

ADDITIONAL FILE 1

Table S1 Characteristics of International Developmental Assessments for Infants Aged 12 Months.

Tool	Subscales measured	Characteristics of normative sample	Cut-off score for neurodevelopmental delay	Assessment type	Assessors' qualifications	Administration time	Psychometric information
The Bayley Scales of Infant Development – III edition (BSID III) ^a	Cognitive, expressive and receptive communication, fine and gross motor, adaptive behaviour, social emotional.	Country: USA Year: January and October 2004 Sample Size: 12-month normative sample for cognitive, language and motor scales: 100 children (totally 1700 children aged 16 days through 43 months 15 days divided into 17 age groups of 100 children each). Normative sample for the social-emotional scale was based on 456 children and the adaptive behaviour scale was based on 1,350 children. Sample characteristics: Sample selected to match the 2000 United States census.	Multiple criteria: 25% delay in functioning when compared to same age peers; based on SD (< -1 SD i.e. cut-off thresholds of 85 for moderate impairment; < -2 SD i.e. cut-off thresholds of 70 for severe impairment) or performance of a certain number of months below the child's chronological age ¹ .	Structured direct administration	Child health specialists and healthcare professionals with training & credentials	60–90 minutes	Test-re test and inter-rater reliability: acceptable to high Internal consistency: high Predictive validity: high
The Malawi Developmental Assessment Tool (MDAT) ^b	Gross motor, fine motor, language and social.	Country: Malawi Year: June 2006 to July 2007 Sample size: 1426 normal healthy children aged 0 to 6 years Sample characteristics: Those born preterm at <32 weeks' gestation, or with significant malnutrition using WHO criteria, medical problems or significant neurodisability were excluded.	<25% (upper limit of lowest quartile); Plot children on MDAT normal reference ranges, corresponding to 25%, 50% and 75% of children passing each item.	Direct administration	Trained professionals and para-professionals	35 minutes	Test-re test and inter-rater reliability: Good, $k > 0.4$ Sensitivity: 97% Specificity: 82%
The Griffiths Mental Development Scales ^c	Locomotor, expressive and receptive language, personal-social, hand and eye coordination, performance, practical reasoning.	Country: UK and Ireland Year: 1960s, revised in 2015 Sample size: 1026 children Sample characteristics: National representative sample of children in UK; stratified according to geographical region and proportionate to the 1997 ONS population ratios ¹ .	Centiles and z -scores; z score < -2 or SD < -2 indicates significant developmental delay on that subscale.	Structured direct administration	Healthcare professionals who have completed a 5 day accredited training course	60 minutes	The publishers state that, except for Scale E (performance), in children aged less than 48 months the coefficients 'all comfortably exceed the minimum acceptable value of 0.70'
Pre-school version of	Problem and syndrome scales	Country: USA Year: 1979, 2000	>93^d centile is abnormal (norms vary according to	Caregiver report	Parents or teachers,	10 to 20 minutes	Test-re test and inter-rater reliability:

Child Behavior Checklist (CBCL) ^d	for emotionally reactive; anxious/depressed; somatic complaints; withdrawn; sleep problems; attention problems; aggressive behaviour.	Sample size: Originally normed on 1728 US children. Sample characteristics: - Note: Multicultural norms available.	societies, and map onto the Diagnostic and Statistical Manual for Mental Disorders). Any score that falls below the 93 rd centile is considered normal, scores between the 93-97 th centile are borderline clinical, and any score above the 97 th centile is considered to be in the clinical range.		scoring restricted to qualified professionals		kappa=0.8; 0.61 Internal consistency: good Predictive validity: high Sensitivity: 84%
The Rapid Neurodevelopmental Assessment (RNDA) ^e	Primitive reflexes, gross motor, fine motor, vision, hearing, speech, cognition, behaviour, and seizures.	Country: Bangladesh Year: 2010 Sample size: 81 children aged ≥3 to 24 months in urban (n = 47) and rural (n = 34) community-based populations Sample characteristics: 15% did not 'look' properly nourished, parental concerns regarding child development were expressed in 8% and for 50% at least one parent was illiterate.	< -2 SD: threshold for severe impairment; < -1 SD: threshold for mild impairment. If low scores in >1 domain; the child is classified as having 'any' neurodevelopmental impairment.	Direct administration	Trained professionals and para-professionals who have completed an accredited training course	30 minutes	Inter-rater reliability: Good to excellent k=0.63-1.00 Validity: The authors state "good concurrent validity (ie, significantly lower mean Mental Development Index and Psychomotor Development Index scores) for children with >1 neurodevelopmental impairment and for children with impairments in most functional domains, compared with children with no impairments".
The Ages and Stages Questionnaire III edition (ASQ III) ^f	Communication, gross motor, fine motor, problem solving and personal-social.	Country: USA Year: June 2008 Sample Size: 15,138 children. Sample characteristics: 76% of the sample had one or no known risk factor, 19% had 2 risk factors and 4% 3 or more risk factors. Risk factors were defined as extreme poverty, maternal age ≤19 years, maternal education <12 th grade; involvement of child protective services with the family for abuse and/or neglect; medical risk, including prematurity; and infant's birth weight	< -2SD	Caregiver report	Parents, caregivers, home visitors; requires a 6 th grade reading level.	15-20 minutes	Test-re test and inter-rater reliability: high Internal consistency: moderate Predictive validity: moderate Sensitivity: 38% to 90% Specificity: 81 to 90%

