GSED measures, Sessions on the consenting process, Rapport building with caregiver and child, Techniques to maintain privacy and confidentiality, especially while asking sensitive questions and Upholding a distraction-free environment for children, especially during the GSED LF administration
	Video clips of trainers administering difficult items	These aid supervisors for site training. For example, an item from the GSED LF asks the "child to run up to the ball and kick it without stopping running"; since assessors found this item challenging to score, a video from the Bangladesh team was later shared with all teams for better understanding of exactly what the scoring of this item entailed
	List of foreseeable practical challenges as "scenarios" for discussion	For example, in one scenario, "An assessor tells you that prior to the start of a practice administration, she observed the child running with her sibling. When the assessor administers the running items of GSED LF, however, the child does not run, similarly, the assessor wants to know if she can score YES for crawling, as she saw the child crawl earlier. How would you manage these situations? (Where a child may demonstrate behavior before or after administration, but not during administration)?" Such scenarios were discussed in smaller groups to elicit more understanding of the items
Interactive session	Role plays, practice sessions among trainees	
Training materials	Practice sessions with children at the study site	
	Instruction manuals for each tool Standard operating procedures SOPs	With brief yet clearly laid visit schedules, administration timelines, instructions in exceptional cases, e.g., if the child is sick or if the mother is divorced
Certification	Checklists before starting administrations	
	The same guidelines for certification to be followed (see Methods)	

primary assessor was found to be preferable to audio and video recording. Therefore, during the virtual technical meeting held after the feasibility phase, it was decided, with agreement from SMEs and site investigators, to adopt the more traditional method of parallel coding by the assessor and supervision for assessing inter-rater reliability for the main validation phase.

2) Evaluation of the acceptability of

- The GSED measures and the supplementary battery of contextual measures to be administered.

Overall, assessors and caregivers across all sites for the GSED SF considered the tool acceptable in their contexts. The GSED SF, which includes media files composed of pictures, audio, and animation clips, was found to enhance the assessors' experience and maintain the caregivers' interest. Assessors shared

that caregivers felt excited to see the media files during the GSED SF administration. For example, for the item "*While holding onto furniture, does your child squat with control?*" an animation clip proved to be extremely useful in helping with task comprehension. Assessors gave feedback that, at times, some mothers had difficulty understanding an item, but as soon as a picture, video, or audio clip was played, they immediately understood and gave a confident response. One concern that assessors shared in the feedback was the disappointment shown by caregivers when a chain of seven "no" answers were needed to stop the GSED SF assessment (as per measure administration instructions). This was part of more challenging scenarios discussed in the training package teaching that assessors should explain to caregivers that since it is a validation study, the start and stop rules are conservative to allow enough data to be collected. These will be revised when the package will be launched for use at scale.

Table 9 Summary of responses during the GSED LF and PF exit interviews

GSED long form (LF) exit interview (total N=71)	Yes n (%)	No n (%)	No response n (%)
1. Do you feel that any of the activities administered to your child were not suitable or appropriate for your child's age?	5 (7.0%)	66 (93.0%)	0 (0.0%)
2. Was there anything during the administration of the tests with your child that you did not feel comfortable with?	1 (1.4%)	70 (98.6%)	0 (0.0%)
3. Do you feel the tasks we asked your child to do are relevant and easy to observe in your community?	48 (67.6%)	23 (32.4%)	0 (0.0%)
4. Do you feel the materials/objects/toys used to assess the skills/actions/behaviors of your child are things that your child is used to seeing or using?	39 (54.9%)	32 (45.1%)	0 (0.0%)
5. Do you feel the pictures used to assess the skills/actions/behaviors of your child are things that your child is used to seeing or using?	33 (46.5%)	22 (31.0%)	16 (22.5%)
6. Did you think the test was too short, too long, or just about right?	Just about right 63 (88.7%)	Too long 7 (9.9%)	Too short 1 (1.4%)
GSED psychosocial form (PF) exit interview (total N=72)	Yes n (%)	No n (%)	No response n (%)
1. Were there any questions that you did not understand?	1 (1.4%)	71 (98.6%)	0 (0.0%)
2. Did you feel uncomfortable answering any of these questions?	0 (0.0%)	72 (100.0%)	0 (0.0%)
3. Did any of these questions seem inappropriate to ask? That is, was any question not relevant, culturally inappropriate or offensive?	0 (0.0%)	56 (77.8%)	16 (22.2%)
4. Were you concerned that other family members, neighbors or study staff might hear your responses to these items?	1 (1.4%)	71 (98.6%)	0 (0.0%)
5. Did you think the questionnaire was too short, too long, or just about right?	Just about right 66 (91.6%)	Too long 4 (5.6%)	Too short 2 (2.8%)

