BMJ Open

Psychometric properties of early childhood development assessment tools in low-and-middle-income countries: a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-096365
Article Type:	Original research
Date Submitted by the Author:	09-Nov-2024
Complete List of Authors:	Bliznashka, Lilia; International Food Policy Research Institute, Nutrition, Diets, and Health Unit; University of Edinburgh, Global Academy of Agriculture and Food Systems Hentschel, Elizabeth; Harvard T H Chan School of Public Health, Department of Global Health and Population Ali, Nazia Binte; Harvard T H Chan School of Public Health, Department of Global Health and Population Hunt, Xanthe; Stellenbosch University, Department of Global Health; Africa Health Research Institute (AHRI) Neville, Sarah ; Brown University School of Public Health, International Health Institute Olney, D.; International Food Policy Research Institute, Nutrition, Diets, and Health Unit Pitchik, Helen; University of California Berkeley, School of Public Health Roy, Aditi; Centre for Chronic Disease Control Seiden, Jonathan; Harvard Graduate School of Education Solís-Cordero, Katherine ; University of Costa Rica, School of Nursing Thapa, Aradhana; Emory University Hubert Department of Global Health Jeong, Joshua; Emory University Hubert Department of Global Health
Keywords:	PAEDIATRICS, Child, Community child health < PAEDIATRICS, Patient Reported Outcome Measures, Psychometrics

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Psychometric properties of early childhood development assessment tools in low-and-middle-income countries: a systematic review

Lilia Bliznashka, PhD[1,2,†], Elizabeth Hentschel, PhD [3], Nazia Binte Ali, MPH[3], Xanthe Hunt, PhD [4,5], Sarah Elizabeth Neville, PhD[6], Deanna Olney, PhD[1], Helen O. Pitchik, PhD[7], Aditi Roy, PhD[8], Jonathan Seiden, EdM[9], Katherine Solís Cordero, PhD[10], Aradhana Thapa, MPH[11], Joshua Jeong, ScD[11]

[1] International Food Policy Research Institute, Washington, DC, 20001

[2] Global Academy of Agriculture and Food Systems, University of Edinburgh, UK

[3]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA

[4] Institute for Life Course Health Research, Department of Global Health, Stellenbosch University, South Africa

[5] Africa Health Research Institute (AHRI), Somkele, South Africa

[6] Brown University School of Public Health, International Health Institute, Providence, RI, USA

[7] Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California, USA

[8] Centre for Chronic Disease Control, New Delhi, India

[9] Harvard Graduate School of Education, Cambridge, MA, USA

[10] School of Nursing, University of Costa Rica, San José, Costa Rica

[10] Hubert Department of Global Health, Emory University, Atlanta, GA, USA

† After the first, second, and last authors, all authors are listed in alphabetical order by last name.

Corresponding author:

Lilia Bliznashka

l.bliznashka@cgiar.org

1201 I St NW, Washington, DC 20001, US

Abstract

Background: Valid and reliable measurement of early childhood development (ECD) is critical for monitoring and evaluating ECD-related policies and programmes.

Objective: To systematically appraised available evidence on the psychometric properties of tools used to measure ECD in low- and middle-income countries (LMICs).

Design: We searched MEDLINE, Embase, PubMed, PsycInfo, SciELO, and BVS. We included studies that examined the reliability, validity, and measurement invariance of tools assessing ECD in children 0-6 years of age living in LMICs. Each study was independently screened by two researchers and data extracted by one randomly assigned researcher.

Results: A total of 153 articles covering 114 tools met inclusion criteria. Most reported psychometric properties were internal consistency reliability (n=114, 65%), concurrent validity (n=78, 45%), test-retest reliability (n=72, 41%), convergent validity (n=72, 41%), and structural validity (n=69, 39%). Measurement invariance was least commonly reported (n=17, 10%). Most articles came from Brazil, China, India, and South Africa. Most psychometric evidence was from urban (n=89, 51%) or urban-rural (n=39, 22%) contexts. Study samples focused on children aged 6-17.9 or 48-59.9 months. The most assessed developmental domain was language (n=105, 60%), followed by motor (n=98, 56%) and cognitive (n=77, 44%).

Conclusions: Psychometric evidence is fragmented, limited, and heterogeneous. More rigorous psychometric analyses, especially on measurement invariance, are needed to establish the quality and accuracy of ECD tools for use in LMICs.

PROSPERO registration: CRD42022372305

Key messages

What is already known on this topic

- A critical aspect to monitoring and evaluating interventions and policies to improve early childhood development (ECD) in children <6 years of age in low- and middle-income countries (LMICs) is the availability of reliable and valid assessment tools.
- Existing literature largely focuses on individual ECD tools, ECD domains, or psychometric properties.

What this study adds

- This systematic review provides a comprehensive systematic appraisal of available psychometric evidence of ECD tools used in LMICs.
- The analysis is unique in that it disaggregates articles by those developing or creating new ECD tools and those using or adapting existing tools. It further provides the first disaggregation by urban and rural setting.
- The analysis additionally looks at psychometric evidence for specific child age groups to provide a more in-depth understanding of adequate tools to measure ECD in the early years.

How this study might affect research, practice or policy

- The findings highlight the need for rigorous research and psychometric analyses in diverse child populations to ensure improved population level assessment and monitoring and evaluation of ECD policies and programs.

INTRODUCTION

The Sustainable Development Goals (SDGs) recognise the importance of early childhood development (ECD),[1] which sets the foundation for children's later learning and economic outcomes.[2–4] Rigorous ECD measurement is critical for accurate comparability of children's skills across populations and time, evaluating the effectiveness of ECD interventions, and tracking ECD-related policies.

In low-and-middle-income countries (LMICs), >140 tools (60% of which originated in high-income countries (HICs)) have been used to assess ECD in children 0-8 years old.[5] Directly applying ECD tools from HICs in LMICs can be problematic without psychometric evidence for new cultures and contexts. Few studies to-date have synthesised the evidence on the psychometric properties of ECD tools.

Evidence pertaining to a tool's reliability, validity, and measurement invariance in a given context is critical for selecting an ECD outcome or indicator. Other factors include the purpose of measurement, the population and age-range of interest, and the developmental domain(s) of interest.[6] Evidence of reliability and validity ensures consistent and accurate ECD measurement,[7] whereas evidence of measurement invariance guarantees assessment of the same construct across countries, sub-groups, and time.[7,8]

Prior reviews have provided guidance for selecting ECD tools for use in LMICs[5,9–11] and underscored that evidence on tool reliability and validity is fundamental.[9,10] However, prior work has often focused on individual ECD domains,[12–14] which has limited use for population level assessment or tracking of SDG-related policies, or a subset of psychometric properties,[10] which is not evidence of reliability and validity as a whole. Further, prior reviews do not disaggregate use of ECD tools by HICs vs LMICs or urban vs rural settings,[12–14] despite known and persistent

disparities in ECD by country income level and urban/rural residence[15,16] and the fact that young child populations are increasingly diverse due to migration and urbanization.[13] These are important distinctions given that psychometric properties can vary by population characteristics.

We reviewed available evidence on 11 psychometric properties of tools used to assess ECD in children 0-6 years old living in LMICs. Our findings can assist stakeholders in selection of ECD tools with adequate psychometric evidence for the intended use, and inform what research is needed to improve how we track ECD-related SDGs, programmes, and policies in LMICs.

METHODS

We identified articles in any language through MEDLINE, Embase, PubMed, PsycInfo, SciELO, and BVS (**Supplemental Methods, Supplemental Table 1**). Full-text, peer-reviewed articles were included if: (1) the studies were conducted in a LMIC, (2) included children 0-6 years old, (3) included at least one ECD domain (cognitive, language, motor, social-emotional, attention/executive function, personal-social, and pre-academic/academic) defined in **Supplemental Table 2**, (4) developed a new tool or adapted an existing one, and provided primary evidence of at least one of 11 psychometric properties in terms of reliability, validity, or measurement invariance (**Table 1**), and (5) were published between January 1, 2007 and March 9, 2023. We considered all ECD tools regardless of their intended or actual use (diagnosis, screening, or surveillance). We included both articles that adapted original tools ("adaptation articles" hereafter) and articles that developed new tools ("development articles" hereafter). Cited references were reviewed for potential inclusion. We excluded articles where a tool was used to assess an outcome measure (e.g., trials reporting impacts on ECD outcomes), but the article did not include measurement objectives.[17] Among articles with an explicit measurement focus, we

excluded those reporting only on convergent validity (e.g., correlations with socio-demographic variables) which alone provides limited psychometric evidence.[17]

Search results were imported into Covidence, where duplicates were automatically removed. Two reviewers independently screened titles and abstracts for inclusion and reviewed full texts of retained articles. Disagreements in screening or full text review were resolved through discussion with a third reviewer. Each article was extracted by one reviewer using a standard data extraction sheet (**Supplemental Methods**). For quality assurance, 20% of articles were randomly selected for independent extraction by a second reviewer. Any discrepancies regarding data extraction were resolved by a third reviewer. Authors of included articles were not contacted when information was missing or unclear.

Seven reviewers assessed risk of bias using a checklist created by the authors (**Supplemental Methods**), which contained three categories of bias due to: (1) training/administration (not assessed for tools relying on self-assessment), (2) selective reporting (only assessed for studies reporting on convergent or predictive validity), and (3) missing data. Each article was rated separately on each category using the following ratings: low risk, some concerns, high risk, and very high risk. We also assessed indirectness of populations [18] by assessing whether the sample was limited to a specific setting, the tool covered the entire age range it was intended for, subgroups were generalisable, and results were generalisable. Where insufficient information was provided in the article, we rated the study as "unable to assess". Disagreements for articles assessed by two reviewers (20%) were resolved through discussion with a third reviewer.

Data analysis was conducted at the article-tool level because some articles reported on multiple tools. We created binary indicators for whether evidence on each one of the 11

psychometric properties was reported. In longitudinal studies, we considered the psychometric properties reported at any time point. We then summarised the evidence by country, location, age group at assessment, and type of article (development vs adaptation). Results were synthesised narratively.

Results

Our search strategy identified 6,430 records from six databases and two other sources (**Figure 1**). After removing duplicates, 5,388 records were excluded during title and abstract screening. After full text review of the remaining 250 records, 98 records were excluded. We included 153 articles, covering 114 tools. Four articles reported on multiple tools, resulting in 175 article-tool combinations (referred to as articles for brevity).

Most articles were adaptation article (n=140, 80%) (**Figure 2, Supplemental Table 3**). The most often evaluated psychometric properties were internal consistency reliability (n=114, 65%), concurrent validity (n=78, 45%), test-retest reliability (n=72, 41%) convergent validity (n=72, 41%), and structural validity (n=69, 39%). Measurement invariance was the least frequently reported psychometric property (n=17, 10%) and was primarily evaluated over countries and child sex. The number of articles increased over time (**Supplemental Figure 1**). Between 2007-2010, most included articles were development articles; since 2011, most included articles have been adaptation articles.

Psychometric evidence came from 53 countries (**Figure 3**); one-third of articles came from Brazil (n=22, 12%), China (n=18, 10%), India (n=13, 7%), and South Africa (n=10, 5%). For 40% of countries represented, there was only one article reporting psychometric evidence (**Supplemental Table 4**). Psychometric evidence for each ECD tool included was generally

limited to 1-2 countries with a few notable exceptions: the Global Scales for Early Development (GSED, 32 countries), the Caregiver Reported Early Development Instrument (CREDI, 18 countries), the International Development and Early Learning Assessment (IDELA, 17 countries), the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III, 12 countries), and the Ages and Stages Questionnaire-3 (ASQ-3, 11 countries) (**Supplemental Table 5**).

Most psychometric evidence came from urban (n=89, 51%) or urban-rural (n=39, 22%) settings. In South Asia and Sub-Saharan Africa, a similar number of articles originated from urban, rural, and urban-rural settings (**Supplemental Table 6**).

For development articles, a single article reported psychometric properties for all tools except for the CREDI covered in two articles (**Supplemental Table** 7). For adaptation articles, ASQ-3 and BSID-III were most often studied (13 and 11 articles, respectively). For 75% of tools only a single article provided psychometric evidence (**Supplemental Table 8**).

Although included tools targeted 0-71.9-month-old children, most studies focused on children 12–17.9 month (n=83, 47%), 6-11.9 months (n=78, 45%), or 54-59.9 months (n=74, 42%) old. Articles largely did not cover the full age groups defined here, or the full age range the ECD tool can be used for. Some articles included age groups as narrow as one month (**Supplemental Figures 2-3**).

Most articles reported on language (n=105, 60%), motor (n=98, 56%), and cognitive (n=77, 44%) development (**Supplemental Table 9**). Academic/pre-academic was the least studied domain (n=8, 5%). Within domains, internal consistency reliability was most frequently reported (**Figure 5**).

Bias due to training/administration was assessed for 53 articles (30%): 24% had low risk of bias (n=42) and 6% some concerns (n=11) (**Supplemental Tables 10-11**). Most articles (n=80,

82%) had low risk of bias due to selective reporting. Risk of bias due to missing data was assessed for 130 articles (74%): 28% had some concerns (n=49) and 4% high risk (n=4). For indirectness, 18 articles (10%) were rated as generalisable and 81 (46%) as probably generalisable.

DISCUSSION

Based on 153 articles, available evidence on 11 psychometric properties of 114 ECD tools for children 0-6 years old is fragmented, limited, and heterogeneous. The most frequently provided evidence was on internal consistency reliability, test-retest reliability and/or concurrent, convergent, and structural validity. Psychometric evidence on measurement invariance was the least commonly available. Although evidence came from 53 LMICs, four countries were most represented. Most evidence came from urban or urban-rural settings. The most studied tools were ASQ-3 and BSID-III.

Our findings support ECD measurement trends:[19] much of the work on psychometric properties is recent, with ECD tools being developed and adapted concurrently. Psychometric efforts remain limited to a few ECD tools,[20,21] individual ECD domains,[12–14] and few psychometric properties.[10,12] This fragmentation is evidenced here by included articles focusing on individual countries, limited age ranges, and single developmental domains. The resulting heterogeneous psychometric evidence can hinder comparability and large-scale monitoring of ECD policies and programmes within and across LMICs, which is crucial for identifying and implementing effective approaches to support ECD.

Despite efforts to consolidate ECD measurement through tools like CREDI, GSED, and IDELA, such tools do not fully meet research, programmatic, and policy needs as evidenced by the increase of adaptation and development articles since 2015. This is not surprising given that

no ECD tool is suitable for all populations.[22] Our disaggregation by development and adaptation articles permitted better understanding of the psychometric evidence available and highlighted evidence gaps for both existing and newly developed tools.

This review highlights four important limitations of existing psychometric evidence for ECD tools in LMICs. First, although most tools are designed for a wide age range, the psychometric evidence behind most tools pertained to narrower age ranges and in some cases as narrow as one month. This may have limited applicability to diverse age ranges (given that there is a natural variability in child development in the early years[13]) in these specific contexts. Relatedly, most psychometric evidence pertained to urban contexts. Given existing urban-rural disparities in ECD[15,16] and increasingly diverse young child populations in urban settings,[13] ECD tools whose psychometric properties were examined only in urban settings might be inadequate for rural settings. Those developing or adapting ECD tools should consider establishing psychometric properties across the full intended child age range and across both urban and rural settings. Although this implies longer, more expensive, and logistically difficult studies, it would ensure the tool can be used broadly and in more diverse populations.

A second limitation is the very little evidence on measurement invariance. While several studies reported on samples drawn from multiple countries, few conducted statistical analysis to test for equivalence across countries, thus providing no evidence of measurement invariance.[8] Cross-country invariance, which guarantees assessment of the same construct across countries, is key for tracking global SDG goals. Further, there was a scarcity of articles on measurement invariance over time, a property which guarantees observed changes can be attributed to interventions/policies rather than fluctuations in the measure's stability over time.[8,23] Psychometric evidence of measurement invariance even for the most widely used tools to assess intervention

impact like BSID[24] is limited in LMICs. This may be because establishing measurement invariance over time is particularly challenging due to the natural variability in child development in early life.[13] Measurement non-invariance can lead to incorrect conclusions about intervention/programme effectiveness.[8] More work in this domain is needed particularly for tools widely used for policy making and programme evaluations. Measurement work should be considered relative to competing and more urgent ECD priorities in LMICs.

Third, consistent with existing literature, we found limited psychometric evidence on tools measuring socio-emotional and personal-social development.[13,23] This is surprising given that these two domains are among the most culturally specific,[13] implying they require more comprehensive and rigorous adaptation. Psychometric evidence on tools to assess attention/executive function and academic/preacademic development was most limited. Without additional work to establish a psychometric base, this poses major challenges for those seeking to monitor these domains in early life.

Lastly, our assessment of study quality demonstrated a common lack of reporting and transparency in training of assessors and data management. Risk of bias could not be assessed for many included studies or was considered high. Better reporting standards and guidelines for psychometric studies can help strengthen the field and ensure more critical evaluation of the evidence is possible.