		less than 3 pounds, 5 ounces.					
The Parents' Evaluation of Developmental Status (PEDS) ⁹	General development.	Country: USA & Canada Year: 2012 Sample size: 47531 families, of which 13,523 were aged 0-11 months Sample characteristics: Families from varying backgrounds, including SES and ethnicity.	Table for using scores to identify parental difficulties, non-significant concerns, one significant concern or two or more significant concerns by shading boxes based on scores. These are then used to select associated algorithms for further screening and/or referral.	Caregiver report	Parents	2-10 minutes	Reliability: 94% acceptable Sensitivity: 74% to 80% Specificity: 70 to 80%
Caregiver-reported Early Developmental Instruments (CREDI) ^h	Long Form: Motor, cognitive, language, social-emotional and overall. Short form: Overall development.	Countries: Brazil, Cambodia, Chile, Colombia, Ghana, Guatemala, India, Nepal, Philippines and USA Year: 2017-2018 Sample size: 7807 children aged 0-35 months Sample characteristics: Children with an "ideal home environment" defined through maternal educational attainment (college or higher), and the number of activities done by adults with the child in the last 3 days (at least 4 out of the 6 MICS home stimulation activities); authors acknowledge that "although the data were representative for local populations in Brazil, Ghana, Tanzania, and Zambia, the overall sample is not representative of any country or a global population of children".	Raw scaled scores, norm referenced standardised scores , and z scores .	Caregiver report	Parents	Short Form <5 minutes, Long Form 15 minutes	Test-re test reliability: 0.62 Internal consistency: 0.84 for 12-17 months Correlation with other measures (ASQ, BSID, INTER-NDA, McArthur Bates Inventory, PRIDI): r=0.2-0.4
Denver Development Screening Test II (DDST II) ⁱ	Personal social, fine motor adaptive, language and gross motor.	Country: USA Year: 1980s and 1990 Sample size: 2096 children Sample characteristics: Children from Colorado; based on 1988, 1989 and then the 1990 US census population.	Centile ranks (25th, 50th, 75th, and 90 th) are displayed as bar graphs and reflect the ages at which 25%, 50%, 75% and 90% of typically developing children in the standardisation sample completed the task. Overall categories: Normal, and suspect.	Direct administration	Trained professionals and paraprofessionals	15-20 minutes	Test-re test reliability: 0.9 Inter-rater reliability: 0.99 No validity study available.
Guide for Monitoring Child Development (GMCD) ^j	Parental concerns, expressive language and communication,	Country: Turkey Year: 1980s and 1990 Sample size: 4949 children up to 42 months of age from primary healthcare centres in Argentina, India, South Africa and Turkey	Paper 1: <10th centile for any domain; if a child did not demonstrate ≥ 1 of the age-appropriate milestones, the GMCD interpretation was	Open ended caregiver interview	Trained providers completing the GMCD Provider	7 \pm 2.3 minutes	Inter-rater reliability: 0.83-0.88 Internal consistency: 0.28-0.91 Sensitivity: 0.71-0.94