Regarding the GSED LF, the overall feedback from caregivers and assessors was largely positive. Assessors reported during the FGD that caregivers reacted excitedly toward the GSED LF administration. However, mothers of very young children found it uncomfortable during the administration when they were asked to put the child in a prone position. To address this, a

reassuring brief script was added for all items where the child needed to be put in the prone position. In addition, three items needed tapping wooden blocks on a block picture on the tablet screen, but this was found to damage the screen. After the virtual technical meeting, it was decided that laminated sheets should replace the tablet screen for those items that needed tapping. Additionally,

Table 10 Summary of responses during the comprehensive exit interview

Comprehensive Exit Interview (total N=63)	Yes n (%)	No n (%)	No response n (%)
1. Did you feel uncomfortable with any of the questions?	2 (3.2%)	61 (96.8%)	0 (0.0%)
2. Did you feel uncomfortable with how any questions were administered?	1 (1.6%)	56 (88.9%)	6 (9.5%)
3. Did you feel uncomfortable about where any questions were administered?	2 (3.2%)	55 (87.3%)	6 (9.5%)
4. Did you ever feel like you wanted to stop answering questions?	5 (7.9%)	52 (82.5%)	6 (9.5%)
5. Do you feel that the order in which we administered the various questionnaires to you and your child was acceptable?	53 (84.1%)	4 (6.3%)	6 (9.5%)
6. Do you feel the visits were a burden or significant disruption to your day?	6 (9.5%)	57 (90.5%)	0 (0.0%)
7. Did you ever feel that some of the questions were inappropriate or unnecessary?	3 (4.8%)	54 (85.7%)	6 (9.5%)
8. Did you think the visits were too short, too long, or just about right?	Just about right 54 (85.7%)	Too long 9 (14.3%)	Too short 0 (0.0%)

Table 11 Challenges faced regarding "visit schedules" by assessors during the feasibility study

Challenges	Modifications made to visit schedules for the main validation phase
Mothers not available for a second visit on the scheduled day	<ul style="list-style-type: none"> Teams should plan a revisit before the set time limit ends A dashboard needs to be developed to schedule children for the next day to prevent missing any child
If a child sleeps during or before administration of GSED LF	<ul style="list-style-type: none"> Wait for the child to complete their nap; the caregiver-reported questionnaires can be filled during that time Alternatively, teams can schedule another visit if the time limit allows Visits should always be scheduled after discussing with the caregiver their availability, convenient time, and the child's nap time
Anthropometry took a lot of time in the current sequence of the form Height/length, weight, HC, and MUAC	A change in the sequence was suggested: MUAC, HC, weight, and height/length were agreed to be followed
Hesitance from caregivers while answering CPAS and PHQ9	<ul style="list-style-type: none"> CPAS and PHQ9 are recommended to be administered in complete privacy A short script needs to be added before starting the set of questions. For example, the script below has been added before asking questions related to the conflict at home: <i>"Now, I would like to ask you some questions about your relationship with other people in your home. Even when people in a home get along well, sometimes they disagree with each other, get angry, expect different things from each other, or fight... People have different ways to manage their differences. This is common. You are safe to share these things here, and they are confidential. Please remember that if you do not feel comfortable with any of the questions, you can refuse to answer"</i>
The presence of a camera posed a constant distraction for the child. Additionally, the camera position needed shifting many times, especially during the motor component	Video recording was used to assess inter-rater reliability, but after discussing many disadvantages of video recording, it was decided that inter-rater reliability should be assessed live in-person by the supervisor
Maintaining a quiet and distraction-free environment	Family members were counselled about the study and requested a quiet space for administration. This helped a lot to get the child's attention with minimal distraction
Rapport build-up with child	Two visit schedules allowed the assessor to build rapport with the child
Performing interviews during the COVID-19 pandemic	Teams were instructed to get tested for COVID-19 if members had COVID-like illness (CLI). Teams also asked the participants during and before the visit whether any family members or neighbors were suffering from CLI and, if so, rescheduled the visit
Poor network connection in some rural places	In some rural areas, due to poor network connections, teams faced problems sending the eligibility data, which was needed to know the subsample category of the participant. In these conditions, teams were advised to move to a place with a good network connection
Long duration of visits	Emphasis was placed on further clarifying the time commitment for study participation at the consent stage