Nevertheless, some ECD tools had multiple forms of psychometric properties assessed. Using IDELA as an example: one article examined three types of reliability and one type of validity, drawing on pilots from 12 countries,[19] followed by another article examining measurement invariance across countries.[25] Subsequent articles have examined additional psychometric properties, albeit in individual countries.[26,27] Although such examples of

consecutively examining psychometric properties should be the norm, they are often financially and logistically infeasible and the timing does not always align with programmatic and policy agendas. When prior evidence is unavailable and psychometric studies not possible, statistical analysis should be conducted to verify psychometric properties and at a minimum should include reporting internal consistency reliability and structural validity (where relevant). This can help build evidence across multiple settings and populations and confirm the usefulness of ECD tools with diverse populations.

Several strengths of our review should be noted. We used extensive search filters, made no restrictions on domains, included both single- and multi-domain tools, and had a rigorously trained team with diverse backgrounds in ECD measurement in LMICs. Nevertheless, we acknowledge some limitations. We excluded studies published prior to 2007, the search start year when the inaugural *Lancet* Series on Child Development in Developing Countries was published which fundamentally changed the breadth and scale of ECD research in LMICs. Therefore, the search period captured the most important years for the evolution of psychometric research on ECD tools in LMICs. Further, although we applied no language restrictions, we were limited to the languages spoken by the review team. We were unable to conduct full text review and extraction for articles in Chinese, Farsi, and Turkish. Consequently, results from China, Iran, and Turkey may be underrepresented.

CONCLUSION

Psychometric evidence is fragmented, limited, and heterogeneous. More research is warranted to establish the applicability of existing tools in diverse populations, including urban and rural settings, and on establishing measurement invariance over countries and time.

Nevertheless, the results by country and ECD tool presented here can serve stakeholders in selecting tools with at least some available psychometric evidence. To improve monitoring, evaluation, and accountability for ECD globally, psychometric evidence should be a key consideration when selecting ECD tools. Improved reporting for psychometric studies can help ensure transparency, replication, and adequate ability to assess the quality of evidence.

Acknowledgements

Contributors: LB conceptualised the review, designed the methodology, screened articles, extracted and analysed data, drafted and revised the manuscript. EH extracted data and analysed data, drafted and revised the manuscript. NBA and AR extracted data, and revised the manuscript. XH and KSC screened articles, extracted data, and revised the manuscript. SEN and JS extracted data and revised the manuscript. DO revised the manuscript. HOP screened articles and revised the manuscript. AT curated the data and revised the manuscript. JJ conceptualised the review, designed the methodology, screened articles, and revised the manuscript. All authors read and approved the final manuscript. LB is the guarantor.

Funding: This work was supported by the CGIAR Initiative on Resilient Cities. We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund: https://www.cgiar.org/funders. SEN's time was supported by the National Institute of Mental Health (grant number T32 MH078788).

Competing interests: JS has contributed and continues to contribute to the development of the CREDI, GSED, IDELA and AIM-ECD. The remaining authors declare no conflicts of interest.

Patient consent for publication: Not applicable.

Ethical considerations: Not applicable.

Provenance and peer review: Not commissioned; externally peer reviewed.

Data availability statement: All data are included in the manuscript and appendix.

References

1. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development. Published online 2015. https://sdgs.un.org/sites/default/files/publications/21252030 Agenda for Sustainable Development web.pdf
2. Walker SP, Chang SM, Vera-Hernandez M, Grantham-McGregor S. Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior. *Pediatrics*. 2011;127(5):849-857. doi:10.1542/peds.2010-2231
3. Gertler P, Heckman J, Pinto R, et al. Labor market returns to an early childhood stimulation intervention in Jamaica. *Science (80-)*. 2014;344(6187):998-1001. doi:10.1126/science.1251178
4. Campbell F, Conti G, Heckman JJ, et al. Early Childhood Investments Substantially Boost Adult Health. *Science (80-)*. 2014;343(6178):1478-1485. doi:10.1126/science.1248429
5. Fernald LCH, Prado EL, Kariger PK, Raikes A. A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries. Published online 2017.
6. Pushparatnam A, Seiden JM, Luna Bazaldua DA. Guiding Questions for Choosing the Right Tools to Measure Early Childhood Outcomes: Why, What, Who, and How. Published online 2022. https://openknowledge.worldbank.org/handle/10986/37030
7. American Psychological Association. APA Dictionary of Psychology. Accessed February 26, 2023. https://dictionary.apa.org/
8. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Dev Rev*. 2016;41:71-90. doi:10.1016/j.dr.2016.06.004
9. Semrud-Clikeman M, Romero RAA, Prado EL, Shapiro EG, Bangirana P, John CC. Selecting measures for the neurodevelopmental assessment of children in low- and middle-income countries. *Child Neuropsychol*. Published online September 9, 2016:1-42. doi:10.1080/09297049.2016.1216536
10. Munoz-Chereau B, Ang L, Dockrell J, Outhwaite L, Heffernan C. Measuring early child development across low and middle-income countries: A systematic review. *J Early Child Res*. 2021;19(4):443-470. doi:10.1177/1476718X211020031
11. Boggs D, Milner KM, Chandna J, et al. Rating early child development outcome measurement tools for routine health programme use. *Arch Dis Child*. 2019;104(Suppl 1):S22-S33. doi:10.1136/archdischild-2018-315431
12. Griffiths A, Toovey R, Morgan PE, Spittle AJ. Psychometric properties of gross motor assessment tools for children: a systematic review. *BMJ Open*. 2018;8(10):e021734. doi:10.1136/bmjopen-2018-021734
13. Halle TG, Darling-Churchill KE. Review of measures of social and emotional development. *J Appl Dev Psychol*. 2016;45:8-18. doi:10.1016/j.appdev.2016.02.003
14. Humphrey N, Kalambouka A, Wigelsworth M, Lendrum A, Deighton J, Wolpert M. Measures of Social and Emotional Skills for Children and Young People. *Educ Psychol Meas*. 2011;71(4):617-637. doi:10.1177/0013164410382896
15. Lu C, Cuartas J, Fink G, et al. Inequalities in early childhood care and development in low/middle-income countries: 2010–2018. *BMJ Glob Heal*. 2020;5(2):e002314. doi:10.1136/bmjgh-2020-002314
16. McCoy DC, Peet ED, Ezzati M, et al. Early Childhood Developmental Status in Low- and Middle-Income Countries: National, Regional, and Global Prevalence Estimates Using Predictive Modeling. *PLoS Med*. 2016;13(6):1-18. doi:10.1371/journal.pmed.1002034

17. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). Published online 2018:1-78. https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf
18. GRADE Working Group. *Handbook for Grading the Quality of Evidence and the Strength of Recommendations Using the GRADE Approach. Updated October 2013.* (Schünemann H, Brożek J, Guyatt G, Oxman A, eds.).; 2013. https://gdt.gradepro.org/app/handbook/handbook.html
19. Pisani L, Borisova I, Dowd AJ. Developing and validating the International Development and Early Learning Assessment (IDELA). *Int J Educ Res*. 2018;91(July):1-15. doi:10.1016/j.ijer.2018.06.007
20. Kersten P, Czuba K, McPherson K, et al. A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *Int J Behav Dev*. 2016;40(1):64-75. doi:10.1177/0165025415570647
21. Velikonja T, Edbrooke-Childs J, Calderon A, Sleed M, Brown A, Deighton J. The psychometric properties of the Ages & Stages Questionnaires for ages 2-2.5: a systematic review. *Child Care Health Dev*. 2017;43(1):1-17. doi:10.1111/cch.12397
22. Sabanathan S, Wills B, Gladstone M. Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately? *Arch Dis Child*. 2015;100(5):482-488. doi:10.1136/archdischild-2014-308114
23. Gridley N, Blower S, Dunn A, Bywater T, Bryant M. Psychometric Properties of Child (0–5 Years) Outcome Measures as used in Randomized Controlled Trials of Parent Programs: A Systematic Review. *Clin Child Fam Psychol Rev*. 2019;22(3):388-405. doi:10.1007/s10567-019-00277-1
24. Jeong J, Franchett EE, Ramos de Oliveira C V., Rehmani K, Yousafzai AK. Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLOS Med*. 2021;18(5):e1003602. doi:10.1371/journal.pmed.1003602
25. Halpin PF, Wolf S, Yoshikawa H, et al. Measuring early learning and development across cultures: Invariance of the IDELA across five countries. *Dev Psychol*. 2019;55(1):23-37. doi:10.1037/dev0000626
26. Wolf S, Halpin P, Yoshikawa H, Dowd AJ, Pisani L, Borisova I. Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Child Res Q*. 2017;41(June):21-36. doi:10.1016/j.ecresq.2017.05.001
27. Pisani L, Seiden J, Wolf S. Longitudinal evidence on the predictive validity of the International Development and Early Learning Assessment (IDELA). *Educ Assessment, Eval Account*. 2022;34(2):173-194. doi:10.1007/s11092-021-09374-8

Table 1 Definitions of the psychometric properties used in the selection criteria for included articles

Domain	Measurement property	Example test statistic	Definition
Reliability			The consistency of a test or measurement, i.e., how consistently a measure produces similar results with repeated measures over a short period of time or across assessors at the same time point. This can also be thought of as the correlation between observed scores across replications.
	Test-retest reliability	correlation coefficient	Correlation between scores from the same test from assessments conducted over a short time interval.
	Inter-rater reliability	kappa, Bland-Altman test	The extent to which independent assessors produce similar ratings in judging the same abilities or characteristics in the same target person at the same time.
	Internal consistency reliability	Cronbach's alpha, alpha	Degree of interrelatedness among items on the same tool, i.e., how well the items work together to provide information on an underlying construct.
Validity			The degree to which the tool measures what it is supposed to measure, i.e., the degree to which the tool reflects the underlying construct.
	Content/face validity		The degree to which the content of the tool is adequate for the construct being measured, i.e., assessing the extent to which a tool appears to reflect the underlying construct.
	Concurrent/criterion validity	correlation coefficient; regression estimate	The degree to which scores on one measurement tool are related to scores obtained at about the same point in time from another tool considered the gold standard.
	Convergent validity	correlation coefficient; regression estimate	Evidence that scores on a test or measurement are associated with theoretically related measures or variables.

Domain	Measurement property	Example test statistic	Definition
	Predictive	correlation coefficient; regression estimate	Evidence that a score correlates with a variable that can only be assessed at some point after the test has been administered or the measurement made, e.g., evidence that scores now are correlated with scores at a future timepoint.
	Structural validity (dimensionality)	Exploratory factor analysis: number of factors, eigen values Confirmatory factor analysis: model fit statistics such as Comparative Fit Index (CF), Root mean square error of approximation (RMSEA)	The degree to which the scores of an assessment are an adequate reflection of the dimensionality of the construct to be measured.
Invariance			The property when a scale or construct provides the same results across different samples, populations, settings, or characteristics.
	Measurement invariance over countries	Likelihood ratio chi squared statistic and p-value from freeing parameters across groups	The degree to which an assessment or construct provides the same results across separate samples in different countries.
	Measurement invariance over time	Likelihood ratio chi squared statistic and p-value from freeing parameters across groups	The degree to which an assessment or construct provides the same results across time.
	Measurement invariance over other groups	Likelihood ratio chi squared statistic and p-value from freeing parameters across groups	The degree to which an assessment or construct provides the same results across different groups.

Note: All definitions based on the APA Dictionary of Psychology.[7]

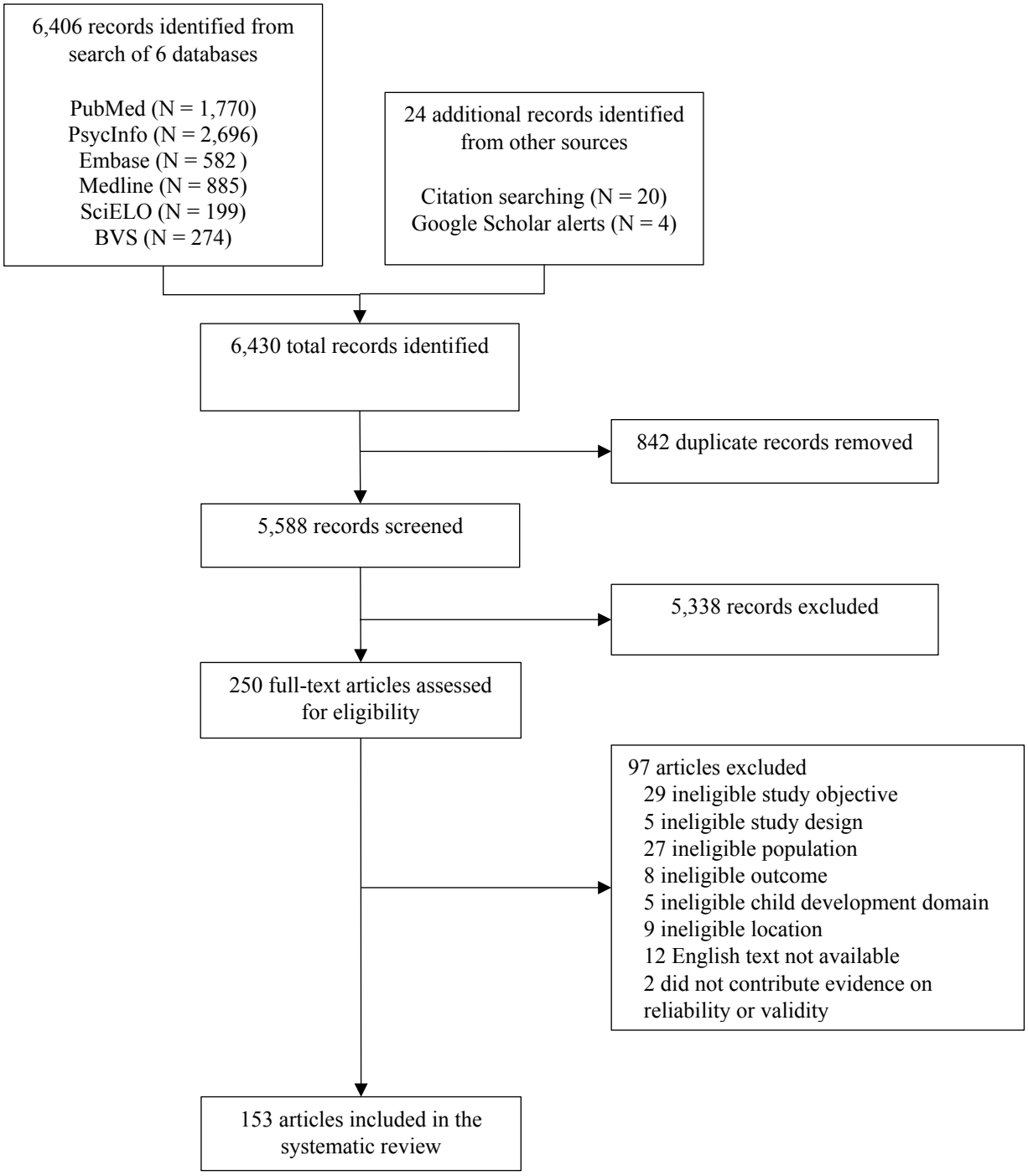
Figure Captions

Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram of search results and included articles

Figure 2 Number of included article-tools and type of psychometric evidence provided

Figure 3 Countries where studies providing evidence on at least one psychometric property were conducted

Figure 4 Number of articles providing psychometric evidence by article type and early childhood development domain



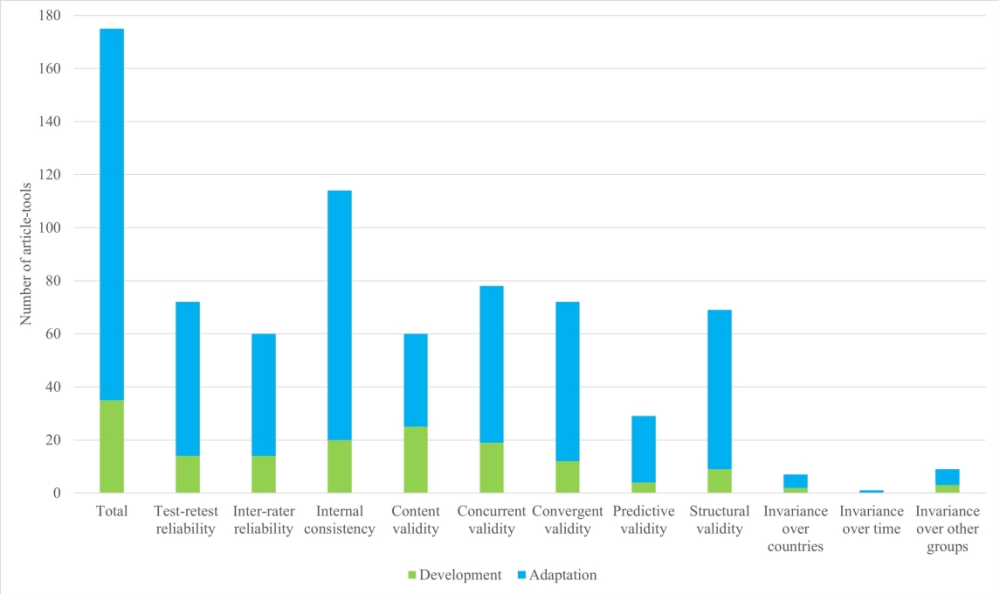


Figure 2 Number of included article-tools and type of psychometric evidence provided

276x164mm (300 x 300 DPI)

