

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

## Protocol for Validation of the Global Scales for Early Development (GSED) for Children under 3 Years of Age in Seven Countries

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-062562
Article Type:	Protocol
Date Submitted by the Author:	04-Mar-2022
Complete List of Authors:	<p>Cavallera, Vanessa; World Health Organization, Department of Mental Health and Substance Use  Lancaster, Gillian; Keele University School of Medicine  Gladstone, Melissa; University of Liverpool Institute of Translational Medicine, Women and Children's Health  Black, Maureen; RTI International; University of Maryland School of Medicine  McCray, Gareth; Keele University School of Medicine  Nizar, Ambreen; The Aga Khan University, Department of Peadiatrics  Ahmed, Salahuddin; Projahnmo Research Foundation  Dutta, Arup; Center for Public Health Kinetics  Anago, Romuald ; Innovations for Poverty Action  Brentani, Alexandra; University of Sao Paulo, Pediatrics  Jiang, Fan; Shanghai Children's Medical Center Affiliated to Shanghai Jiaotong University School of Medicine  Schönbeck, Yvonne; Netherlands Organization for Applied Scientific Research, Department of Child Health  McCoy, Dana; Harvard Graduate School of Education  Kariger, Patricia; University of California Berkeley School of Public Health  Weber, Ann; University of Nevada Reno  Raikes, Abbie; University of Nebraska Medical Center College of Public Health  Waldman, Marcus; University of Nebraska Medical Center College of Public Health  van Buuren, Stef; TNO, Department of Statistics  Kaur, Raghbir; World Health Organization  Maillard, Pérez; World Health Organization  Nisar, Muhammad; The Aga Khan University, Pediatrics and Child Health  Khanam, Rasheda; Johns Hopkins University, International Center for Maternal and Newborn Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health  Sazawal, Sunil ; Centre for Public Health Kinetics  Zongo, Arsène; Innovations for Poverty Action  Pacífico Mercadante, Mariana; University of Sao Paulo, Pediatrics  Zhang, Yunting ; Shanghai Children's Medical Center Affiliated to Shanghai Jiaotong University School of Medicine  Roy, Arunangshu ; Projahnmo Research Foundation  Hepworth, Katelyn; University of Nebraska-Lincoln College of Education and Human Sciences</p>

	<p>Fink, Günther ; Schweizerisches Tropen- und Public Health-Institut, Household Economics and Health System Research Unit</p> <p>Rubio-Codina, Marta; Inter-American Development Bank</p> <p>Tofail, Fahmida; International Centre for Diarrhoeal Disease Research Bangladesh, Centre for Nutrition &amp; Food Security</p> <p>Eekhout, Iris; TNO, Child Health</p> <p>Seiden, Jonathan; Harvard Graduate School of Education</p> <p>Norton, Rebecca; World Health Organization</p> <p>Baqui, Abdullah; Johns Hopkins University Bloomberg School of Public Health</p> <p>Zhao, Jin; Shanghai Children's Medical Center Affiliated to Shanghai Jiaotong University School of Medicine</p> <p>Holzinger, Andreas; Innovations for Poverty Action</p> <p>Detmar, Symone; TNO</p> <p>Kembou, Samuel; Innovations for Poverty Action</p> <p>Begum, Farzana; The Aga Khan University, Department of Paediatrics</p> <p>Jehan, Fyezah; The Aga Khan University, Paediatrics and Child Health</p> <p>Dua, Tarun; World Health Organization</p> <p>Janus, Magdalena; McMaster University, Offord Centre for Child Studies</p>
Keywords:	<p>MENTAL HEALTH, Paediatric neurology &lt; NEUROLOGY, Developmental neurology &amp; neurodisability &lt; PAEDIATRICS, PUBLIC HEALTH, EPIDEMIOLOGY</p>

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## BMJ Open

**Title:** Protocol for Validation of the Global Scales for Early Development (GSED) for Children under 3 Years of Age in Seven Countries

**Authors:** Cavallera V, Lancaster G, Gladstone M, Black MM, McCray G, Nizar A, Ahmed S, Dutta A, Anago R, Brentani A, Jiang F, Schönbeck Y, McCoy D, Kariger P, Weber A, Raikes A, Waldman M, van Buuren S, Kaur R, Pérez Maillard M, Nisar I, Khanam R, Sazawal S, Zongo A, Pacifico Mercadante M, Zhang Y, Roy AD, Hepworth K, Fink G, Rubio Codina M, Tofail F, Eekhout I, Seiden J, Norton R, Baqui AH, Zhao J, Holzinger A, Detmar S, Kembou S N, Begum F, Jehan F, Dua T, Janus M.

**Affiliations:**

Corresponding author: Dr Vanessa Cavallera, MD, MPH, World Health Organization (WHO), Brain Health Unit in Department of Mental Health and Substance Use, Av. Appia 20, 1202 Geneva, Switzerland. [cavallerav@who.int](mailto:cavallerav@who.int)

Prof Gillian Lancaster, PhD, MSc, BSc (Hons), Clinical Trials Unit, School of Medicine, Keele University, UK

Prof Melissa Gladstone, MD, International Child Health and Neurodevelopmental Paediatrics, Department of Women and Children's Health, Institute of Translational Medicine, University of Liverpool, Alder Hey Children's NHS Foundation Trust, UK.

Prof Maureen Black, PhD, Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, USA and RTI International, Research Park, NC, USA

Dr Gareth McCray, MA, MRes, PhD, Research Associate, the School of Medicine, Keele University, Keele, UK.

Ambreen Nizar, BDS, MSc, Department of Paediatrics, Aga Khan University, Karachi Pakistan.

- 1  
2  
3 27 Salahuddin Ahmed, MBBS, Projahnmo Research Foundation, Dhaka, Bangladesh  
4  
5  
6 28 Arup Dutta, MBA, Center for Public Health Kinetics, New Delhi, India  
7  
8  
9 29 Romuald Kouadio E Anago, MA, BA, Innovations for Poverty Action (IPA), Abidjan,  
10  
11 30 Côte d'Ivoire  
12  
13  
14 31 Dr. Alexandra Valeria Maria Brentani, PhD, Dept. of Pediatrics, University of Sao Paulo  
15  
16 32 Medical School, Brazil  
17  
18  
19 33 Prof. Fan Jiang, MD, PhD, Department of Developmental and Behavioral Pediatrics,  
20  
21 34 National Children's Medical Center, Shanghai Children's Medical Center, affiliated to  
22  
23 35 School of Medicine Shanghai Jiao Tong University, China  
24  
25  
26 36 Dr. Yvonne Schönbeck, PhD, Msc, Dept. Child Health, Netherlands Organization for  
27  
28 37 Applied Scientific Research (TNO), Leiden, The Netherlands  
29  
30  
31 38 Dr Dana McCoy, PhD, Harvard Graduate School of Education, Cambridge, MA, USA  
32  
33  
34 39 Dr Patricia Kariger, PhD CEGA (Center for Effective Global Action), School of Public  
35  
36 40 Health, University of California, Berkeley, California, USA.  
37  
38  
39 41 Dr Ann Weber, PhD, MPH, School of Public Health, University of Nevada, Reno, NV,  
40  
41 42 USA  
42  
43  
44 43 Dr Hilary Abigail Raikes, MPH, PhD, University of Nebraska Medical Center, College of  
45  
46 44 Public Health, Omaha, NE, USA  
47  
48  
49 45 Dr. Marcus Waldman, PhD, University of Nebraska Medical Center, College of Public  
50  
51 46 Health, Omaha, NE, USA  
52  
53  
54 47 Prof Stef van Buuren, PhD, Department of Child Health, Netherlands Organization for  
55  
56 48 Applied Scientific Research TNO, Leiden, The Netherlands  
57  
58  
59  
60

- 1  
2  
3 49 Raghbir Kaur, MPH, DMD, MS, Consultant World Health Organization (WHO), Brain  
4  
5 50 Health Unit in Department of Mental Health and Substance Use, Geneva, Switzerland  
6  
7 51 Michelle Pérez Maillard, MSc, Consultant World Health Organization (WHO), Brain  
8  
9 52 Health Unit in Department of Mental Health and Substance Use, Geneva, Switzerland  
10  
11  
12 53 Prof. Muhammad Imran Nisar, MBBS, MSc, Department of Peadiatrics, Aga Khan  
13  
14 54 University, Karachi Pakistan.  
15  
16  
17 55 Dr. Rasheda Khanam, MBBS, MPH, PhD, International Center for Maternal and  
18  
19 56 Newborn Health, Department of International Health, Johns Hopkins University  
20  
21 57 Bloomberg School of Public Health, Baltimore, MD, USA  
22  
23  
24 58 Dr. Sunil Sazawal, MBBS, MPH, PhD, Center for Public Health Kinetics New Delhi,  
25  
26 59 India  
27  
28  
29 60 Mr. Arsène Zongo, MA, BA, Innovations for Poverty Action (IPA), Abidjan, Côte  
30  
31 61 d'Ivoire  
32  
33  
34 62 Dr. Mariana Pacifico Mercadante, MD, Dept. of Pediatrics, University of Sao Paulo  
35  
36 63 Medical School, Brazil  
37  
38  
39 64 Dr. Yunting Zhang, PhD, Child Health Advocacy Institute, National Children's Medical  
40  
41 65 Center, Shanghai Children's Medical Center, affiliated to School of Medicine Shanghai  
42  
43 66 Jiao Tong University, China  
44  
45  
46 67 Dr. Arunangshu D. Roy, MBBS, Projahnmo Research Foundation, Dhaka, Bangladesh  
47  
48  
49 68 Katelyn Hepworth, MA, University of Nebraska-Lincoln, College of Education and  
50  
51 69 Human Sciences, NE, USA  
52  
53  
54 70 Prof Günther Fink, PhD, Swiss Tropical and Public Health Institute, Basel, Switzerland  
55  
56  
57  
58  
59  
60

- 1  
2  
3 71 Dr Marta Rubio-Codina, PhD, Inter-American Development Bank, Washington, D.C.,  
4  
5 72 USA  
6  
7  
8 73 Dr. Fahmida Tofail, MBBS, PhD, International Centre for Diarrhoeal Disease Research  
9  
10 74 Iccdr,b, Dhaka, Bangladesh  
11  
12  
13 75 Dr. Iris Eekhout, PhD, MSc, Department of Child Health, Netherlands Organization for  
14  
15 76 Applied Scientific Research TNO, Leiden, The Netherlands  
16  
17  
18 77 Jonathan Seiden, EdM, Harvard Graduate School of Education, Cambridge, MA, USA  
19  
20  
21 78 Rebecca Norton, Consultant World Health Organization (WHO), Brain Health Unit in  
22  
23 79 Department of Mental Health and Substance Use, Geneva, Switzerland  
24  
25  
26 80 Dr. Abdullah H. Baqui, MBBS, DrPH, Johns Hopkins Bloomberg School of Public  
27  
28 81 Health, Baltimore, Maryland, USA  
29  
30  
31 82 Dr. Jin Zhao, MD, PhD, Department of Developmental and Behavioral Pediatrics,  
32  
33 83 National Children's Medical Center, Shanghai Children's Medical Center, affiliated to  
34  
35 84 School of Medicine Shanghai Jiao Tong University, China  
36  
37  
38 85 Andreas Holzinger, MA, BA, Innovations for Poverty Action (IPA), Abidjan, Côte  
39  
40 86 d'Ivoire  
41  
42  
43 87 Dr. Symone Detmar, PhD, Msc, Department of Child Health, Netherlands Organization  
44  
45 88 for Applied Scientific Research TNO, Leiden, The Netherlands  
46  
47  
48 89 Dr, Samuel Nzale Kembou, PhD, MA, BA, Innovations for Poverty Action (IPA),  
49  
50 90 Abidjan, Côte d'Ivoire  
51  
52  
53 91 Farzana Begum, MSc, Department of Peadiatrics, Aga Khan University, Karachi  
54  
55 92 Pakistan.  
56  
57  
58  
59  
60

1  
2  
3 93 Prof. Fyezah Jehan, MBBS, FCPS, MSc, Department of Peadiatrics, Aga Khan  
4

5 94 University, Karachi Pakistan.  
6  
7

8 95 Dr Tarun Dua, MD, World Health Organization (WHO), Brain Health Unit in  
9

10 96 Department of Mental Health and Substance Use, Geneva, Switzerland  
11  
12

13 97 Prof Magdalena Janus, MBBS, FCPS, MSc, Offord Centre for Child Studies, McMaster  
14

15 98 University, Hamilton ON, Canada.  
16  
17

18 99 **Disclaimers:**

20 100 The author is a member of the World Health Organization. The author alone is  
21  
22 101 responsible for the views expressed in this publication and they do not necessarily  
23  
24 102 represent the decisions, policy or views of the World Health Organization. [Applies to  
25  
26 103 Cavallera V, Dua T, Kaur R, Pérez Maillard M and Norton R]  
27  
28

29 104 The views here presented do not represent the Inter-American Development Bank, its  
30  
31 105 board of directors, or the countries it represents. [Applies to Rubio Codina M]  
32  
33

34 106 **Wordcount:** 4000  
35

36  
37 107 **Keywords:** Early childhood development, global, measurement, index, population,  
38  
39 108 evaluation, validation, protocol, programmes.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 109 **ABSTRACT**  
4  
5

6 110 **Introduction.** Children's early development is affected by caregiving experiences, with  
7  
8 111 life-long health and wellbeing implications. Governments and civil societies need  
9  
10 112 population-based measures to monitor children's early development and ensure that  
11  
12 113 children receive the care needed to thrive. To this end, the World Health Organization  
13  
14 114 (WHO) developed the Global Scales for Early Development (GSED) to measure  
15  
16 115 children's early development (ages 0-3 years). The GSED includes three measures: 1)  
17  
18 116 short form (SF) for population-evaluation (caregiver-report), 2) complementary long  
19  
20 117 form (LF) for programmatic-evaluation (direct assessment), and 3) psychosocial form  
21  
22 118 (PF) for psychosocial development evaluation (caregiver-report). The primary aim of this  
23  
24 119 protocol is to validate the GSED SF and LF. Secondary aims are to create preliminary  
25  
26 120 reference scores for the GSED SF and LF, validate an adaptive testing algorithm, and  
27  
28 121 assess the feasibility and preliminary validity of the GSED PF.  
29  
30

31  
32  
33 122 **Methods and Analysis.** We will conduct the validation in seven countries varying in  
34  
35 123 geography, language, culture and income through a one-year prospective design,  
36  
37 124 combining cross-sectional and longitudinal methods with 1248 children per site, stratified  
38  
39 125 by age and sex. The GSED generates an innovative common metric (Development-score:  
40  
41 126 D-score) using the Rasch model and a development-for-age z-score (DAZ). We will  
42  
43 127 evaluate six psychometric properties of the GSED SF and LF: concurrent validity,  
44  
45 128 predictive validity at six months, convergent and discriminant validity, and test-retest and  
46  
47 129 inter-rater reliability. We will evaluate measurement invariance by comparing differential  
48  
49 130 item functioning (DIF) and differential test functioning (DTF) across sites.  
50  
51

52  
53  
54 131 **Ethics and dissemination.** This study has received ethical approval from the WHO  
55  
56 132 (protocol GSED validation 004583 20.04.2020) and approval in each site. Study results  
57  
58  
59  
60

1  
2  
3 133 will be disseminated through webinars and publications from WHO, international  
4  
5 134 organisations, academic journals, and conference proceedings.  
6  
7

8 135 **Registration details:** Open Science Framework (OSF) <https://osf.io/> on 19/11/2021 (DOI  
9  
10 136 10.17605/OSF.IO/KX5T7; identifier: osf-registrations-kx5t7-v1)  
11  
12

## 13 137 **ARTICLE SUMMARY**

### 14 138 **Strengths and limitations of this study**

- 15  
16  
17 139 • The study collects validation data (n = 8736 children) for the Global Scales for  
18  
19  
20 140 Early Development (GSED) in seven countries that vary in geographic, linguistic,  
21  
22 141 cultural and sociodemographic characteristics. The sampled populations are  
23  
24 142 chosen to be diverse and are not nationally representative.  
25  
26 143 • The methods for the validation of GSED are systematic across sites and follow  
27  
28  
29 144 rigorous standard operating procedures based on the best scientific evidence  
30  
31 145 available.  
32  
33 146 • A tablet-based App is used for data collection to make the administration of the  
34  
35  
36 147 GSED measures user-friendly, to reduce recording and transcribing errors, and to  
37  
38 148 facilitate adaptive testing.  
39  
40 149 • The GSED SF and LF aims to include items that are culturally neutral and fit the  
41  
42  
43 150 Rasch model, which assumes that child development milestones are age-ordinal,  
44  
45 151 to create D-scores. Psychosocial items are included in a separate measure (GSED  
46  
47 152 psychosocial form [PF]) and cultural-specific items can be supplemented by  
48  
49 153 countries.  
50  
51 154 • The three secondary aims (preliminary reference scores, an adaptive testing  
52  
53  
54 155 algorithm, and the feasibility and validity of the GSED PF), are exploratory and  
55  
56 156 will require further research.  
57  
58  
59  
60

## 157 INTRODUCTION

158 Prenatal and early postnatal experiences have significant impacts on early childhood  
159 development (ECD) and can influence the accrual of health, well-being, and productivity  
160 throughout the life-course (1). To promote current and sustainable peace and prosperity,  
161 the United Nations has focused the Sustainable Development Goals (SDG) on improving  
162 children's outcomes in the early years through multiple targets. The most explicit target  
163 for young children is SDG 4 (Education goal), which requires reporting on the  
164 "proportion of children under 5 years of age who are developmentally on track in health,  
165 learning and psychosocial well-being, by sex" (2).

166 There are few valid measures that can be used globally to assess child development for  
167 children under three years of age. Current measures of ECD range from proxy measures  
168 (e.g., prevalence of country-level stunting and poverty) to detailed measures of individual  
169 performance on developmental tasks (3). The Early Childhood Development Index 2030  
170 (ECDI 2030) (4) does not include children below two years of age. A recent review has  
171 identified the creation and validation of population-based instruments for assessing very  
172 young children as a global priority (5).

173 The Global Scales for Early Development (GSED) build on advances made by analyses  
174 of existing global datasets (6), and new data collection (7) that demonstrated the cross  
175 cultural applicability of items that measure young children's development. Three research  
176 teams (8) joined efforts to develop the GSED in response to the pressing need for  
177 instruments and metrics to measure ECD at population and programmatic levels across  
178 diverse parts of the world.

## 179 **The Global Scales for Early Development (GSED)**

180 The GSED consist of three open-access measures developed by a WHO-led team<sup>1</sup> to  
181 provide a standardized methodology for measuring the development of children aged 0-3  
182 years (0-36 months) across diverse cultures and contexts (9, 10). They are developed for  
183 three objectives: 1) for population-level evaluation based on caregiver-report, GSED  
184 Short Form (SF); 2) for programmatic evaluation in combination with SF, direct child  
185 assessment, GSED Long Form (LF); and 3) for measuring psychosocial behaviours,  
186 caregiver-reported GSED Psychosocial Form (PF). The development and piloting of the  
187 GSED SF, LF, and PF are described elsewhere (9).

188 The GSED SF and LF produce metrics on the same age-ordinal scale and quantify the  
189 same latent construct. The Developmental Score (D-score) (see Box 1) underlies both  
190 measures and reflects children's overall development across multiple domains typically  
191 demonstrated in this age group (e.g., cognitive, motor, language, social-emotional) (6).

192 The GSED PF items, designed to measure non-normative developmental patterns,  
193 including behavioural or regulatory challenges, are not age-ordinal and do not use the D-  
194 Score metric.

### **Box 1: The Developmental score**

The Developmental score (15), or D-score, is a unidimensional latent variable measuring child development during the first three years across multiple domains. The milestones that make up the D-score conform to the Rasch model (25), thus yielding a scale with interval properties with a fixed unit (Figure 1). It is therefore possible to calculate a meaningful difference between two D-scores. Similar to height-for-age Z-score, given suitable age-conditional references, the D-score can be transformed to a Z-

<sup>1</sup> The full team and contributors are listed in the Acknowledgments.

score that accounts for children's age (i.e., Development for Age Z-score, or DAZ).

The DAZ facilitates comparisons across children of different ages.

## 195 AIMS

196 The primary aim of this study is to validate the GSED measures (11), through testing for  
197 measurement invariance and evaluation of the psychometric properties to measure  
198 development among children aged 0-3 years (0-36 months) globally (including creation  
199 of D-scores and Development for Age Z-score [DAZ]).