some children were unfamiliar with particular objects in the kit, including blocks, a peg board, and a shape board. It was later added to the SOP to present kit objects and toys to children before starting the GSED LF during the rapport-building stage. Since it was found that younger children were attracted to objects in the GSED LF administration kit (see Additional file 2), it was suggested by the sites to have a small car or ball as a takeaway gift for the child at the end of the administration.

For the GSED PF, feedback on its acceptability was elicited from cognitive testing and the exit interview. However, caregivers found the basic structure of cognitive test questions themselves challenging to understand. Only a few caregivers could interpret the items asked. Instead of interpreting the item itself, they mostly remained silent or responded to what their child did. One of the items

removed from the GSED PF based on a lack of comprehensibility was item PS12: *Does your child seem to look through or past people as if they were not there?* This was removed as almost all caregivers from the three sites misunderstood this item, interpreting it incorrectly, *as their child ignored people*. Another example was asked to describe: *"After you have been separated, does your child seem upset (e.g., angry or withdrawn) when you are reunited?"* Many caregivers could only partially rephrase the item, and some caregivers had trouble describing what the behavior would look like. After a consensus meeting with SMEs, this item was rewritten as *"When reuniting after being separated, does your child get upset with you (e.g., angry or withdrawn)"*. Cognitive testing had incomplete responses for many other items, and hence, they

Table 12 Items from GSED LF, SF, and PF that were revised or removed

Item no	Original item/object/picture	Change/adaptation/removal (of any item) suggested after feasibility study	Reason
Long-form item C 26	Finds two objects. Three trials were given for each object, and the child must at least find two times out of three trials for both objects	Revised item is: Finds one object instead of two Hence, the number of total trials reduced from 6 to 3	This item was found difficult and lengthy to administer
Long-form item C27	Understands "more" and "less" (2 cups). Initially, asked the child to indicate which of the two cups contained more water	This item was adapted by asking the child to indicate which of the two piles of blocks had more blocks (in place of the cups with water) Additionally, item sequence has now changed to C34 No. of items reduced from 6 to 3. In addition, tapping on the tablet was replaced with tapping on a laminated sheet	This item was reported as tricky for the children to understand and complete as some children wanted to play with the water or drink from the cups
Long form item B50, B50-A, Tapping blocks B51, B51-A, B52 and B52-A		Only sequence has changed; now this is C44 Only sequence changed; now this is A5 Only sequence changed; now this is A35 Item removed	For age appropriateness Its order was late Its order was late Assessors and caregivers both found this item challenging to understand
Long form C43	Understand adjective faster		
Long form A11	Balances head when sitting		
Long form A48	Moves from sitting to standing		
Psychosocial form PS12	Does your child look through or past people as if they were not there?		
Psychosocial form PS21 HOME CPAS	Does your child display any strange behavior No change in the original items. Only the item sequence was changed. All interview and observation items were grouped separately No. of items reduced from 10 to 7 from the section on child discipline. Items where it asked about hitting the child more harshly were removed, as caregivers did not find these questions appropriate. Caregiver explained that the option of Not applicable could be used where the caregiver finds her child is very young and so the question does not apply		

were kept in the measure to track their performance in the main validation analysis.

Table 12 lists items that were revised after receiving specific item feedback for the GSED LF, PF, HOME, and CPAS.

b) Using a tablet-based GSED app for administration

The feasibility study trialed data collection using a custom-made GSED app. Following the development of the GSED app, and before commencing the feasibility study, the GSED team carefully checked the app's robustness on an iterative basis, fixing any issues flagged at each iteration. Each site tested the app for all forms, checking the functionality of start and stop rules, the appearance of age-specific questions for a particular tool, and the screen layout. Extensive written feedback was received from all the sites, after which various changes were made to the app that included (i) adding more skip patterns and specifying field values for sociodemographic information and (ii) correcting the placement of some media files in the GSED SF.