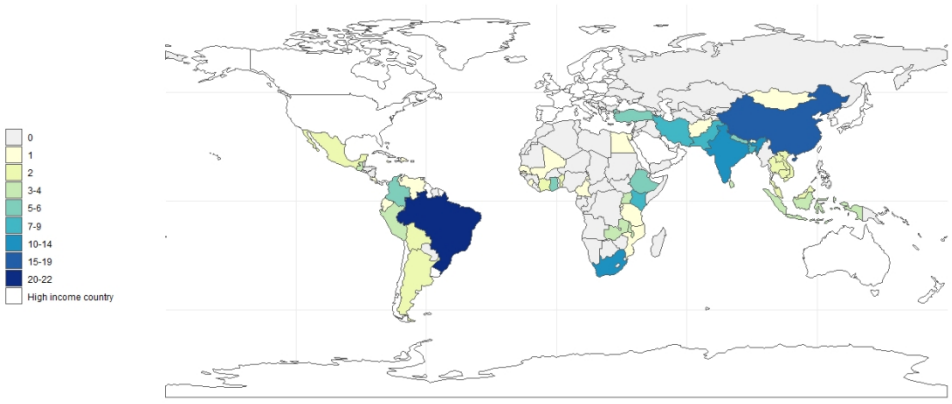


Figure 3 Countries where studies providing evidence on at least one psychometric property were conducted

345x211mm (96 x 96 DPI)

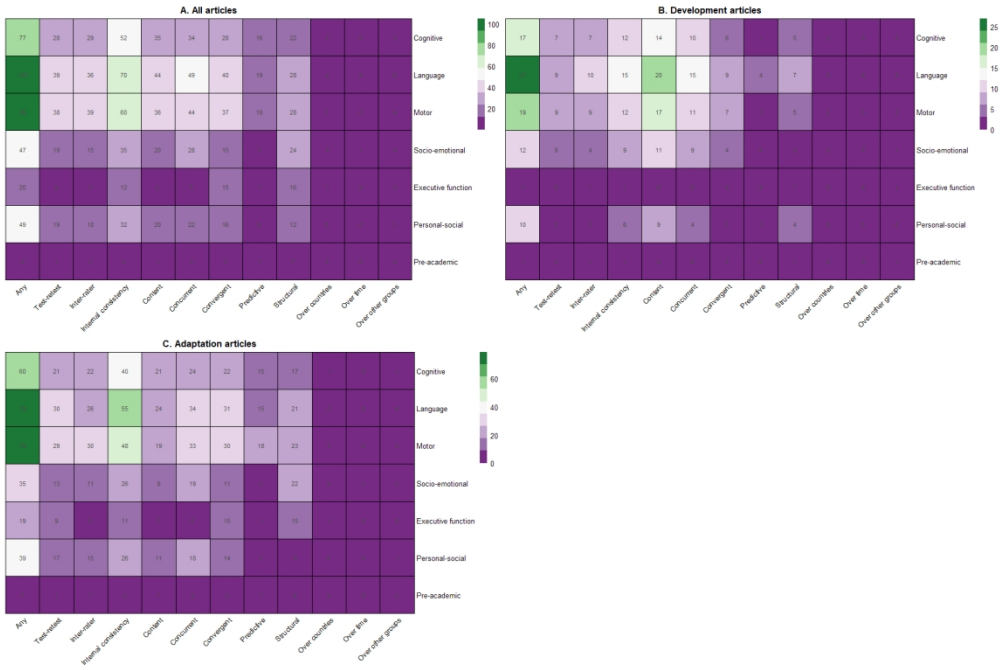


Figure 4 Number of articles providing psychometric evidence by article type and early childhood development domain

476x317mm (96 x 96 DPI)

Supplemental appendix

Table of Contents

Supplemental Methods

Search strategy and selection criteria

The search strategy used medical subject headings (MeSH), keywords, and free text words along four key elements: population, construct, measurement properties, and location. Search terms were combined using Boolean operators. Truncation wildcards were used to include variations of the search terms. The search strategy used a combination of searches through titles, abstracts, and keywords. The COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) measurement properties filter was used for the third search element.[1] The search strategy was piloted in PubMed and then adapted for each database.

Articles were included if at least 50% of the study sample was within the range 0-6 years and the average age in the sample was <6 years or unspecified. If information on the child age range was not available or clear, the article was included. Multi-country articles were included if ≥50% of the countries were LMICs. If an article reported that measurement properties were reported elsewhere (e.g., referencing another study and thus providing secondary evidence), the article was ineligible. We included prospective, retrospective, cross-sectional, and longitudinal quantitative study designs.

We excluded articles studying children with developmental disabilities and disorders (e.g., autism spectrum disorder, cerebral palsy), and children with physical disabilities that impair performance on ECD measures (e.g., deafness, blindness). We excluded the following study designs: animal studies, simulation studies, case studies, opinions, letters, preprints, protocols, conference abstracts, ecological studies, dissertations/theses, reviews or systematic reviews, and meta-analyses.

1. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009; **18**: 1115–23.

Data analysis

One reviewer developed the data extraction sheet, and two reviewers piloted it. Revisions were made through discussion between the two reviewers. Seven reviewers were trained on using the data extraction sheet through pilot extractions of included articles. Further revisions were made to the data extraction sheet based on discussions of the piloting process. Data extraction included information on publication details, study meta-data, characteristics of the ECD tool, type of administration, ECD domains assessed, and psychometric properties with respect to reliability, validity, and measurement invariance.

One reviewer created a risk of bias checklist by adapting items from the COSMIN[1] and Cochrane's ROBINS-E[2] risk of bias tools. The checklist was then refined via discussions with two other reviewers and after piloting by four reviewers. Seven reviewers were trained and conducted the risk of bias assessment at the article level. Disagreements for articles assessed by two reviewers (20%) were resolved through discussion with a third reviewer. The ROBINS-E risk of bias ratings were used: low risk (little-to-no risk of bias in this domain), some concerns (some risk of bias in this domain, but not clear it is an important risk of bias), high risk (important problems in this domain), and very high risk (the study is very problematic in this domain).

For all tools, child age at assessment was converted to months based on the information provided in the article. For longitudinal studies, age at first assessment was used. For multi-site/multi-country studies, the full age range at assessment was used across the sites/countries. Because the objective the paper was to take stock of the available evidence rather than to report the adequacy and relevance of specific psychometric properties, we did not summarize evidence on the psychometric properties themselves.

1. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). 2018; : 1–78.
2. ROBINS-E Development Group. Risk Of Bias In Non-randomized Studies - of Exposure (ROBINS-E). Launch version, 1 June 2022. 2022. https://www.riskofbias.info/welcome/robins-e-tool.

Supplemental Table 1 Search strategy as executed in PubMed

("Child"[MeSH] OR "Child, Preschool"[MeSH] OR "child*"[tiab] OR "Infant"[MeSH] OR "infant*"[tiab] OR "paediatric"[tiab] OR "pediatric"[tiab] OR "toddler*"[tiab])

AND

("Child Development"[MeSH] OR "child development*"[tiab] OR "early child development"[tiab] OR "early childhood development"[tiab] OR "ECD"[tiab] OR "Child Behavior"[MeSH] OR "child behavior"[tiab] OR "child behaviour"[tiab] OR "Cognition"[MeSH] OR "Executive Function"[MeSH] OR "executive function*"[tiab] OR "Emotional Intelligence"[MeSH] OR "attachment"[tiab] OR "attention"[tiab] OR "cognition"[tiab] OR "cognitive development"[tiab] OR "cognitive skills"[tiab] OR "neurocognitive"[tiab] OR "communication"[tiab] OR "language"[tiab] OR "language skills"[tiab] OR "language development"[tiab] OR "receptive language"[tiab] OR "expressive language"[tiab] OR "speech development"[tiab] OR "communication skills"[tiab] OR "Motor Skills"[MeSH] OR "motor skills"[tiab] OR "fine motor"[tiab] OR "gross motor"[tiab] OR "motor"[tiab] OR "prosocial"[tiab] OR "socialisation"[tiab] OR "socialization"[tiab] OR "socioemotion*"[tiab] OR "socio-emotion*"[tiab] OR "neurodevelopment"[tiab] OR "neurobehavior"[tiab] OR "neurobehaviour"[tiab] OR "neurobehavioral"[tiab] OR "neurobehavioural"[tiab] OR "Intelligence"[MeSH] OR "intelligence"[tiab] OR "IQ"[tiab] OR "psychomotor"[tiab] OR "sensorimotor"[tiab] OR "attention"[tiab] OR "inhibition"[tiab] OR "impulse control"[tiab] OR "working memory"[tiab] OR "problem solving"[tiab] OR "self-regulation"[tiab] OR "prosocial"[tiab] OR "psychosocial development"[tiab] OR "internali*"[tiab] OR "externali*"[tiab] OR "empathy"[tiab] OR "neuropsychology*"[tiab] OR "school readiness"[tiab] or "numeracy"[tiab] or "literacy"[tiab] OR "pre-academic"[tiab] OR "early learning"[tiab] or "personal-social"[tiab] OR "personal social"[tiab])

AND

(("developing countr*"[tiab] OR "under developed countr*"[tiab] OR lmic*[tiab] OR "less developed"[tiab] OR "low income"[tiab] OR "lower income"[tiab] OR "low- and middle-income"[tiab] OR "low middle income"[tiab] OR "resource poor"[tiab] OR "resource constrained"[tiab] OR "low resource"[tiab] OR "limited resource*"[tiab] OR "resource limited"[tiab]) AND (Afghan*[tiab] OR Africa[tiab] OR African[tiab] OR Algeria*[tiab] OR "American Samoa*"[tiab] OR Angola*[tiab] OR Argentin*[tiab] OR Bangladesh*[tiab] OR Barbad*[tiab] OR Belorussian[tiab] OR Beliz*[tiab] OR Benin*[tiab] OR Bhutan*[tiab] OR Bolivia*[tiab] OR Botswan*[tiab] OR Brazil*[tiab] OR "Burkina Faso"[tiab] OR Burkinabe[tiab] OR Burund*[tiab] OR Cambodia*[tiab] OR Cameroon*[tiab] OR "Cape Verde"[tiab] OR "Cape Verdean"[tiab] OR "Central African Republic"[tiab] OR Chad*[tiab] OR Chile*[tiab] OR China[tiab] OR Chinese[tiab] OR Colombia*[tiab] OR Comoros[tiab] OR Comorian[tiab] OR Congo[tiab] OR Congolese[tiab] OR Costa Rica*[tiab] OR "Côte d'Ivoire"[tiab] OR "Ivory Coast"[tiab] OR Ivorian[tiab] OR Croatia*[tiab] OR Croat[tiab] OR

Cuba*[tiab] OR Djibouti*[tiab] OR Dominica*[tiab] OR "Dominican Republic"[tiab] OR Ecuador*[tiab] OR Egypt*[tiab] OR "El Salvador"[tiab] OR Salvadorian[tiab] OR "Equatorial Guinea"[tiab] OR Guinean[tiab] OR Eritrea*[tiab] OR Ethiopia*[tiab] OR Fiji*[tiab] OR Gabon*[tiab] OR Gambia*[tiab] OR Gaza[tiab] OR Gazan[tiab] OR Ghana[tiab] OR Ghanaian[tiab] OR Grenad*[tiab] OR Guatemala*[tiab] OR Guinea[tiab] OR Guyan*[tiab] OR Haiti*[tiab] OR Hondura*[tiab] OR Hungar*[tiab] OR India[tiab] OR Indian[tiab] OR Indonesia*[tiab] OR Iran*[tiab] OR Iraq*[tiab] OR Jamaica*[tiab] OR Jordan*[tiab] OR Kenya[tiab] OR Kenyan[tiab] OR Kiribati[tiab] OR Korea*[tiab] OR Kyrgy*[tiab] OR Laos[tiab] OR Laotian*[tiab] OR Lebanon[tiab] OR Lebanese[tiab] OR Lesotho[tiab] OR Liberia*[tiab] OR Libya*[tiab] OR Macedonia*[tiab] OR Madagasca*[tiab] OR Malawi*[tiab] OR Malaysia*[tiab] OR Maldives[tiab] OR Maldivian[tiab] OR Mali[tiab] OR Malian*[tiab] OR "Marshall Islands"[tiab] OR Mauritania*[tiab] OR Mauritius[tiab] OR Mauritian[tiab] OR Mayotte[tiab] OR Mexic*[tiab] OR Micronesia*[tiab] OR Moldov*[tiab] OR Mongolia*[tiab] OR Morocc*[tiab] OR Mozambique[tiab] OR Mozambican[tiab] OR Myanmar[tiab] OR Namibia*[tiab] OR Nepal*[tiab] OR Nevis[tiab] OR Nicaragua*[tiab] OR Niger*[tiab] OR "Northern Mariana Islands"[tiab] OR Oman*[tiab] OR Pakistan*[tiab] OR Palau*[tiab] OR Panama*[tiab] OR "Papua New Guinea"[tiab] OR Paraguay*[tiab] OR Peru*[tiab] OR Philippine*[tiab] OR Filipino*[tiab] OR Poland[tiab] OR Polish[tiab] OR Rwanda*[tiab] OR Samoa*[tiab] OR Sao Tome*[tiab] OR Principe[tiab] OR Senegal*[tiab] OR Seychell*[tiab] OR Sierra Leon*[tiab] OR Solomon Island*[tiab] OR Somali*[tiab] OR South Africa*[tiab] OR Sri Lanka*[tiab] OR "Saint Kitts"[tiab] OR "St Kitts"[tiab] OR "Saint Lucia"[tiab] OR "St Lucia"[tiab] OR "Saint Vincent" [tiab] OR "St Vincent"[tiab] OR Sudan*[tiab] OR Suriname*[tiab] OR Swaziland Or Swazi[tiab] OR Syria*[tiab] OR Tajik*[tiab] OR Tanzania*[tiab] OR Thailand[tiab] OR Thai[tiab] OR "Timor Leste"[tiab] OR Togo*[tiab] OR Tonga*[tiab] OR Trinidad[tiab] OR Trinidadian[tiab] OR Tobago[tiab] OR Tobagonian[tiab] OR Tunisia*[tiab] OR Turky[tiab] OR Turkish[tiab] OR Uganda*[tiab] OR Uruguay*[tiab] OR Vanuat*[tiab] OR Venezuela*[tiab] OR Vietnam*[tiab] OR "West Bank"[tiab] OR Yemen*[tiab] OR Zambia*[tiab] OR Zimbabwe*))

AND

(instrumentation[sh] OR methods[sh] OR “Validation Studies”[pt] OR “Comparative Study”[pt] OR “psychometrics”[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR “outcome assessment (health care)”[MeSH] OR “outcome assessment”[tiab] OR “outcome measure*”[tw] OR “observer variation”[MeSH] OR “observer variation”[tiab] OR “Health Status Indicators”[Mesh] OR “reproducibility of results”[MeSH] OR reproducib*[tiab] OR “discriminant analysis”[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR “coefficient of variation”[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR “internal consistency”[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR “precise values”[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND (test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR

intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab] OR "measurement invariance"[tiab] OR "invariance"[tiab] OR assessment[tiab] OR adapt*[tiab])

NOT

("editorial"[Publication Type] OR "comment"[Publication Type] OR "case reports"[Publication Type] OR "congress"[Publication Type] OR "consensus development conference"[Publication Type] OR "interview"[Publication Type] OR "lecture"[Publication Type] OR "legal case"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "Observational Study, Veterinary"[Publication Type] OR "Patient Education Handout"[Publication Type] OR "practice guideline"[Publication Type] OR "preprint"[Publication Type])

Supplemental Table 2 Definitions of the seven early childhood domains of interest

Domain	Definition
Cognitive	Cognitive development includes abilities such as general intellectual ability, problem solving, reasoning, visual-spatial ability, memory, learning, etc.
Language	Language development includes the ability to understand and express verbal communication, and vocabulary development and acquisition.
Motor	Moto development includes gross motor development (the ability to control and coordinate gross movements of the legs and arms (e.g., jumping, throwing)) and fine motor development (the ability to control and coordinate fine movements of the fingers and toes.
Socio-emotional/temperament	Socio-emotional development is the ability to regulate emotional responses and social interactions. It includes behaviour problems, social competency, emotional competency, and self-regulation. Temperament includes biological influences on the experience and expression of emotion: extraversion (positive affect, activity level, impulsivity, risk-taking), negative affectivity (fear, anger, sadness, discomfort) and effortful control (attention shifting and focusing, perceptual sensitivity, inhibitory and activational control). Socio-emotional development and temperament are often overlapping constructs, particularly in the early years.
Attention/Executive function	Executive function, including attention, entails international control over behaviour and cognition. It covers inhibitory control, cognitive flexibility, planning, and working memory.
Personal-social/adaptive	Personal-social and adaptive behaviour comprises the ability to perform daily-life and self-help skills, such as feeding, dressing, toilet training, recognising others, interacting with others, and adjusting to new situations.
Academic/pre-academic	Preacademic skills are skills needed to learn reading and math, such as counting and letters.

Note: Definitions based on (Boggs et al., 2019; Fernald et al., 2017).