200 Specific Objectives:

- 201 a) Fit a Rasch model to the item data to calculate the D-scores and DAZ.
- 202 b) Investigate differential item functioning (DIF) and differential test functioning  
203 (DTF) across sites to determine measurement invariance.
- 204 c) Examine psychometric properties of the GSED SF and LF:
  - 205 • Test-retest and inter-rater reliability (score and item level),
  - 206 • Concurrent validity (association between scores on GSED and Bayley Scales  
207 of Infant and Toddler Development (Bayley-III) or Griffiths Scales of Child  
208 Development administered concurrently) (12),
  - 209 • Convergent validity (strength of association between GSED D-scores and  
210 other theoretically-relevant constructs)
  - 211 • Predictive validity (association between GSED scores six months after initial  
212 assessment).

213 The secondary aims are to: 1) establish preliminary reference scores<sup>2</sup> for optimal  
214 development on the D-score (GSED SF and LF), 2) develop and validate an adaptive

---

<sup>2</sup> The population has not been selected as a representative sample of all children aged 0-3 in each site (as would happen in a countrywide population census). Selection and recruitment of a representative sample

1  
2  
3 215 testing algorithm, and 3) obtain preliminary validity data on the psychometric properties  
4  
5 216 of the GSED PF.  
6  
7

## 8 9 217 **METHODS**

### 10 11 218 **Design and study sites**

12  
13 219 The GSED validation study uses a prospective cross-sectional design with a longitudinal  
14  
15 220 component of age and sex stratified samples of children in seven countries. The countries  
16  
17 221 are culturally, linguistically and geographically diverse, representing low-income  
18  
19 222 (Bangladesh, Côte d'Ivoire, Pakistan, United Republic of Tanzania), middle-income  
20  
21 223 (Brazil and The Republic of China), and high-income (The Netherlands) settings.  
22  
23 224 Samples in each site are not nationally representative; however, they are diverse, e.g.  
24  
25 225 covering both rural/urban settings.  
26  
27

28  
29 226 Preparation and feasibility phases are described elsewhere (11), and assess feasibility of  
30  
31 227 administration of GSED and associated measures including processes for translating and  
32  
33 228 culturally adapting GSED and other study measures, creating data management systems,  
34  
35 229 and training teams in data collection procedures.  
36  
37

### 38 39 230 **Patient and Public involvement**

40  
41 231 Participants were involved in the study design as the burden of the assessment was  
42  
43 232 discussed with them in a pilot stage through qualitative data collection. We intend to  
44  
45 233 disseminate the main results to trial participants and will seek patient and public  
46  
47 234 involvement in the development of an appropriate method of dissemination.  
48  
49  
50  
51  
52  
53  
54  
55  
56

---

57 was beyond the scope of this study and not required for validation purposes. We are therefore developing  
58 'reference scores' which should not be interpreted as population-sampled norms.  
59  
60

235 **Study sample**

236 The study sample includes children between 0 to 41 months of age (inclusive) living in  
 237 study areas (see Table 1 for inclusion and exclusion criteria).

238 **Table 1. Study sample inclusion and exclusion criteria**

Sample	Inclusion criteria	Exclusion criteria
Total per site N=1248 (as described in sample size section below)	1. Age 0-41 months 2. Family speaks to the child in same language as GSED translation 3. Primary caregiver available to participate	1. Missing gestational age (ultrasound or last menstrual period LMP) 2. Missing birth weight data 3. Acutely unwell at time of assessment (temporary exclusion: to be rescheduled after 7 days)

240 ***Recruitment and consent***

241 In each site the sampling frame consists of a list of potentially eligible caregiver-child  
 242 dyads residing in the defined study area. Lists of potential participants vary by site and  
 243 may include: participants in local pregnancy surveillance systems, families who have  
 244 previously agreed to be contacted for participation, birth registries from hospital/health  
 245 centers, or families with children attending local child health/care centers.  
 246 Eligible children are sampled from this list using the GSED sampling scheme (Figure 2).  
 247 To minimize clustering of correlated scores within households, one child per caregiver  
 248 and in multi-family household is selected, guided by age and sex quotas. For siblings or  
 249 twins, one is chosen randomly. Target children's primary caregiver (person most familiar  
 250 with the child and spends most time with them) is approached for consent and enrollment.  
 251 A non-technically worded information sheet is shared and consent to participate is  
 252 obtained at first visit. In the Netherlands, participants provide consent online, confirmed  
 253 by study staff at first visit. Refusals to participate and dropouts are registered and  
 254 replaced.

1  
2  
3 255 ***Sampling frame and schemes***  
4

5 256 Sample size for recruitment within each site is 1248 children (total 8736 children) across  
6  
7 257 seven countries. After consent is provided, children are allocated by sex and age groups  
8  
9  
10 258 using a randomization procedure to one of several sampling schema (e.g., predictive,  
11  
12 259 reference-score, reliability; Figure 2). See sampling Table S1 in Supplementary file S1  
13  
14 260 for sampling frame. Out of the full site sample of 1248 children, 504 children per site are  
15  
16  
17 261 randomly selected for re-evaluation 6 months later to assess predictive validity (primary  
18  
19 262 aim). A second scheme indicates the minimum sub-sample of children needed to  
20  
21 263 calculate preliminary reference scores (secondary aim) that will facilitate cross-country  
22  
23 264 comparisons. To maximize precision of parameter estimates, larger quotas are kept for  
24  
25 265 the youngest age brackets where rates of development are accelerated. A third scheme  
26  
27 266 addresses inter-rater reliability for 90 children per site using two assessors who  
28  
29 267 independently assess the same child sequentially or within 24 hours (12)<sup>3</sup>. Test-retest  
30  
31 268 (intra-rater reliability) is performed by inviting 50 children per site to return for repeat  
32  
33 269 assessment with the same rater within 7-10 days.  
34  
35 270 In the Netherlands, the GSED SF and PF are administered online. A subset of participants  
36  
37 271 (N=32) are interviewed face-to-face to compare method of administration. To determine  
38  
39 272 test-retest reliability (intra-rater reliability), the primary caregiver completes the SF and  
40  
41 273 PS form online and then a second time 7-10 days later.  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54

---

55  
56 <sup>3</sup> We note that this procedure differs from typical inter-rater reliability (IRR) designs which involve  
57 simultaneous scoring of a single assessment. This sequential design was necessitated by logistical  
58 constraints. Given that this design captures both variance due to differences in raters *and* differences in  
59 occasions, the observed IRR represents a lower bound for the true inter-rater reliability of the assessments.  
60



1  
2  
3 274 **Data collection**  
4

5  
6 275 *Measures*  
7

8  
9 276 *GSED*  
10

11  
12 277 **GSED SF and LF.** The creation of the GSED SF and LF is described elsewhere (10).  
13

14 278 Briefly, we constructed an item bank from previously gathered data and compiled cross-  
15

16 279 sectional and longitudinal data from 31 countries representing over 73,000 anonymized  
17

18 280 children with 109,079 assessments (using 22 established ECD instruments) (6, 13, 14).  
19

20 281 Using subject matter expert input and statistical modelling (15), we developed a  
21

22 282 caregiver-reported measure intended to capture child development at population-level  
23

24 283 (GSED SF), and a complementary direct-assessment measure to evaluate programmatic  
25

26 284 impacts (GSED LF) (10). The measures are created paper-based and app-based (GSED  
27

28 285 App) with built-in administration rules and supporting media-files (see below).  
29

30 286 The GSED SF includes 139 items representing emerging skills and behaviors within  
31

32 287 cognitive, motor, language and social-emotional domains. All items are presented as  
33

34 288 questions to the caregiver, with binary response options (Yes/No and “Don’t Know”) that  
35

36 289 use start rules based on the child’s age, and stop rules based on age and performance.  
37

38 290 Assessors record caregiver’s responses, regardless of the assessor’s observations. In the  
39

40 291 Netherlands only, the GSED SF is completed online by caregivers. The GSED SF  
41

42 292 administration includes sounds, images, and short video clips that assist in understanding,  
43

44 293 interpreting and administering the items.  
45

46 294 The GSED LF includes 155 items capturing similar domains to the SF but, observed by  
47

48 295 the assessor following start and stop rules based on the child’s age and responses. Items  
49

50 296 are organized into three grids (A, B and C) that enable assessors to measure the child’s  
51

52 297 performance on similar tasks in succession, making the administration easier for both  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 298 assessors and children. To further facilitate administration, icons are placed next to each  
4  
5 299 item that inform the assessor whether the item is observed, demonstrated to or by the  
6  
7 300 child, listened for or spoken to the child. The GSED LF uses a locally constructed and  
8  
9 301 low cost kit with basic materials that the child interacts with to demonstrate abilities. The  
10  
11 302 kit is created by local teams with detailed guidance from WHO. Responses of all LF  
12  
13 303 items are binary (skill observed/not observed).

14  
15 304 The items in both measures are ordered by difficulty reflecting children's emerging skills.  
16  
17 305 Based on the analyses from the validation, we will select the items to be included in the  
18  
19 306 final GSED SF and LF versions available for use.

20  
21 307 **Psychosocial Form (PF).** Unlike the SF and the LF, the GSED PF has been developed to  
22  
23 308 index non-normative developmental patterns that provide a window into early  
24  
25 309 manifestations of children's mental health challenges, including internalizing and  
26  
27 310 externalizing behaviour problems and dysregulation (e.g., eating and sleeping). Items  
28  
29 311 capturing developmentally normative information about socio-emotional competencies  
30  
31 312 are included in the GSED SF and LF, as the SDG 4.2 includes children's psychosocial  
32  
33 313 well-being. Because few instruments have been developed to capture psychosocial  
34  
35 314 difficulties for children under 3 years, little existing data are available and the  
36  
37 315 development of the GSED PF is exploratory. The PF initial prototype was created  
38  
39 316 through a review of existing measures of infant and toddler mental health and consensus  
40  
41 317 by subject matter experts. The GSED PF includes 47 items and reflects caregiver  
42  
43 318 perceptions of the behaviors' frequency, using response options: Often; Sometimes;  
44  
45 319 Never/almost never. Items are divided into two age groups: 0-6 and 6-36 months.

#### 320 *Contextual and demographic measures*

321 In addition to the GSED, the validation study includes measures of children's growth and  
322 nutrition, health, environmental and contextual information (see Table 2 for measures and

323 sources). The selection of measures was based on known biological and social  
 324 determinants of development (16), the demonstrated validity of the contextual measures  
 325 in at least one low- and middle-income country (LMIC), and efficiency for data  
 326 collection. See Supplementary File S2 for visit schedules (Tables S2a and S2b).  
 327 In three sites (Côte d'Ivoire, The Netherlands and The Republic of China) where  
 328 administration of the Home Observation for Measurement of the Environment Inventory  
 329 (HOME) is not feasible, household stimulation data and caregiver-child activities are  
 330 collected using Family Care Indicators (FCI). In all sites, a concurrent measure of child  
 331 development (Bayley-III or Griffiths Mental Development Scales) is administered in a  
 332 subsample of children to determine concurrent validity of GSED to a well-established  
 333 measure of the same construct.

334 **Table 2: Study measures in addition to GSED**

Construct	What the Measure Captures	Measure	Administration Mode	Time for Administer (Minutes)
Child health and household socioeconomic status (SES)	<ul style="list-style-type: none"> <li>• Eligibility (exclusion - criteria)</li> <li>• Demographic information</li> <li>• Information about acute child health</li> <li>• Delivery and Perinatal conditions</li> <li>• Breastfeeding</li> <li>• Child's health history</li> <li>• Household socio-economic status*</li> <li>• Caregiver education</li> <li>• Maternal health/ chronic illness</li> <li>• COVID-19 exposure</li> </ul>	Eligibility and Contextual Form [Specifically developed for the study]	Caregiver Report	35
Anthropometry	<ul style="list-style-type: none"> <li>• Weight at time of assessment</li> </ul>	Anthropometry Form	Child Assessment	15

	<ul style="list-style-type: none"> <li>• Infant Length/ Child Height at time of assessment</li> <li>• Child's Mid-upper arm circumference at time of assessment</li> <li>• Child's head circumference at time of assessment</li> </ul>			
<i>Family / home environment</i>	<ul style="list-style-type: none"> <li>• Home Environment (HOME only)</li> <li>• Play/ Stimulation / linteractions between the child and other family members in the home (HOME and FCI)</li> </ul>	Home Observation for Measurement of the Environment Inventory (HOME) (26) OR Family Care Indicators (FCI) (27) †	<i>HOME</i> : Caregiver report & Observation <i>FCI</i> : Caregiver Report	<i>HOME</i> : 45 <i>FCI</i> : 15
	<ul style="list-style-type: none"> <li>• Child neglect/abuse</li> <li>• Exposure to violence or conflict</li> </ul>	Childhood Psychosocial Adversity Scale (CPAS)(28) †	Caregiver Report	15
	<ul style="list-style-type: none"> <li>• Family resilience</li> </ul>	Brief Resilience Scale (BRS) (29) †	Caregiver Report	1
	<ul style="list-style-type: none"> <li>• Family social support</li> </ul>	Family Support Scale (FSS) (30) †	Caregiver Report	5
Caregiver health and well-being	<ul style="list-style-type: none"> <li>• Caregiver Depressive Symptoms</li> </ul>	The Patient Health Questionnaire-9 (PHQ-9) (31)	Caregiver Report	5
Child development	<ul style="list-style-type: none"> <li>• Global child development (0-41 months)</li> </ul>	Bayley Scales of Infant and Toddler Development (Bayley-III) (32) OR Griffiths Mental Development Scales(33) ‡	Direct child Assessment	45-60

	<ul style="list-style-type: none"> <li>Global child development (24-41 months)</li> </ul>	Early Childhood Development Index 2030 (ECDI2030) (4) <sup>§</sup>	Caregiver Report	10
--	---	--	------------------	----

335 \* Socioeconomic information on this form comes from the standard DHS multiple assets  
 336 index; however, some sites have adapted the socio-economic status items to better fit  
 337 their contexts

338 † these measures have been minorly adapted for the purpose of the study

339 ‡ in a sub-sample (N=150)

340 § in a sub-sample (all children of 24 to 41-months within the predictive validity  
 341 subsamples in three countries)

342

### 343 ***Schedule***

344 Data collection is scheduled over one to three visits depending on the study site to  
 345 accommodate rules of measure administration order and location. The first  
 346 administration of the GSED SF and PF is completed at home (or online in the  
 347 Netherlands) to test it in the setting intended for future use (e.g., Multiple Indicator  
 348 Cluster Surveys MICS or Demographic and Health Surveys DHS) and prior to  
 349 administration of the GSED LF. The GSED LF is administered in a controlled  
 350 environment (e.g., clinic) to match the required concurrent validity testing protocols. For  
 351 the concurrent validation, the GSED and concurrent measures are administered in the  
 352 same location on different days and counter-balanced in order of administration.

### 353 ***Training and Quality Control***

354 Training of local master trainers is performed by the WHO team for the GSED SF, PF  
 355 and LF, using slide presentations, discussion forums, audio-visual aids, and practice  
 356 exercises. Local master trainers are responsible for training local field teams using  
 357 materials adapted and translated to local languages. Reliable administration of the GSED  
 358 measures must be met (inter-rater agreement with a master trainer of  $\geq 90\%$ ) for  
 359 certification.

1  
2  
3 360 To ensure quality assurance, 10% of all the study visits are observed by the study  
4  
5 361 supervisor in person (or through video-recording in the Netherlands), covering each child  
6  
7 362 age band and certified assessors. Supervisors independently complete questionnaires  
8  
9 363 being administered by the assessor and complete a fidelity checklist. Assessors are given  
10  
11 364 feedback based on checklist score. Supervisors review quality assurance findings with the  
12  
13 365 WHO biweekly, along with discussions with the subject matter experts for further  
14  
15 366 resolution, as needed.

16  
17 367 The GSED application software for data collection has built-in data range and  
18  
19 368 consistency checks. Data managers review and resolve issues daily in consultation with  
20  
21 369 the local field team and/or WHO team.

### 22 370 **Sample size**

23 371 Sample size determination was based on the primary aim of assessing the psychometric  
24 372 properties of the GSED. To have sufficient power to estimate measurement parameters  
25 373 (abilities and difficulties) needed to calculate the D-score and DAZ scores at baseline and  
26 374 to detect DIF of 1 logit with a power of  $1 - \beta = 0.90$  and a two-sided significance level  
27 375 of  $\alpha = 0.05$ , a sample of  $N=1248$  per site is required. This sample size was calculated via  
28 376 optimization of the sample size at i) each age/sex stratum and ii) overall on 1000  
29 377 simulated datasets generated from parameters suggested by the Rasch GSED model. See  
30 378 Supplementary file S1 for additional details.

### 31 379 **Statistical Analysis**

32 380 To construct the scores for the GSED SF and LF, a Rasch model will be fitted and the  
33 381 item fit statistics (infit and outfit) will be assessed. Any items with unacceptable fit  
34 382 levels will be removed. Items will be screened for whether they exhibit unacceptable  
35 383 levels of measurement non-invariance (i.e., they have approximately equal difficulties)  
36 384 across countries and other contextual variables. Items exhibiting unacceptable DIF (using

1  
2  
3 385 the logistic regression method) will be discarded sequentially, and the item response  
4  
5 386 models will be refit using the remaining items. The expected a posteriori (EAP) method  
6  
7 387 (17) will be applied to the final model to estimate the latent ability parameter (the D-  
8  
9 388 score). Systematic deviations from unidimensionality will be tested by performing a  
10  
11 389 principal components analysis on the residuals of the Rasch model. The ability estimates  
12  
13 390 will be used to estimate preliminary developmental percentile curves against age using a  
14  
15 391 Generalized Additive Model for Location Scale and Shape (GAMLSS). Following  
16  
17 392 previous methodology (18) software will be written to calculate DAZ-scores based on the  
18  
19 393 final dataset in R, and a user-friendly front-end version created in R (ShinyApp) and/or  
20  
21 394 Excel.  
22  
23  
24  
25  
26 395 Reliability (inter-rater and test-retest) for all GSED measures will be analyzed using ICC  
27  
28 396 (at the score level) and Gwet's AC1 agreement (at the item level) statistics with 95%  
29  
30 397 confidence intervals to determine whether items perform reliably within and between  
31  
32 398 assessors (19). A cut-off value of 0.4 and above will be used to flag items as adequately  
33  
34 399 reliable. Those items with agreement between 0.4 and 0.5 will be discussed to determine  
35  
36 400 if modifications can be made to improve their administration and/or comprehension.  
37  
38  
39  
40  
41 401 DAZ scores from the GSED SF and LF will be used to conduct validity analyses to  
42  
43 402 ensure that the measures are capturing the construct they are purported to measure  
44  
45 403 (construct validity). Concurrent validity will be assessed by correlating age-corrected  
46  
47 404 Bayley-III or Griffiths Mental Development Scales scores with GSED DAZ scores. We  
48  
49 405 anticipate that these scores will have low to moderate positive correlations. Convergent  
50  
51 406 validity will be supported by statistically significant positive correlations (with 95%  
52  
53 407 confidence interval) between the GSED scores and continuous contextual measures with  
54  
55 408 prior evidence of association with child development. Comparisons between "known  
56  
57 409 groups" will be made using the following variables: maternal education, home learning  
58  
59  
60

1  
2  
3 410 opportunities, home environment, socioeconomic status (SES), maternal mental health  
4  
5 411 and child anthropometry, and stunting to determine if scores discriminate between high  
6  
7 412 and low categories for each variable using mean DAZ scores.  
8  
9  
10 413 GSED scores at baseline and follow up will be correlated for predictive validity (positive  
11  
12 414 association between baseline and at 6 months) and mixed-effects linear regression used to  
13  
14 415 adjust for other contextual covariates and baseline scores.  
15  
16  
17

## 18 416 **Secondary (Exploratory) Aims**

### 19 417 *Reference Scores*

20  
21 418 We plan to develop a set of preliminary reference scores to facilitate comparison of DAZ  
22  
23 419 scores across countries. From the full validation study sample, a sub-sample of children  
24  
25 420 who have not experienced prior exposure to major known biological and environmental  
26  
27 421 risk factors is selected (i.e., “reference sub-sample”) (Table 3). Such an approach relies  
28  
29 422 on the assumption that the attainment of basic developmental milestones captured by the  
30  
31 423 GSED of children who are free of major risk factors is relatively similar globally (20).  
32  
33 424 To develop the reference scores, we will fit GAMLSS (21) to flexibly model both  
34  
35 425 conditional means, conditional standard deviations of scores, and, if necessary,  
36  
37 426 conditional skewness and kurtosis. We will test our assumption that the distribution of  
38  
39 427 scores is equivalent across sites by adding a site indicator at each moment of the  
40  
41 428 distribution, and testing site effects for their statistical significance. Where possible, we  
42  
43 429 will conduct standardization of scores to assist with the interpretation of scores by  
44  
45 430 pooling data across countries. We will report the corresponding parameters of the  
46  
47 431 GAMLSS model at appropriate ages.  
48  
49  
50  
51  
52  
53  
54  
55

56 432 **Table 3. “Reference” sub-sample exclusion criteria (healthy sub-sample)**

Sample	Exclusion criteria
--------	--------------------



<p>Minimum sub-sample of “reference” children per site N=522</p>	<ol style="list-style-type: none"> <li>1. Below secondary maternal education (&lt;6 years of schooling)</li> <li>2. Birthweight less than 2500 gr</li> <li>3. Gestational age &lt; 37 completed weeks (259 days) and <math>\geq</math> 42 completed weeks (294 days) [assessed by ultrasound]</li> <li>4. Undernutrition (weight for age, length for age, OR weight for height Z score of less than <math>-2</math> on the WHO Child Growth Standards) at the time of developmental assessment</li> <li>5. Known severe congenital birth defect</li> <li>6. History of birth asphyxia OR neonatal sepsis requiring hospitalization</li> <li>7. Known neurodevelopmental disorder/ disability (Severe visual problems, seizures, hearing impairment) OR other chronic health problems (that is congenital heart disease)</li> </ol>
--	--

433

434 ***Adaptive testing***

435 We will determine whether adaptive testing is a feasible and valid option to measure  
 436 child development within the GSED (Box 2). Adaptive testing (22) is an administration  
 437 method that continually adapts to the level of the child’s performance, thereby reducing  
 438 test administration time. Previous simulations (23) indicated that theoretically substantial  
 439 gains in the precision of scores are possible when using adaptive testing even if  
 440 administering fewer items.