During FGDs, all assessors appreciated that application-based data collection was more efficient. The built-in algorithm for skip patterns and start/stop rules for the administration of the study measures simplified the data collection process, as it facilitated assessor tasks and helped ensure standardized administration.

Since the feasibility phase needed enrolling only approximately 32 participants per country, data collection, storage, and transfer were performed manually. However, as the main validation study would require a more robust data storage system, it was decided that real-time data collection should be adopted for the main validation phase during the virtual technical meeting.

Discussion

The overarching aim of the feasibility study was to test the integrity of the study protocol by trialing the preparatory, administrative, and field logistics that would be needed to implement the GSED battery of measures in three culturally diverse countries before its rollout in a large-scale validation study. The set of GSED measures includes a caregiver-reported questionnaire short form and a directly administered measure long form, both providing a single Development-for-Age Z score (DAZ) that represents the age-adjusted child's level of development. The third measure is a newly created measure, the GSED PF, assessing early precursors of behavior problems and regulatory issues, whose items do not display a developmental trajectory in the same way as the GSED SF and LF. While the GSED LF and SF have items taken from the previous work of the team that are well established,

many items in the GSED PF are new. The tool although have evidenced feasibility and acceptability in USA [41] is still under review to be used in other participating countries of GSED. Once validated, these GSED measures will allow program personnel, researchers, and policymakers to measure global levels of child development for 0–3 years that are comparable across countries.

Overall, the implementation of the processes worked well, and the administration of the measures over two visits was found to be acceptable. However, valuable lessons were learned that were critical for the success of the main study. For example, the collaborative work of translation and back translation among SMEs and site supervisors aided in finalizing translations. Additionally, the meaningful feedback from caregivers and assessors prompted some items to be revised, reworded, and hence retranslated for local adaptations so that the items retained the intent yet were easy for caregivers and children to understand [42]. Another example was finding the need of including a second language Sindhi, in addition to Urdu, for the Pakistan site to ensure inclusiveness in participation.

Similarly, the feasibility study showed that training played a pivotal role in assuring the quality of data collection. The comprehensiveness of both ToT and site training, based on clear objectives and led by SMEs, proved helpful in preparing teams for data collection during the feasibility phase. After the data collection phase, the feedback gained during FGDs (from site assessors) helped refine the training module for the main validation study. A longer training agenda based on comprehensive classroom and interactive sessions with practice in the field was then developed to allow assessors to fully prepare themselves for administration of all measures across the age range of children and gain accreditation [43]. Additionally, it was advised by SMEs that data collection should begin soon after the training.

Testing the feasibility of visit schedules was another important objective that was achieved in the study. Many practical challenges were faced during the feasibility phase, and different approaches were tested. These findings informed solutions to be implemented in the main validation study, thus ensuring that it would run more smoothly. The feasibility phase also assessed and ascertained the acceptability of the GSED and other measures. This was achieved through important feedback received from caregivers and assessors that helped SMEs revise items or change the order in which they were asked where necessary. The media files part of the GSED SF assessment enhanced the assessor's and caregiver's experience. The files were found to be helpful in understanding items. GSED LF was also received positively by caregivers and assessors except for a few items for very young children

requiring a prone position, which were addressed by adding a reassuring script for caregivers. Additionally, allowing children to play with GSED materials before the assessment helped children become familiarized with the kit. Feedback on the GSED PF from cognitive testing and exit interviews inferred that caregivers found few items difficult to understand.

Our feasibility study demonstrated that the GSED app was successful in ensuring smooth data collection with fewer chances of errors, missing values, and entries of illogical values. Start and stop rules for the GSED SF and LF, informed by the experiences in the feasibility phase, were incorporated into the app. Since the feasibility phase needed enrolling a smaller number of participants than the numbers needed in the main validation study, we were able to focus on the app's robustness and its ability to track data collection. Several suggestions were discussed about having real-time data collection built into the app, allowing data to be monitored or viewed by anyone at any time. More details about the app functionality will be discussed in a separate paper.