Supplemental Table 3 Number of included articles by type of psychometric evidence provided

Type of article	Total number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		
		Test-retest	Inter-rater	Internal consistency	Content	Concurrent	Convergent	Predictive	Structural	Invariance over countries	Invariance over time	Invariance over other groups
All types of articles	175	72	60	114	60	78	72	29	69	7	1	9
Development articles	35	14	14	20	25	19	12	4	9	2	0	3
Adaptation articles	140	58	46	94	35	59	60	25	60	5	1	6

Supplemental Table 4 List of countries and assessment tools used in included articles

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Afghanistan	International Development and Early Learning Assessment (IDELA)	1	0	0	0	0	0	0	0	1	0	0	0	Halpin 2019
Argentina	Bayley Infant Neurodevelopmental Screener (BINS)	1	1	1	0	0	0	0	0	0	0	0	0	McCarthy 2012
Argentina	Escala de Habildades Sociales (Scale of Social Abilities)	1	0	0	1	0	0	0	0	1	0	0	0	Lacunza 2009
Bangladesh	Ages & Stages Questionnaire Inventory (ASQ:I)	1	0	1	1	0	1	1	0	0	0	0	0	Pitchik 2023
Bangladesh	Bayley Scales of Infant Development, Second Edition (BSID-II)	1	0	0	0	1	0	0	0	0	0	0	0	Parveen 2014
Bangladesh	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	1	0	0	0	1	0	0	0	Pendergast 2018
Bangladesh	Dimensions of Mastery Questionnaire (DMQ-18)	1	1	1	1	1	1	0	0	1	0	0	0	Shaoli 2019
Bangladesh	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Bangladesh	Rapid Neurodevelopmental Assessment (RNDA)	2	0	2	0	1	2	0	1	0	0	0	0	Khan 2010; Khan 2013
Bangladesh	The language inventory	1	1	0	0	1	1	1	1	0	0	0	0	Hamadani 2010
Bangladesh	WHO Gross Motor Milestones	1	0	1	0	0	1	1	1	0	0	0	0	Hamadani 2013
Benin	Mullen Scales of Early Learning (MSEL)	2	0	1	0	0	1	1	1	0	0	0	0	Koura 2013; Boivin 2021
Bhutan	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018
Bolivia	Bayley Infant Neurodevelopmental Screener (BINS)	1	1	1	0	0	0	0	0	0	0	0	0	McCarthy 2012
Bolivia	International Development and Early Learning Assessment (IDELA)	1	0	0	0	0	0	0	0	1	0	0	0	Halpin 2019
Brazil	Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	1	0	0	0	0	0	0	0	1	0	0	0	Anunciação 2019
Brazil	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	0	1	0	0	0	0	1	0	0	1	Filgueiras 2013
Brazil	Ages & Stages Questionnaires: Social Emotional (ASQ:SE-BR)	1	0	0	0	0	0	0	0	1	0	1	0	Chen 2017

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Brazil	Alberta Infant Motor Scale (AIMS)	4	3	3	1	1	3	0	1	0	0	0	0	Saccani 2016; Siegle 2018; de Albuquerque 2018; Valentini 2012
Brazil	Battelle Developmental Inventory Second Edition, BDI-2	1	0	0	0	1	0	0	0	0	0	0	0	de Albuquerque 2022
Brazil	Bayley Infant Neurodevelopment Screener (BINS)	2	1	1	1	1	1	0	1	1	0	0	0	Guedes 2011; McCarthy 2012
Brazil	Bayley Scales of Infant Development, Third Edition (BSID-III)	2	1	0	1	1	1	0	0	2	0	0	0	Pendergast 2018; Madaschi 2016
Brazil	Caregiver Reported Early Development Instruments (CREDI), short-form	1	1	0	1	1	1	1	0	0	0	0	0	Altafim 2018
Brazil	Denver Developmental Screening Test (Denver-II)	1	0	1	1	1	0	1	0	1	0	0	0	Lopez Boo 2020
Brazil	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	0	1	Sincovich 2019
Brazil	Motor Development Scale	1	0	0	1	0	0	0	0	1	0	0	0	Okuda 2020
Brazil	Oxford Neurodevelopment Assessment (OX-NDA)	1	1	1	1	0	1	0	0	0	0	0	0	Fernandes 2022

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Brazil	Preschool and Kindergarten Behavior Scale (PKBS)	1	0	0	1	0	0	0	0	1	0	0	0	Dias 2011
Brazil	Questionário para Avaliação do Desenvolvimento Infantil (QAD-PIPAS)	1	0	0	0	1	0	0	0	0	0	0	0	Venancio 2020
Brazil	Screening for Identification of Oral Language Difficulties by Preschool Teachers (SIOLD)	1	0	0	0	0	0	0	0	1	0	0	0	Puglisi 2020
Brazil	Self-ordered Pointing Task (SOPT)	1	0	0	0	0	1	1	0	0	0	0	0	Sallum 2017
Brazil	Test of Gross Motor Development (TGMD-2)	1	0	1	1	0	0	1	0	1	0	0	0	Bandeira 2020
Cambodia	Angkor Hospital for Children Developmental Milestone Assessment Tool (AHC DMAT)	1	0	0	0	1	0	0	0	0	0	0	0	Ngoun 2012
Cambodia	Developmental Milestone Chart (DMC)	1	0	0	0	1	0	0	0	0	0	0	0	Scherzer 2009
Cameroon	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	0	0	0	1	0	1	0	0	Lohaus 2014
China	Ages & Stages Questionnaire Inventory (ASQ:I)	1	1	0	1	1	1	0	0	0	0	0	0	Xie 2017

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
China	Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	1	1	1	1	1	0	0	0	1	0	0	0	Bian 2017
China	Ages & Stages Questionnaire: Social-Emotional, First Edition (ASQ:SE-1)	1	1	1	1	0	1	0	0	0	0	0	0	Xie 2019 (Journal of Child & Dev Studies)
China	Ages & Stages Questionnaire-3 (ASQ-3)	2	1	1	1	0	2	1	0	1	0	0	0	Yue 2019; Bian 2012
China	Caregiver Reported Early Development Instruments (CREDI), long-form	1	0	0	1	0	1	1	0	0	0	0	0	Li 2020
China	Child Behavior Checklist/1.5-5 (CBCL/1.5-5) and the Caregiver-Teacher Report Form (C-TRF)	1	0	0	0	0	0	0	0	1	0	0	0	Liu 2011
China	Child Social Preference Scale	1	0	0	1	0	0	0	1	1	0	0	1	Li 2016
China	Chinese Preschool Readiness Scale (CPRS)	1	0	0	1	1	1	0	0	1	0	0	0	Xie 2019 (Early Educ & Dev)
China	Chinese version of Urban Infant-Toddler Social and Emotional Assessment	1	1	1	1	1	1	0	0	1	0	0	0	Jianduan 2009
China	Chinese Vocabulary Checklist	1	0	0	0	0	0	1	0	0	0	0	0	Hao 2008

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
China	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	0	1	Sincovich 2019
China	Emotional and Behavioral Screener (EBS)	1	0	1	1	0	1	0	0	1	0	0	0	Zhang 2018
China	Movement Assessment Battery for Children- second edition (MBAC-2)	1	1	1	1	1	1	0	0	1	0	0	0	Hua 2013
China	Preschool Learning Behavior Scale (PLBS)	1	0	0	1	0	1	0	0	1	0	0	0	Wu 2019 (JPA)
China	Social Skills domain of System-Rating Scales (SSIS-RS)	1	0	0	1	0	0	1	0	1	0	0	0	Wu 2019 (PS)
China	The Chinese Early Language and Communication Questionnaire	1	0	0	1	0	1	0	0	0	0	0	0	Zhang 2021
China	The Chinese Intelligence Scale for Young Children	1	0	0	0	0	0	0	0	1	0	0	1	Guo 2009
Colombia	Ages & Stages Questionnaire-3 (ASQ-3)	2	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	Battelle Developmental Inventory Second Edition, BDI-2	2	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	Bayley Infant Neurodevelopmental Screener (BINS)		1	1	0	0	0	0	0	0	0	0	0	McCarthy 2012

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Colombia	Bayley Scales of Infant Development, Third Edition (BSID-III)	2	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	Denver Developmental Screening Test (Denver-II)	2	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	MacArthur-Bates Communicative Development Inventories CDI	1	0	0	1	1	0	1	0	0	0	0	0	Lara Díaz 2011
Colombia	MacArthur-Bates Communicative Development Inventories Short-Forms I vocabulary checklist	2	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	MacArthur-Bates Communicative Development Inventories Short-Forms II vocabulary checklist	2	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	Preschool Language Scale-3 (PLS- 3)	1	0	0	1	0	0	0	0	1	0	0	0	Romero 2013
Colombia	WHO Gross Motor Milestones	2	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Costa Rica	EDIN II (Evaluación del Desarrollo Integral de la niña y el niño)	1	0	0	0	1	0	1	0	0	0	0	0	Cordero 2019

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Côte d'Ivoire	Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	1	0	0	1	0	0	1	0	1	0	0	0	Barthel 2021
Côte d'Ivoire	Kilifi Developmental Inventory	1	0	1	1	0	0	0	1	1	0	0	0	Barthel 2022
Dominican Republic	Malawi Developmental Assessment Tool (MDAT), renamed Dominican Child Development Screening Tool	1	0	1	1	0	0	1	0	0	0	0	0	Sánchez-Vincitore 2019
Ecuador	Bayley Infant Neurodevelopmental Screener (BINS)	1	1	1	0	0	0	0	0	0	0	0	0	McCarthy 2012
Egypt	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018
Ethiopia	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	1	1	0	0	1	1	1	0	0	0	Hanlon 2016
Ethiopia	Denver Developmental Screening Test (Denver-II)	1	1	1	0	1	0	0	0	0	0	0	0	Abessa 2016
Ethiopia	International Development and Early Learning Assessment (IDELA)	3	1	1	1	1	2	2	0	2	1	0	1	Halpin 2019; Pisani 2018; Wolf 2017

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Georgia	Ages & Stages Questionnaire-3 (ASQ-3)	1	1	0	1	1	0	1	0	0	0	0	0	Zirakashvili 2018
Ghana	Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	1	0	0	1	0	0	1	0	1	0	0	0	Barthel 2021
Ghana	International Development and Early Learning Assessment (IDELA)	1	0	0	1	0	0	0	1	1	0	0	0	Pisani 2022
Ghana	Kilifi Developmental Inventory	1	0	1	1	0	0	0	1	1	0	0	0	Barthel 2022
Ghana	Rapid Assessment of Cognitive and Emotional Regulation (RACER)	1	1	1	0	0	1	1	0	0	0	0	0	Yuan 2022
Ghana	Vocabulary Checklists	1	0	0	0	0	1	0	1	0	0	0	0	Prado 2018
Grenada	INTER- GROWTH-21st Project Neurodevelopment Assessment (INTER-NDA)	1	1	1	1	0	0	0	0	0	0	0	0	Waechter 2022
Guatemala	Ages & Stages Questionnaire-3 (ASQ-3)	1	1	0	1	0	1	0	1	0	0	0	0	Colbert 2021
Guatemala	Mullen Scales of Early Learning (MSEL)	1	1	1	1	0	0	1	1	0	0	0	0	Colbert 2020

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Guatemala	Rapid Neurodevelopmental Assessment (RNDA)	1	0	1	0	1	1	1	0	0	0	0	0	Thompson 2015
India	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	1	1	1	0	0	0	0	0	0	0	Kvestad 2013
India	Bayley Scales of Infant Development, Second Edition (BSID-II)	1	0	0	0	0	1	0	0	0	0	0	0	Biasini 2015
India	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	1	0	0	0	1	0	0	0	Pendergast 2018
India	Caregiver Reported Early Development Instruments (CREDI), long-form	1	0	0	1	0	1	1	0	0	0	0	0	Alderman 2021
India	DEvelopmental assessment on an E-Platform (DEEP)	2	0	0	0	1	1	0	0	0	0	0	0	Bhavnani 2019; Mukherjee 2020
India	Developmental Assessment Tool for Anganwadis (DATA)	1	0	0	1	1	0	0	0	1	0	0	0	Nair 2009
India	Developmental Assessment Tool for Anganwadis (DATA-II)	1	0	0	1	1	0	0	0	1	0	0	0	Nair 2012
India	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
India	Parents' Evaluation of Developmental Status (PEDS)	1	0	0	0	0	1	0	0	0	0	0	0	Mukherjee 2022
India	Screening Test of Early Language Development-Test version (STELD-T)	1	0	0	1	0	1	0	0	0	0	0	0	Pathak 2023
India	Bracken School Readiness Assessment, Third Edition (BSRA-3)	1	0	0	1	1	0	0	0	0	0	0	0	Shah 2013
India	Vineland Adaptive Behaviour Scale (VABS II)	1	1	1	0	0	0	1	0	0	0	0	0	Kumar 2016
Indonesia	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	1	1	1	1	0	0	0	0	0	0	Yunilda 2022
Indonesia	Early Development Instrument	1	0	1	1	1	1	0	1	0	0	0	0	Brinkman 2017
Indonesia	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018
Iran	Ages & Stages Questionnaire (ASQ)	1	0	1	1	1	0	0	0	1	0	0	0	Vameghi 2013
Iran	Bayley Scales of Infant and Toddler Development (BSID)	1	1	1	1	1	0	1	0	1	0	0	0	Azari 2017
Iran	Childhood Nonverbal Communication Scale (CNCS)	1	0	0	1	1	1	1	0	1	0	0	0	Oryadi-Zanjani 2020

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Iran	Children's Behavior Questionnaire (CBQ)	1	1	0	1	1	1	1	0	1	0	0	0	Golmohammadi 2022
Iran	Elicited Imitation Test (EIT)	1	0	1	0	0	1	0	0	0	0	0	1	Mehrani 2018
Iran	Farsi Narrative Norms Instrument	1	1	1	0	1	0	0	0	0	0	0	0	Soleymani 2016
Iran	Infant Neurological International Battery	1	0	1	0	0	1	0	0	0	0	0	0	Soleimani 2007
Jamaica	Intelligibility in Context Scale (ICS) / ICS-Jamaican Creole	1	1	1	1	0	1	1	0	1	0	0	0	Washington 2017
Kenya	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	1	0	1	1	0	1	0	1	0	0	0	McHenry 2021
Kenya	Child behaviour checklist (CBCL)	1	1	0	1	0	0	1	0	1	0	0	0	Kariuki 2016
Kenya	Communicative Development Inventory (CDI)	1	1	0	1	0	1	0	0	0	0	0	0	Alcock 2015
Kenya	Developmental Milestones Checklist (DMC)	1	1	0	1	1	1	1	0	1	0	0	0	Abubakar 2010
Kenya	Executive Function Touch (EF Touch)	2	0	1	1	1	1	2	0	1	0	0	0	Willoughby 2019 (Child Neuropsych); Willoughby 2019 (Dev Science)
Kenya	Kilifi Developmental Inventory	1	1	1	1	1	1	1	0	0	0	0	0	Abubakar 2008

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Kenya	Receptive and expressive vocabulary test battery (English, Luo and Swahili)	1	0	1	1	0	1	1	1	0	0	0	0	Knauer 2019
Kiribati	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	0	1	Sincovich 2019
Laos	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	0	1	Sincovich 2019
Laos	MELQO Measure of Development and Learning Outcomes	1	0	0	0	0	0	1	0	1	0	0	1	Gomez 2022
Malawi	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018
Malawi	Malawi Developmental Assessment Tool (MDAT)	1	1	1	0	1	0	1	0	0	0	0	0	Gladstone 2010
Malawi	Vocabulary Checklists	1	0	0	0	0	1	0	1	0	0	0	0	Prado 2018
Malaysia	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	1	0	0	0	0	1	0	0	0	Zakaria 2012
Malaysia	Malay Preschool Language Assessment Tool (MPLAT)	1	1	1	1	1	0	0	0	0	0	0	0	Razak 2010

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Mali	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018
Mexico	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	1	1	0	0	0	0	1	0	0	0	Ortiz-Leon 2018
Mexico	Spanish-language MacArthur–Bates Communicative Development Inventories short forms (S-CDI SFI and SFII)	1	1	0	0	0	0	1	0	0	0	0	0	Jackson-Maldonado 2013
Mongolia	Mongolian Rapid Baby Scale (MORBAS)	1	0	0	0	1	1	0	0	0	0	0	0	Dagvadorj 2015
Mozambique,	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018
Nepal	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	1	1	0	0	1	0	0	0	0	0	Shrestha 2019
Nepal	Bayley Scales of Infant Development, Third Edition (BSID-III)	3	0	2	1	1	0	0	0	1	0	0	0	Pendergast 2018; Ranjitkar 2018; Manandhar 2016
Nepal	The Test of Infant Motor Performance (TIMP)	1	0	1	1	0	1	0	1	0	0	0	0	Kvestad 2023