**Box 2: Adaptive testing validation methodology**

We investigate the feasibility by applying adaptive testing in addition to the traditional “fixed” GSED administration methods in the sub-sample designated for predictive validity analyses (N=502 per site) in three sites. The adaptive test is executed using tablets that are specially programmed to continually adjust child’s score after each item is administered, and to suggest the next item based on the answers already received (e.g., a more difficult item for a child with a higher score, an easier item for a child with a lower score). Once the program establishes a reliable score, the administration is

1  
2  
3 terminated. Both the adaptive test and the fixed test are administered with the same  
4 sub-sample during two separate visits alternating the order of administration to  
5 investigate the difference between the two modes of administration. We will  
6 investigate the following: the variance of user experience as a function of the average  
7 difficulty of milestones (leniency); the comparison of the D-score distribution under  
8 the adaptive testing procedure with the D-score distribution under the fixed GSED  
9 administration (using a z-test to assess the equivalence of the two modalities and  
10 plotting the results to show the level of concordance); and relation of the difference  
11 between the two D-scores to background variables.  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23

441

#### 442 ***Psychosocial Form***

443 The PF measure is in an early stage and will undergo exploratory and confirmatory factor  
444 analyses to assess the internal scale structure. Associations between items and factor  
445 scores with variables suggesting a high risk of psychosocial stress, such as family  
446 resilience, social support, and family and community violence, in addition to GSED SF  
447 and LF scores (concurrent validity measures) will be examined.

#### 448 **ETHICS AND DISSEMINATION**

449 The study complies with the International Ethical Guidelines for Biomedical Research  
450 Involving Human Subjects (24) and received ethical approval from the appropriate body  
451 in each site and within WHO (protocol GSED validation 004583 approved on  
452 20.04.2020). The findings of the study will be disseminated following a comprehensive  
453 dissemination strategy to reach a diverse range of stakeholders at the local, national and  
454 international level.

#### 455 **DISCUSSION**

1  
2  
3 456 The validation of the GSED SF and LF is a meticulous and systematic global process that  
4  
5 457 introduces an innovative common metric (the D-score) that countries can use to track the  
6  
7 458 progress of child development among populations of young children and to measure the  
8  
9 459 impact of programmatic interventions. Additional attention is required on understanding  
10  
11 460 young children's responses to psychosocial challenges within global contexts. The  
12  
13 461 exploration of the GSED PF introduces an important opportunity to capture the non-  
14  
15 462 normative developmental patterns among young children that are potential precursors to  
16  
17 463 behaviour and psychiatric problems. The GSED validation has several important design,  
18  
19 464 methodological and implementation characteristics that illustrate the rigour required to  
20  
21 465 validate instruments to measure child development globally. First, it is conducted in  
22  
23 466 seven countries with multiple linguistic, cultural and socioeconomic backgrounds.  
24  
25 467 Second, GSED is implemented through an app-based data collection system that  
26  
27 468 facilitates the implementation by reducing recording and transcribing errors and other  
28  
29 469 common pitfalls of paper-based instruments. Third, this study builds on the best practices  
30  
31 470 in validation by including a broad spectrum of psychometric methodologies (concurrent,  
32  
33 471 predictive, convergent, and discriminant validity, test-retest and inter-rater reliability,  
34  
35 472 differential item functioning, and differential test functioning). Fourth, a secondary aim  
36  
37 473 builds the evidence for the creation of preliminary reference scores for the SF and LF,  
38  
39 474 based on a sub-sample with minimal exposure to major biological risk factors and to the  
40  
41 475 extent possible, minimal social and environmental risk factors. Fifth, we are validating an  
42  
43 476 adaptive testing design that can streamline administration by tailoring and reducing the  
44  
45 477 number of items required to obtain a valid score. Sixth, we are testing a new measure of  
46  
47 478 young children's non-normative psychosocial development.  
48  
49 479 One notable difference between the GSED SF and LF measures and other instruments of  
50  
51 480 early child development is that the GSED measures are based on a unidimensional model  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 481 of development through measurement approaches that are universally applicable across  
4  
5 482 cultures. The measures do not follow the common multidimensional approach with  
6  
7 483 separate scores for different domains or contexts. Our validation study intends to  
8  
9 484 demonstrate that this model provides valid, reliable, and interpretable data globally. The  
10  
11 485 GSED SF and LF may exclude some items that measure development in cultural or  
12  
13 486 setting-specific ways, because the focus is on selecting items that are meaningful for  
14  
15 487 understanding child development within any given setting. If specific aspects need to be  
16  
17 488 captured locally, to increase cultural relevance we suggest that the GSED measures are  
18  
19 489 lightly adapted with country or culture-specific item props (in agreement with WHO)  
20  
21 490 and/or through the administration of additional measures.  
22  
23  
24  
25  
26

27 491 There are several limitations to our study. Although we are validating the GSED in seven  
28  
29 492 countries, including one high income setting, three sites are resource-limited  
30  
31 493 (Bangladesh, Pakistan and United Republic of Tanzania). Additional evidence may be  
32  
33 494 needed in high income countries to expand the validity and reliability of the GSED to  
34  
35 495 population-representative samples in additional countries. Second, the GSED has been  
36  
37 496 created using items that fit a Rasch model demonstrating developmental progress across  
38  
39 497 ages 0-3 years (9). This univariate model makes strict assumptions designed for global  
40  
41 498 population estimates and may exclude items that do not show strong age gradients or  
42  
43 499 items that measure development in a culturally-specific ways. Third, GSED was  
44  
45 500 developed to address population and programmatic level evaluations of early child  
46  
47 501 development globally. The GSED is presently not being validated for screening or  
48  
49 502 diagnosing individual children. Finally, our three secondary aims are exploratory, and  
50  
51 503 will require further research, including developing global standards to replace our  
52  
53 504 preliminary reference scores with more specific global norms, as in the Multi-country  
54  
55 505 Growth Reference Standards for children's weight and height. In the future we plan to  
56  
57  
58  
59  
60

1  
2  
3 506 collect additional data from countries using strict inclusion/exclusion criteria (e.g.,  
4  
5 507 additional considerations around environmental risk and protective factors) to further  
6  
7  
8 508 validate our initial reference scores. Similarly, we plan to conduct further work to explore  
9  
10 509 the functionality, reliability, validity, and invariance of the PF.

## 11 510 **Conclusion**

12  
13  
14  
15 511 The validation of the GSED will enable countries to implement the measures with  
16  
17 512 confidence. With valid, reliable, and invariant measures in hand, countries can advance  
18  
19 513 initiatives to ensure that children reach their developmental potential, while reducing or  
20  
21 514 eliminating disparities. Ensuring that all communities have access to policies and  
22  
23 515 programs that provide nurturing care to children and families, with additional support for  
24  
25 516 regions in need, promotes equity and increases the likelihood of achieving the SDGs.  
26  
27 517 Once validated, the GSED measures will enable countries to adapt, modify, and evaluate  
28  
29 518 their policies and programs to ensure that young children are effectively and equitably  
30  
31 519 reaching their development potential and building the human capital needed for  
32  
33 520 sustainable development.  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 521 **LIST OF ABBREVIATIONS**  
4

5 522  
6 523 Bayley-III Bayley Scales of Infant and Toddler Development  
7 524 BRS Brief Resilience Scale  
8  
9 525 CI Confidence interval  
10  
11 526 CPAS Childhood Psychosocial Adversity Scale  
12 527 DAZ Development for Age Z-score  
13  
14 528 DHS Demographic and Health Surveys  
15 529 DIF Differential item functioning  
16  
17 530 DTF Differential test functioning  
18  
19 531 D-score Developmental Score  
20 532 EAP Expected a posteriori  
21  
22 533 ECD Early child development  
23  
24 534 ECDI 2030 Early Childhood Development Index 2030  
25 535 FCI Family Care Indicators  
26  
27 536 FSS Family Support Scale  
28  
29 537 GAMLSS Generalized Additive Model for Location Scale and Shape  
30 538 GSED Global Scales for Early Development  
31  
32 539 HAZ Height-for-age z-score  
33  
34 540 HOME Home Observation for Measurement of the Environment Inventory  
35 541 ICC Intraclass correlation coefficient  
36  
37 542 LF Long Form  
38  
39 543 LMIC Low- and middle-income country  
40 544 MICS Multiple Indicator Cluster Surveys  
41  
42 545 OSF Open Science Framework  
43  
44 546 PF Psychosocial Form  
45 547 PHQ-9 The Patient Health Questionnaire-9  
46  
47 548 SDG Sustainable Development Goals  
48  
49 549 SES Socioeconomic status  
50 550 SF Short Form  
51  
52 551 SOPs Standard operating procedures  
53  
54 552 WAZ Weight-for-age z-score  
55  
56  
57  
58  
59  
60

1  
2  
3 553 **STATEMENTS**  
4

5 554 **Ethics approval:** The study complies with the International Ethical Guidelines for  
6  
7  
8 555 Biomedical Research Involving Human Subjects and received ethical approval from the  
9  
10 556 appropriate Ethics Review Committee (ERC) at the World Health Organization (WHO)  
11  
12 557 (protocol ID GSED validation 004583 approved on 20.04.2020) and in each site.

13  
14 558 **Patient consent:** Written consent is gathered by all study participants. Informed consent  
15  
16 559 forms are written to be easily understood by lay persons, enabling them to understand the  
17  
18 560 aims, procedures and potential risks of participation and have been approved by the  
19  
20 561 WHO ERC. For participants who are illiterate, culturally acceptable options including  
21  
22 562 witnessed oral consent and a thumbprint in lieu of a signature are accepted by the WHO  
23  
24 563 and local ERCs.

25  
26 564 **Data sharing:** Not applicable (protocol paper)

27  
28 565 **Author Contributions:** All authors contributed substantively to this work. VC, TD, MG,  
29  
30 566 MMB, MJ, and PK conceptualized the study protocol and drafted the manuscript; GL,  
31  
32 567 GMc, DMc, MW, SvB, IE, and JS drafted the statistical analysis plan and determined the  
33  
34 568 sample size calculations; AN, AR, KH, and AW drafted pieces of the manuscript. SA,  
35  
36 569 AD, RA, AB, FJ, YS, IN, RK, SS, AZ, MPM, YZ, FT, ARD, AB, JZ, AH, GF, SD, NSK,  
37  
38 570 FB, FJ, and MRC contributed to the adaptation of the study protocol for feasibility and  
39  
40 571 on-the-ground implementation. All above authors, in addition to RK, MMP, and RN  
41  
42 572 reviewed and edited the study protocol and the manuscript. All authors read and approved  
43  
44 573 the final manuscript submission.

45  
46 574 **Acknowledgements:** None  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 575 **FUNDING**  
4

5 576 The authors declare that they have no competing interests.  
6

7 577 **COMPETING INTERESTS**  
8

9  
10 578 This work was supported (alphabetical order) by Bernard van Leer Foundation, Bill &  
11

12 579 Melinda Gates Foundation, Children's Investment Fund Foundation, Jacobs Foundation  
13

14 580 and King Baudouin Foundation, United States. The funders provided financial support.  
15

16  
17 581 The design, implementation, and writing of the manuscript were led by the World Health  
18

19 582 Organization.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



583 **REFERENCES:**

- 584 1. Clark H, Coll-Seck AM, Banerjee A, Peterson S, Dalglish SL, Ameratunga S, et  
585 al. A Future for The World's Children? A WHO–UNICEF–Lancet Commission. The  
586 Lancet. 2020;395(10224):605-58.
- 587 2. UnitedNations. <https://sdgs.un.org/goals/goal4> [cited 2021 November].  
588 Department of Economic and Social Affairs, Sustainable Development].
- 589 3. Fernald LC, Prado E, Kariger P, Raikes A. A Toolkit for Measuring Early  
590 Childhood Development in Low- and Middle-Income Countries. Washington DC 20433:  
591 International Bank for Reconstruction and Development / The World Bank; 2017.
- 592 4. UNICEF. [https://data.unicef.org/resources/early-childhood-development-index-  
593 2030-ecdi2030/](https://data.unicef.org/resources/early-childhood-development-index-2030-ecdi2030/) [updated January 2021].
- 594 5. Faruk T, King C, Muhit M, Islam MK, Jahan I, Baset KU, et al. Screening tools  
595 for early identification of children with developmental delay in low- and middle-income  
596 countries: a systematic review. BMJ Open. 2020;10(11):e038182.
- 597 6. Weber AM, Rubio-Codina M, Walker SP, Van Buuren S, Eekhout I, Grantham-  
598 McGregor SM, et al. The D-Score: A Metric for Interpreting The Early Development of  
599 Infants and Toddlers across Global Settings. BMJ Global Health. 2019;4(6):e001724.
- 600 7. McCoy DC, Waldman M, Team CF, Fink G. Measuring Early Childhood  
601 Development at a Global Scale: Evidence from the Caregiver-Reported Early  
602 Development Instruments. Early Childhood Research Quarterly. 2018;45:58-68.
- 603 8. GSEDTeam. A New Measure of Development in Children from Birth to Age 3 at  
604 Population Level: The Global Scale for Early Development (GSED). Early Childhood  
605 Matters. 2019.
- 606 9. Van Buuren S, Eekhout I, McCray G, Waldman M, Lancaster G, Black M, et al.  
607 A Scale for Tracking Child Development from Directly Observed and Caregiver-  
608 Reported Data. [Manuscript in preparation]
- 609 10. McCray G, McCoy D, Kariger P, Janus M, Black M, Chang-Lopez S, et al. The  
610 Creation of the Global Scales of Early Development (GSED) for 0-3 year old Children:  
611 Combining Subject Matter Expert judgements with big data. [Manuscript in preparation]
- 612 11. Nizar A, Kaur R, Gladstone M, Lancaster G, Baqui A, Cavallera V, et al.  
613 Assessment of Acceptability and Feasibility of the Global Scales for Early Development  
614 (GSED) Across Three Settings for 0 to 3-year-old Children. [Manuscript in preparation]
- 615 12. Streiner DL, Norman GR, Cairney J. Health measurement scales: a practical guide  
616 to their development and use: Oxford University Press, USA; 2015.
- 617 13. Lancaster GA, Kariger P, McCray G, Janus M, Gladstone M, Cavallera V, et al.  
618 Conducting a Feasibility Study in a Global Health Setting for Constructing a Caregiver-  
619 Reported Measurement Tool: An Example in Infant and Young Child Development:  
620 SAGE Publications Ltd; 2020.
- 621 14. McCoy DC, Sudfeld CR, Bellinger DC, Muhihi A, Ashery G, Weary TE, et al.  
622 Development and Validation of an Early Childhood Development Scale for Use in Low-  
623 Resourced Settings. Population Health Metrics. 2017;15(1):1-18.
- 624 15. Van Buuren S, Eekhout I. Child development with the D-score: turning  
625 milestones into measurement [version 1; peer review: 1 approved with reservations].  
626 Gates Open Research. 2021;5(81):1-75.

- 1  
2  
3 627 16. Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, Pollitt E, et al.  
4 628 Child Development: Risk Factors for Adverse Outcomes in Developing Countries. *The*  
5 629 *Lancet*. 2007;369(9556):145-57.
- 6  
7 630 17. Bock RD, Mislavy RJ. Adaptive EAP Estimation of Ability in a Microcomputer  
8 631 Environment. *Applied Psychological Measurement*. 1982;6(4):431-44.
- 9  
10 632 18. Gladstone M, Lancaster G, McCray G, Cavallera V, Alves CR, Maliwichi L, et al.  
11 633 Validation of the Infant and Young Child Development (IYCD) Indicators in Three  
12 634 Countries: Brazil, Malawi and Pakistan. *International Journal of Environmental Research*  
13 635 *and Public Health*. 2021;18(11):6117.
- 14  
15 636 19. Gwet KL. Computing Inter-Rater Reliability and its Variance in the Presence of  
16 637 High Agreement. *British Journal of Mathematical and Statistical Psychology*.  
17 638 2008;61(1):29-48.
- 18  
19 639 20. Ertem IO, Krishnamurthy V, Mulaudzi MC, Sguassero Y, Balta H, Gulumser O,  
20 640 et al. Similarities and Differences in Child Development from Birth to Age 3 years by  
21 641 Sex and Across Four Countries: A Cross-sectional, Observational Study. *The Lancet*  
22 642 *Global Health*. 2018;6(3):e279-e91.
- 23  
24 643 21. Rigby RA, Stasinopoulos DM. Generalized Additive Models for Location, Scale  
25 644 and Shape. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.  
26 645 2005;54(3):507-54.
- 27  
28 646 22. Wainer H, Dorans NJ, Flaugher R, Green BF, Mislavy RJ. *Computerized*  
29 647 *Adaptive Testing: A Primer*: Routledge; 2000.
- 30  
31 648 23. Jacobusse G, Van Buuren S. Computerized Adaptive Testing for Measuring  
32 649 Development of Young Children. *Statistics in Medicine*. 2007;26(13):2629-38.
- 33  
34 650 24. Singh AN. Current Status of Research Ethics and Projecting Future Initiatives.  
35 651 *International Medical Journal*. 2018;25(6).
- 36  
37 652 25. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*.  
38 653 MESA Press, 5835 S. Kimbark Ave., Chicago: ERIC; 1993.
- 39  
40 654 26. Jones PC, Pendergast LL, Schaefer BA, Rasheed M, Svensen E, Scharf R, et al.  
41 655 Measuring Home Environments Across Cultures: Invariance of the HOME Scale Across  
42 656 Eight International Sites from the MAL-ED Study. *Journal of School Psychology*.  
43 657 2017;64:109-27.
- 44  
45 658 27. Kariger P, Frongillo EA, Engle P, Britto PMR, Sywulka SM, Menon P. Indicators  
46 659 of Family Care for Development for Use in Multicountry Surveys. *Journal of Health,*  
47 660 *Population, and Nutrition*. 2012;30(4):472.
- 48  
49 661 28. Berens AE, Kumar S, Tofail F, Jensen SK, Alam M, Haque R, et al. Cumulative  
50 662 Psychosocial Risk and Early Child Development: Validation and Use of the Childhood  
51 663 Psychosocial Adversity Scale in Global Health Research. *Pediatric Research*.  
52 664 2019;86(6):766-75.
- 53  
54 665 29. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The Brief  
55 666 Resilience Scale: Assessing the Ability to Bounce Back. *International Journal of*  
56 667 *Behavioral Medicine*. 2008;15(3):194-200.
- 57  
58 668 30. Dunst C. The Family Support Scale: Reliability and Validity. *Journal of*  
59 669 *Individual, Family, and Community Wellness*. 1984;1(4):45-52.
- 60

- 1  
2  
3 670 31. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and Case Finding for  
4 671 Major Depressive Disorder Using the Patient Health Questionnaire (PHQ-9): A Meta-  
5 672 Analysis. *General Hospital Psychiatry*. 2015;37(6):567-76.
- 7 673 32. Bayley N. *Bayley Scales of Infant and Toddler Development: Administration*  
8 674 *Manual*,. 3rd edition ed. United States of America: Psychorp; 2006.
- 10 675 33. Jacklin L, Cockcroft K. The Griffiths Mental Developmental Scales: An  
11 676 Overview and a Consideration of their Relevance for South Africa. In: Laher S,  
12 677 Cockcroft K, editors. *Psychological Assessment in South Africa. Research and*  
13 678 *applications*: Wits University Press; 2013. p. 169-85.