The feasibility phase allowed us to determine the best way to undertake inter-rater reliability testing. Live observation was compared with audio-video recordings, which showed that the camera's fixed location could not capture all actions, and high levels of lighting were needed. Some caregivers were hesitant to provide consent for making video and audio recordings. Hence, the traditional method of parallel scoring was adopted for the main validation phase. This decision was made with agreement from SMEs and site investigators.

In an effort to include samples of children from more and diverse regions of the world in the validation of GSED, a subsequent second phase of validation will include four more countries (Brazil, China, Ivory coast, and the Netherlands). The feasibility study will also be carried out in these countries to ensure that processes and measures are relevant, well understood, and appropriate for their contexts.

The results of this feasibility study have direct implications not only for the design and implementation of the main GSED study but also for the field of global early childhood development more generally. Our findings reinforce several key lessons, including the importance of careful translation and back-translation processes, the critical role of training in promoting data quality, and the importance of designing data collection to reflect the needs, comfort, and cultural priorities of research participants [43]. This study also identifies several new insights for the field, including how to leverage technology-based data collection tools (e.g., the app) to streamline data collection and reduce measurement error [44] as well as how to design validation studies that generate data that

are comparable across diverse cultural and linguistic contexts.

After being validated in a large-scale study, the GSED measures will allow us to monitor child development globally and compare child development across countries. Furthermore, the GSED measures aim to allow assessment of the impact of programs, policies, and changes in the environment at the macro level on the development of groups of children. This study contributes to these overall goals by providing key insights regarding the opportunities and challenges in implementing validation studies in global contexts.

Abbreviations

APP	Application
AKU	Aga Khan University
AMANHI	Alliance for Maternal and Newborn Health Improvement
ACTION	Antenatal Corticosteroids for Improving Outcomes in Preterm Newborns
BRS	Brief Resilience Scale
CPAS	Child Psychosocial Adversity Scale
CSV	Comma-separated values
CREDI	Caregiver Reported Early Developmental Instrument
CLI	COVID-like illness
DAZ	Development-for-Age Z score
ECD	Early childhood development
ERC	Ethical review committee
FSS	Family Support Scale
FGD	Focus group discussion
GSED	Global Scale of Early Child Development
HOME	Home Observation and Measurement of the Environment
HC	Head circumference
IYCD	Infant and Young Child Development
LF	Long form
MOE	Margin of error
MUAC	Mid-upper arm circumference
ODK	Open Data Kit
PF	Psychosocial form
PHC	Primary Health Centre
PHQ9	Patient Health Questionnaire 9
SDG	Strengths and Difficulties Questionnaire
SME	Subject matter experts
SF	Short form
SOP	Standard operating procedure
TOT	Training of Trainers
UN	United Nations
WHO	World Health Organization

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40814-024-01583-4>.

Supplementary Material 1.

Acknowledgements

We would like to acknowledge the data collectors and data managers from all three country sites as well as the local institutions for their valuable support and collaboration.

Authors' contributions

AN drafted all sections of the manuscript, GM and RK made major contributions to the methods and results section, GL edited all sections of the manuscript, VC and TD conceived the idea for the study, and AW, MG, MJ, AR, and KH designed the study procedures. The rest of the authors contributed to data collection and oversaw the conduct of the study. All the authors have read and approved the final manuscript.

Funding

The study was funded by the Bill and Melinda Gates Foundation (BMGF).