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Pakistan	Bayley Scales of Infant Development, Second Edition (BSID-II)	1	0	0	0	0	1	0	0	0	0	0	0	Biasini 2015
Pakistan	Bayley Scales of Infant Development, Third Edition (BSID-III)	2	0	0	0	1	1	0	1	1	0	0	0	Pendergast 2018; Rasheed 2023
Pakistan	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018
Pakistan	Executive function battery	1	0	0	1	1	0	1	1	1	0	0	0	Obradovic 2019
Pakistan	Executive Function Touch (EF Touch)	1	0	0	0	0	0	1	0	1	0	0	0	Toor 2022
Pakistan	MacArthur Communicative Development Inventory Level II	1	1	1	1	1	1	0	0	0	0	0	0	Rasheed 2017
Pakistan	Strengths and Difficulties Questionnaire (SDQ)	1	1	1	1	1	1	1	0	1	0	0	0	Finch 2018
Pakistan	Wechsler Preschool and Primary Scale of Intelligence, thid edition (WPPSI-III)	1	0	1	1	1	0	1	0	1	0	0	0	Rasheed 2018
Peru	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	1	0	0	0	1	1	0	0	Pendergast 2018

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Peru	Child Behavior Scale (CBS)	1	0	0	1	0	0	0	0	1	0	0	0	Meyer 2011
Peru	Escala de Conductas de Aprendizaje Preescolar (ECAP), a Spanish translation of the Preschool Learning Behaviors Scale (PLBS)	1	0	0	0	0	0	0	0	1	0	0	0	Hahn 2009
Rwanda	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018
Samoa	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	0	1	Sincovich 2019
Senegal	Milestones Checklist; Vocabulatory Inventory	1	0	0	1	1	1	1	0	0	0	0	0	Weber 2018
Sierra Leone	Saccadic reaction time (SRTs)	1	1	0	0	0	0	1	0	0	0	0	0	Leppänen 2022
South Africa	Ages & Stages Questionnaire (ASQ)	1	0	0	0	1	0	0	0	0	0	0	0	Bornman 2010
South Africa	Ages & Stages Questionnaire-3 (ASQ-3)	2	0	0	2	0	0	1	0	1	0	0	0	van Heerden 2017; Hsiao 2016
South Africa	Bayley Infant Neurodevelopmental Screener (BINS)	1	0	0	1	0	0	1	0	1	0	0	0	Rodriguez 2020

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
South Africa	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	1	0	0	0	1	0	0	0	Pendergast 2018
South Africa	Early Learning Outcomes Measure (ELOM)	1	0	1	1	0	0	0	0	1	0	0	1	Snelling 2019
South Africa	Little Developmental Coordination Disorder Questionnaire (Little DCDQ)	1	0	0	1	0	1	0	0	0	0	0	0	Venter 2015
South Africa	Mullen Scales of Early Learning (MSEL)	1	0	0	0	1	0	0	0	0	0	0	0	Bornman 2010
South Africa	Parents' Evaluation of Developmental status (PEDS)	1	0	0	0	0	1	0	0	0	0	0	0	Abdoola 2021
South Africa	The infant gross motor screening test (IGMST)	1	1	1	0	1	1	0	0	0	0	0	0	Hilburn 2011
South Africa	The Road to Health Booklet (RTHB)	1	0	0	0	0	1	0	0	0	0	0	0	van der Linde 2015
Sri Lanka	Child Behaviour Assessment Instrument (CBAI)	1	1	0	1	1	1	0	0	0	0	0	0	Samarakkody 2010
Sri Lanka	Denver Developmental Screening Test for Sri Lankan Children (DDST-SL)	1	0	1	0	0	0	0	0	0	0	0	0	Wijedasa 2012

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Sri Lanka	Ragama Early Assessment for Children (REACh)	1	0	1	1	1	1	0	0	0	0	0	0	Caldera 2023
Tanzania	Caregiver Reported Early Development Instruments (CREDI)	1	1	1	1	1	1	1	0	0	0	0	0	McCoy 2017
Thailand	Parents' Evaluation of Developmental Status (PEDS)	1	0	0	0	0	1	0	0	0	0	0	0	Chunsuwan 2016
Thailand	Thai Speech and Language Test (TSLT)	1	0	1	1	1	0	0	0	0	0	0	0	Prathanee 2008
The Gambia	Mullen Scales of Early Learning (MSEL)	1	0	0	0	0	0	1	1	0	0	0	0	Milosavljevic 2019
Tonga	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	0	1	Sincovich 2019
Turkey	Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	1	1	1	1	0	1	0	0	0	0	0	0	Kucuker 2011
Turkey	Bayley Infant Neurodevelopmental Screener (BINS)	1	0	0	0	0	0	0	1	0	0	0	0	Soysal 2014
Turkey	Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	1	1	1	1	0	1	0	0	0	0	0	0	Karabekiroglu 2009
Turkey	Emotional Regulation Checklist	1	1	0	1	1	1	0	0	1	0	0	0	Danişman 2016

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Turkey	Kaufman Survey of Early Academic and Language Skills (K-SEALS)	1	1	1	1	1	0	0	0	1	0	0	0	Uyanik 2014
Turkey	Strengths and Difficulties Questionnaire (SDQ)	1	0	0	1	0	1	1	0	0	0	0	0	Dursun 2020
Tuvalu	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	0	1	Sincovich 2019
Uganda	Bayley Scales of Infant Development, bead threding task	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	British Ability Scales – 3rd ed., block design task	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	International Development and Early Learning Assessment (IDELA)	1	0	0	0	0	0	0	0	1	0	0	0	Halpin 2019
Uganda	Kilifi Developmental Inventory, coin box task	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	Kilifi Picture Vocabulary Test	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	Modified Fagan Test for Infant Intelligence (FTII)	1	0	0	0	0	0	0	1	0	0	0	0	Familiar-Lopez 2022
Uganda	Movement Assessment Battery for Children, balancing on one leg	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Uganda	NEPSY verbal fluence task	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	NEPSY, knock tap game	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	NEPSY, sentence repetition task	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	Sky Search in the Tests of Everyday Attention for Children, Sky Search Task	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	Tap Once Tap Twice Task	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Uganda	Tobii Professional Studio eye tracking programming	1	0	0	0	0	1	0	0	0	0	0	0	Boivin 2017
Uganda	Wisconsin Card Sort Test	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Venezuela	Bayley Infant Neurodevelopmental Screener (BINS)	1	1	1	0	0	0	0	0	0	0	0	0	McCarthy 2012
Viet Nam	Vietnamese version of the Intelligibility in Context Scale (ICS-VN).	1	0	0	1	0	1	1	0	0	0	0	0	Phạm 2017
Vietnam	International Development and Early Learning Assessment (IDELA)	1	0	0	0	0	0	0	0	1	1	0	0	Halpin 2019

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Zambia	Ages & Stages Questionnaire-3 (ASQ-3)	2	0	0	2	0	0	1	0	1	0	0	0	van Heerden 2017; Hsiao 2016
Zambia	Bayley Scales of Infant Development, Second Edition (BSID-II)	1	0	0	0	0	1	0	0	0	0	0	0	Biasini 2015
Zambia	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	0	Pisani 2018

Supplemental Table 5 List of child development assessment tools and countries where they were applied

Name assessment tool	Number of countries where the tool was applied	List of countries where the tool was applied	Number of articles where tool was applied	Age range at assessment (in months)	Type of administration	References (first author & year)
Ages & Stages Questionnaire (ASQ)	2	Iran; South Africa	1	4-60	Indirect/reported	Vameghi 2013; Bornman 2010
Ages & Stages Questionnaire Inventory (ASQ:I)	2	Bangladesh; China	1	1-27	Indirect/reported	Pitchik 2023; Xie 2017
Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	3	Brazil; China; Turkey	3	4-72	Indirect/reported	Anunciação 2019; Bian 2017; Kucuker 2011
Ages & Stages Questionnaire: Social-Emotional, First Edition (ASQ:SE-1)	1	China	1	15-63	Indirect/reported	Xie 2019 (Journal of Child & Dev Studies)
Ages & Stages Questionnaire-3 (ASQ-3)	11	Brazil; China; Colombia; Georgia; Guatemala; India; Indonesia; Mexico; Nepal; South Africa; Zambia	13	1-66	Indirect/reported	Filgueiras 2013; Yue 2019; Rubio-Codina 2016; Rubio-Codina 2020; Zirakashvili 2018; Colbert 2021; Kvestad 2013; Yunilda 2022; Ortiz-Leon 2018; Shrestha 2019; van Heerden 2017; Hsiao 2016; Bian 2012
Ages & Stages Questionnaires: Social Emotional (ASQ:SE-BR)	1	Brazil	1	21-66	Indirect/reported	Chen 2017
Alberta Infant Motor Scale (AIMS)	1	Brazil	4	0-19	Direct assessment	Saccani 2016; Siegle 2018; de Albuquerque 2018; Valentini 2012
Angkor Hospital for Children Developmental Milestone Assessment Tool (AHC DMAT)	1	Cambodia		1-72	Direct assessment, indirect/reported	Ngoun 2012
Battelle Developmental Inventory Second Edition, BDI-2	2	Brazil; Colombia	3	0-95	Direct assessment, indirect/reported	de Albuquerque 2022; Rubio-Codina 2016; Rubio-Codina 2020
Bayley Infant Neurodevelopmental Screener (BINS)	8	Argentina, Brazil, Ecuador, Bolivia, Colombia, Venezuela; Turkey; South Africa	4	3-24	Direct assessment	McCarthy 2012; Soysal 2014; Rodriguez 2020; Guedes 2011
Bayley Scales of Infant and Toddler Development (BSID)	1	Iran	1	1-42	Not reported	Azari 2017
Bayley Scales of Infant Development, bead threding task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010

Bayley Scales of Infant Development, Second Edition (BSID-II)	4	Bangladesh, India, Pakistan, Zambia	2	1-42	Direct assessment	Parveen 2014
Bayley Scales of Infant Development, Third Edition (BSID-III)	12	Bangladesh, Brazil, India, Nepal, Pakistan, Peru, South Africa; Colombia; Ethiopia; Cameroon; Kenya; Malaysia	11	6-42	Direct assessment	Pendergast 2018; Madaschi 2016; Rubio-Codina 2016; Rubio-Codina 2020; Hanlon 2016; Lohaus 2014; McHenry 2021; Zakaria 2012; Ranjitkar 2018; Manandhar 2016; Rasheed 2023
Bracken School Readiness Assessment, Third Edition (BSRA-3)	1	India	1	36-60	Direct assessment	Shah 2013
Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	3	Côte d'Ivoire; Ghana; Turkey	2	23-32	Indirect/reported	Barthel 2021; Karabekiroglu 2009
British Ability Scales – 3rd ed., block design task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Caregiver Reported Early Development Instruments (CREDI)	18	17 countries not listed, Tanzania	2	0-35	Indirect/reported	Waldman 2021; McCoy 2017
Caregiver Reported Early Development Instruments (CREDI), long-form	2	China; India	2	5-36	Indirect/reported	Li 2020; Alderman 2021
Caregiver Reported Early Development Instruments (CREDI), short-form	18	17 countries not listed; Brazil	2	0-35	Indirect/reported	McCoy 2018; Altafim 2018
Child Behavior Checklist/1.5-5 (CBCL/1.5-5) and the Caregiver-Teacher Report Form (C-TRF)	1	China	1	50-71	Indirect/reported	Liu 2011
Child Behavior Scale (CBS)	1	Peru	1	24-72	Indirect/reported	Meyer 2011
Child Behaviour Assessment Instrument (CBAI)	1	Sri Lanka	1	48-72	Indirect/reported	Samarakkody 2010
Child behaviour checklist (CBCL)	1	Kenya	1	12-72	Indirect/reported	Kariuki 2016
Child Social Preference Scale	1	China	1	36-60	Indirect/reported	Li 2016
Childhood Nonverbal Communication Scale (CNCS)	1	Iran	1	3-18	Indirect/reported	Oryadi-Zanjani 2020
Children's Behavior Questionnaire (CBQ)	1	Iran	1	48-72	Indirect/reported	Golmohammadi 2022
Chinese Preschool Readiness Scale (CPRS)	1	China	1	31-62	Indirect/reported	Xie 2019 (Early Educ & Dev)
Chinese version of Urban Infant-Toddler Social and Emotional Assessment	1	China	1	12-36	Indirect/reported	Jianduan 2009
Chinese Vocabulary Checklist	1	China	1	12-30	Indirect/reported	Hao 2008
Communicative Development Inventory (CDI)	1	Kenya	1	6-36	Indirect/reported	Alcock 2015
Denver Developmental Screening Test for Sri Lankan Children (DDST-SL)	1	Sri Lanka	1	0-80	Direct assessment	Wijedasa 2012

Denver Developmental Screening Test (Denver-II)	3	Brazil; Colombia; Ethiopia	4	0-70.6	Direct assessment, indirect/reported	Lopez Boo 2020; Rubio-Codina 2016; Rubio-Codina 2020; Abessa 2016
DEvelopmental assessment on an E-Platform (DEEP)	1	India	2	34-40	Direct assessment	Bhavnani 2019; Mukherjee 2020
Developmental Assessment Tool for Anganwadis (DATA)	1	India	1	19-36	Direct assessment	Nair 2009
Developmental Assessment Tool for Anganwadis (DATA-II)	1	India	1	36-48	Direct assessment	Nair 2012
Developmental Milestone Chart (DMC)	1	Cambodia	1	1-96	Direct assessment, indirect/reported	Scherzer 2009
Developmental Milestones Checklist (DMC)	1	Kenya	1	2.53–12.06	Indirect/reported	Abubakar 2010
Dimensions of Mastery Questionnaire (DMQ-18)	1	Bangladesh	1	36-72	Indirect/reported	Shaoli 2019
Early Development Instrument	1	Indonesia	1	48-96	Indirect/reported	Brinkman 2017
Early Human Capability Index (eHCI)	7	Brazil, Kiribati, China, Laos, Samoa, Tonga, Tuvalu	1	24-72	Indirect/reported	Sincovich 2019
Early Learning Outcomes Measure (ELOM)	1	South Africa	1	50-69	Direct assessment	Snelling 2019
EDIN II (Evaluación del Desarrollo Integral de la niña y el niño)	1	Costa Rica	1	0-72	Direct assessment	Cordero, 2019
Elicited Imitation Test (EIT)	1	Iran	1	36-83	Direct assessment	Mehrani 2018
Emotional and Behavioral Screener (EBS)	1	China	1	48-84	Indirect/reported	Zhang 2018
Emotional Regulation Checklist	1	Turkey	1	48-60	Indirect/reported	Danişman 2016
Escala de Conductas de Aprendizaje Preescolar (ECAP), a Spanish translation of the Preschool Learning Behaviors Scale (PLBS)	1	Peru	1	24-72	Indirect/reported	Hahn 2009
Escala de Habildades Sociales (Scale of Social Abilities)	1	Argentina	1	36-60	Indirect/reported	Lacunza 2009
Executive function battery	1	Pakistan	1		Direct assessment	Obradovic 2019
Executive Function Touch (EF Touch)	2	Kenya	3	36-72	Self-administration	Willoughby 2019 (Child Neuropsych); Willoughby 2019 (Dev Science); Toor 2022
Farsi Narrative Norms Instrument	1	Iran	1	60-72	Direct assessment	Soleymani 2016
Global Scales for Early Development	32	Not listed	1		Direct assessment, indirect/reported	McCray 2023

Infant Neurological International Battery	1	Iran	1	4-18	Direct assessment	Soleimani 2007
Intelligibility in Context Scale (ICS) / ICS-Jamaican Creole	1	Jamaica	1	39-75	Indirect/reported	Washington 2017
INTER- GROWTH-21st Project Neurodevelopment Assessment (INTER-NDA)	1	Grenada	1	22-30	Direct assessment, indirect/reported	Waechter 2022
International Development and Early Learning Assessment (IDELA)	17	Afghanistan, Ethiopia, Uganda, Vietnam, Bolivia, Bangladesh, Bhutan, Egypt, India, Indonesia, Mali, Malawi, Mozambique, Pakistan, Rwanda, Zambia, Ghana	4	36-96	Direct assessment	Halpin 2019; Pisani 2018; Wolf 2017; Pisani 2022
Kaufman Survey of Early Academic and Language Skills (K-SEALS)	1	Turkey		61-72	Direct assessment	Uyanik 2014
Kilifi Developmental Inventory	2	Ghana, Côte d'Ivoire, Kenya	2	6-35	Direct assessment	Barthel 2022; Abubakar 2008
Kilifi Developmental Inventory, coin box task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Kilifi Picture Vocabulary Test	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Little Developmental Coordination Disorder Questionnaire (Little DCDQ)	1	South Africa	1	36-49	Indirect/reported	Venter 2015
MacArthur Communicative Development Inventory Level II	1	Pakistan	1		Indirect/reported	Rasheed 2017
MacArthur-Bates Communicative Development Inventories CDI	1	Colombia	1	8-30	Indirect/reported	Lara Díaz 2011
MacArthur-Bates Communicative Development Inventories Short-Forms I vocabulary checklist	1	Colombia	2	6-42	Indirect/reported	Rubio-Codina 2016; Rubio-Codina 2020
MacArthur-Bates Communicative Development Inventories Short-Forms II vocabulary checklist	1	Colombia	2	6-42	Indirect/reported	Rubio-Codina 2016; Rubio-Codina 2020
Malawi Developmental Assessment Tool (MDAT)	1	Malawi	1	0-84	Direct assessment, indirect/reported	Gladstone 2010
Malawi Developmental Assessment Tool (MDAT), renamed Dominican Child Development Screening Tool	1	Dominican Republic	1	2-24	Direct assessment, indirect/reported	Sánchez-Vincitore 2019