15 679

16 680

For peer review only

1  
2  
3 681 **FIGURE LEGEND**  
4

5 682 **Figure 1. Development chart**  
6

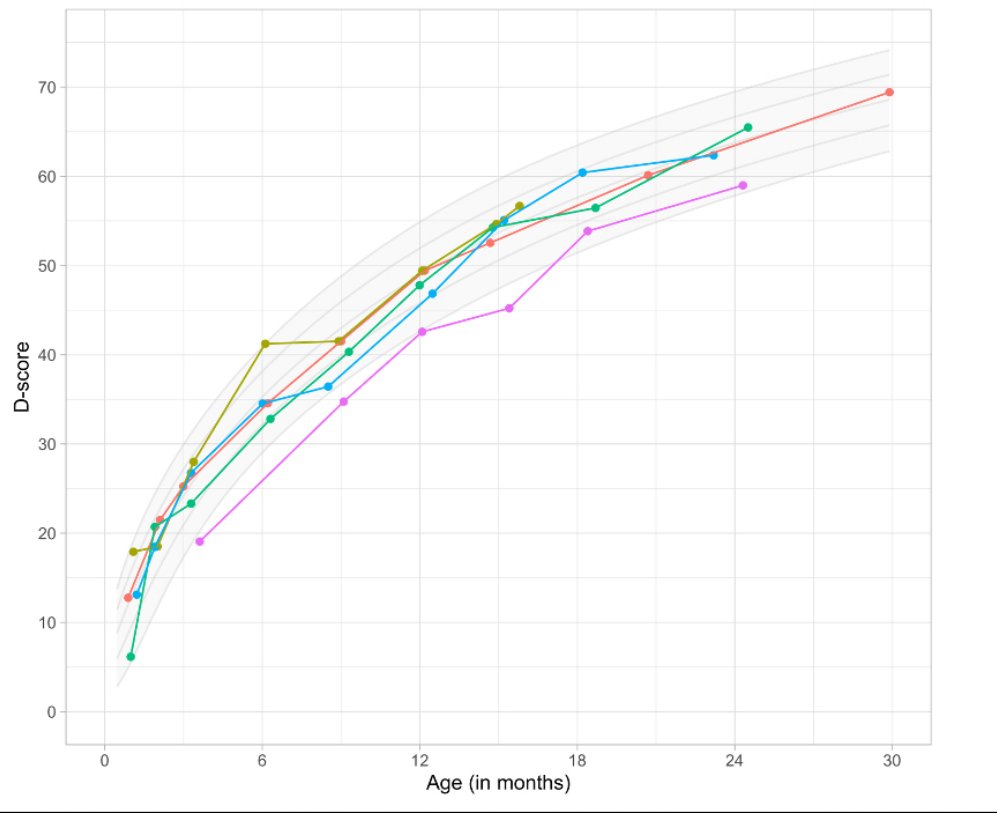
7  
8 683 Reproduced with permission from van Buuren S and Eekhout I (2021) (15)  
9

10 684  
11

12 685 **Figure 2. Study Sampling schema diagram**  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

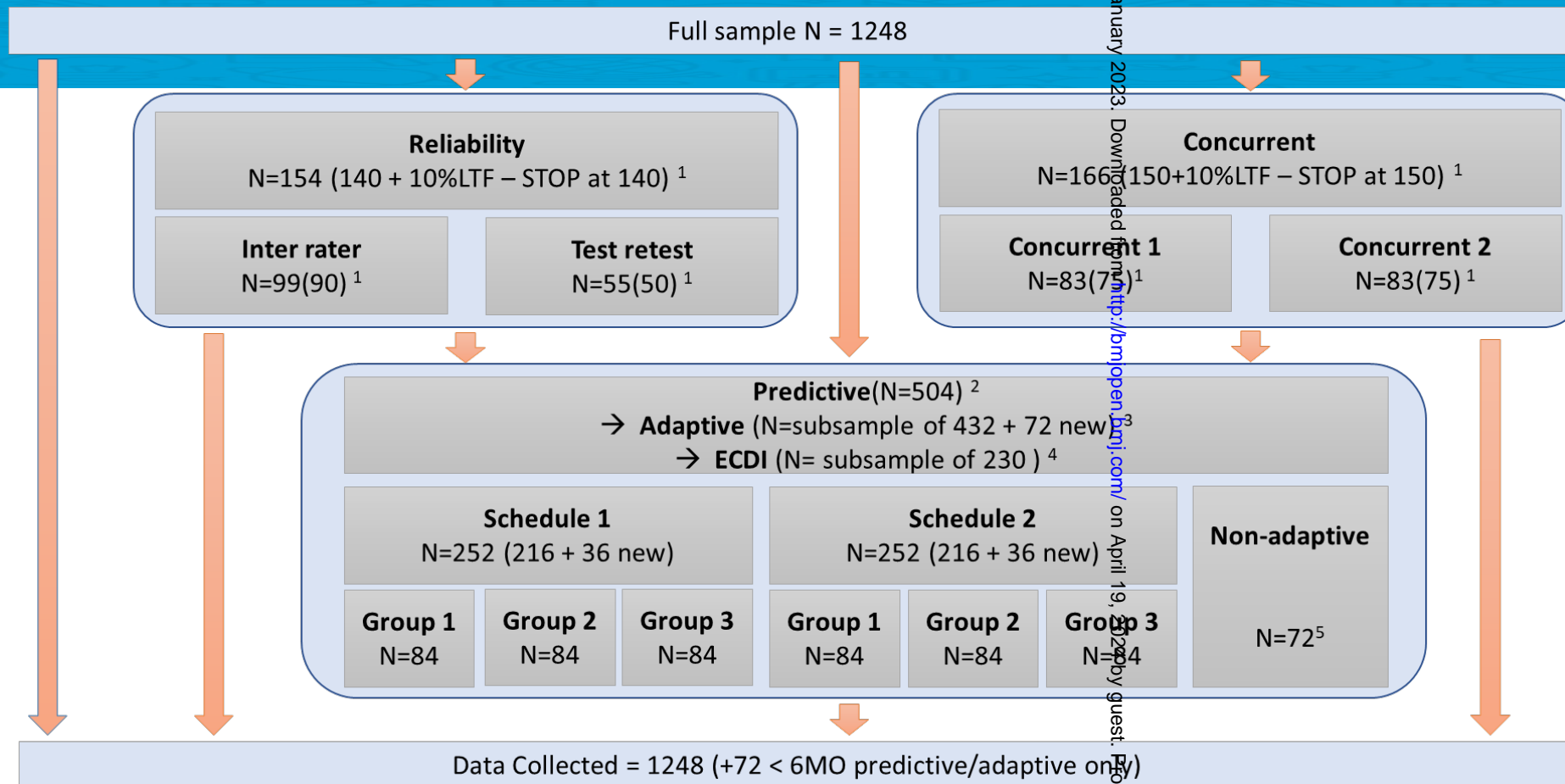
For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



152x125mm (144 x 144 DPI)

# GSED study sample



[1] The number inside parentheses is the number collected and the number outside is the number randomised to account for loss to follow-up

[2] Two additional participants have been added to the predictive to have equal numbers in each experimental group

[3] 72 new children between 2 weeks and 6 months of age have been added to the adaptive sample to ensure coverage at the lower ages.

[4] ECDI will only be done on N=230 Children between the ages of 2+ years at the time of the predictive data collection.

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

[5] The 72 oldest children (36-41 months) from the predictive sample will not be part of the adaptive sample.

2-062562 on 24 January 2023. Downloaded from <http://bmjopen.bmj.com/> on April 19, 2023 by guest. Protected by copyright.

## Supplementary file 1 – Sample size calculations and sampling frame

The sample size calculation for reliability is based on a confidence interval (CI) approach and the desired accuracy for the lower bound of the CI for the ICC estimates. In an Analysis of Variance (ANOVA) with 2-way random effects on a single score with 2 observations per subject (following Shrout and Fleiss, 1979) (27) and with a two-sided 95% confidence interval and an expected ICC of 0.875, the lower confidence interval for the inter-rater reliability sample (N=90) = 0.852. With the same parameters but an expected ICC of 0.90 the lower confidence interval for the test-retest reliability sample (N=50) is 0.830. We expect the ICC to be higher for the test-retest reliability than the inter-rater reliability as inter-rater estimates contain all the sources of error in the test-retest estimates, plus additional error between assessors (14).

To assess concurrent validity, a sample size of 150 per site produces a two-sided 95% CI 0.15-0.44, when the estimate of Pearson's product-moment correlation is 0.30, with an equal spread of participants tested across age and sex. The CI will be narrower when the data are combined across all seven countries. To assess predictive validity a sample size of 404 produces a two-sided 95% CI 0.65-0.75 when the estimate of Pearson's product-moment correlation is 0.70 between individual scores at baseline and at the 6-month follow-up. Allowing 20% dropout at follow up, a sample size of approximately 500 participants is required.

**Table S1. Sampling Frame**

Sample size per site by age and sex for total population (n=1248) which includes a minimum subsample of healthy 'reference' children (n=522)

Age (Days)	Sex	Total Sample size	Minimum sub-sample of reference children	Predictive validity sample (6-month follow-up; age at baseline)	Reliability: Inter-rater	Reliability: Test-Retest	Concurrent validity
15-30	Male	40	20	8	2	1	4
	Female	40	20	8	2	1	2
31-61	Male	40	12	8	1	1	2
	Female	40	12	8	2	1	2
62-91	Male	40	10	8	2	1	2
	Female	40	10	8	1	0	4
92-122	Male	36	9	8	2	1	2
	Female	36	9	8	2	1	2
123-152	Male	32	8	8	1	1	2
	Female	32	8	8	2	1	2

153-183	Male	28	8	8	1	0	4
	Female	28	8	8	1	1	2
184-213	Male	25	7	8	2	1	2
	Female	25	7	8	1	0	2
214-244	Male	23	7	8	1	1	2
	Female	23	7	8	2	1	4
245-274	Male	21	6	8	1	1	2
	Female	21	6	8	1	1	2
275-304	Male	19	6	8	2	0	2
	Female	19	6	8	1	1	2
305-335	Male	17	6	8	1	1	4
	Female	17	6	8	2	0	2
336-365	Male	16	6	7	1	1	2
	Female	16	6	7	1	1	2
366-396	Male	14	6	7	2	1	2
	Female	14	6	7	1	1	4
397-426	Male	13	6	7	1	0	2
	Female	13	6	7	2	1	2
427-457	Male	12	5	7	1	1	2
	Female	12	5	7	1	0	2
458-487	Male	11	5	7	2	1	4
	Female	11	5	7	1	1	2
488-517	Male	11	5	7	1	1	2
	Female	11	5	7	2	1	2
518-548	Male	10	5	7	1	0	2
	Female	10	5	7	1	1	4
549-578	Male	9	5	7	2	1	2
	Female	9	5	7	1	0	2
579-609	Male	9	5	7	1	1	2
	Female	9	5	7	2	1	2
610-639	Male	9	5	7	1	1	4
	Female	9	5	7	1	1	2
640-670	Male	9	5	7	2	0	2
	Female	9	5	7	1	1	2
671-700	Male	9	5	7	1	1	2
	Female	9	5	7	2	0	4
701-730	Male	9	5	7	1	1	2
	Female	9	5	7	1	1	2
731-761	Male	9	5	7	2	1	2
	Female	9	5	7	1	1	2
762-791	Male	9	5	6	1	0	4
	Female	9	5	6	2	1	2
792-822	Male	9	5	6	1	1	2
	Female	9	5	6	1	0	2
823-852	Male	9	5	6	2	1	2
	Female	9	5	6	1	1	2



853-883	Male	9	5	6	1	1	2
	Female	9	5	6	2	1	2
884-913	Male	9	5	6	1	0	2
	Female	9	5	6	1	1	2
914-944	Male	9	5	6	2	1	2
	Female	9	5	6	1	0	2
945-974	Male	9	5	6	1	1	2
	Female	9	5	6	2	1	2
975-1004	Male	9	5	6	1	1	2
	Female	9	5	6	1	1	2
1005-1035	Male	9	5	6	2	0	2
	Female	9	5	6	1	1	2
1036-1065	Male	9	5	6	1	1	2
	Female	9	5	6	2	0	2
1066-1096	Male	9	5	6	1	1	2
	Female	9	5	6	1	1	2
1097-1126	Male	9	5	0	0	0	0
	Female	9	5	0	0	0	0
1127-1157	Male	9	5	0	0	0	0
	Female	9	5	0	0	0	0
1158-1187	Male	9	5	0	0	0	0
	Female	9	5	0	0	0	0
1188-1218	Male	9	6	0	0	0	0
	Female	9	6	0	0	0	0
1219-1248	Male	9	6	0	0	0	0
	Female	9	6	0	0	0	0
1249-1279	Male	9	7	0	0	0	0
	Female	9	7	0	0	0	0
<b>TOTAL</b>		1248	522	504	*99	**55	***166

\*90 + ~10% Loss to follow up = 99; \*\*50 + ~10% Loss to follow up = 55; \*\*\*150 + ~10% Loss to follow up = 166

## Supplementary file 2 – Visit schedule

Table S2a. Visit Schedule for the GSED Validation Study (all sites except the Netherlands)

Main Study Only [No Sub-sample]	Inter- Rater Reliability Sub- Sample	Test- Retest Reliability Sub- Sample	Concurrent Sub- Sample 1 [LF First]	Concurrent Sub- Sample 2 [BSID III First]
<b>Visit 1 [At Home]</b>				
Eligibility and Consent	Eligibility and Consent	Eligibility and Consent	Eligibility and Consent	Eligibility and Consent
COVID Questionnaire	COVID Questionnaire	COVID Questionnaire	COVID Questionnaire	COVID Questionnaire
Contextual	Contextual	Contextual	Contextual	Contextual
GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]
GSED Psychosocial form [PF]	GSED Psychosocial form [PF]	GSED Psychosocial form [PF]	GSED Psychosocial form [PF]	GSED Psychosocial form [PF]
HOME Inventory or Family Care Indicators (FCI)	HOME Inventory or Family Care Indicators (FCI)	HOME Inventory or Family Care Indicators (FCI)	HOME Inventory or Family Care Indicators (FCI)	HOME Inventory or Family Care Indicators (FCI)
Anthropometrics*	Anthropometrics*	Anthropometrics*	Anthropometrics*	Anthropometrics*
<b>Visit 2 [At home, clinic, or other setting within 48 hours of visit 1] Note: For Concurrent Sample, the Visit is at the Clinic setting</b>				
Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
GSED Long form [LF]	GSED Long form [LF]	GSED Long form [LF]	<b>GSED Long form [LF]</b>	<b>BSID III</b>
CPAS	CPAS	CPAS	CPAS	-----
PHQ9	PHQ9	PHQ9	PHQ9	-----
Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale	-----
<b>Visit 3 [Setting and timing vary by sub-sample]</b>				
Visit 3 not required	Visit 3 [At home, clinic or other setting where the LF was completed- <b>within 24 hours</b> of the LF]	Visit 3 [At home, clinic or other setting where the LF was completed- this should happen <b>7 to 10 days after</b> LF]	Visit 3 [Clinic setting within <b>24- 72 hours</b> of the LF- can be done at same time as Visit 2 – taking child fatigue into consideration]	Visit 3 [Clinic setting within <b>24- 72 hours</b> of the BSID III - can be done at same time as Visit 2 – taking child fatigue into consideration]
	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
	GSED Short form [SF]	GSED Short form [SF]	<b>BSID III</b>	<b>GSED Long form [LF]</b>
	GSED Psychosocial form [PF]	GSED Psychosocial form [PF]	-----	CPAS
	GSED Long form [LF]	GSED Long form [LF]	-----	PHQ9
-----	-----	-----	-----	Family support & Resilience Scale

\* Anthropometrics may be done either at visit 1 or visit 2

**Table S2b: Visit Schedule for the GSED Validation Study (the Netherlands only)**

Main Study Only [No Sub-sample]	Inter-Rater Reliability Sub-Sample	Test-Retest Reliability Sub-Sample	Concurrent Sub-Sample 1 [LF First]	Concurrent Sub-Sample 2 [BSID III First]
<b>Session 1 [Online]</b>				
Eligibility and Consent	Eligibility and Consent	Eligibility and Consent	Eligibility and Consent	Eligibility and Consent
Contextual	Contextual	Contextual	Contextual	Contextual
GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]
GSED Psychosocial scale [PS]	GSED Psychosocial scale [PS]	GSED Psychosocial scale [PS]	GSED Psychosocial scale [PS]	GSED Psychosocial scale [PS]
<b>Visit 1 [At clinic within 48 hours of session1]</b>				
Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
GSED Long form [LF]	GSED Long form [LF]	GSED Long form [LF]	<b>GSED Long form [LF]</b>	<b>BSID III</b>
Anthropometrics	Anthropometrics	Anthropometrics	Anthropometrics	Anthropometrics
<b>Session 2 [Online, Test-Retest of SF/PSY within 7 to 10 days of online session 1]</b>				
Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
COVID Questionnaire	COVID Questionnaire	COVID Questionnaire	COVID Questionnaire	COVID Questionnaire
-----	-----	GSED Short form [SF]	-----	-----
-----	-----	GSED Psychosocial scale [PS]	-----	-----
CPAS	CPAS	CPAS	CPAS	CPAS
PHQ9	PHQ9	PHQ9	PHQ9	PHQ9
Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale
Family Care Indicators (FCI)	Family Care Indicators (FCI)	Family Care Indicators (FCI)	Family Care Indicators (FCI)	Family Care Indicators (FCI)
<b>Visit 2 [At clinic, timing varies by sub-sample]</b>				
Visit 2 not required	Visit 2 [within 24 hours of the LF]	Visit 2 [7 to 10 days after LF]	Visit 2 [within 24- 72 hours of the LF- can be done at same time as Visit 1 – taking child fatigue into consideration]	Visit 2 [within 24- 72 hours of the BSID III - can be done at same time as Visit 1 – taking child fatigue into consideration]
	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
	GSED Long form [LF]	GSED Long form [LF]	<b>BSID III</b>	<b>GSED Long form [LF]</b>

# BMJ Open

## Protocol for Validation of the Global Scales for Early Development (GSED) for Children under 3 Years of Age in Seven Countries

Journal:	<i>BMJ Open</i>
Manuscript ID	bmjopen-2022-062562.R1
Article Type:	Protocol
Date Submitted by the Author:	23-Sep-2022
Complete List of Authors:	<p>Cavallera, Vanessa; World Health Organization, Department of Mental Health and Substance Use  Lancaster, Gillian; Keele University School of Medicine  Gladstone, Melissa; University of Liverpool Institute of Translational Medicine, Women and Children's Health  Black, Maureen; RTI International; University of Maryland School of Medicine  McCray, Gareth; Keele University School of Medicine  Nizar, Ambreen; The Aga Khan University, Department of Peadiatrics  Ahmed, Salahuddin; Projahnmo Research Foundation  Dutta, Arup; Center for Public Health Kinetics  Anago, Romuald ; Innovations for Poverty Action  Brentani, Alexandra; University of Sao Paulo, Pediatrics  Jiang, Fan; Shanghai Children's Medical Center Affiliated to Shanghai Jiaotong University School of Medicine  Schönbeck, Yvonne; Netherlands Organization for Applied Scientific Research, Department of Child Health  McCoy, Dana; Harvard Graduate School of Education  Kariger, Patricia; University of California Berkeley School of Public Health  Weber, Ann; University of Nevada Reno  Raikes, Abbie; University of Nebraska Medical Center College of Public Health  Waldman, Marcus; University of Nebraska Medical Center College of Public Health  van Buuren, Stef; TNO, Department of Statistics  Kaur, Raghbir; World Health Organization  Maillard, Pérez; World Health Organization  Nisar, Muhammad; The Aga Khan University, Pediatrics and Child Health  Khanam, Rasheda; Johns Hopkins University, International Center for Maternal and Newborn Health, Department of International Health, Johns Hopkins Bloomberg School of Public Health  Sazawal, Sunil ; Centre for Public Health Kinetics  Zongo, Arsène; Innovations for Poverty Action  Pacífico Mercadante, Mariana; University of Sao Paulo, Pediatrics  Zhang, Yunting ; Shanghai Children's Medical Center Affiliated to Shanghai Jiaotong University School of Medicine  Roy, Arunangshu ; Projahnmo Research Foundation  Hepworth, Katelyn; University of Nebraska-Lincoln College of Education and Human Sciences</p>

	Fink, Günther ; Schweizerisches Tropen- und Public Health-Institut, Household Economics and Health System Research Unit Rubio-Codina, Marta; Inter-American Development Bank Tofail, Fahmida; International Centre for Diarrhoeal Disease Research Bangladesh, Centre for Nutrition & Food Security Eekhout, Iris; TNO, Child Health Seiden, Jonathan; Harvard Graduate School of Education Norton, Rebecca; World Health Organization Baqui, Abdullah; Johns Hopkins University Bloomberg School of Public Health Zhao, Jin; Shanghai Children's Medical Center Affiliated to Shanghai Jiaotong University School of Medicine Holzinger, Andreas; Innovations for Poverty Action Detmar, Symone; TNO Kembou, Samuel; Innovations for Poverty Action Begum, Farzana; The Aga Khan University, Department of Paediatrics Jehan, Fyezah; The Aga Khan University, Paediatrics and Child Health Dua, Tarun; World Health Organization Janus, Magdalena; McMaster University, Offord Centre for Child Studies
<b>Primary Subject Heading</b>:	Public health
Secondary Subject Heading:	Paediatrics
Keywords:	MENTAL HEALTH, Paediatric neurology < NEUROLOGY, Developmental neurology & neurodisability < PAEDIATRICS, PUBLIC HEALTH, EPIDEMIOLOGY, International health services < HEALTH SERVICES ADMINISTRATION & MANAGEMENT

SCHOLARONE™  
Manuscripts

## BMJ Open

**Title:** Protocol for Validation of the Global Scales for Early Development (GSED) for Children under 3 Years of Age in Seven Countries

**Authors:** Cavallera V, Lancaster G, Gladstone M, Black MM, McCray G, Nizar A, Ahmed S, Dutta A, Anago R, Brentani A, Jiang F, Schönbeck Y, McCoy D, Kariger P, Weber A, Raikes A, Waldman M, van Buuren S, Kaur R, Pérez Maillard M, Nisar I, Khanam R, Sazawal S, Zongo A, Pacifico Mercadante M, Zhang Y, Roy AD, Hepworth K, Fink G, Rubio Codina M, Tofail F, Eekhout I, Seiden J, Norton R, Baqui AH, Zhao J, Holzinger A, Detmar S, Kembou S N, Begum F, Jehan F, Dua T, Janus M.