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study complies with the International Ethical Guidelines for Biomedical Research Involving Human Subjects [45] and received ethical approval from the WHO Ethics Board (Ref 004583), followed by ethical approval from institutional ERCs of individual study sites. From Pakistan, approval was sought from the National Bioethics Committee NBC (Ref 4-87/NBC-/422/19/1170) and Aga Khan University AKU (Ref. 1567). For the Bangladesh site, approval was obtained from the Institutional Review Board (IRB) of the Projahnmo Research Foundation (PR-190002) and Johns Hopkins Bloomberg School of Public Health (IRB No.: 00009615). In the United Republic of Tanzania-Pemba, the study was approved by the Zanzibar Health Research Ethics Committee (Ref: ZAHREC/03/PR/Sept/2019/02).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Pediatrics and Child Health, The Aga Khan University, Karachi, Pakistan. ²Department of Mental Health and Substance Use, World Health Organization, Geneva, Switzerland. ³School of Medicine, Keele University, Keele, UK. ⁴School of Public Health, University of Nevada Reno, Reno, NV, USA. ⁵Department of Women and Children's Health, Institute of Life Course and Medical Sciences, University of Liverpool, Liverpool, UK. ⁶Offord Centre for Child Studies, Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, ON, Canada. ⁷Center for Effective Global Action, University of California Berkeley School of Public Health, Berkeley, CA, USA. ⁸Center for Public Health Kinetics, CPHK Global, Pemba, Zanzibar, Tanzania. ⁹Research, Projahnmo Research Foundation, Dhaka, Bangladesh. ¹⁰Child Health Advocacy Institute, National Children's Medical Center, Shanghai Children's Medical Center Affiliated to Shanghai Jiao Tong University School of Medicine, Shanghai, People's Republic of China. ¹¹Department of Pediatrics, University of São Paulo Medical School, São Paulo, Brazil. ¹²Innovations for Poverty Action, IPA Côte d'Ivoire, Abidjan, Côte d'Ivoire. ¹³Department of Child Health, Netherlands Organization for Applied Scientific Research, Leiden, Netherlands. ¹⁴Nutrition and Clinical Services Division (NCSD), International Centre for Diarrhoeal Disease Research Bangladesh, Dhaka, Bangladesh. ¹⁵International Education, RTI International, Research Triangle Park, NC, USA. ¹⁶Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, USA. ¹⁷International Center for Maternal and Newborn Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, USA. ¹⁸Education Policy and Program Evaluation, Harvard Graduate School of Education, Cambridge, MA, USA. ¹⁹Health Promotion, University of Nebraska Medical Center College of Public Health, Omaha, NE, USA. ²⁰Department of Methodology and Statistics, Faculty of Social and Behavioural Sciences, University of Utrecht, Utrecht, Netherlands. ²¹Social Protection and Health Division, Inter-American Development Bank, Washington, DC, USA.