	receptive language, relationship (social-emotional), play (social-emotional and cognitive) and self-help.	Sample characteristics: Children were born healthy singletons with birth weight ≥ 2500 g and gestational age ≥ 37 weeks, with postnatal growth was between the 5 th and 95 th centiles since birth; free from chronic health and developmental problems or anemia (Hemoglobin < 10.5 g/dL) at recruitment	classified as "requires follow-up evaluation with or without intervention." Paper 2: thresholds based on BSID-III cut-offs		Training Program		Specificity: 0.69-0.82
Development Assessment Scale for Indian Infants (DASII) ^k	Mental and Motor Development	Country: India Year: 1996 Sample size: 513 aged 1 to 30 months Sample characteristics: Certified as normal and healthy by a pediatrician	Developmental quotients calculated from motor age (MoA) and mental age (MeA), percentile ranks for developmental quotients	Structured direct administration	Child health specialists and healthcare professionals with training & credentials	30-40 minutes	Inter correlation between motor and mental performance: 0.24 to 0.62 Median reliability Index: 0.88 (Motor), 0.91 (Mental)
Test de Aprendizaje y Desarrollo Infantil (TADI) ^l	Cognition, motor, language, and social-emotional.	Country: Ecuador Year: 2016-2020 Sample size: 411 children aged 12 to 36 months Sample characteristics: Convenience sample	Ecuador norm based z score	Direct administration, caregiver reported	Trained specialists	20-30 minutes	Test re test reliability: Good, z score 0.19 Internal consistency: 0.97 Validity: Concordance of 90% with expert judgment of the child's abilities
Indicators of Infant and Young Child Development (IYCD) ^m	Fine motor, gross motor, receptive language, expressive language, and socio-emotional.	Country: Meta-data analysis of datasets from Costa Rica, Nicaragua, Paraguay, Peru, Bangladesh, India, Indonesia, Kenya, Malawi, and Tanzania. Year: 2014-2018 Sample size: 14 cross sectional datasets representing data from 7 tools from 21 083 children from 10 low/middle- income countries Sample characteristics: authors were in broad agreement as to the important domains of development, which aligned with the review findings; sampled children came from diverse socioeconomic backgrounds with the highest proportions classified as 'normal' by height for age z-score (HAZ) and weight for age z-score (WAZ); and most mothers had been educated to at least primary level.	Item response theory curves from fitted models depicting the ages at which 10% (lower limit), 50% (diamond) and 90% (upper limit) of children pass the item for each item group.	Caregiver report	Not specified	Not reported	Feasibility study ongoing, validity study imminent.
Battelle's	Cognitive,	Country: USA	Basal and ceiling level for	Structured	Child health	90	Test-re test and inter-

Development Inventory (BDI) ⁿ	adaptive (self-help), motor, communication, and personal-social development	Year: c1988 Sample size: 50 children aged 12-17 months Sample characteristics: 75% urban, 25% rural, 50% girls, 50% boys, 84% Caucasian, 16% ethnic minorities.	domain scores, developmental quotients , expressed as age equivalents (in months)	administration, observation, and interviews with parents or other sources	specialists and healthcare professionals with training & credentials	minutes	rater reliability: 0.99 Internal consistency: >0.85 Construct validity: High Concurrent validity: r=0.78-0.93 with the Vineland Social Maturity Scale and the Developmental Activities Screening Inventory; r=0.40-0.61 with the the Stanford-Binet Intelligence Test
Mullen Scales of Early Learning (MSEL) ^o	Gross Motor and four "cognitive scales": Visual Reception, Fine Motor, Expressive Language, and Receptive Language	Country: USA Year: 1981-1989 Sample size: 1849 children aged 0-68 months Sample characteristics: not mentioned	Raw domain scores compared against age equivalents . "Cognitive" scores summarized into an Early Learning Composite . Scores mapped to obtain percentile ranks and age equivalent scores.	Structured direct administration	Child health specialists and healthcare professionals with training & credentials	30-45 minutes	Limited evidence for concurrent, content and construct validity.
Comprehensive Developmental Inventory for Infants and Toddlers (CDIIT) ^p	Cognition, motor skills, language, self help and social development	Country: Taiwan Year: c1998 Sample size: 3703 aged 3 to 72 months Sample characteristics: The authors state that the sample was "randomly selected according to age, sex, urbanization, and geographic regions in Taiwan"	Items are scored 0 or 1 (fail or pass), developmental quotients (DQs) calculated based on Taiwanese norms. DQ ≥ 85 normal, 70-84 borderline, <70 indicates delay.	Structured direct administration	Trained specialists	Not reported	Test-re test and inter-rater reliability: ICC 0.76-1.00 Internal consistency: 0.75-0.99 Sensitivity: 87% Specificity: 97%

ICCs = interclass correlations.

^aBSID-III: Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III); ©Pearson. Available from: <http://www.pearsonclinical.co.uk>

^bMDAT: The Malawi Developmental Assessment Tool; Gladstone M, Lancaster GA, Umar E, Nyirenda M, Kayira E, van den Broek NR, Smyth RL. The Malawi Developmental Assessment Tool (MDAT): the creation, validation, and reliability of a tool to assess child development in rural African settings. *PLoS medicine* 2010; **7**: e1000273.