**Affiliations:**

Corresponding author: Dr Vanessa Cavallera, MD, MPH, World Health Organization (WHO), Brain Health Unit in Department of Mental Health and Substance Use, Av. Appia 20, 1202 Geneva, Switzerland. [cavallerav@who.int](mailto:cavallerav@who.int)

Prof Gillian Lancaster, PhD, MSc, BSc (Hons), Clinical Trials Unit, School of Medicine, Keele University, UK

Prof Melissa Gladstone, MD, International Child Health and Neurodevelopmental Paediatrics, Department of Women and Children's Health, Institute of Translational Medicine, University of Liverpool, Alder Hey Children's NHS Foundation Trust, UK.

Prof Maureen Black, PhD, Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, USA and RTI International, Research Park, NC, USA

Dr Gareth McCray, MA, MRes, PhD, Research Associate, the School of Medicine, Keele University, Keele, UK.

Ambreen Nizar, BDS, MSc, Department of Paediatrics, Aga Khan University, Karachi Pakistan.

- 1  
2  
3 27 Salahuddin Ahmed, MBBS, Projahnmo Research Foundation, Dhaka, Bangladesh  
4  
5  
6 28 Arup Dutta, MBA, Center for Public Health Kinetics, New Delhi, India  
7  
8  
9 29 Romuald Kouadio E Anago, MA, BA, Innovations for Poverty Action (IPA), Abidjan,  
10  
11 30 Côte d'Ivoire  
12  
13  
14 31 Dr. Alexandra Valeria Maria Brentani, PhD, Dept. of Pediatrics, University of Sao Paulo  
15  
16 32 Medical School, Brazil  
17  
18  
19 33 Prof. Fan Jiang, MD, PhD, Department of Developmental and Behavioral Pediatrics,  
20  
21 34 National Children's Medical Center, Shanghai Children's Medical Center, affiliated to  
22  
23 35 School of Medicine Shanghai Jiao Tong University, China  
24  
25  
26 36 Dr. Yvonne Schönbeck, PhD, Msc, Dept. Child Health, Netherlands Organization for  
27  
28 37 Applied Scientific Research (TNO), Leiden, The Netherlands  
29  
30  
31 38 Dr Dana McCoy, PhD, Harvard Graduate School of Education, Cambridge, MA, USA  
32  
33  
34 39 Dr Patricia Kariger, PhD CEGA (Center for Effective Global Action), School of Public  
35  
36 40 Health, University of California, Berkeley, California, USA.  
37  
38  
39 41 Dr Ann Weber, PhD, MPH, School of Public Health, University of Nevada, Reno, NV,  
40  
41 42 USA  
42  
43  
44 43 Dr Hilary Abigail Raikes, MPH, PhD, University of Nebraska Medical Center, College of  
45  
46 44 Public Health, Omaha, NE, USA  
47  
48  
49 45 Dr. Marcus Waldman, PhD, University of Nebraska Medical Center, College of Public  
50  
51 46 Health, Omaha, NE, USA  
52  
53  
54 47 Prof Stef van Buuren, PhD, Department of Child Health, Netherlands Organization for  
55  
56 48 Applied Scientific Research TNO, Leiden, The Netherlands  
57  
58  
59  
60

- 1  
2  
3 49 Raghbir Kaur, MPH, DMD, MS, Consultant World Health Organization (WHO), Brain  
4  
5 50 Health Unit in Department of Mental Health and Substance Use, Geneva, Switzerland  
6  
7 51 Michelle Pérez Maillard, MSc, Consultant World Health Organization (WHO), Brain  
8  
9 52 Health Unit in Department of Mental Health and Substance Use, Geneva, Switzerland  
10  
11  
12 53 Prof. Muhammad Imran Nisar, MBBS, MSc, Department of Peadiatrics, Aga Khan  
13  
14 54 University, Karachi Pakistan.  
15  
16  
17 55 Dr. Rasheda Khanam, MBBS, MPH, PhD, International Center for Maternal and  
18  
19 56 Newborn Health, Department of International Health, Johns Hopkins University  
20  
21 57 Bloomberg School of Public Health, Baltimore, MD, USA  
22  
23  
24 58 Dr. Sunil Sazawal, MBBS, MPH, PhD, Center for Public Health Kinetics New Delhi,  
25  
26 59 India  
27  
28  
29 60 Mr. Arsène Zongo, MA, BA, Innovations for Poverty Action (IPA), Abidjan, Côte  
30  
31 61 d'Ivoire  
32  
33  
34 62 Dr. Mariana Pacifico Mercadante, MD, Dept. of Pediatrics, University of Sao Paulo  
35  
36 63 Medical School, Brazil  
37  
38  
39 64 Dr. Yunting Zhang, PhD, Child Health Advocacy Institute, National Children's Medical  
40  
41 65 Center, Shanghai Children's Medical Center, affiliated to School of Medicine Shanghai  
42  
43 66 Jiao Tong University, China  
44  
45  
46 67 Dr. Arunangshu D. Roy, MBBS, Projahnmo Research Foundation, Dhaka, Bangladesh  
47  
48  
49 68 Katelyn Hepworth, MA, University of Nebraska-Lincoln, College of Education and  
50  
51 69 Human Sciences, NE, USA  
52  
53  
54 70 Prof Günther Fink, PhD, Swiss Tropical and Public Health Institute, Basel, Switzerland  
55  
56  
57  
58  
59  
60



- 1  
2  
3 71 Dr Marta Rubio-Codina, PhD, Inter-American Development Bank, Washington, D.C.,  
4  
5 72 USA  
6  
7  
8 73 Dr. Fahmida Tofail, MBBS, PhD, International Centre for Diarrhoeal Disease Research  
9  
10 74 Iccdr,b, Dhaka, Bangladesh  
11  
12  
13 75 Dr. Iris Eekhout, PhD, MSc, Department of Child Health, Netherlands Organization for  
14  
15 76 Applied Scientific Research TNO, Leiden, The Netherlands  
16  
17  
18 77 Jonathan Seiden, EdM, Harvard Graduate School of Education, Cambridge, MA, USA  
19  
20  
21 78 Rebecca Norton, Consultant World Health Organization (WHO), Brain Health Unit in  
22  
23 79 Department of Mental Health and Substance Use, Geneva, Switzerland  
24  
25  
26 80 Dr. Abdullah H. Baqui, MBBS, DrPH, Johns Hopkins Bloomberg School of Public  
27  
28 81 Health, Baltimore, Maryland, USA  
29  
30  
31 82 Dr. Jin Zhao, MD, PhD, Department of Developmental and Behavioral Pediatrics,  
32  
33 83 National Children's Medical Center, Shanghai Children's Medical Center, affiliated to  
34  
35 84 School of Medicine Shanghai Jiao Tong University, China  
36  
37  
38 85 Andreas Holzinger, MA, BA, Innovations for Poverty Action (IPA), Abidjan, Côte  
39  
40 86 d'Ivoire  
41  
42  
43 87 Dr. Symone Detmar, PhD, Msc, Department of Child Health, Netherlands Organization  
44  
45 88 for Applied Scientific Research TNO, Leiden, The Netherlands  
46  
47  
48 89 Dr, Samuel Nzale Kembou, PhD, MA, BA, Innovations for Poverty Action (IPA),  
49  
50 90 Abidjan, Côte d'Ivoire  
51  
52  
53 91 Farzana Begum, MSc, Department of Peadiatrics, Aga Khan University, Karachi  
54  
55 92 Pakistan.  
56  
57  
58  
59  
60

1  
2  
3 93 Prof. Fyezah Jehan, MBBS, FCPS, MSc, Department of Peadiatrics, Aga Khan  
4

5 94 University, Karachi Pakistan.  
6  
7

8 95 Dr Tarun Dua, MD, World Health Organization (WHO), Brain Health Unit in  
9

10 96 Department of Mental Health and Substance Use, Geneva, Switzerland  
11  
12

13 97 Prof Magdalena Janus, MBBS, FCPS, MSc, Offord Centre for Child Studies, McMaster  
14

15 98 University, Hamilton ON, Canada.  
16  
17

18 99 **Disclaimers:**

20 100 The author is a member of the World Health Organization. The author alone is  
21  
22 101 responsible for the views expressed in this publication and they do not necessarily  
23  
24 102 represent the decisions, policy or views of the World Health Organization. [Applies to  
25  
26 103 Cavallera V, Dua T, Kaur R, Pérez Maillard M and Norton R]  
27  
28

30 104 The views here presented do not represent the Inter-American Development Bank, its  
31  
32 105 board of directors, or the countries it represents. [Applies to Rubio Codina M]  
33  
34

35 106 **Wordcount:** 4000  
36

37 107 **Keywords:** Early childhood development, global, measurement, index, population,  
38  
39 108 evaluation, validation, protocol, programmes.  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 109 **ABSTRACT**  
4  
5

6 110 **Introduction.** Children's early development is affected by caregiving experiences, with  
7  
8 111 life-long health and wellbeing implications. Governments and civil societies need  
9  
10 112 population-based measures to monitor children's early development and ensure that  
11  
12 113 children receive the care needed to thrive. To this end, the World Health Organization  
13  
14 114 (WHO) developed the Global Scales for Early Development (GSED) to measure  
15  
16 115 children's early development (ages 0-3 years). The GSED includes three measures: 1)  
17  
18 116 short form (SF) for population-evaluation (caregiver-report), 2) complementary long  
19  
20 117 form (LF) for programmatic-evaluation (direct assessment), and 3) psychosocial form  
21  
22 118 (PF) for psychosocial development evaluation (caregiver-report). The primary aim of this  
23  
24 119 protocol is to validate the GSED SF and LF. Secondary aims are to create preliminary  
25  
26 120 reference scores for the GSED SF and LF, validate an adaptive testing algorithm, and  
27  
28 121 assess the feasibility and preliminary validity of the GSED PF.  
29  
30

31  
32  
33 122 **Methods and Analysis.** We will conduct the validation in seven countries (Bangladesh,  
34  
35 123 Brazil, Côte d'Ivoire, Pakistan, The Netherlands, The Republic of China, United  
36  
37 124 Republic of Tanzania), varying in geography, language, culture and income through a  
38  
39 125 one-year prospective design, combining cross-sectional and longitudinal methods with  
40  
41 126 1248 children per site, stratified by age and sex. The GSED generates an innovative  
42  
43 127 common metric (Development-score: D-score) using the Rasch model and a  
44  
45 128 development-for-age z-score (DAZ). We will evaluate six psychometric properties of the  
46  
47 129 GSED SF and LF: concurrent validity, predictive validity at six months, convergent and  
48  
49 130 discriminant validity, and test-retest and inter-rater reliability. We will evaluate  
50  
51 131 measurement invariance by comparing differential item functioning (DIF) and differential  
52  
53 132 test functioning (DTF) across sites.  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 133 **Ethics and dissemination.** This study has received ethical approval from the WHO  
4  
5 134 (protocol GSED validation 004583 20.04.2020) and approval in each site . Study results  
6  
7 135 will be disseminated through webinars and publications from WHO, international  
8  
9 136 organisations, academic journals, and conference proceedings.

10  
11  
12 137 **Registration details:** Open Science Framework (OSF) <https://osf.io/> on 19/11/2021 (DOI  
13  
14 138 10.17605/OSF.IO/KX5T7; identifier: osf-registrations-kx5t7-v1)

## 17 18 139 **ARTICLE SUMMARY**

### 19 140 **Strengths and limitations of this study**

- 21 141 • The study collects validation data (n = 8736 children) for the Global Scales for  
22 142 Early Development (GSED) in seven countries that vary in geographic, linguistic,  
23 143 cultural and sociodemographic characteristics.
- 24 144 • The methods for the validation of GSED are systematic across sites and follow  
25 145 rigorous standard operating procedures based on the best scientific evidence  
26 146 available.
- 27 147 • A tablet-based App is used for data collection to make the administration of the  
28 148 GSED measures user-friendly, to reduce recording and transcribing errors, and to  
29 149 facilitate adaptive testing.
- 30 150 • The GSED SF and LF aims to include items that are culturally neutral and fit the  
31 151 Rasch model, which assumes that child development milestones are age-ordinal,  
32 152 to create D-scores while psychosocial items are included in a separate measure  
33 153 (GSED psychosocial form [PF]) and cultural-specific items can be supplemented  
34 154 by countries.
- 35 155 • The three secondary aims (preliminary reference scores, an adaptive testing  
36 156 algorithm, and the feasibility and validity of the GSED PF), are exploratory and  
37 157 will require further research.

## 158 INTRODUCTION

159 Prenatal and early postnatal experiences have significant impacts on early childhood  
160 development (ECD) and can influence the accrual of health, well-being, and productivity  
161 throughout the life-course (1). To promote current and sustainable peace and prosperity,  
162 the United Nations has focused the Sustainable Development Goals (SDG) on improving  
163 children's outcomes in the early years through multiple targets. The most explicit target  
164 for young children is SDG 4 (Education goal), which requires reporting on the  
165 "proportion of children under 5 years of age who are developmentally on track in health,  
166 learning and psychosocial well-being, by sex" (2).

167 There are few valid measures that can be used globally to assess child development for  
168 children under three years of age. Current measures of ECD range from proxy measures  
169 (e.g., prevalence of country-level stunting and poverty) to detailed measures of individual  
170 performance on developmental tasks (3). The Early Childhood Development Index 2030  
171 (ECDI 2030) (4) does not include children below two years of age. A recent review has  
172 identified the creation and validation of population-based instruments for assessing very  
173 young children as a global priority (5).

174 The Global Scales for Early Development (GSED) build on advances made by analyses  
175 of existing global datasets (6), and new data collection (7) that demonstrated the cross  
176 cultural applicability of items that measure young children's development. Three research  
177 teams (8) joined efforts to develop the GSED in response to the pressing need for  
178 instruments and metrics to measure ECD at population and programmatic levels across  
179 diverse parts of the world.

## 180 **The Global Scales for Early Development (GSED)**

181 The GSED consist of three open-access measures developed by a WHO-led team<sup>1</sup> to  
182 provide a standardized methodology for measuring the development of children aged 0-3  
183 years (0-36 months) across diverse cultures and contexts (9, 10). They are developed for  
184 three objectives: 1) for population-level evaluation based on caregiver-report, GSED  
185 Short Form (SF); 2) for programmatic evaluation in combination with SF, direct child  
186 assessment, GSED Long Form (LF); and 3) for measuring psychosocial behaviours,  
187 caregiver-reported GSED Psychosocial Form (PF). The development and piloting of the  
188 GSED SF, LF, and PF are described elsewhere (9).

189 The GSED SF and LF produce metrics on the same age-ordinal scale and quantify the  
190 same latent construct. The Developmental Score (D-score) (see Box 1) underlies both  
191 measures and reflects children's overall development across multiple domains typically  
192 demonstrated in this age group (e.g., cognitive, motor, language, social-emotional) (6).

193 The GSED PF items, designed to measure non-normative developmental patterns,  
194 including behavioural or regulatory challenges, are not age-ordinal and do not use the D-  
195 Score metric.

### **Box 1: The Developmental score**

The Developmental score (11), or D-score, is a unidimensional latent variable measuring child development during the first three years across multiple domains. The milestones that make up the D-score conform to the Rasch model (12), thus yielding a scale with interval properties with a fixed unit (Figure 1). It is therefore possible to calculate a meaningful difference between two D-scores. Similar to height-for-age Z-score, given suitable age-conditional references, the D-score can be transformed to a Z-

<sup>1</sup> The full team and contributors are listed in the Acknowledgments.

score that accounts for children's age (i.e., Development for Age Z-score, or DAZ).

The DAZ facilitates comparisons across children of different ages.

## 196 **AIMS**

197 The primary aim of this study is to validate the GSED measures (13), through testing for  
198 measurement invariance and evaluation of the psychometric properties to measure  
199 development among children aged 0-3 years (0-36 months) globally (including creation  
200 of D-scores and Development for Age Z-score [DAZ]).

201 Specific Objectives:

- 202 a) Fit a Rasch model to the item data to calculate the D-scores and DAZ.
- 203 b) Investigate differential item functioning (DIF) and differential test functioning  
204 (DTF) across sites to determine measurement invariance.
- 205 c) Examine psychometric properties of the GSED SF and LF:
  - 206 • Test-retest and inter-rater reliability (score and item level),
  - 207 • Concurrent validity (association between scores on GSED and Bayley Scales  
208 of Infant and Toddler Development (Bayley-III) or Griffiths Scales of Child  
209 Development administered concurrently) (14),
  - 210 • Convergent validity (strength of association between GSED D-scores and  
211 other theoretically-relevant constructs)
  - 212 • Predictive validity (association between GSED scores six months after initial  
213 assessment).

214 The secondary aims are to: 1) establish preliminary reference scores<sup>2</sup> for optimal  
215 development on the D-score (GSED SF and LF), 2) develop and validate an adaptive

---

<sup>2</sup> The population has not been selected as a representative sample of all children aged 0-3 in each site (as would happen in a countrywide population census). Selection and recruitment of a representative sample

1  
2  
3 216 testing algorithm, and 3) obtain preliminary validity data on the psychometric properties  
4  
5 217 of the GSED PF.  
6  
7

## 8 218 **METHODS**

### 9 219 **Design and study sites**

10  
11  
12 220 The GSED validation study uses a prospective cross-sectional design with a longitudinal  
13  
14 221 component of age and sex stratified samples of children in seven countries. The countries  
15  
16 222 are culturally, linguistically and geographically diverse, representing low-income  
17  
18 223 (Bangladesh, Côte d'Ivoire, Pakistan, United Republic of Tanzania), middle-income  
19  
20 224 (Brazil and The Republic of China), and high-income (The Netherlands) settings.  
21  
22  
23 225 Samples in each site are not nationally representative; however, they are diverse, e.g.  
24  
25 226 covering both rural/urban settings.  
26  
27

28  
29 227 Preparation and feasibility phases are described elsewhere (13), and assess feasibility of  
30  
31 228 administration of GSED and associated measures including processes for translating and  
32  
33 229 culturally adapting GSED and other study measures, creating data management systems,  
34  
35 230 and training teams in data collection procedures.  
36  
37

### 38 231 **Patient and Public involvement**

39  
40 232 Caregivers of children 0-41 month-olds were involved in the study design as the burden  
41  
42 233 of the assessment was discussed with them in a pilot stage through qualitative data  
43  
44 234 collection. We intend to disseminate the main results to trial participants and will seek  
45  
46 235 patient and public involvement in the development of an appropriate method of  
47  
48 236 dissemination.  
49  
50  
51  
52  
53  
54  
55  
56

---

57 was beyond the scope of this study and not required for validation purposes. We are therefore developing  
58 'reference scores' which should not be interpreted as population-sampled norms.  
59  
60



237 **Study sample**

238 The study sample includes children between 0 to 41 months of age (inclusive) living in  
 239 study areas (see Table 1 for inclusion and exclusion criteria). The small sample of  
 240 children from 36-41 months aims to ensure that parameters are estimated with adequate  
 241 precision for children at the top of our age range (36 months).