Received: 12 December 2023 Accepted: 24 December 2024

Published: 14 February 2025

References

- Clark H, Coll-Seck AM, Banerjee A, Peterson S, Dalglish SL, Ameratunga S, et al. A future for the world's children? A WHO-UNICEF Lancet Commission. *Lancet*. 2020;395(10224):605–58. [https://doi.org/10.1016/S0140-6736\(19\)32540-1](https://doi.org/10.1016/S0140-6736(19)32540-1).
- Shonkoff J, Richmond J, Levitt P, Bunge S, Cameron J, Duncan G, et al. From best practices to breakthrough impacts a science-based approach to building a more promising future for young children and families. Cambridge: Harvard University, Center on the Developing Child; 2016. p. 747–56.
- Forrest CB, Riley AW. Childhood origins of adult health: a basis for life-course health policy. *Health Aff*. 2004;23(5):155–64.
- Grantham-McGregor S, Cheung YB, Cueto S, Glewwe P, Richter L, Strupp B. Developmental potential in the first 5 years for children in developing countries. *Lancet*. 2007;369(9555):60–70.
- Hertzman C, Boyce T. How experience gets under the skin to create gradients in developmental health. *Annu Rev Public Health*. 2010;31:329–47.
- Chan M. Linking child survival and child development for health, equity, and sustainable development. *Lancet*. 2013;381(9877):1514.
- Richter LM, Norris SA, De Wet T. Transition from Birth to Ten to Birth to Twenty: the South African cohort reaches 13 years of age. *Paediatr Perinat Epidemiol*. 2004;18(4):290–301.
- Daelmans B, Darmstadt GL, Lombardi J, Black MM, Britto PR, Lye S, et al. Early childhood development: the foundation of sustainable development. *Lancet*. 2017;389(100064):9–11.
- Ellingsen KM. Standardized assessment of cognitive development: instruments and issues. *Early childhood assessment in school and clinical child psychology*. 2016;25–49. https://doi.org/10.1007/978-1-4939-6349-2_2.
- Medicine Io, Council NR. From neurons to neighborhoods: the science of early childhood development. Shonkoff JP, Phillips DA. Washington, DC: The National Academies Press; 2000. p. 608. <https://nap.nationalacademies.org/catalog/9824/from-neurons-to-neighborhoods-the-science-of-early-childhood-development>.
- Fernald LCHP, Elizabeth Leah; Kariger, Patricia Karol; Raikes, Abbie. A toolkit for measuring early childhood development in low and middle income countries (English). Washington, D.C: World Bank Group. <http://documents.worldbank.org/curated/en/384681513101293811/A-toolkit-for-measuring-early-childhood-development-in-low-and-middle-income-countries>.
- van Buuren S, Eekhout I. Child development with the D-score: turning milestones into measurement. *Gates Open Res*. 2022;5(81):81.
- Cavallera V, Lancaster G, Gladstone M, Black MM, McCray G, Nizar 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.
- Gladstone M, Lancaster G, McCray G, Cavallera V, Alves CR, Malivichi L, et al. Validation of the infant and young child development (IYCD) indicators in three countries: Brazil, Malawi and Pakistan. *Int J Environ Res Public Health*. 2021;18(11):6117.
- McCoy DC, Sudfeld CR, Bellinger DC, Muhihi A, Ashery G, Weary TE, et al. Development and validation of an early childhood development scale for use in low-resourced settings. *Popul Health Metrics*. 2017;15(1):1–18.
- Weber AM, Rubio-Codina M, Walker SP, Van Buuren S, Eekhout I, Grantham-McGregor SM, et al. The D-score: a metric for interpreting the early development of infants and toddlers across global settings. *BMJ Glob Health*. 2019;4(6):e001724.
- Stef van Buuren, Iris Eekhout, Gareth McCray, Gillian A. Lancaster, Marcus R. Waldman, Dana C. McCoy, et al. D-score: a scale to compare child development across ages, samples and instruments. In preparation.
- McCray G, McCoy D, Kariger P, Janus M, Black MM, Chang SM, 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):e009827. <https://doi.org/10.1136/bmjgh-2022-009827>.
- Organization WH. Global Scales for Early Development (GSED) v1.0, Technical report 2023. Available from: <https://www.who.int/publications/item/WHO-MSD-GSED-package-v1.0-2023.1>.
- McCoy DC, Peet ED, Ezzati M, Danaei G, Black MM, Sudfeld CR, et al. 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.
- Cappa C, Petrowski N, De Castro EF, Geisen E, LeBaron P, Allen-Leigh B, et al. Identifying and minimizing errors in the measurement of early childhood development: lessons learned from the cognitive testing of the ECDI2030. *Int J Environ Res Public Health*. 2021;18(22):12181.

22. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. *Pilot Feasibility Stud.* 2019;5(1):114. <https://doi.org/10.1186/s40814-019-0499-1>.
23. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, et al. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. *BMJ.* 2016;355:i5239. <https://doi.org/10.1136/bmj.i5239>.
24. Council for International Organizations of Medical Sciences. International ethical guidelines for biomedical research involving human subjects. *Bull Med Ethics.* 2002;(182):17–23. PMID: 14983848.
25. Aftab F, Ahmed S, Ali SM, Ame SM, Bahl R, Baqui AH, et al. Cohort profile: the Alliance for Maternal and Newborn Health Improvement (AMANHI) biobanking study. *Int J Epidemiol.* 2021;50(6):1780–1. <https://doi.org/10.1093/ije/dyab124>.
26. The World Health Organization ACTION-I (Antenatal CorTicosteroids for Improving Outcomes in preterm Newborns) Trial: a multi-country, multi-centre, two-arm, parallel, double-blind, placebo-controlled, individually randomized trial of antenatal corticosteroids for women at risk of imminent birth in the early preterm period in hospitals in low-resource countries. *Trials.* 2019;20(1):507. <https://doi.org/10.1186/s13063-019-3488-z>.
27. Naem K, Ilyas M, Fatima U, Kazi M, Jehan F, Shafiq Y, et al. Profile: Karachi Health and Demographic Surveillance System of Pakistan (KHDSS). *Online J Public Health Inform.* 2018;10(1). <https://doi.org/10.5210/ojphi.v10i1.8953>.
28. Lancaster G, Kariger P, McCray G, Janus M. Conducting a feasibility study in a global health setting for constructing a caregiver-reported measurement tool: an example in infant and young child development. London; 2020. Available from: <https://methods.sagepub.com/case/feasibility-study-global-health-caregiver-reported-measurement-tool-lycd>.
29. de Onis M, Garza C, Victora CG, Onyango AW, Frongillo EA, Martines J. The WHO Multicentre Growth Reference Study: planning, study design, and methodology. *Food Nutr Bull.* 2004;25(1 Suppl):S15–26. <https://doi.org/10.1177/15648265040251s103>.
30. Jones PC, Pendegast LL, Schaefer BA, Rasheed M, Svensen E, Scharf R, et al. Measuring home environments across cultures: invariance of the HOME scale across eight international sites from the MAL-ED study. *J Sch Psychol.* 2017;64:109–27.
31. Berens AE, Kumar S, Tofail F, Jensen SK, Alam M, Haque R, et al. Cumulative psychosocial risk and early child development: validation and use of the Childhood Psychosocial Adversity Scale in global health research. *Pediatr Res.* 2019;86(6):766–75.
32. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The Brief Resilience Scale: assessing the ability to bounce back. *Int J Behav Med.* 2008;15(3):194–200.
33. Dunst C, Jenkins V, Trivette CM. Family Support Scale: reliability and validity. *J Individ Fam Commun Wellness.* 1984;1:45–52.
34. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and case finding for major depressive disorder using the Patient Health Questionnaire (PHQ-9): a meta-analysis. *Gen Hosp Psychiatry.* 2015;37(6):567–76.
35. (WHO) WHO. Global Scales for Early Development v1.0: Adaptation and translation guide [English]. 2023. Available from: <https://apps.who.int/iris/bitstream/handle/10665/366278/WHO-MSD-GSEDpackage-v1.0-2023.9-eng.pdf>.
36. Wild D, Grove A, Martin M, Eremenco S, McElroy S, Verjee-Lorenz A, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR task force for translation and cultural adaptation. *Value Health.* 2005;8(2):94–104.
37. de Onis M, Onyango AW, Van den Broeck J, Chumlea WC, Martorell R. Measurement and standardization protocols for anthropometry used in the construction of a new international growth reference. *Food Nutr Bull.* 2004;25(1 Suppl):S27–36. <https://doi.org/10.1177/15648265040251s104>.
38. De Maio TJ, Rothgeb JM. Cognitive interviewing techniques: in the lab and in the field. Answering questions: methodology for determining cognitive and communicative processes in survey research. Hoboken, NJ: Jossey-Bass/Wiley; 1996. p. 177–95.
39. Willis G. Cognitive Interviewing. Thousand Oaks: SAGE Publications, Inc.; 2005. Available from: <https://methods.sagepub.com/book/mono/cognitive-interviewing/toc>.
40. Huynh J. Media Review: Qualitative and mixed methods data analysis using Dedoose: a practical approach for research across the social sciences. *J Mixed Methods Res.* 2021;15(2):284–6. <https://doi.org/10.1177/1558689820977627>.
41. Waldman MR, Raikes A, Hepworth K, Black MM, Cavallera V, Dua T, et al. Psychometrics of psychosocial behavior items under age 6 years: evidence from Nebraska, USA. *Infant Ment Health J.* 2024;45(1):56–78. <https://doi.org/10.1002/imhj.22090>.
42. Peña ED. Lost in translation: methodological considerations in cross-cultural research. *Child Dev.* 2007;78(4):1255–64. <https://doi.org/10.1111/j.1467-8624.2007.01064.x>.
43. Fernald LC, Prado E, Kariger P, Raikes A. A toolkit for measuring early childhood development in low and middle-income countries. 2017.
44. Frank MC, Sugarman E, Horowitz AC, Lewis ML, Yurovsky D. Using tablets to collect data from young children. *J Cogn Dev.* 2016;17(1):1–17. <https://doi.org/10.1080/15248372.2015.1061528>.
45. Organization WH. International ethical guidelines for biomedical research involving human subjects. International ethical guidelines for biomedical research involving human subjects. 1993. p. 63.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.