Malay Preschool Language Assessment Tool (MPLAT)	1	Malaysia	1	48-83	Direct assessment	Razak 2010
MELQO Measure of Development and Learning Outcomes	1	Laos	1	48-49	Direct assessment	Gomez 2022
Milestones Checklist; Vocabulatory Inventory	1	Senegal	1	4-30	Indirect/reported	Weber 2018
Modified Fagan Test for Infant Intelligence (FTII)	1	Uganda	1	6-12	Direct assessment	Familiar-Lopez 2022
Mongolian Rapid Baby Scale (MORBAS)	1	Mongolia	1	0-42	Indirect/reported	Dagvadorj 2015
Motor Development Scale	1	Brazil	1	48-72	Not reported	Okuda 2020
Movement Assessment Battery for Children-second edition (MBAC-2)	1	China	1	36-72	Direct assessment	Hua 2013
Movement Assessment Battery for Children, balancing on one leg	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Mullen Scales of Early Learning (MSEL)	4	Guatemala; South Africa; The Gambia	5	0-68	Direct assessment	Koura 2013; Boivin 2021; Colbert 2020; Bornman 2010; Milosavljevic 2019
NEPSY verbal fluence task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
NEPSY, knock tap game	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
NEPSY, sentence repetition task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Oxford Neurodevelopment Assessment (OX-NDA)	1	Brazil	1		Direct assessment; indirect/reported	Fernandes 2022
Parents' Evaluation of Developmental Status (PEDS)	3	India; South Africa; Thailand	3	0-30	Indirect/reported	Mukherjee 2022; Abdoola 2021; Chunsuwan 2016
Preschool and Kindergarten Behavior Scale (PKBS)	1	Brazil	1	36-72	Indirect/reported	Dias 2011
Preschool Language Scale-3 (PLS- 3)	1	Colombia	1	48-84	Direct assessment	Romero 2013
Preschool Learning Behavior Scale (PLBS)	1	China	1		Indirect/reported	Wu 2019 (JPA)
Questionário para Avaliação do Desenvolvimento Infantil (QAD-PIPAS)	1	Brazil	1	0-59	Indirect/reported	Venancio 2020
Ragama Early Assessment for Children (REACh)	1	Sri Lanka	1	24-60	Indirect/reported	Caldera 2023
Rapid Assessment of Cognitive and Emotional Regulation (RACER)	1	Ghana	1	48-72	Self-administration	Yuan 2022
Rapid Neurodevelopmental Assessment (RNDA)	2	Bangladesh; Guatemala	3	0-60	Direct assessment	Khan 2010; Khan 2013; Thompson 2015
Receptive and expressive vocabulary test battery (English, Luo and Swahili)	1	Kenya	1	24-83	Direct assessment	Knauer 2019

Saccadic reaction time (SRTs)	1	Sierra Leone	1	7-11.9	Self-administration	Leppänen 2022
Screening for Identification of Oral Language Difficulties by Preschool Teachers (SIOLD)	1	Brazil	1	51-65	Indirect/reported	Puglisi 2020
Screening Test of Early Language Development-Test version (STELD-T)	1	India	1	6-21	Indirect/reported	Pathak 2023
Self-ordered Pointing Task (SOPT)	1	Brazil	1	42-60	Self-administration	Sallum 2017
Sky Search in the Tests of Everyday Attention for Children, Sky Search Task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Social Skills domain of System-Rating Scales (SSIS-RS)	1	China	1	36-60	Indirect/reported	Wu 2019 (PS)
Spanish-language MacArthur–Bates Communicative Development Inventories short forms (S-CDI SFI and SFII)	1	Mexico	1	8-30	Indirect/reported	Jackson-Maldonado 2013
Strengths and Difficulties Questionnaire (SDQ)	2	Pakistan; Turkey	2	20-51	Indirect/reported	Finch 2018; Dursun 2020
Tap Once Tap Twice Task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Test of Gross Motor Development (TGMD-2)	1	Brazil	1	36-60	Direct assessment	Bandeira 2020
Thai Speech and Language Test (TSLT)	1	Thailand	1	3-24	Direct assessment	Prathanee 2008
The Chinese Early Language and Communication Questionnaire	1	China	1	2-48	Indirect/reported	Zhang 2021
The Chinese Intelligence Scale for Young Children	1	China	1	36-84	Direct assessment	Guo 2009
The infant gross motor screening test (IGMST)	1	South Africa	1	6-18	Direct assessment	Hilburn 2011
The language inventory	1	Bangladesh	1	12-18	Indirect/reported	Hamadani 2010
The Road to Health Booklet (RTHB)	1	South Africa	1	6-12	Indirect/reported	van der Linde 2015
The Test of Infant Motor Performance (TIMP)	1	Nepal	1	2-3	Direct assessment	Kvestad 2023
Tobii Professional Studio eye tracking programming	1	Uganda	1	44-65	Direct assessment	Boivin 2017
Vietnamese version of the Intelligibility in Context Scale (ICS-VN).	1	Viet Nam	1	24-71	Indirect/reported	Phạm 2017
Vineland Adaptive Behaviour Scale (VABS II)	1	India	1	36-108	Indirect/reported	Kumar 2016
Vocabulary Checklists	2	Malawi, Ghana	1	12-17; 48-72	Indirect/reported	Prado 2018

Wechsler Preschool and Primary Scale of Intelligence, thid edition (WPPSI-III)	1	Pakistan	1		Direct assessment	Rasheed 2018
WHO Gross Motor Milestones	2	Bangladesh; Colombia	3	3-42	Direct assessment; indirect/reported	Hamadani 2013; Rubio-Codina 2016; Rubio-Codina 2020
Wisconsin Card Sort Test	1	Uganda	1	55-67	Direct assessment	Nampijja 2010

Supplemental Table 6 Number of articles providing psychometric evidence by world region and urban/rural location

Region	Urban	Rural	Urban-rural
Caribbean	2	0	0
Central America	1	3	0
East Asia	10	1	8
Oceania	0	0	1
North Africa and the Middle East	0	0	1
North America	1	0	0
South America	18	0	7
South Asia	11	13	9
Southeast Asia	5	1	3
Sub-Saharan Africa	10	15	13
West Asia	8	0	2

Supplemental Table 7 Number of articles providing psychometric evidence by child development assessment tools (development articles)

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Angkor Hospital for Children Developmental Milestone Assessment Tool (AHC DMAT)	1	1	0	0	0	1	0	0	0	0	0	0	0	Ngoun 2012
Caregiver Reported Early Development Instruments (CREDI)	2	18	2	1	2	2	2	2	0	1	0	0	0	McCoy 2017; Waldman 2021
Caregiver Reported Early Development Instruments (CREDI), short-form	1	17	1	0	1	1	1	1	0	0	0	0	1	McCoy 2018
Child Behaviour Assessment Instrument (CBAI)	1	1	1	0	1	1	1	0	0	0	0	0	0	Samarakkody 2010
Childhood Nonverbal Communication Scale (CNCS)	1	1	0	0	1	1	1	1	0	1	0	0	0	Oryadi-Zanjani 2020
Chinese Preschool Readiness Scale (CPRS)	1	1	0	0	1	1	1	0	0	1	0	0	0	Xie 2019 (Early Educ & Dev)
Chinese Vocabulary Checklist	1	1	0	0	0	0	0	1	0	0	0	0	0	Hao 2008
DEvelopmental assessment on an E-Platform (DEEP)	1	1	0	0	0	1	0	0	0	0	0	0	0	Bhavnani 2019
Developmental Assessment Tool for Anganwadis (DATA)	1	1	0	0	1	1	0	0	0	1	0	0	0	Nair 2009
Developmental Assessment Tool for Anganwadis (DATA-II)	1	1	0	0	1	1	0	0	0	1	0	0	0	Nair 2012
Developmental Milestone Chart (DMC)	1	1	0	0	0	1	0	0	0	0	0	0	0	Scherzer 2009
Developmental Milestones Checklist (DMC)	1	1	1	0	1	1	1	1	0	1	0	0	0	Abubakar 2010
Early Learning Outcomes Measure (ELOM)	1	1	0	1	1	0	0	0	0	1	0	0	1	Snelling 2019
Elicited Imitation Test (EIT)	1	1	0	1	0	0	1	0	0	0	0	0	1	Mehrani 2018
Escala de Habildades Sociales (Scale of Social Abilities)	1	1	0	0	1	0	0	0	0	1	0	0	0	Lacunza 2009

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Farsi Narrative Norms Instrument	1	1	1	1	0	1	0	0	0	0	0	0	0	Soleymani 2016
Global Scales for Early Development	1	32	0	0	1	1	0	0	0	0	0	0	0	McCray 2023
International Development and Early Learning Assessment (IDELA)	1	12	1	1	1	1	1	1	0	0		0	0	Pisani 2018
Kilifi Developmental Inventory	1	1	1	1	1	1	1	1	0	0		0	0	Abubakar 2008
Malawi Developmental Assessment Tool (MDAT)	1	1	1	1	0	1	0	1	0	0		0	0	Gladstone 2010
Malay Preschool Language Assessment Tool (MPLAT)	1	1	1	1	1	1	0	0	0	0		0	0	Razak 2010
Mongolian Rapid Baby Scale (MORBAS)	1	1	0	0	0	1	1	0	0	0		0	0	Dagvadorj 2015
Oxford Neurodevelopment Assessment (OX-NDA)	1	1	1	1	1	0	1	0	0	0		0	0	Fernandes 2022
Questionário para Avaliação do Desenvolvimento Infantil (QAD-PIPAS)	1	1	0	0	0	1	0	0	0	0	0	0	0	Venancio 2020
Ragama Early Assessment for Children (REACh)	1	1	0	1	1	1	1	0	0	0	0	0	0	Caldera 2023
Rapid Neurodevelopmental Assessment (RNDA)	1	1	0	1	0	1	1	0	1	0	0	0	0	Khan 2010
Receptive and expressive vocabulary test battery (English, Luo and Swahili)	1	1	0	1	1	0	1	1	1	0	0	0	0	Knauer 2019
Saccadic reaction time (SRTs)	1	1	1	0	0	0	0	1	0	0	0	0	0	Leppänen 2022
Screening for Identification of Oral Language Difficulties by Preschool Teachers (SIOLD)	1	1	0	0	0	0	0	0	0	1	0	0	0	Puglisi 2020
Screening Test of Early Language Development-Test version (STELD-T)	1	1	0	0	1	0	1	0	0	0	0	0	0	Pathak 2023
Thai Speech and Language Test (TSLT)	1	1	0	1	1	1	0	0	0	0	0	0	0	Prathanee 2008
The infant gross motor screening test (IGMST)	1	1	1	1	0	1	1	0	0	0	0	0	0	Hilburn 2011
The language inventory	1	1	1	0	0	1	1	1	1	0	0	0	0	Hamadani 2010

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu -ral	Over countries	Over time	Over other groups	First author & year
Vocabulary Checklists	1	1	0	0	0	0	1	0	1	0	0	0	0	Prado 2018

Supplemental Table 8 Number of articles providing psychometric evidence by child development assessment tools (adaptation articles)

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Ages & Stages Questionnaire (ASQ)	2	2	0	1	1	2	0	0	0	1	0	0	0	Vameghi 2013; Bornman 2010
Ages & Stages Questionnaire Inventory (ASQ:I)	2	2	1	1	2	1	2	1	0	0	0	0	0	Xie 2017; Pitchik 2023
Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	3	3	2	2	2	1	1	0	0	2	0	0	0	Kucuker 2011; Anunciação 2019; Bian 2017
Ages & Stages Questionnaire: Social-Emotional, First Edition (ASQ:SE-1)	1	1	1	1	1	0	1	0	0	0	0	0	0	Xie 2019 (Journal of Child & Dev Studies)
Ages & Stages Questionnaire-3 (ASQ-3)	13	11	5	5	12	3	5	5	2	4	0	0	1	Zirakashvili 2018; Colbert 2021; Rubio-Codina 2016; Rubio-Codina 2020; Shrestha 2019; Hsiao 2016; Filgueiras 2013; van Heerden 2017; Ortiz-Leon 2018; Kvestad 2013; Yue 2019; Yunilda 2022; Bian 2012
Ages & Stages Questionnaires: Social Emotional (ASQ:SE-BR)	1	1	0	0	0	0	0	0	0	1	0	1	0	Chen 2017
Alberta Infant Motor Scale (AIMS)	4	1	3	3	2	1	3	1	1	0	0	0	0	Saccani 2016; Siegle 2018; de Albuquerque 2018; Valentini 2012
Battelle Developmental Inventory Second Edition, BDI-2	3	2	2	0	2	1	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020; de Albuquerque 2022
Bayley Infant Neurodevelopment Screener (BINS)	4	8	1	1	2	1	1	1	2	2	0	0	0	Guedes 2011; McCarthy 2012; Soysal 2014; Rodriguez 2020
Bayley Scales of Infant and Toddler Development (BSID)	1	1	1	1	1	1	0	1	0	1	0	0	0	Azari 2017

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Bayley Scales of Infant Development, bead threding task	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Bayley Scales of Infant Development, Second Edition (BSID-II)	2	4	0	0	0	1	1	0	0	0	0	0	0	Biasini 2015; Parveen 2014
Bayley Scales of Infant Development, Third Edition (BSID-III)	11	12	4	3	7	2	3	3	4	5	2	0	0	Rubio-Codina 2016; Rubio-Codina 2020; Hanlon 2016; Ranjitkar 2018; Madaschi 2016; Pendergast 2018; Manandhar 2016; Zakaria 2012; Lohaus 2014; McHenry 2021; Rasheed 2023
Bracken School Readiness Assessment, Third Edition (BSRA-3)	1	1	0	0	1	1	0	0	0	0	0	0	0	Shah 2013
Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	2	3	1	1	2	0	1	1	0	1	0	0	0	Karabekiroglu 2009; Barthel 2021
British Ability Scales – 3rd ed., block design task	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Caregiver Reported Early Development Instruments (CREDI), long-form	2	2	0	0	2	0	2	2	0	0	0	0	0	Alderman 2021; Li 2020
Caregiver Reported Early Development Instruments (CREDI), short-form	1	1	1	0	1	1	1	1	0	0	0	0	0	Altafim 2018
Child Behavior Checklist/1.5-5 (CBCL/1.5-5) and the Caregiver-Teacher Report Form (C-TRF)	1	1	0	0	0	0	0	0	0	1	1	0	0	Liu 2011
Child Behavior Scale (CBS)	1	1	0	0	1	0	0	0	0	1	0	0	0	Meyer 2011
Child behaviour checklist (CBCL)	1	1	1	0	1	0	0	1	0	1	0	0	0	Kariuki 2016
Child Social Preference Scale	1	1	0	0	1	0	0	0	1	1	0	0	1	Li 2016
Children's Behavior Questionnaire (CBQ)	1	1	1	0	1	1	1	1	0	1	0	0	0	Golmohammadi 2022
Chinese version of Urban Infant-Toddler Social and Emotional Assessment	1	1	1	1	1	1	1	0	0	1	0	0	0	Jianduan 2009

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Communicative Development Inventory (CDI)	1	1	1	0	1	0	1	0	0	0	0	0	0	Alcock 2015
Denver Developmental Screening Test for Sri Lankan Children (DDST-SL)	1	1	0	1	0	0	0	0	0	0	0	0	0	Wijedasa 2012
Denver Developmental Screening Test (Denver-II)	4	3	3	2	3	2	1	2	1	1	0	0	0	Abessa 2016; Rubio-Codina 2016; Rubio-Codina 2020; Lopez Boo 2020
DEvelopmental assessment on an E-Platform (DEEP)	1	1	0	0	0	0	1	0	0	0	0	0	0	Mukherjee 2020
Dimensions of Mastery Questionnaire (DMQ-18)	1	1	1	1	1	1	1	0	0	1	0	0	0	Shaoli 2019
Early Development Instrument	1	1	0	1	1	1	1	0	1	0	0	0	0	Brinkman 2017
Early Human Capability Index (eHCI)	1	7	0	0	1	0	0	0	0	1	0	0	1	Sincovich 2019
EDIN II (Evaluación del Desarrollo Integral de la niña y el niño)	1	1	0	0	0	1	0	1	0	0	0	0	0	Cordero, 2019
Emotional and Behavioral Screener (EBS)	1	1	0	1	1	0	1	0	0	1	0	0	0	Zhang 2018
Emotional Regulation Checklist	1	1	1	0	1	1	1	0	0	1	0	0	0	Danişman 2016
Escala de Conductas de Aprendizaje Preescolar (ECAP), a Spanish translation of the Preschool Learning Behaviors Scale (PLBS)	1	1	0	0	0	0	0	0	0	1	0	0	0	Hahn 2009
Executive function battery	1	1	0	0	1	1	0	1	1	1	0	0	0	Obradovic 2019
Executive Function Touch (EF Touch)	3	2	0	1	1	1	1	3	0	2	0	0	0	Willoughby 2019 (Child Neuropsych); Willoughby 2019 (Dev Science); Toor 2022
Infant Neurological International Battery	1	1	0	1	0	0	1	0	0	0	0	0	0	Soleimani 2007
Intelligibility in Context Scale (ICS) / ICS-Jamaican Creole	1	1	1	1	1	0	1	1	0	1	0	0	0	Washington 2017