242 **Table 1. Study sample inclusion and exclusion criteria**

Sample	Inclusion criteria	Exclusion criteria
Total per site N=1248 (as described in sample size section below)	1. Age 0-41 months 2. Family speaks to the child in same language as GSED translation 3. Primary caregiver available to participate	1. Missing gestational age (ultrasound or last menstrual period LMP) 2. Missing birth weight data 3. Acutely unwell at time of assessment (temporary exclusion: to be rescheduled after 7 days)

244 ***Recruitment and consent***

245 In each site the sampling frame consists of a list of potentially eligible caregiver-child  
 246 dyads residing in the defined study area. Lists of potential participants are created in  
 247 compliance with ethical review boards approved processes; they vary by site and may  
 248 include: participants in local pregnancy surveillance systems, families who have  
 249 previously agreed to be contacted for participation, from hospital/health center registries,  
 250 or families with children attending local child health/care centers. Sites using registries  
 251 will rely on hospital or health center staff (unaffiliated with GSED) to contact families  
 252 and obtain consent for sharing their information with the GSED team. A sample listing of  
 253 the pre-consented families will be provided to the GSDE team for recruitment. Sites  
 254 recruiting families from local child health/care centers will rely on advertisements or  
 255 flyers with information about the project, participation requirements, GSED team contact

1  
2  
3 256 information for questions, and a scan code or website link for interested families to  
4  
5 257 provide basic eligibility information and consent to be contacted for enrollment.  
6  
7  
8 258 Eligible children are sampled from this list using the GSED sampling scheme (Figure 2).  
9  
10 259 To minimize clustering of correlated scores within households, one child per caregiver  
11  
12 260 and in multi-family household is selected, guided by age and sex quotas. For siblings or  
13  
14 261 twins, one is chosen randomly. Target children's primary caregiver (person most familiar  
15  
16 262 with the child and spends most time with them) is approached for consent and enrollment.  
17  
18  
19 263 A non-technically worded information sheet is shared and consent to participate is  
20  
21 264 obtained at first visit. In the Netherlands, participants provide consent online, confirmed  
22  
23  
24 265 by study staff at first visit. Refusals to participate and dropouts are registered and  
25  
26 266 replaced.

### 29 267 *Sampling frame and schemes*

30  
31 268 Sample size for recruitment within each site is 1248 children (total 8736 children) across  
32  
33 269 seven countries. After consent is provided, children are allocated by sex and age groups  
34  
35 270 using a randomization procedure to one of several sampling schema (e.g., predictive,  
36  
37 271 reference-score, reliability; Figure 2). See sampling Table S1 in Supplementary file S1  
38  
39 272 for sampling frame. Out of the full site sample of 1248 children, 504 children per site are  
40  
41 273 randomly selected for re-evaluation 6 months later to assess predictive validity (primary  
42  
43 274 aim). A second scheme indicates the minimum sub-sample of children needed to  
44  
45 275 calculate preliminary reference scores (secondary aim) that will facilitate cross-country  
46  
47 276 comparisons. To maximize precision of parameter estimates, larger quotas are kept for  
48  
49 277 the youngest age brackets where rates of development are accelerated. A third scheme  
50  
51 278 addresses inter-rater reliability for 90 children per site using two assessors who  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 279 independently assess the same child sequentially or within 24 hours (14)<sup>3</sup>. Test-retest  
4  
5 280 (intra-rater reliability) is performed by inviting 50 children per site to return for repeat  
6  
7 281 assessment with the same rater within 7-10 days. For concurrent validity, to assess the  
8  
9 282 GSED against the Bayley-III, a sample size of N = 150 per country produces a two-sided  
10  
11 283 95% confidence interval 0.15-0.44, when the estimate of Pearson's product-moment  
12  
13 284 correlation is 0.3, with an equal spread of participants tested across age and sex.  
14  
15 285 In the Netherlands, the GSED SF and PF are administered online. A subset of participants  
16  
17 286 (N=32) are interviewed face-to-face to compare method of administration. To determine  
18  
19 287 test-retest reliability (intra-rater reliability), the primary caregiver completes the SF and  
20  
21 288 PS form online and then a second time 7-10 days later.  
22  
23  
24  
25

## 26 289 **Data collection**

### 27 290 *Measures*

#### 28 291 *GSED*

29 292 **GSED SF and LF.** The creation of the GSED SF and LF is described elsewhere (10).  
30  
31 293 Briefly, we constructed an item bank from previously gathered data and compiled cross-  
32  
33 294 sectional and longitudinal data from 31 countries representing over 73,000 anonymized  
34  
35 295 children with 109,079 assessments (using 22 established ECD instruments) (6, 15, 16).  
36  
37 296 Using subject matter expert input and statistical modelling (11), we developed a  
38  
39 297 caregiver-reported measure intended to capture child development at population-level  
40  
41 298 (GSED SF), and a complementary direct-assessment measure to evaluate programmatic  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55

---

56 <sup>3</sup> We note that this procedure differs from typical inter-rater reliability (IRR) designs which involve  
57 simultaneous scoring of a single assessment. This sequential design was necessitated by logistical  
58 constraints. Given that this design captures both variance due to differences in raters *and* differences in  
59 occasions, the observed IRR represents a lower bound for the true inter-rater reliability of the assessments.  
60

1  
2  
3 299 impacts (GSED LF) (10). The measures are created paper-based and app-based (GSED  
4  
5 300 App) with built-in administration rules and supporting media-files (see below).  
6  
7 301 The GSED SF includes 139 items representing emerging skills and behaviors within  
8  
9 302 cognitive, motor, language and social-emotional domains. All items are presented as  
10  
11 303 questions to the caregiver, with binary response options (Yes/No and “Don’t Know”) that  
12  
13 304 use start rules based on the child’s age, and stop rules based on age and performance.  
14  
15 305 Assessors record caregiver’s responses, regardless of the assessor’s observations. In the  
16  
17 306 Netherlands only, the GSED SF is completed online by caregivers. The GSED SF  
18  
19 307 administration includes sounds, images, and short video clips that assist in understanding,  
20  
21 308 interpreting and administering the items.  
22  
23 309 The GSED LF includes 155 items capturing similar domains to the SF but, observed by  
24  
25 310 the assessor following start and stop rules based on the child’s age and responses. LF  
26  
27 311 items must either be observed incidentally or by eliciting the behaviour or both,  
28  
29 312 depending on the item. Items are organized into three grids (A, B and C) that enable  
30  
31 313 assessors to measure the child’s performance on similar tasks in succession, making the  
32  
33 314 administration easier for both assessors and children. To further facilitate administration,  
34  
35 315 icons are placed next to each item that inform the assessor whether the item is observed,  
36  
37 316 demonstrated to or by the child, listened for or spoken to the child. The GSED LF uses a  
38  
39 317 locally constructed and low cost kit with basic materials that the child interacts with to  
40  
41 318 demonstrate abilities. The kit is created by local teams with detailed guidance from  
42  
43 319 WHO. Responses of all LF items are binary (skill observed/not observed).  
44  
45 320 The items in both measures are ordered by difficulty reflecting children’s emerging skills.  
46  
47 321 Based on the analyses from the validation, we will select the items to be included in the  
48  
49 322 final GSED SF and LF versions available for use.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 323 **Psychosocial Form (PF).** Unlike the SF and the LF, the GSED PF has been developed to  
4  
5 324 index non-normative developmental patterns that provide a window into early  
6  
7 325 manifestations of children's mental health challenges, including internalizing and  
8  
9 326 externalizing behaviour problems and dysregulation (e.g., eating and sleeping). Items  
10  
11 327 capturing developmentally normative information about socio-emotional competencies  
12  
13 328 are included in the GSED SF and LF, as the SDG 4.2 includes children's psychosocial  
14  
15 329 well-being. Because few instruments have been developed to capture psychosocial  
16  
17 330 difficulties for children under 3 years, little existing data are available and the  
18  
19 331 development of the GSED PF is exploratory. The PF initial prototype was created  
20  
21 332 through a review of existing measures of infant and toddler mental health and consensus  
22  
23 333 by subject matter experts. The GSED PF includes 47 items and reflects caregiver  
24  
25 334 perceptions of the behaviors' frequency, using response options: Often; Sometimes;  
26  
27 335 Never/almost never. Items are divided into two age groups: 0 to <6 and 6 to <36 months.

336 *Contextual and demographic measures*

337 In addition to the GSED, the validation study includes measures of children's growth and  
338 nutrition, health, environmental and contextual information (see Table 2 for measures and  
339 sources). The selection of measures was based on known biological and social  
340 determinants of development (17), the demonstrated validity of the contextual measures  
341 in at least one low- and middle-income country (LMIC), and efficiency for data  
342 collection. See Supplementary File S2 for visit schedules (Tables S2a and S2b).  
343 In three sites (Côte d'Ivoire, The Netherlands and The Republic of China) where  
344 administration of the Home Observation for Measurement of the Environment Inventory  
345 (HOME) is not feasible, household stimulation data and caregiver-child activities are  
346 collected using Family Care Indicators (FCI). In all sites, a concurrent measure of child  
347 development (Bayley-III or Griffiths Mental Development Scales) is administered in a

348 subsample of children to determine concurrent validity of GSED to a well-established  
 349 measure of the same construct.

350 **Table 2: Study measures in addition to GSED**

Construct	What the Measure Captures	Measure	Administration Mode	Time for Administer (Minutes)
Child health and household socioeconomic status (SES)	<ul style="list-style-type: none"> <li>• Eligibility (exclusion - criteria)</li> <li>• Demographic information</li> <li>• Information about acute child health</li> <li>• Delivery and Perinatal conditions</li> <li>• Breastfeeding</li> <li>• Child's health history</li> <li>• Household socio-economic status*</li> <li>• Caregiver education</li> <li>• Maternal health/ chronic illness</li> <li>• COVID-19 exposure</li> </ul>	Eligibility and Contextual Form [Specifically developed for the study]	Caregiver Report	35
Anthropometry	<ul style="list-style-type: none"> <li>• Weight at time of assessment</li> <li>• Infant Length/ Child Height at time of assessment</li> <li>• Child's Mid-upper arm circumference at time of assessment</li> <li>• Child's head circumference at time of assessment</li> </ul>	Anthropometry Form	Child Assessment	15
<i>Family / home environment</i>	<ul style="list-style-type: none"> <li>• Home Environment (HOME only)</li> <li>• Play/ Stimulation / interactions between the child and other family members in the</li> </ul>	Home Observation for Measurement of the Environment Inventory (HOME) (18) OR Family Care Indicators (FCI) (19) †	<i>HOME</i> : Caregiver report & Observation <i>FCI</i> : Caregiver Report	<i>HOME</i> : 45 <i>FCI</i> : 15

	home (HOME and FCI)			
	<ul style="list-style-type: none"> <li>Child neglect/abuse</li> <li>Exposure to violence or conflict</li> </ul>	Childhood Psychosocial Adversity Scale (CPAS)(20) <sup>†</sup>	Caregiver Report	15
	<ul style="list-style-type: none"> <li>Family resilience</li> </ul>	Brief Resilience Scale (BRS) (21) <sup>†</sup>	Caregiver Report	1
	<ul style="list-style-type: none"> <li>Family social support</li> </ul>	Family Support Scale (FSS) (22) <sup>†</sup>	Caregiver Report	5
Caregiver health and well-being	<ul style="list-style-type: none"> <li>Caregiver Depressive Symptoms</li> </ul>	The Patient Health Questionnaire-9 (PHQ-9) (23)	Caregiver Report	5
Child development	<ul style="list-style-type: none"> <li>Global child development (0-41 months)</li> </ul>	Bayley Scales of Infant and Toddler Development (Bayley-III) (24) OR Griffiths Mental Development Scales(25) <sup>‡</sup>	Direct child Assessment	45-60
	<ul style="list-style-type: none"> <li>Global child development (24-41 months)</li> </ul>	Early Childhood Development Index 2030 (ECDI2030) (4) <sup>§</sup>	Caregiver Report	10

351 \* Socioeconomic information on this form comes from the standard DHS multiple assets  
 352 index; however, some sites have adapted the socio-economic status items to better fit  
 353 their contexts

354 <sup>†</sup> these measures have been minorly adapted for the purpose of the study

355 <sup>‡</sup> in a sub-sample (N=150)

356 <sup>§</sup> in a sub-sample (all children of 24 to 41-months within the predictive validity  
 357 subsamples in three countries)

358

### 359 ***Schedule***

360 Data collection is scheduled over one to three visits depending on the study site to

361 accommodate rules of measure administration order and location. The first

1  
2  
3 362 administration of the GSED SF and PF is completed at home (or online in the  
4  
5 363 Netherlands) to test it in the setting intended for future use (e.g., Multiple Indicator  
6  
7 364 Cluster Surveys MICS or Demographic and Health Surveys DHS) and prior to  
8  
9 365 administration of the GSED LF. The GSED LF is administered in a controlled  
10  
11 366 environment (e.g., clinic) to match the required concurrent validity testing protocols. For  
12  
13 367 the concurrent validation, the GSED and concurrent measures are administered in the  
14  
15 368 same location on different days and counter-balanced in order of administration.

16  
17  
18  
19 369 ***Training and Quality Control***

20  
21 370 Training of local master trainers is performed by the WHO team for the GSED SF, PF  
22  
23 371 and LF, using slide presentations, discussion forums, audio-visual aids, and practice  
24  
25 372 exercises. Local master trainers are responsible for training local field teams using  
26  
27 373 materials adapted and translated to local languages. Reliable administration of the GSED  
28  
29 374 measures must be met (inter-rater agreement with a master trainer of  $\geq 90\%$ ) for  
30  
31 375 certification.

32  
33  
34  
35 376 To ensure quality assurance, 10% of all the study visits are observed by the study  
36  
37 377 supervisor in person (or through video-recording in the Netherlands), covering each child  
38  
39 378 age band and certified assessors. Supervisors independently complete questionnaires  
40  
41 379 being administered by the assessor and complete a fidelity checklist. Assessors are given  
42  
43 380 feedback based on checklist score. Supervisors review quality assurance findings with the  
44  
45 381 WHO biweekly, along with discussions with the subject matter experts for further  
46  
47 382 resolution, as needed.

48  
49  
50  
51 383 The GSED application software for data collection has built-in data range and  
52  
53 384 consistency checks. Data managers review and resolve issues daily in consultation with  
54  
55 385 the local field team and/or WHO team.  
56  
57  
58  
59  
60



### 386 **Sample size**

387 Sample size determination was based on the primary aim of assessing the psychometric  
388 properties of the GSED. To have sufficient power to estimate measurement parameters  
389 (abilities and difficulties) needed to calculate the D-score and DAZ scores at baseline and  
390 to detect DIF of 1 logit with a power of  $1 - \beta = 0.90$  and a two-sided significance level  
391 of  $\alpha = 0.05$ , a sample of  $N=1248$  per site is required. Given the rapidity of development of  
392 children at this age, the latent trait is longer than tends to be found in educational tests which  
393 focus on a narrower ability range. The easiest item in our tool “Does your child smile?” has a  
394 difficulty of -13.2 logits (1.1 on the D-score scale) and the most difficult item has a difficulty of  
395 8.4 logits (88.86 on the D-Score scale), a 21.6 logit span. Thus, a one logit difference is not  
396 particularly large, given the length of the latent trait. This sample size was calculated via  
397 optimization of the sample size at i) each age/sex stratum and ii) overall on 1000  
398 simulated datasets generated from parameters suggested by the Rasch GSED model. See  
399 Supplementary file S1 for additional details.

### 400 **Statistical Analysis**

401 To construct the scores for the GSED SF and LF, a Rasch model will be fitted and the  
402 item fit statistics (infit and outfit) will be assessed (26). Any items with unacceptable fit  
403 levels will be removed. Items will be screened for whether they exhibit unacceptable  
404 levels of measurement non-invariance (i.e., they have approximately equal difficulties)  
405 across countries and other contextual variables. Items exhibiting unacceptable DIF (using  
406 the logistic regression method) will be discarded sequentially, and the item response  
407 models will be refit using the remaining items. The expected a posteriori (EAP) method  
408 (27) will be applied to the final model to estimate the latent ability parameter (the D-  
409 score). Systematic deviations from unidimensionality will be tested by performing a  
410 principal components analysis on the residuals of the Rasch model. The method uses a

1  
2  
3 411 prior normal distribution with a mean set equal to the average proficiency at the child's  
4  
5 412 age and a standard deviation of 5. The ability estimates will be used to estimate  
6  
7 413 preliminary developmental percentile curves against age using a Generalized Additive  
8  
9 414 Model for Location Scale and Shape (GAMLSS). Note that this application of EAP  
10  
11 415 estimates underestimates the true variability in the population because EAP estimates – as  
12  
13 416 any measurement – are always imprecise. In daily practice, analysts will compare other  
14  
15 417 EAP estimates to the reference. To support this type of application, we create the  
16  
17 418 references from the EAP estimates and accept a (perhaps slight) underestimate of the true  
18  
19 419 variability in child development in the population. Following previous methodology (28)  
20  
21 420 software will be written to calculate DAZ-scores based on the final dataset in R (29), and  
22  
23 421 a user-friendly front-end version created in R (ShinyApp) (30) and/or Excel.  
24  
25 422 Reliability (inter-rater and test-retest) for all GSED measures will be analyzed using ICC  
26  
27 423 (at the score level) and Gwet's AC1 agreement (at the item level) statistics with 95%  
28  
29 424 confidence intervals to determine whether items perform reliably within and between  
30  
31 425 assessors (31). A cut-off value of 0.4 and above will be used to flag items as adequately  
32  
33 426 reliable. Those items with agreement between 0.4 and 0.5 will be discussed to determine  
34  
35 427 if modifications can be made to improve their administration and/or comprehension.  
36  
37 428 DAZ scores from the GSED SF and LF will be used to conduct validity analyses to  
38  
39 429 ensure that the measures are capturing the construct they are purported to measure  
40  
41 430 (construct validity). Concurrent validity will be assessed by correlating age-corrected  
42  
43 431 Bayley-III or Griffiths Mental Development Scales scores with GSED DAZ scores. We  
44  
45 432 anticipate that these scores will have low to moderate positive correlations. Convergent  
46  
47 433 validity will be supported by statistically significant positive correlations (with 95%  
48  
49 434 confidence interval) between the GSED scores and continuous contextual measures with  
50  
51 435 prior evidence of association with child development. Comparisons between "known  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 436 groups” will be made using the following variables: maternal education, home learning  
4  
5 437 opportunities, home environment, socioeconomic status (SES), maternal mental health  
6  
7 438 and child anthropometry, and stunting to determine if scores discriminate between high  
8  
9 439 and low categories for each variable using mean DAZ scores.

10  
11  
12 440 GSED scores at baseline and follow up will be correlated for predictive validity (positive  
13  
14 441 association between baseline and at 6 months) and mixed-effects linear regression used to  
15  
16 442 adjust for other contextual covariates and baseline scores.

### 20 443 **Secondary (Exploratory) Aims**

#### 23 444 *Reference Scores*

24  
25 445 We plan to develop a set of preliminary reference scores to facilitate comparison of DAZ  
26  
27 446 scores across countries. From the full validation study sample, a sub-sample of children  
28  
29 447 who have not experienced prior exposure to major known biological and environmental  
30  
31 448 risk factors is selected (i.e., “reference sub-sample”) (Table 3). Such an approach relies  
32  
33 449 on the assumption that the attainment of basic developmental milestones captured by the  
34  
35 450 GSED of children who are free of major risk factors is relatively similar globally (32).  
36  
37 451 To develop the reference scores, we will fit GAMLSS (33) to flexibly model both  
38  
39 452 conditional means, conditional standard deviations of scores, and, if necessary,  
40  
41 453 conditional skewness and kurtosis. We will test our assumption that the distribution of  
42  
43 454 scores is equivalent across sites by adding a site indicator at each moment of the  
44  
45 455 distribution, and testing site effects for their statistical significance. Where possible, we  
46  
47 456 will conduct standardization of scores to assist with the interpretation of scores by  
48  
49 457 pooling data across countries. We will report the corresponding parameters of the  
50  
51 458 GAMLSS model at appropriate ages.