^cGriffiths: Griffiths Mental Development Scales; Luiz D, Barnard A, Knoesen N, Kotras N, Horrocks S, McAlinden P, Challis D, O'Connell R. Griffiths Mental Development Scales—Extended Revised: Two to Eight Years: Administration Manual. Hogrefe, Oxford, UK 2006.

^dCBCL: Preschool Child Behavior Checklist; Achenbach TM. Manual for the Child Behavior Checklist/4-18 and 1991 profile. University of Vermont, Department of Psychiatry 1991.

^eRNDA: Rapid Neurodevelopmental Assessment; Khan NZ, Muslima H, Begum D, Shilpi AB, Akhter S, Bilkis K, Begum N, Parveen M, Ferdous S, Morshed R, Batra M. Validation of rapid neurodevelopmental assessment instrument for under-two-year-old children in Bangladesh. *Pediatrics*. 2010 Apr 1;125(4):e755-62.

^fASQ-III: Ages & Stages Questionnaires[®], Third Edition; <https://agesandstages.com/products-pricing/asq3/>

⁹PEDS: Parents' Evaluation of Developmental Status. Glascoe FP. Collaborating with parents: Using Parents' Evaluation of Developmental Status to detect and address developmental and behavioral problems. Ellsworth & Vandermeer Press; 1998. Available at: <http://forepath.org/>.

¹CREDI: Caregiver Reported Early Developmental Instruments; McCoy DC, Waldman M, Team CF, Fink G. Measuring early childhood development at a global scale: evidence from the Caregiver-Reported early development instruments. *Early childhood research quarterly*. 2018 Oct 1;45:58-68. <https://sites.sph.harvard.edu/credi/>

¹DDST-II: The Denver Developmental Screening Test IIInd edition; Frankenburg WK, Dodds J, Archer P, Shapiro H, Bresnick B. The Denver II: a major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics* 1992; 89: 91-7.

¹GMCD: A Guide for Monitoring Child Development; Ozturk Ertem, I., Krishnamurthy, V., Mulaudzi, M.C., Sguassero, Y., Bilik, B., Srinivasan, R., Balta, H., Gulumser, O., Gan, G., Calvocoressi, L., Johnson, B., Shabanova, V. and Forsyth, B.W. (2019), Validation of the International Guide for Monitoring Child Development demonstrates good sensitivity and specificity in four diverse countries. *Acta Paediatr*, 108: 1074-1086.

¹DASII: Developmental Assessment Scales for Indian Infants; Phatak P. Mental and Motor Growth of Indian Babies (1 Month-30 Months).(Longitudinal Growth of Indian Children). Final Report. <https://www.manashakti.org/tests/developmental-assessment-scales-indian-infants>

¹TADI: Test de Aprendizaje y Desarrollo Infantil; López Vanegas, N. y Peñafiel Aguirre, T. (2020). Adaptación y validación del test de Aprendizaje y desarrollo infantil "TADI" en el GAD de Calderón. Trabajo de titulación previo a la obtención del Título de Psicólogo Infantil y Psicorehabilitación. Carrera de Psicología Infantil y Psicorehabilitación. Quito: UCE. 190 p. <http://www.dspace.uce.edu.ec/bitstream/25000/20829/1/T-UCE-0007-CPS-245.pdf>

¹IYCD: WHO Indicators of Infant and Young Child Development; Lancaster GA, McCray G, Kariger P, *et al.* Creation of the WHO Indicators of Infant and Young Child Development (IYCD): metadata synthesis across 10 countries. *BMJ Glob Health* 2018;3:e000747.

¹BDI-2: Battelle's Development Inventory 2; Newborg, J. , Stock, J.R. , Wnek, L. , Guidabaldi, J. , & Svinicki, J. (1984). Battelle Developmental Inventory Screening Test. Allen Texas: DLM-Teaching Resources. Abbey Tyner Berls, Irene R McEwen, Battelle Developmental Inventory, *Physical Therapy*, Volume 79, Issue 8, 1 August 1999, Pages 776–783.

¹MSEL: Mullen's Scales of Early Learning; Mullen, E. *Mullen Scales of Early Learning*. Circle Pines, NM: American Guidance Service 1995.

¹CDIIT: Comprehensive Developmental Inventory for Infants and Toddlers; Hwang AW, Weng LJ, Liao HF. Construct validity of the comprehensive developmental inventory for infants and toddlers. *Pediatrics international*. 2010 Aug;52(4):598-606.