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over time	Over other groups	First author & year
INTER- GROWTH-21st Project Neurodevelopment Assessment (INTER-NDA)	1	1	1	1	1	0	0	0	0	0	0	0	0	Waechter 2022
International Development and Early Learning Assessment (IDELA)	3	17	1	1	1	0	1	1	1	2	1	0	1	Halpin 2019; Pisani 2022; Wolf 2017
Kaufman Survey of Early Academic and Language Skills (K-SEALS)	1	1	1	1	1	1	0	0	0	1	0	0	0	Uyanik 2014
Kilifi Developmental Inventory	1	2	0	1	1	0	0	0	1	1	0	0	0	Barthel 2022
Kilifi Developmental Inventory, coin box task	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Kilifi Picture Vocabulary Test	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Little Developmental Coordination Disorder Questionnaire (Little DCDQ)	1	1	0	0	1	0	1	0	0	0	0	0	0	Venter 2015
MacArthur Communicative Development Inventory Level II	1	1	1	1	1	1	1	0	0	0	0	0	0	Rasheed 2017
MacArthur-Bates Communicative Development Inventories CDI	1	1	0	0	1	1	0	1	0	0	0	0	0	Lara Díaz 2011
MacArthur-Bates Communicative Development Inventories Short-Forms I vocabulary checklist	2	1	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
MacArthur-Bates Communicative Development Inventories Short-Forms II vocabulary checklist	2	1	2	0	2	0	1	1	1	0	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Malawi Developmental Assessment Tool (MDAT), renamed Dominican Child Development Screening Tool	1	1	0	1	1	0	0	1	0	0	0	0	0	Sánchez-Vincitore 2019
MELQO Measure of Development and Learning Outcomes	1	1	0	0	0	0	0	1	0	1	1	0	1	Gomez 2022
Milestones Checklist; Vocabulatory Inventory	1	1	0	0	1	1	1	1	0	0	0	0	0	Weber 2018
Modified Fagan Test for Infant Intelligence (FTII)	1	1	0	0	0	0	0	0	1	0	0	0	0	Familiar-Lopez 2022

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Motor Development Scale	1	1	0	0	1	0	0	0	0	1	0	0	0	Okuda 2020
Movement Assessment Battery for Children- second edition (MBAC-2)	1	1	1	1	1	1	1	0	0	1	0	0	0	Hua 2013
Movement Assessment Battery for Children, balancing on one leg	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Mullen Scales of Early Learning (MSEL)	5	4	1	2	1	1	1	3	3	0	0	0	0	Colbert 2020; Koura 2013; Milosavljevic 2019; Bornman 2010; Boivin 2021
NEPSY verbal fluence task	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
NEPSY, knock tap game	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
NEPSY, sentence repetition task	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Parents' Evaluation of Developmental status (PEDS)	3	3	0	0	0	0	3	0	0	0	0	0	0	Abdoola 2021; Chunsuwan 2016; Mukherjee 2022
Preschool and Kindergarten Behavior Scale (PKBS)	1	1	0	0	1	0	0	0	0	1	0	0	0	Dias 2011
Preschool Language Scale-3 (PLS- 3)	1	1	0	0	1	0	0	0	0	1	0	0	0	Romero 2013
Preschool Learning Behavior Scale (PLBS)	1	1	0	0	1	0	1	0	0	1	0	0	0	Wu 2019 (JPA)
Rapid Assessment of Cognitive and Emotional Regulation (RACER)	1	1	1	1	0	0	1	1	0	0	0	0	0	Yuan 2022
Rapid Neurodevelopmental Assessment (RNDA)	2	2	0	2	0	1	2	1	0	0	0	0	0	Thompson 2015; Khan 2013
Self-ordered Pointing Task (SOPT)	1		0	0	0	0	1	1	0	0	0	0	0	Sallum 2017
Sky Search in the Tests of Everyday Attention for Children, Sky Search Task	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Social Skills domain of System-Rating Scales (SSIS-RS)	1	1	0	0	1	0	0	1	0	1	0	0	0	Wu 2019 (PS)
Spanish-language MacArthur–Bates Communicative Development Inventories short forms (S-CDI SFI and SFII)	1	1	1	0	0	0	0	1	0	0	0	0	0	Jackson-Maldonado 2013

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance			References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over time	Over other groups	First author & year
Strengths and Difficulties Questionnaire (SDQ)	2	2	1	1	2	1	2	2	0	1	0	0	0	Finch 2018; Dursun 2020
Tap Once Tap Twice Task	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010
Test of Gross Motor Development (TGMD-2)	1	1	0	1	1	0	0	1	0	1	0		0	Bandeira 2020
The Chinese Early Language and Communication Questionnaire	1	1	0	0	1	0	1	0	0	0	0		0	Zhang 2021
The Chinese Intelligence Scale for Young Children	1	1	0	0	0	0	0	0	0	1	0		1	Guo 2009
The Road to Health Booklet (RTHB)	1	1	0	0	0	0	1	0	0	0	0		0	van der Linde 2015
The Test of Infant Motor Performance (TIMP)	1	1	0	1	1	0	1	0	1	0	0		0	Kvestad 2023
Tobii Professional Studio eye tracking programming	1	1	0	0	0	0	1	0	0	0	0		0	Boivin 2017
Vietnamese version of the Intelligibility in Context Scale (ICS-VN).	1	1	0	0	1	0	1	1	0	0	0	0	0	Phạm 2017
Vineland Adaptive Behaviour Scale (VABS II)	1	1	1	1	0	0	0	1	0	0	0	0	0	Kumar 2016
Wechsler Preschool and Primary Scale of Intelligence, thid edition (WPPSI-III)	1	1	0	1	1	1	0	1	0	1	0	0	0	Rasheed 2018
WHO Gross Motor Milestones	3	2	2	1	2	0	2	2	2	0	0	0	0	Hamadani 2013; Rubio-Codina 2016; Rubio-Codina 2020
Wisconsin Card Sort Test	1	1	1	0	1	0	0	1	0	1	0	0	0	Nampijja 2010

Supplemental Table 9 Number of articles providing psychometric evidence by article type and early childhood development domain

Child development domain	Total number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		
		Test-retest	Inter-rater	Internal consistency	Content/ face	Concurrent /Criterion	Convergent	Predictive	Structural	Over countries	Over time	Over other groups
All articles (N=175 article-tool dyads)												
Cognitive	77	28	29	52	35	34	28	16	22	0	0	3
Language	105	39	36	70	44	49	40	19	28	0	0	6
Motor	98	38	39	60	36	44	37	19	28	0	0	4
Socioemotional/ temperament	47	19	15	35	20	28	15	3	24	0	1	4
Attention/ executive function	20	9	3	12	2	5	15	1	16	0	0	3
Personal-social/ adaptive	49	19	18	32	20	22	16	6	12	0	0	1
Academic/ preacademic	8	1	3	4	1	2	2	1	7	0	0	4
Development articles (N=35 article-tool dyads)												
Cognitive	17	7	7	12	14	10	6	1	5	0	0	2
Language	27	9	10	15	20	15	9	4	7	0	0	2
Motor	19	9	9	12	17	11	7	1	5	0	0	2
Socioemotional/ temperament	12	6	4	9	11	9	4	0	2	0	0	1
Attention/ executive function	1	0	1	1	0	0	0	0	1	0	0	1
Personal-social/ adaptive	10	2	3	6	9	4	2	1	4	0	0	0
Academic/ preacademic	1	0	1	1	0	0	0	0	1	0	0	1
Adaptation articles (N=140 article-tool dyads)												
Cognitive	60	21	22	40	21	24	22	15	17	1	0	1
Language	78	30	26	55	24	34	31	15	21	2	0	4
Motor	79	29	30	48	19	33	30	18	23	3	0	2
Socioemotional/ temperament	35	13	11	26	9	19	11	3	22	2	1	3

Attention/ executive function	19	9	2	11	2	5	15	1	15		0	2
Personal-social/ adaptive	39	17	15	26	11	18	14	5	8		0	1
Academic/ preacademic	7	1	1	3	1	2	2	1	6		0	3

Supplemental Table 10 Number of articles by category of risk of bias

	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data
Low risk	42	80	74
Some concerns	11	8	49
High risk	0	1	7
Unable to assess	122	8	44
Not applicable	0	78	1

	Indirectness (are results generalisable?)
Yes	18
Probably yes	81
Probably no	49
No	24
No information	3

Supplemental Table 11 Risk of bias assessment by included article

First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Abdoola 2021	unable to assess	N/A	unable to assess	probably no
Abessa 2016	low risk	N/A	unable to assess	probably yes
Abubakar 2008	unable to assess	Low risk	low risk	probably yes
Abubakar 2010	unable to assess	Some concerns	High risk	No
Alcock 2015	unable to assess	N/A	low risk	Probably yes
Alderman 2021	Low risk	unable to assess	Low risk	Yes
Altafim 2018	low risk	unable to assess	some concerns	probably yes
Anunciação 2019	unable to assess	n/a	some concerns	Yes
Azari 2017	unable to assess	high risk	unable to assess	probably no
Bandeira 2020	unable to assess	low risk	Low risk	probably yes
Barthel 2021	unable to assess	low risk	low risk	probably yes
Barthel 2022	Some concerns	Low risk	Low risk	Probably no
Bhavnani 2019	unable to assess	N/A	low risk	Yes
Bian 2012	unable to assess	N/A	Low risk	no
Bian 2017	unable to assess	n/a	unable to assess	Yes
Biasini 2015	unable to assess	N/A	Some concerns	Probably yes
Boivin 2017	unable to assess	Low risk	Low risk	No
Boivin 2021	unable to assess	low risk	low risk	probably yes
Bornman 2010	unable to assess	N/A	Low risk	no
Bornman 2010	unable to assess	N/A	low risk	no
Brinkman 2017	some concerns	Low risk	unable to assess	probably yes
Caldera 2023	unable to assess	N/A	unable to assess	Yes
Chen 2017	unable to assess	N/A	some concerns	yes
Chunsuwan 2016	unable to assess	N/A	unable to assess	probably no
Colbert 2020	low risk	low risk	unable to assess	probably yes
Colbert 2021	unable to assess	low risk	Low risk	No information
Cordero, 2019	unable to assess	N/A	unable to assess	Yes
Dagvadorj 2015	unable to assess	N/A	Low risk	Probably no
Danişman 2016	unable to assess	N/A	high risk	no
de Albuquerque 2018	unable to assess	N/A	Low risk	probably yes
de Albuquerque 2022	unable to assess	N/A	Low risk	probably yes
Dias 2011	unable to assess	N/A	low risk	Probably yes
Dursun 2020	unable to assess	low risk	Some concerns	probably no
Familiar-Lopez 2022	unable to assess	Some concerns	Some concerns	No
Fernandes 2022	unable to assess	low risk	low risk	probably yes
Filgueiras 2013	unable to assess	N/A	unable to assess	probably yes
Finch 2018	low risk	Low risk	low risk	probably no
Gladstone 2010	unable to assess	low risk	some concerns	Probably yes

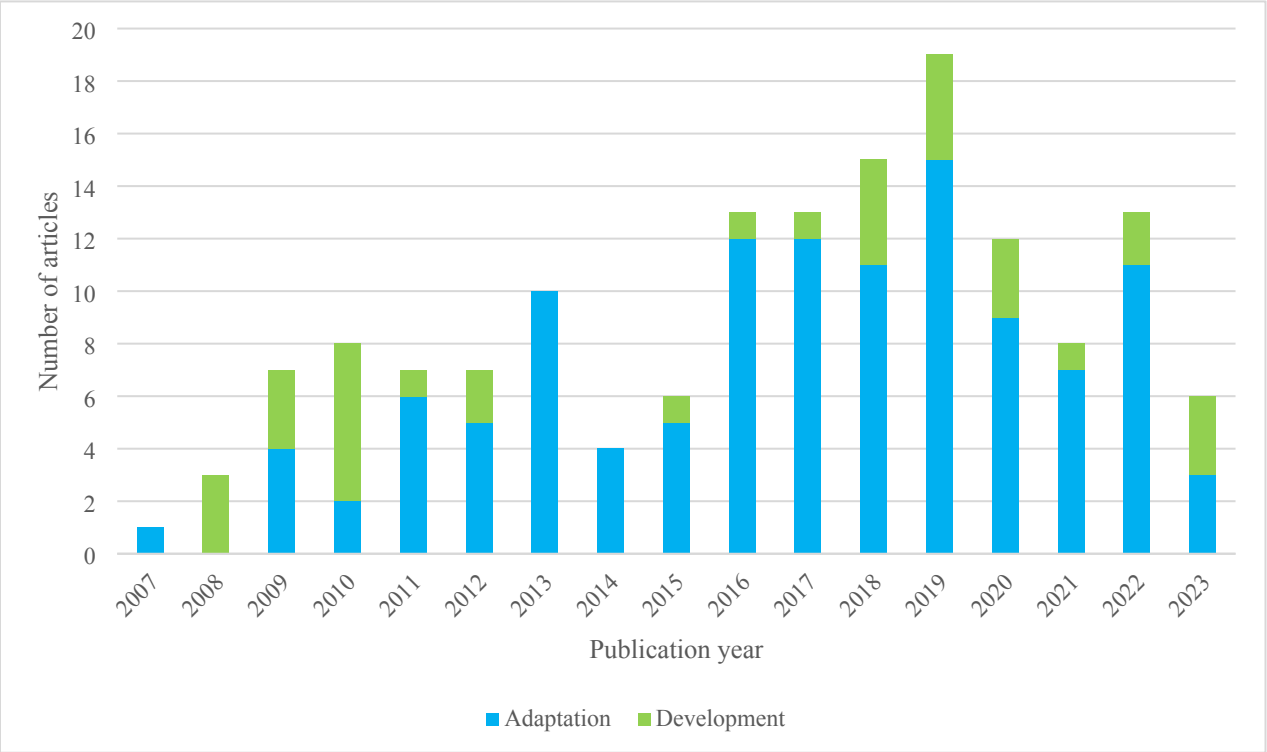
First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Golmohammadi 2022	unable to assess	Low risk	low risk	probably no
Gomez 2022	unable to assess	Low risk	Low risk	Probably yes
Guedes 2011	unable to assess	low risk	unable to assess	probably no
Guo 2009	unable to assess	N/A	unable to assess	Probably yes
Hahn 2009	unable to assess	N/A	high risk	probably yes
Halpin 2019	unable to assess	N/A	unable to assess	yes
Hamadani 2010	low risk	low risk	high risk	probably yes
Hamadani 2013	unable to assess	low risk	low risk	Probably yes
Hanlon 2016	Some concerns	Some concerns	Low risk	Probably no
Hao 2008	unable to assess	low risk	unable to assess	no
Hilburn 2011	low risk	N/A	unable to assess	Probably yes
Hsiao 2016	some concerns	low risk	low risk	probably yes
Hua 2013	unable to assess	N/A	Some concerns	probably yes
Jackson-Maldonado 2013	unable to assess	low risk	Low risk	Yes
Jianduan 2009	unable to assess	N/A	unable to assess	probably yes
Karabekiroglu 2009	unable to assess	N/A	low risk	no
Kariuki 2016	Some concerns	Low risk	Some concerns	Probably yes
Khan 2010	unable to assess	Some concerns	low risk	Probably no
Khan 2013	unable to assess	N/A	unable to assess	probably yes
Knauer 2019	unable to assess	low risk	some concerns	probably yes
Koura 2013	low risk	unable to assess	unable to assess	probably no
Kucuker 2011	unable to assess	N/A	some concerns	Probably yes
Kumar 2016	unable to assess	some concerns	low risk	probably no
Kvestad 2013	Low risk	N/A	some concerns	Probably yes
Kvestad 2023	low risk	low risk	some concerns	probably no
Lacunza 2009	unable to assess	N/A	unable to assess	probably no
Lara Díaz 2011	unable to assess	low risk	Low risk	Probably yes
Leppänen 2022	unable to assess	low risk	high risk	probably no
Li 2016	unable to assess	Low risk	Some concerns	No
Li 2020	unable to assess	unable to assess	low risk	probably yes
Liu 2011	unable to assess	N/A	Some concerns	probably yes
Lohaus 2014	unable to assess	Low risk	Some concerns	No
Lopez Boo 2020	Some concerns	low risk	some concerns	probably yes
Madaschi 2016	unable to assess	N/A	Some concerns	probably yes
Manandhar 2016	low risk	N/A	unable to assess	probably no
McCarthy 2012	low risk	N/A	unable to assess	probably yes
McCoy 2017	Low risk	low risk	low risk	yes
McCoy 2018	unable to assess	low risk	low risk	yes
McCray 2023	unable to assess	N/A	Low risk	Probably yes
McHenry 2021	low risk	low risk	some concerns	Probably no
Mehrani 2018	unable to assess	N/A	Low risk	Probably no
Meyer 2011	unable to assess	N/A	unable to assess	Probably no