#### 52 459 **Table 3. “Reference” sub-sample exclusion criteria (healthy sub-sample)**

Sample	Exclusion criteria
Minimum sub-sample of “reference” children per site N=522	<ol style="list-style-type: none"> <li>1. Below secondary maternal education (&lt;6 years of schooling)</li> <li>2. Birthweight less than 2500 gr</li> <li>3. Gestational age &lt; 37 completed weeks (259 days) and <math>\geq</math> 42 completed weeks (294 days) [assessed by ultrasound]</li> <li>4. Undernutrition (weight for age, length for age, OR weight for height Z score of less than <math>-2</math> on the WHO Child Growth Standards) at the time of developmental assessment</li> <li>5. Known severe congenital birth defect</li> <li>6. History of birth asphyxia OR neonatal sepsis requiring hospitalization</li> <li>7. Known neurodevelopmental disorder/ disability (Severe visual problems, seizures, hearing impairment) OR other chronic health problems (that is congenital heart disease)</li> </ol>

460

461 ***Adaptive testing***

462 We will determine whether adaptive testing is a feasible and valid option to measure  
 463 child development within the GSED (Box 2). Adaptive testing (34) is an administration  
 464 method that continually adapts to the level of the child’s performance, thereby reducing  
 465 test administration time. Previous simulations (35) indicated that theoretically substantial  
 466 gains in the precision of scores are possible when using adaptive testing even if  
 467 administering fewer items.

**Box 2: Adaptive testing validation methodology**

We investigate the feasibility by applying adaptive testing in addition to the traditional “fixed” GSED administration methods in the sub-sample designated for predictive validity analyses (N=502 per site) in three sites. The adaptive test is executed using tablets that are specially programmed to continually adjust child’s score after each item is administered, and to suggest the next item based on the answers already received (e.g., a more difficult item for a child with a higher score, an easier item for a child

with a lower score). Once the program establishes a reliable score, the administration is terminated. Both the adaptive test and the fixed test are administered with the same sub-sample during two separate visits alternating the order of administration to investigate the difference between the two modes of administration. We will investigate the following: the variance of user experience as a function of the average difficulty of milestones (leniency); the comparison of the D-score distribution under the adaptive testing procedure with the D-score distribution under the fixed GSED administration (using a z-test to assess the equivalence of the two modalities and plotting the results to show the level of concordance); and relation of the difference between the two D-scores to background variables.

468

#### 469 *Psychosocial Form*

470 The PF measure is in an early stage and will undergo exploratory and confirmatory factor  
471 analyses to assess the internal scale structure. Associations between items and factor  
472 scores with variables suggesting a high risk of psychosocial stress, such as family  
473 resilience, social support, and family and community violence, in addition to GSED SF  
474 and LF scores (concurrent validity measures) will be examined.

#### 475 **ETHICS AND DISSEMINATION**

476 The study complies with the International Ethical Guidelines for Biomedical Research  
477 Involving Human Subjects (36) and received ethical approval from the appropriate body  
478 in each site [Bangladesh – Projahnmo Research Foundation Institutional Review Board;  
479 Brazil – University Hospital, São Paulo (HU-USP); Cote d’Ivoire – Comité National  
480 D’Ethique des Sciences de la Vie et de la Sante (CNESVS); Pakistan – The Aga Khan  
481 University Ethics Review Committee; The Netherlands – Institutional Review Board  
482 TNO, Netherlands Organisation for Applied Scientific Research; The Republic of China

1  
2  
3 483 – IRB of Shanghai Children’s Medical Center Affiliated to Shanghai Jiao Tong University  
4  
5 484 School of Medicine; United Republic of Tanzania – Zanzibar Health Research Institute]  
6  
7  
8 485 and within WHO (protocol GSED validation 004583 approved on 20.04.2020). The  
9  
10 486 findings of the study will be disseminated following a comprehensive dissemination  
11  
12 487 strategy to reach a diverse range of stakeholders at the local, national and international  
13  
14  
15 488 level.

## 17 18 489 **DISCUSSION**

19  
20  
21 490 The validation of the GSED SF and LF is a meticulous and systematic global process that  
22  
23 491 introduces an innovative common metric (the D-score) that countries can use to track the  
24  
25 492 progress of child development among populations of young children and will enable  
26  
27 493 countries to adapt, modify, and evaluate their policies and programs to ensure that young  
28  
29 494 children are effectively and equitably reaching their development potential and building  
30  
31 495 the human capital needed for sustainable development. Additional attention is required on  
32  
33 496 understanding young children’s responses to psychosocial challenges within global  
34  
35 497 contexts. The exploration of the GSED PF introduces an important opportunity to capture  
36  
37 498 the non-normative developmental patterns among young children that are potential  
38  
39 499 precursors to behaviour and psychiatric problems. The GSED validation has several  
40  
41 500 important design, methodological and implementation characteristics that illustrate the  
42  
43 501 rigour required to validate instruments to measure child development globally. First, it is  
44  
45 502 conducted in seven countries with multiple linguistic, cultural and socioeconomic  
46  
47 503 backgrounds. Second, GSED is implemented through an app-based data collection  
48  
49 504 system that facilitates the implementation by reducing recording and transcribing errors  
50  
51 505 and other common pitfalls of paper-based instruments. Third, this study builds on the best  
52  
53 506 practices in validation by including a broad spectrum of psychometric methodologies  
54  
55 507 (concurrent, predictive, convergent, and discriminant validity, test-retest and inter-rater  
56  
57  
58  
59  
60

1  
2  
3 508 reliability, differential item functioning, and differential test functioning). Fourth, a  
4  
5 509 secondary aim builds the evidence for the creation of preliminary reference scores for the  
6  
7 510 SF and LF, based on a sub-sample with minimal exposure to major biological risk factors  
8  
9 511 and to the extent possible, minimal social and environmental risk factors. Fifth, we are  
10  
11 512 validating an adaptive testing design that can streamline administration by tailoring and  
12  
13 513 reducing the number of items required to obtain a valid score. Sixth, we are testing a new  
14  
15 514 measure of young children's non-normative psychosocial development.

16  
17 515 One notable difference between the GSED SF and LF measures and other instruments of  
18  
19 516 early child development is that the GSED measures are based on a unidimensional model  
20  
21 517 of development through measurement approaches that are universally applicable across  
22  
23 518 cultures. The measures do not follow the common multidimensional approach with  
24  
25 519 separate scores for different domains or contexts. Our validation study intends to  
26  
27 520 demonstrate that this model provides valid, reliable, and interpretable data globally. The  
28  
29 521 GSED SF and LF may exclude some items that measure development in cultural or  
30  
31 522 setting-specific ways, because the focus is on selecting items that are meaningful for  
32  
33 523 understanding child development within any given setting. If specific aspects need to be  
34  
35 524 captured locally, to increase cultural relevance we suggest that the GSED measures are  
36  
37 525 lightly adapted with country or culture-specific item props (in agreement with WHO)  
38  
39 526 and/or through the administration of additional measures.

40  
41 527 There are several limitations to our study. Although we are validating the GSED in seven  
42  
43 528 countries, including one high income setting, three sites are resource-limited  
44  
45 529 (Bangladesh, Pakistan and United Republic of Tanzania). Additional evidence may be  
46  
47 530 needed in high income countries to expand the validity and reliability of the GSED to  
48  
49 531 population-representative samples in additional countries. Second, the GSED has been  
50  
51 532 created using items that fit a Rasch model demonstrating developmental progress across  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 533 ages 0-3 years (9). This univariate model makes strict assumptions and may exclude  
4  
5 534 items that do not show strong age gradients or items that measure development in a  
6  
7 535 culturally-specific ways. Third, GSED was developed to address population and  
8  
9 536 programmatic level evaluations of early child development globally. The GSED is  
10  
11 537 presently not being validated for screening or diagnosing individual children. Finally, our  
12  
13 538 three secondary aims are exploratory, and will require further research, including  
14  
15 539 developing global standards to replace our preliminary reference scores with more  
16  
17 540 specific global norms, as in the Multi-country Growth Reference Standards for children's  
18  
19 541 weight and height. In the future we plan to collect additional data from countries using  
20  
21 542 strict inclusion/exclusion criteria (e.g., additional considerations around environmental  
22  
23 543 risk and protective factors) to further validate our initial reference scores. Similarly, we  
24  
25 544 plan to conduct further work to explore the functionality, reliability, validity, and  
26  
27 545 invariance of the PF. Lastly, as the GSED SF and LF scores are meant to be interpreted  
28  
29 546 and used for population-level measurement, we plan to expand the work towards  
30  
31 547 understanding of how the GSED package could be modified and validated to be able to  
32  
33 548 identify individual children at risk of developmental delays and disorders.  
34  
35  
36  
37  
38  
39  
40 549  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



1  
2  
3 550 **LIST OF ABBREVIATIONS**

4 551

5 552 Bayley-III Bayley Scales of Infant and Toddler Development

6

7 553 BRS Brief Resilience Scale

8

9 554 CI Confidence interval

10

11 555 CPAS Childhood Psychosocial Adversity Scale

12

13 556 DAZ Development for Age Z-score

14

15 557 DHS Demographic and Health Surveys

16

17 558 DIF Differential item functioning

18

19 559 DTF Differential test functioning

20

21 560 D-score Developmental Score

22

23 561 EAP Expected a posteriori

24

25 562 ECD Early child development

26

27 563 ECDI 2030 Early Childhood Development Index 2030

28

29 564 FCI Family Care Indicators

30

31 565 FSS Family Support Scale

32

33 566 GAMLSS Generalized Additive Model for Location Scale and Shape

34

35 567 GSED Global Scales for Early Development

36

37 568 HAZ Height-for-age z-score

38

39 569 HOME Home Observation for Measurement of the Environment Inventory

40

41 570 ICC Intraclass correlation coefficient

42

43 571 LF Long Form

44

45 572 LMIC Low- and middle-income country

46

47 573 MICS Multiple Indicator Cluster Surveys

48

49 574 OSF Open Science Framework

50

51 575 PF Psychosocial Form

52

53 576 PHQ-9 The Patient Health Questionnaire-9

54

55 577 SDG Sustainable Development Goals

56

57 578 SES Socioeconomic status

58

59 579 SF Short Form

60

580 SOPs Standard operating procedures

581 WAZ Weight-for-age z-score

1  
2  
3 582 **STATEMENTS**  
4

5 583 **Ethics approval:** The study complies with the International Ethical Guidelines for  
6  
7 584 Biomedical Research Involving Human Subjects and received ethical approval from the  
8  
9 585 appropriate Ethics Review Committee (ERC) at the World Health Organization (WHO)  
10  
11 586 (protocol ID GSED validation 004583 approved on 20.04.2020) and in each site  
12  
13 587 [Bangladesh – Projahnmo Research Foundation Institutional Review Board; Brazil –  
14  
15 588 University Hospital, São Paulo (HU-USP); Cote d’Ivoire – Comite National D’Ethique  
16  
17 589 des Sciences de la Vie et de la Sante (CNESVS); Pakistan – The Aga Khan University  
18  
19 590 Ethics Review Committee; The Netherlands – Institutional Review Board TNO,  
20  
21 591 Netherlands Organisation for Applied Scientific Research; The Republic of China – IRB  
22  
23 592 of Shanghai Children’s Medical Center Affiliated to Shangai Jiao Tong University  
24  
25 593 School of Medicine; United Republic of Tanzania – Zanzibar Health Research Institute].  
26  
27  
28  
29

30 594 **Patient consent:** Written consent is gathered by all study participants. Informed consent  
31  
32 595 forms are written to be easily understood by lay persons, enabling them to understand the  
33  
34 596 aims, procedures and potential risks of participation and have been approved by the  
35  
36 597 WHO ERC. For participants who are illiterate, culturally acceptable options including  
37  
38 598 witnessed oral consent and a thumbprint in lieu of a signature are accepted by the WHO  
39  
40 599 and local ERCs.  
41  
42  
43

44 600 **Data sharing:** Not applicable (protocol paper)  
45

46 601 **Author Contributions:** All authors contributed substantively to this work. VC was the  
47  
48 602 lead author in drafting the manuscript in addition to the technical contributions to the  
49  
50 603 study protocol conceptualization and development. TD led the conceptualization of the  
51  
52 604 study, MG, MMB, MJ, and PK contributed significantly to the conceptualization of the  
53  
54 605 study design and methodology, drafted sections of the protocol and related manuscript;  
55  
56 606 GL, GMc, [focus on psychometric properties], DMc, JS [focus on preliminary reference  
57  
58  
59  
60

1  
2  
3 607 scores] MW [focus on testing of psychosocial form], SvB and IE [focus on adaptive  
4  
5 608 testing] came to consensus on statistical analysis plan, determined the sample size  
6  
7  
8 609 calculations, and drafted the related parts of the manuscript relevant to their specific  
9  
10 610 expertise ; AN, AR, KH, and AW drafted substantial pieces of the manuscript related to  
11  
12 611 sampling frame, study measures and implementation. SA, AD, RA, AB, FJ, YS, IN, RK,  
13  
14 612 SS, AZ, MPM, YZ, FT, ARD, AB, JZ, AH, GF, SD, NSK, FB, FJ, and MRC contributed  
15  
16 613 to the adaptation of the study protocol for feasibility and on-the-ground implementation,  
17  
18 614 focusing on manuscript write up related to site-specific descriptions. All above authors, in  
19  
20 615 addition to RK, MMP, and RN reviewed and edited the study protocol and the  
21  
22 616 manuscript. All authors read and approved the final manuscript submission.  
23  
24  
25

26 617 **Acknowledgements:** None  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 618 **COMPETING INTERESTS**  
4

5 619 The authors declare that they have no competing interests.  
6

7  
8 620 **FUNDING**  
9

10 621 This work was supported (alphabetical order) by Bernard van Leer Foundation, Bill &  
11

12 622 Melinda Gates Foundation, Children's Investment Fund Foundation, Jacobs Foundation  
13

14 623 and King Baudouin Foundation, United States. The funders provided financial support.  
15

16 624 The design, implementation, and writing of the manuscript were led by the World Health  
17

18  
19 625 Organization.  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

626 **REFERENCES:**

- 627 1. Clark H, Coll-Seck AM, Banerjee A, Peterson S, Dalglish SL, Ameratunga S, et al. A Future  
628 for The World's Children? A WHO–UNICEF–Lancet Commission. *The Lancet*.  
629 2020;395(10224):605-58.
- 630 2. United Nations. <https://sdgs.un.org/goals/goal4> [cited 2021 November]. Department of  
631 Economic and Social Affairs, Sustainable Development].
- 632 3. Fernald LC, Prado E, Kariger P, Raikes A. A Toolkit for Measuring Early Childhood  
633 Development in Low- and Middle-Income Countries. Washington DC 20433: International Bank  
634 for Reconstruction and Development / The World Bank; 2017.
- 635 4. UNICEF. [https://data.unicef.org/resources/early-childhood-development-index-2030-  
636 ecdi2030/](https://data.unicef.org/resources/early-childhood-development-index-2030-ecdi2030/) [updated January 2021].
- 637 5. Faruk T, King C, Muhit M, Islam MK, Jahan I, Baset KU, et al. Screening tools for early  
638 identification of children with developmental delay in low- and middle-income countries: a  
639 systematic review. *BMJ Open*. 2020;10(11):e038182.
- 640 6. Weber AM, Rubio-Codina M, Walker SP, Van Buuren S, Eekhout I, Grantham-McGregor  
641 SM, et al. The D-Score: A Metric for Interpreting The Early Development of Infants and Toddlers  
642 across Global Settings. *BMJ Global Health*. 2019;4(6):e001724.
- 643 7. McCoy DC, Waldman M, Team CF, Fink G. Measuring Early Childhood Development at a  
644 Global Scale: Evidence from the Caregiver-Reported Early Development Instruments. *Early  
645 Childhood Research Quarterly*. 2018;45:58-68.
- 646 8. GSED Team. A New Measure of Development in Children from Birth to Age 3 at  
647 Population Level: The Global Scale for Early Development (GSED). *Early Childhood Matters*.  
648 2019.
- 649 9. Van Buuren S, Eekhout I, McCray G, Waldman M, Lancaster G, Black M, et al. A Scale for  
650 Tracking Child Development from Directly Observed and Caregiver-Reported Data. [Manuscript  
651 in preparation]
- 652 10. McCray G, McCoy D, Kariger P, Janus M, Black M, Chang-Lopez S, et al. The Creation of  
653 the Global Scales of Early Development (GSED) for 0-3 year old Children: Combining Subject  
654 Matter Expert judgements with big data. [Manuscript in preparation]
- 655 11. Van Buuren S, Eekhout I. Child development with the D-score: turning milestones into  
656 measurement [version 1; peer review: 1 approved with reservations]. *Gates Open Research*.  
657 2021;5(81):1-75.
- 658 12. Rasch G. Probabilistic Models for Some Intelligence and Attainment Tests. MESA Press,  
659 5835 S. Kimbark Ave., Chicago: ERIC; 1993.
- 660 13. Nizar A, Kaur R, Gladstone M, Lancaster G, Baqui A, Cavallera V, et al. Assessment of  
661 Acceptability and Feasibility of the Global Scales for Early Development (GSED) Across Three  
662 Settings for 0 to 3-year-old Children. [Manuscript in preparation]
- 663 14. Streiner DL, Norman GR, Cairney J. *Health measurement scales: a practical guide to their  
664 development and use*: Oxford University Press, USA; 2015.
- 665 15. Lancaster GA, Kariger P, McCray G, Janus M, Gladstone M, Cavallera V, et al. Conducting  
666 a Feasibility Study in a Global Health Setting for Constructing a Caregiver-Reported  
667 Measurement Tool: An Example in Infant and Young Child Development: SAGE Publications Ltd;  
668 2020.
- 669 16. McCoy DC, Sudfeld CR, Bellinger DC, Muhihi A, Ashery G, Weary TE, et al. Development  
670 and Validation of an Early Childhood Development Scale for Use in Low-Resourced Settings.  
671 *Population Health Metrics*. 2017;15(1):1-18.
- 672 17. Walker SP, Wachs TD, Gardner JM, Lozoff B, Wasserman GA, Pollitt E, et al. Child  
673 Development: Risk Factors for Adverse Outcomes in Developing Countries. *The Lancet*.  
674 2007;369(9556):145-57.

- 1  
2  
3 675 18. Jones PC, Pendergast LL, Schaefer BA, Rasheed M, Svensen E, Scharf R, et al. Measuring  
4 676 Home Environments Across Cultures: Invariance of the HOME Scale Across Eight International  
5 677 Sites from the MAL-ED Study. *Journal of School Psychology*. 2017;64:109-27.  
6 678 19. Kariger P, Frongillo EA, Engle P, Britto PMR, Sywulka SM, Menon P. Indicators of Family  
7 679 Care for Development for Use in Multicountry Surveys. *Journal of Health, Population, and  
8 680 Nutrition*. 2012;30(4):472.  
9 681 20. Berens AE, Kumar S, Tofail F, Jensen SK, Alam M, Haque R, et al. Cumulative Psychosocial  
11 682 Risk and Early Child Development: Validation and Use of the Childhood Psychosocial Adversity  
12 683 Scale in Global Health Research. *Pediatric Research*. 2019;86(6):766-75.  
13 684 21. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The Brief Resilience  
14 685 Scale: Assessing the Ability to Bounce Back. *International Journal of Behavioral Medicine*.  
15 686 2008;15(3):194-200.  
16 687 22. Dunst C. The Family Support Scale: Reliability and Validity. *Journal of Individual, Family,  
17 688 and Community Wellness*. 1984;1(4):45-52.  
18 689 23. Moriarty AS, Gilbody S, McMillan D, Manea L. Screening and Case Finding for Major  
19 690 Depressive Disorder Using the Patient Health Questionnaire (PHQ-9): A Meta-Analysis. *General  
20 691 Hospital Psychiatry*. 2015;37(6):567-76.  
21 692 24. Bayley N. Bayley Scales of Infant and Toddler Development: Administration Manual, 3rd  
22 693 edition ed. United States of America: Psychorp; 2006.  
23 694 25. Jacklin L, Cockcroft K. The Griffiths Mental Developmental Scales: An Overview and a  
24 695 Consideration of their Relevance for South Africa. In: Laher S, Cockcroft K, editors. *Psychological  
25 696 Assessment in South Africa. Research and applications*: Wits University Press; 2013. p. 169-85.  
26 697 26. Wright BD, Masters GN. *Rating scale analysis: Rasch Measurement*. Chicago, IL.: MESA  
27 698 press; 1982.  
28 699 27. Bock RD, Mislevy RJ. Adaptive EAP Estimation of Ability in a Microcomputer  
29 700 Environment. *Applied Psychological Measurement*. 1982;6(4):431-44.  
30 701 28. Gladstone M, Lancaster G, McCray G, Cavallera V, Alves CR, Maliwichi L, et al. Validation  
31 702 of the Infant and Young Child Development (IYCD) Indicators in Three Countries: Brazil, Malawi  
32 703 and Pakistan. *International Journal of Environmental Research and Public Health*.  
33 704 2021;18(11):6117.  
34 705 29. Team RC. R: A language and environment for statistical computing. R Foundation for  
35 706 Statistical Computing, Vienna, Austria. <http://wwwR-project.org/>. 2016.  
36 707 30. Winston Chang, Joe Cheng, JJ Allaire, Carson Sievert, Barret Schloerke, Yihui Xie, Jeff  
37 708 Allen, Jonathan McPherson, Alan Dipert, Barbara Borges. shiny: Web Application Framework for  
38 709 R. R package version 1.7.2 2022 [Available from: <https://CRAN.R-project.org/package=shiny>. .  
39 710 31. Gwet KL. Computing Inter-Rater Reliability and its Variance in the Presence of High  
40 711 Agreement. *British Journal of Mathematical and Statistical Psychology*. 2008;61(1):29-48.  
41 712 32. Ertem IO, Krishnamurthy V, Mulaudzi MC, Sguassero Y, Balta H, Gulumser O, et al.  
42 713 Similarities and Differences in Child Development from Birth to Age 3 years by Sex and Across  
43 714 Four Countries: A Cross-sectional, Observational Study. *The Lancet Global Health*.  
44 715 2018;6(3):e279-e91.  
45 716 33. Rigby RA, Stasinopoulos DM. Generalized Additive Models for Location, Scale and Shape.  
46 717 *Journal of the Royal Statistical Society: Series C (Applied Statistics)*. 2005;54(3):507-54.  
47 718 34. Wainer H, Dorans NJ, Flaugher R, Green BF, Mislevy RJ. *Computerized Adaptive Testing:  
48 719 A Primer*: Routledge; 2000.  
49 720 35. Jacobusse G, Van Buuren S. Computerized Adaptive Testing for Measuring Development  
50 721 of Young Children. *Statistics in Medicine*. 2007;26(13):2629-38.  
51 722 36. Singh AN. Current Status of Research Ethics and Projecting Future Initiatives.  
52 723 *International Medical Journal*. 2018;25(6).  
53  
54  
55  
56  
57  
58  
59 724  
60

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

725

For peer review only

1  
2  
3 726 **FIGURE LEGEND**  
4

5 727 **Figure 1. Development chart**  
6

7  
8 728 Reproduced with permission from van Buuren S and Eekhout I (2021) (11)  
9

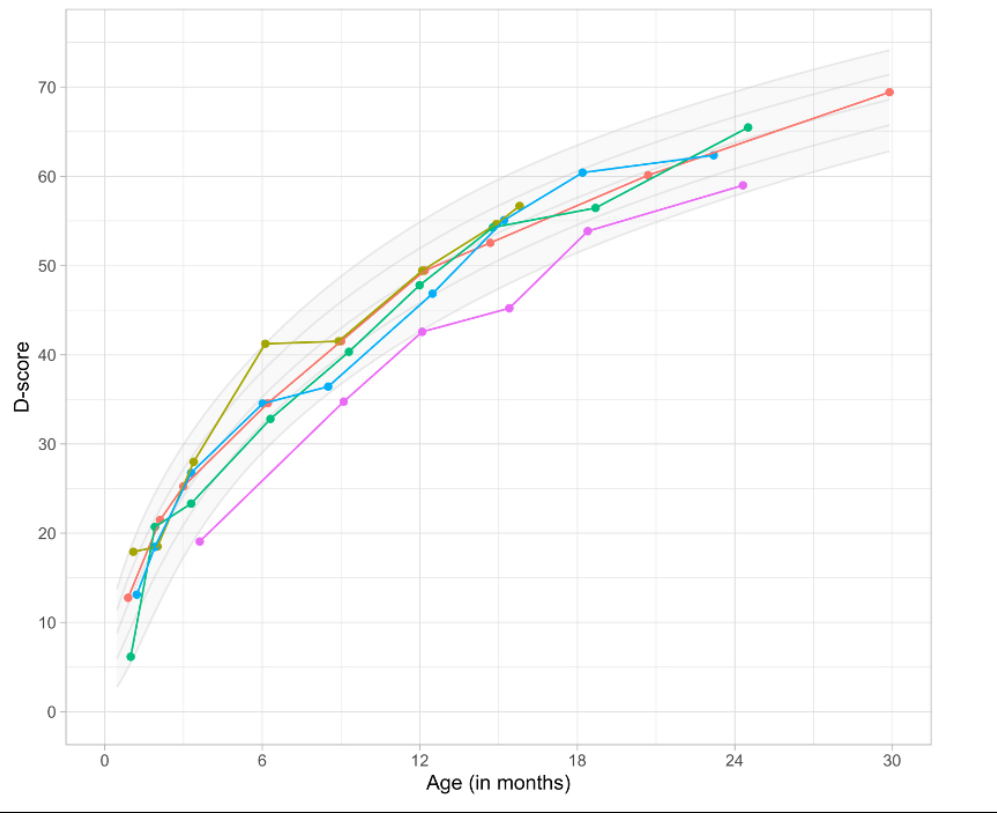
10 729  
11

12 730 **Figure 2. Study Sampling schema diagram**  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

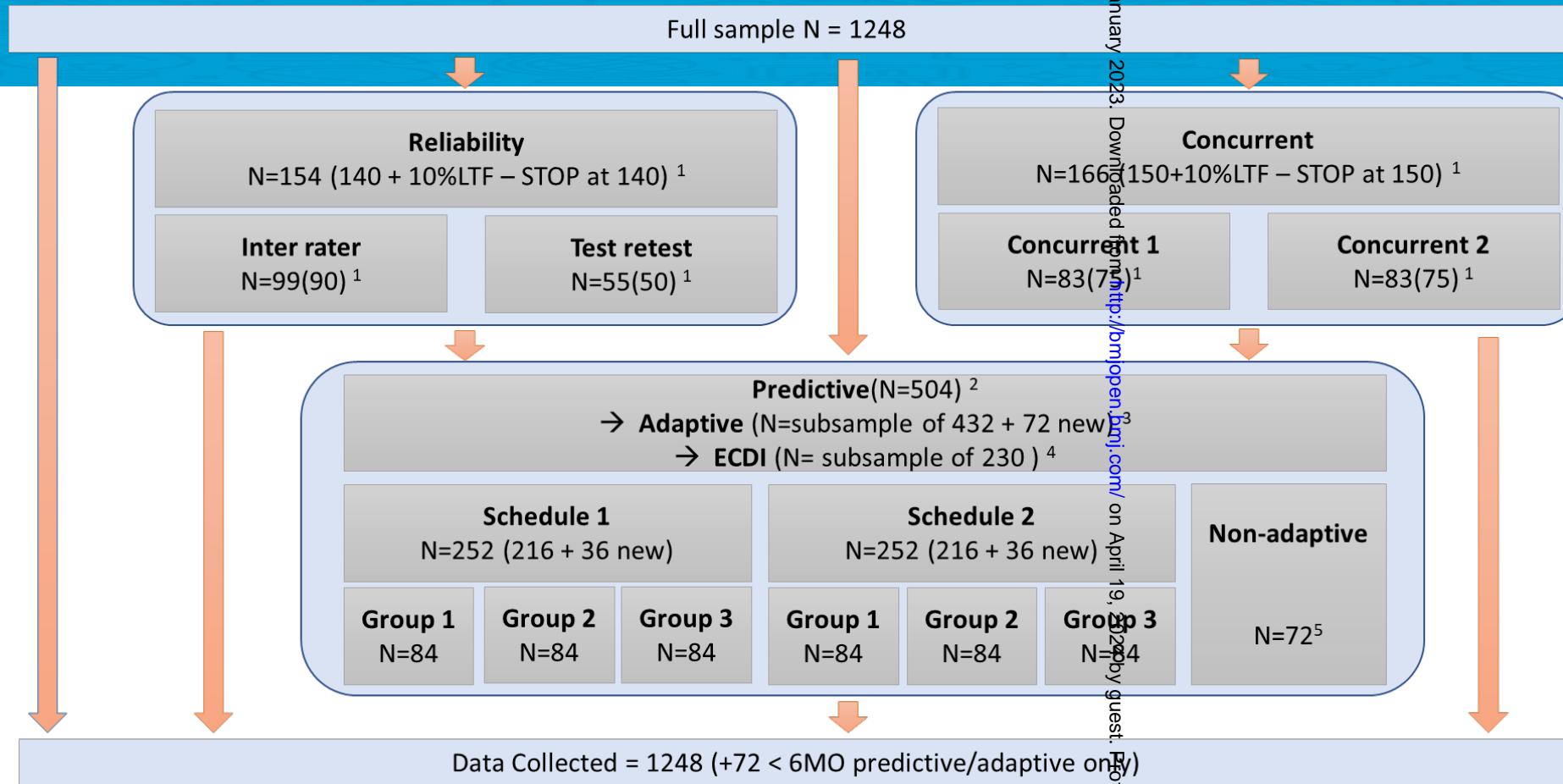


1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



152x125mm (144 x 144 DPI)

# GSED study sample



[1] The number inside parentheses is the number collected and the number outside is the number randomised to account for loss to follow-up

[2] Two additional participants have been added to the predictive to have equal numbers in each experimental group

[3] 72 new children between 2 weeks and 6 months of age have been added to the adaptive sample to ensure coverage at the lower ages.

[4] ECDI will only be done on N=230 Children between the ages of 2+ years at the time of the predictive data collection.

For peer review only - <http://bmjopen.bmj.com/site/about/guidelines.xhtml>

[5] The 72 oldest children (36-41 months) from the predictive sample will not be part of the adaptive sample.

2-062562 on 24 January 2023. Downloaded from <http://bmjopen.bmj.com/> on April 19, 2023 by guest. Protected by copyright.

## Supplementary file 1 – Sample size calculations and sampling frame

The sample size calculation for reliability is based on a confidence interval (CI) approach and the desired accuracy for the lower bound of the CI for the ICC estimates. In an Analysis of Variance (ANOVA) with 2-way random effects on a single score with 2 observations per subject (following Shrout and Fleiss, 1979) (27) and with a two-sided 95% confidence interval and an expected ICC of 0.875, the lower confidence interval for the inter-rater reliability sample (N=90) = 0.852. With the same parameters but an expected ICC of 0.90 the lower confidence interval for the test-retest reliability sample (N=50) is 0.830. We expect the ICC to be higher for the test-retest reliability than the inter-rater reliability as inter-rater estimates contain all the sources of error in the test-retest estimates, plus additional error between assessors (14).

To assess concurrent validity, a sample size of 150 per site produces a two-sided 95% CI 0.15-0.44, when the estimate of Pearson's product-moment correlation is 0.30, with an equal spread of participants tested across age and sex. The CI will be narrower when the data are combined across all seven countries. To assess predictive validity a sample size of 404 produces a two-sided 95% CI 0.65-0.75 when the estimate of Pearson's product-moment correlation is 0.70 between individual scores at baseline and at the 6-month follow-up. Allowing 20% dropout at follow up, a sample size of approximately 500 participants is required.

**Table S1. Sampling Frame**

Sample size per site by age and sex for total population (n=1248) which includes a minimum subsample of healthy 'reference' children (n=522)

Age (Days)	Sex	Total Sample size	Minimum sub-sample of reference children	Predictive validity sample (6-month follow-up; age at baseline)	Reliability: Inter-rater	Reliability: Test-Retest	Concurrent validity
15-30	Male	40	20	8	2	1	4
	Female	40	20	8	2	1	2
31-61	Male	40	12	8	1	1	2
	Female	40	12	8	2	1	2
62-91	Male	40	10	8	2	1	2
	Female	40	10	8	1	0	4
92-122	Male	36	9	8	2	1	2
	Female	36	9	8	2	1	2
123-152	Male	32	8	8	1	1	2
	Female	32	8	8	2	1	2

153-183	Male	28	8	8	1	0	4
	Female	28	8	8	1	1	2
184-213	Male	25	7	8	2	1	2
	Female	25	7	8	1	0	2
214-244	Male	23	7	8	1	1	2
	Female	23	7	8	2	1	4
245-274	Male	21	6	8	1	1	2
	Female	21	6	8	1	1	2
275-304	Male	19	6	8	2	0	2
	Female	19	6	8	1	1	2
305-335	Male	17	6	8	1	1	4
	Female	17	6	8	2	0	2
336-365	Male	16	6	7	1	1	2
	Female	16	6	7	1	1	2
366-396	Male	14	6	7	2	1	2
	Female	14	6	7	1	1	4
397-426	Male	13	6	7	1	0	2
	Female	13	6	7	2	1	2
427-457	Male	12	5	7	1	1	2
	Female	12	5	7	1	0	2
458-487	Male	11	5	7	2	1	4
	Female	11	5	7	1	1	2
488-517	Male	11	5	7	1	1	2
	Female	11	5	7	2	1	2
518-548	Male	10	5	7	1	0	2
	Female	10	5	7	1	1	4
549-578	Male	9	5	7	2	1	2
	Female	9	5	7	1	0	2
579-609	Male	9	5	7	1	1	2
	Female	9	5	7	2	1	2
610-639	Male	9	5	7	1	1	4
	Female	9	5	7	1	1	2
640-670	Male	9	5	7	2	0	2
	Female	9	5	7	1	1	2
671-700	Male	9	5	7	1	1	2
	Female	9	5	7	2	0	4
701-730	Male	9	5	7	1	1	2
	Female	9	5	7	1	1	2
731-761	Male	9	5	7	2	1	2
	Female	9	5	7	1	1	2
762-791	Male	9	5	6	1	0	4
	Female	9	5	6	2	1	2
792-822	Male	9	5	6	1	1	2
	Female	9	5	6	1	0	2
823-852	Male	9	5	6	2	1	2
	Female	9	5	6	1	1	2

853-883	Male	9	5	6	1	1	2
	Female	9	5	6	2	1	2
884-913	Male	9	5	6	1	0	2
	Female	9	5	6	1	1	2
914-944	Male	9	5	6	2	1	2
	Female	9	5	6	1	0	2
945-974	Male	9	5	6	1	1	2
	Female	9	5	6	2	1	2
975-1004	Male	9	5	6	1	1	2
	Female	9	5	6	1	1	2
1005-1035	Male	9	5	6	2	0	2
	Female	9	5	6	1	1	2
1036-1065	Male	9	5	6	1	1	2
	Female	9	5	6	2	0	2
1066-1096	Male	9	5	6	1	1	2
	Female	9	5	6	1	1	2
1097-1126	Male	9	5	0	0	0	0
	Female	9	5	0	0	0	0
1127-1157	Male	9	5	0	0	0	0
	Female	9	5	0	0	0	0
1158-1187	Male	9	5	0	0	0	0
	Female	9	5	0	0	0	0
1188-1218	Male	9	6	0	0	0	0
	Female	9	6	0	0	0	0
1219-1248	Male	9	6	0	0	0	0
	Female	9	6	0	0	0	0
1249-1279	Male	9	7	0	0	0	0
	Female	9	7	0	0	0	0
<b>TOTAL</b>		1248	522	504	*99	**55	***166

\*90 + ~10% Loss to follow up = 99; \*\*50 + ~10% Loss to follow up = 55; \*\*\*150 + ~10% Loss to follow up = 166

## Supplementary file 2 – Visit schedule

Table S2a. Visit Schedule for the GSED Validation Study (all sites except the Netherlands)

Main Study Only [No Sub-sample]	Inter- Rater Reliability Sub- Sample	Test- Retest Reliability Sub- Sample	Concurrent Sub- Sample 1 [LF First]	Concurrent Sub- Sample 2 [BSID III First]
<b>Visit 1 [At Home]</b>				
Eligibility and Consent	Eligibility and Consent	Eligibility and Consent	Eligibility and Consent	Eligibility and Consent
COVID Questionnaire	COVID Questionnaire	COVID Questionnaire	COVID Questionnaire	COVID Questionnaire
Contextual	Contextual	Contextual	Contextual	Contextual
GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]
GSED Psychosocial form [PF]	GSED Psychosocial form [PF]	GSED Psychosocial form [PF]	GSED Psychosocial form [PF]	GSED Psychosocial form [PF]
HOME Inventory or Family Care Indicators (FCI)	HOME Inventory or Family Care Indicators (FCI)	HOME Inventory or Family Care Indicators (FCI)	HOME Inventory or Family Care Indicators (FCI)	HOME Inventory or Family Care Indicators (FCI)
Anthropometrics*	Anthropometrics*	Anthropometrics*	Anthropometrics*	Anthropometrics*
<b>Visit 2 [At home, clinic, or other setting within 48 hours of visit 1] Note: For Concurrent Sample, the Visit is at the Clinic setting</b>				
Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
GSED Long form [LF]	GSED Long form [LF]	GSED Long form [LF]	<b>GSED Long form [LF]</b>	<b>BSID III</b>
CPAS	CPAS	CPAS	CPAS	-----
PHQ9	PHQ9	PHQ9	PHQ9	-----
Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale	-----
<b>Visit 3 [Setting and timing vary by sub-sample]</b>				
Visit 3 not required	Visit 3 [At home, clinic or other setting where the LF was completed- <b>within 24 hours</b> of the LF]	Visit 3 [At home, clinic or other setting where the LF was completed- this should happen <b>7 to 10 days after</b> LF]	Visit 3 [Clinic setting within <b>24- 72 hours</b> of the LF- can be done at same time as Visit 2 – taking child fatigue into consideration]	Visit 3 [Clinic setting within <b>24- 72 hours</b> of the BSID III - can be done at same time as Visit 2 – taking child fatigue into consideration]
	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
	GSED Short form [SF]	GSED Short form [SF]	<b>BSID III</b>	<b>GSED Long form [LF]</b>
	GSED Psychosocial form [PF]	GSED Psychosocial form [PF]	-----	CPAS
	GSED Long form [LF]	GSED Long form [LF]	-----	PHQ9
-----	-----	-----	-----	Family support & Resilience Scale

\* Anthropometrics may be done either at visit 1 or visit 2

**Table S2b: Visit Schedule for the GSED Validation Study (the Netherlands only)**

Main Study Only [No Sub-sample]	Inter-Rater Reliability Sub-Sample	Test-Retest Reliability Sub-Sample	Concurrent Sub-Sample 1 [LF First]	Concurrent Sub-Sample 2 [BSID III First]
<b>Session 1 [Online]</b>				
Eligibility and Consent	Eligibility and Consent	Eligibility and Consent	Eligibility and Consent	Eligibility and Consent
Contextual	Contextual	Contextual	Contextual	Contextual
GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]	GSED Short form [SF]
GSED Psychosocial scale [PS]	GSED Psychosocial scale [PS]	GSED Psychosocial scale [PS]	GSED Psychosocial scale [PS]	GSED Psychosocial scale [PS]
<b>Visit 1 [At clinic within 48 hours of session1]</b>				
Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
GSED Long form [LF]	GSED Long form [LF]	GSED Long form [LF]	<b>GSED Long form [LF]</b>	<b>BSID III</b>
Anthropometrics	Anthropometrics	Anthropometrics	Anthropometrics	Anthropometrics
<b>Session 2 [Online, Test-Retest of SF/PSY within 7 to 10 days of online session 1]</b>				
Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
COVID Questionnaire	COVID Questionnaire	COVID Questionnaire	COVID Questionnaire	COVID Questionnaire
-----	-----	GSED Short form [SF]	-----	-----
-----	-----	GSED Psychosocial scale [PS]	-----	-----
CPAS	CPAS	CPAS	CPAS	CPAS
PHQ9	PHQ9	PHQ9	PHQ9	PHQ9
Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale	Family support & Resilience Scale
Family Care Indicators (FCI)	Family Care Indicators (FCI)	Family Care Indicators (FCI)	Family Care Indicators (FCI)	Family Care Indicators (FCI)
<b>Visit 2 [At clinic, timing varies by sub-sample]</b>				
Visit 2 not required	Visit 2 [within 24 hours of the LF]	Visit 2 [7 to 10 days after LF]	Visit 2 [within 24- 72 hours of the LF- can be done at same time as Visit 1 – taking child fatigue into consideration]	Visit 2 [within 24- 72 hours of the BSID III - can be done at same time as Visit 1 – taking child fatigue into consideration]
	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]	Abbreviated Eligibility [Coversheet]
	GSED Long form [LF]	GSED Long form [LF]	<b>BSID III</b>	<b>GSED Long form [LF]</b>