First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Milosavljevic 2019	low risk	low risk	some concerns	no
Mukherjee 2020	unable to assess	N/A	low risk	No
Mukherjee 2022	unable to assess	N/A	low risk	probably yes
Nair 2009	unable to assess	N/A	unable to assess	probably yes
Nair 2012	unable to assess	N/A	unable to assess	probably yes
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Ngoun 2012	unable to assess	N/A	unable to assess	yes
Obradovic 2019	unable to assess	Low risk	low risk	probably yes
Okuda 2020	unable to assess	N/A	unable to assess	probably yes
Ortiz-Leon 2018	unable to assess	N/A	some concerns	probably no
Oryadi-Zanjani 2020	some concerns	N/A	low risk	probably yes
Parveen 2014	unable to assess	N/A	N/A	no
Pathak 2023	unable to assess	N/A	low risk	probably no
Pendergast 2018	low risk	N/A	Some concerns	yes
Phạm 2017	unable to assess	low risk	low risk	probably no
Pisani 2018	unable to assess	low risk	unable to assess	yes
Pisani 2022	unable to assess	low risk	some concerns	probably yes
Pitchik 2023	low risk	low risk	low risk	probably yes
Prado 2018	Low risk	Low risk	low risk	No
Prathanee 2008	low risk	N/A	low risk	probably yes
Puglisi 2020	Some concerns	N/A	low risk	Probably yes
Ranjitkar 2018	low risk	low risk	low risk	probably yes
Rasheed 2017	some concerns	N/A	unable to assess	probably no
Rasheed 2018	low risk	Low risk	unable to assess	probably yes
Rasheed 2023	unable to assess	Low risk	unable to assess	no information
Razak 2010	unable to assess	N/A	low risk	probably no
Rodriguez 2020	low risk	unable to assess	unable to assess	probably yes
Romero 2013	unable to assess	N/A	unable to assess	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes

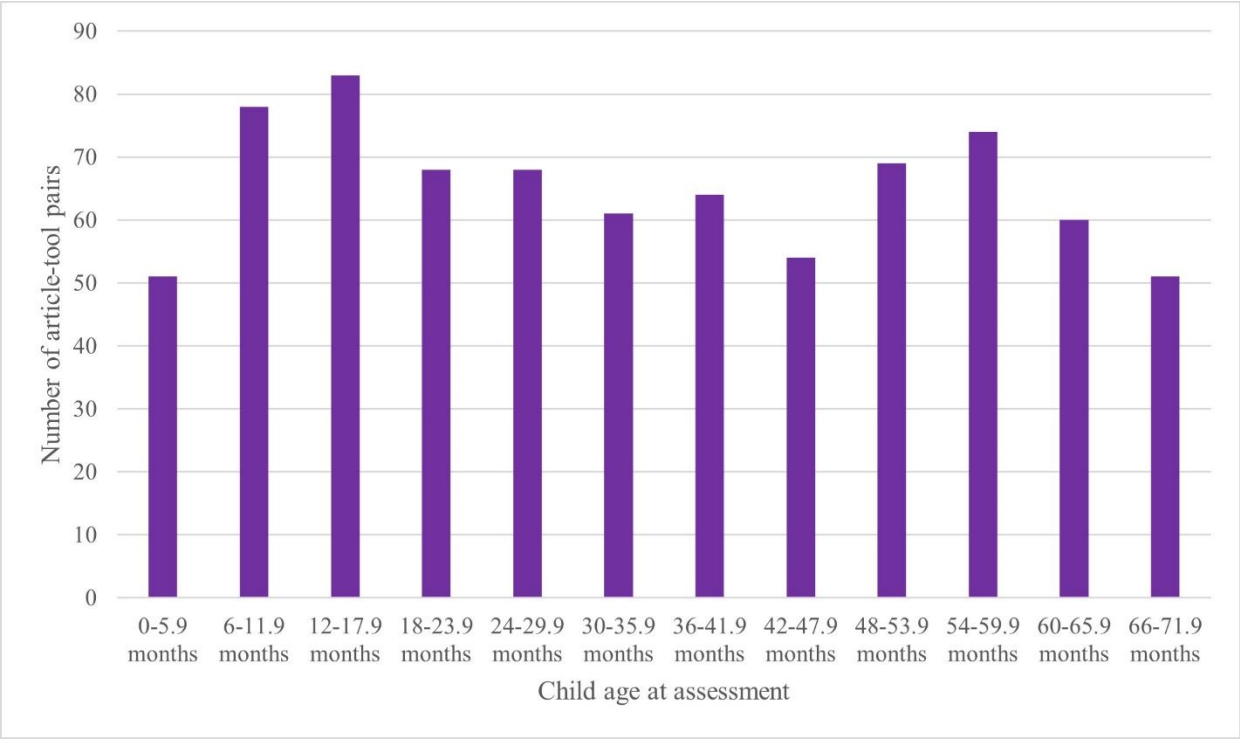
First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	Some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	Some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	low risk	Low risk	some concerns	Probably yes
Saccani 2016	low risk	N/A	unable to assess	yes
Sallum 2017	unable to assess	low risk	some concerns	probably yes
Samarakkody 2010	unable to assess	N/A	unable to assess	yes
Sánchez-Vincitore 2019	low risk	Low risk	low risk	probably no
Scherzer 2009	unable to assess	N/A	unable to assess	probably yes
Shah 2013	unable to assess	n/a	unable to assess	no
Shaoli 2019	unable to assess	n/a	low risk	probably yes
Shrestha 2019	some concerns	low risk	some concerns	probably no
Siegle 2018	unable to assess	N/A	some concerns	probably no
Sincovich 2019	unable to assess	N/A	some concerns	probably no
Snelling 2019	unable to assess	N/A	some concerns	probably yes
Soleimani 2007	low risk	N/A	Some concerns	no information
Soleymani 2016	unable to assess	N/A	Low risk	No
Soysal 2014	unable to assess	some concerns	some concerns	no
Thompson 2015	low risk	unable to assess	unable to assess	probably no
Toor 2022	unable to assess	Low risk	some concerns	No
Uyanik 2014	unable to assess	N/A	Some concerns	Probably no
Valentini 2012	unable to assess	Low risk	unable to assess	Probably no
Vameghi 2013	unable to assess	N/A	high risk	probably no
van der Linde 2015	unable to assess	low risk	low risk	probably no
van Heerden 2017	low risk	N/A	unable to assess	no
Venancio 2020	unable to assess	N/A	Some concerns	Probably yes
Venter 2015	unable to assess	N/A	high risk	probably no
Waechter 2022	low risk	N/A	unable to assess	probably yes
Waldman 2021	unable to assess	Some concerns	Low risk	Probably yes
Washington 2017	unable to assess	Low risk	Some concerns	No
Weber 2018	unable to assess	low risk	some concerns	probably yes
Wijedasa 2012	low risk	N/A	low risk	yes
Willoughby 2019 (Child Neuropsych)	low risk	low risk	low risk	probably yes

First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Willoughby 2019 (Dev Science)	Low risk	Low risk	Some concerns	Probably yes
Wolf 2017	Some concerns	low risk	unable to assess	no
Wu 2019 (JPA)	unable to assess	n/a	unable to assess	Probably yes
Wu 2019 (PS)	unable to assess	Low risk	low risk	probably yes
Xie 2017	low risk	N/A	some concerns	no
Xie 2019 (Early Educ & Dev)	unable to assess	N/A	Some concerns	Probably no
Xie 2019 (Journal of Child & Dev Studies)	unable to assess	N/A	some concerns	probably yes
Yuan 2022	low risk	low risk	low risk	probably yes
Yue 2019	low risk	unable to assess	low risk	yes
Yunilda 2022	unable to assess	unable to assess	low risk	probably no
Zakaria 2012	unable to assess	N/A	unable to assess	no
Zhang 2018	unable to assess	N/A	some concerns	Probably no
Zhang 2021	unable to assess	n/a	unable to assess	Probably yes
Zirakashvili 2018	unable to assess	Some concerns	Some concerns	Probably yes

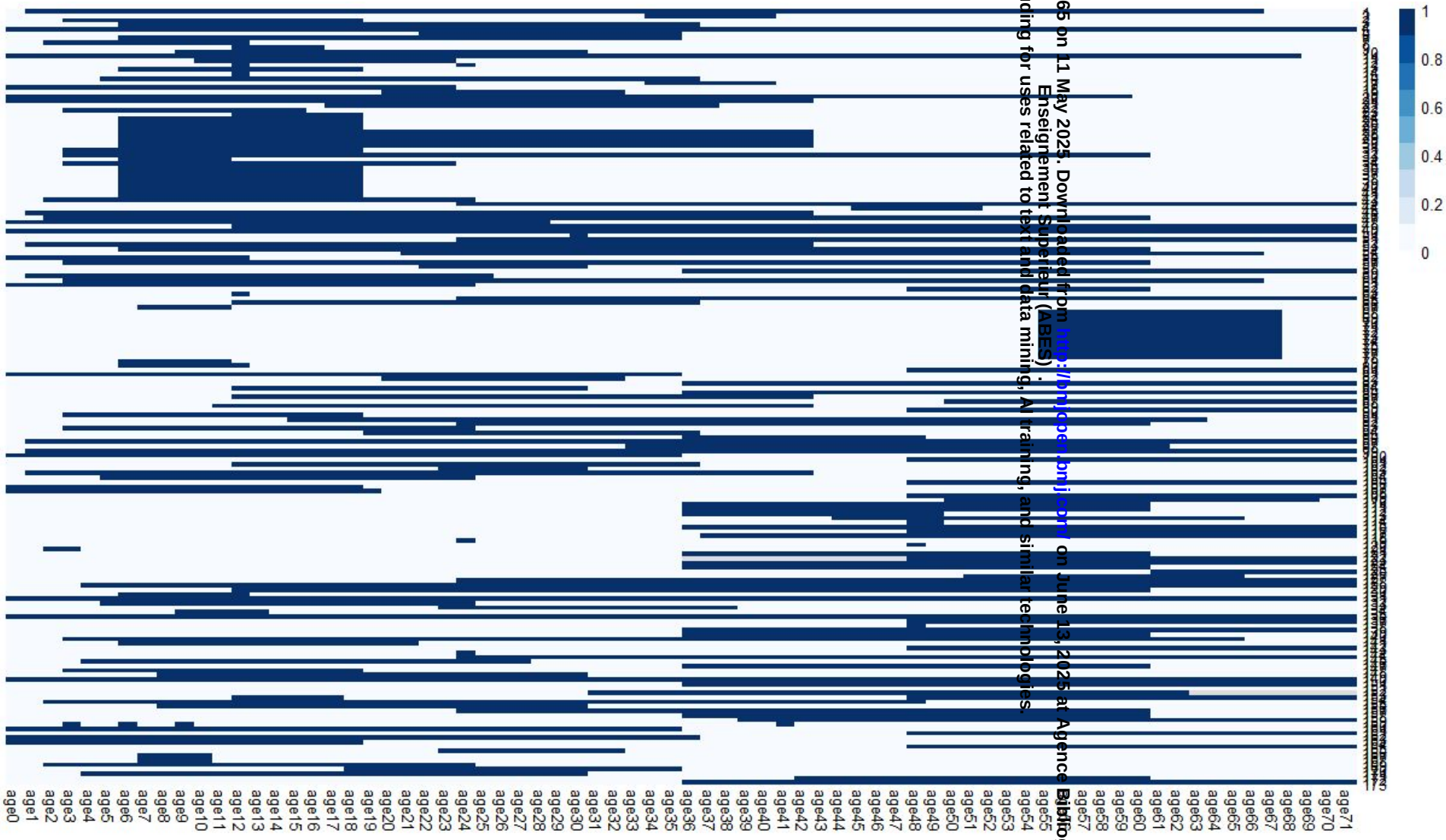
Supplemental Figure 1 Number of articles providing evidence on psychometrics by type of article



Supplemental Figure 2 Number of articles and tools providing psychometric evidence by child age group



Supplemental Figure 3 Number of articles providing evidence on psychometrics by child age at assessment



Note: Y-axis lists each article-tool combination included in the systematic review ordered by tool in alphabetical order.

References of all included articles

Abdoola, S., Swanepoel, D. W., Van Der Linde, J., & Glascoe, F. P. (2021). Detecting developmental delays in infants from a low-income South African community: comparing the BSID-III and PEDS tools. *Early Child Development and Care*, *191*(4), 545–554. https://doi.org/10.1080/03004430.2019.1628027

Abessa, T. G., Worku, B. N., Kibebew, M. W., Valy, J., Lemmens, J., Thijs, H., Yimer, W. K., Kolsteren, P., & Granitzer, M. (2016). Adaptation and standardization of a Western tool for assessing child development in non-Western low-income context. *BMC Public Health*, *16*(1), 1–13. https://doi.org/10.1186/s12889-016-3288-2

Abubakar, A., Holding, P., Van De Vijver, F., Bomu, G., & Van Baar, A. (2010). Developmental monitoring using caregiver reports in a resource-limited setting: The case of Kilifi, Kenya. *Acta Paediatrica, International Journal of Paediatrics*, *99*(2), 291–297. https://doi.org/10.1111/j.1651-2227.2009.01561.x

Abubakar, A., Holding, P., van Baar, A., Newton, C. R. J. C., & van de Vijver, F. J. R. (2008). Monitoring psychomotor development in a resource-limited setting: An evaluation of the Kilifi Developmental Inventory. *Annals of Tropical Paediatrics*, *28*(3), 217–226. https://doi.org/10.1179/146532808X335679

Albuquerque, P. L. de, Guerra, M. Q. de F., Lima, M. de C., & Eickmann, S. H. (2018). Concurrent validity of the Alberta Infant Motor Scale to detect delayed gross motor development in preterm infants: A comparative study with the Bayley III. *Developmental Neurorehabilitation*, *21*(6), 408–414. https://doi.org/10.1080/17518423.2017.1323974

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative Development Inventory (CDI) adaptation for two Kenyan languages. *Journal of Child Language*, *42*(4), 763–785. https://doi.org/10.1017/S0305000914000403

Alderman, H., Friedman, J., Ganga, P., Kak, M., & Rubio-Codina, M. (2021). Assessing the performance of the Caregiver Reported Early Development Instruments (CREDI) in rural India. *Annals of the New York Academy of Sciences*, *1492*(1), 58–72. https://doi.org/10.1111/nyas.14543

Altafim, E. R. P., McCoy, D. C., Brentani, A., Escobar, A. M. de U., Grisi, S. J. F. E., & Fink, G. (2020). Measuring early childhood development in Brazil: validation of the Caregiver Reported Early Development Instruments (CREDI). *Jornal de Pediatria*, *96*(1), 66–75. https://doi.org/10.1016/j.jped.2018.07.008

Anunciação, L., Squires, J., Clifford, J., & Landeira-Fernandez, J. (2019). Confirmatory analysis and normative tables for the Brazilian Ages and Stages Questionnaires: Social–Emotional. *Child: Care, Health and Development*, *45*(3), 387–393. https://doi.org/10.1111/cch.12649

Azari, N., Soleimani, F., Vameghi, R., Sajedi, F., Shahshahani, S., Karimi, H., Kraskian, A., Shahrokhi, A., Teymouri, R., & Gharib, M. (2017). A psychometric study of the Bayley scales of infant and toddler development in Persian language children. *Iranian Journal of Child Neurology*, *11*(1), 50–56.

Bandeira, P. F. R., Duncan, M., Pessoa, M. L., Soares, Í., Da Silva, L., Mota, J., & Martins, C. (2020). Tgmd-2 short version: Evidence of validity and associations with sex, age, and bmi in preschool children. *Journal of Motor Learning and Development*, *8*(3), 528–543. https://doi.org/10.1123/JMLD.2019-0040

Barthel, D., Kriston, L., Guo, N., Yao, E. D., Armel, K. E. J., Fordjour, D., Mohammed, Y., Kotchi, C. E. B., Eberhardt, K. A., Hinz, R., Ehrhardt, S., & Bindt, C. (2022). Psychometric properties of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) in two-year-old Ivorian and Ghanaian children. *Applied Developmental Science*, *26*(3), 567–577. https://doi.org/10.1080/10888691.2021.1931216

Barthel, D., Kriston, L., Yao, E. D., Fordjour, D., Armel, K. E. J., Eberhardt, K. A., Hinz, R., Ehrhardt, S., & Bindt, C. (2022). Psychomotor development in two-year-old Ivorian and Ghanaian children – Psychometric properties of the Kilifi Developmental Inventory. *British Journal of Developmental Psychology*, *40*(4), 471–486. https://doi.org/10.1111/bjdp.12428

Bhavnani, S., Mukherjee, D., Dasgupta, J., Verma, D., Parameshwaran, D., Divan, G., Sharma, K. K., Thiagarajan, T., & Patel, V. (2019). Development, feasibility and acceptability of a gamified cognitive DEvelopmental assessment on an E-Platform (DEEP) in rural Indian pre-schoolers–a pilot study. *Global Health Action*, *12*(1). https://doi.org/10.1080/16549716.2018.1548005

Bian, X., Xie, H., Squires, J., & Chen, C. Y. (2017). Adapting a Parent-Completed, Socioemotional Questionnaire in China: the Ages & Stages Questionnaires: Social-Emotional. *Infant Mental Health Journal*, *38*(2), 258–266. https://doi.org/10.1002/imhj.21626

Bian, X., Yao, G., Squires, J., Hoselton, R., Chen, C. I., Murphy, K., Wei, M., & Fang, B. (2012). Translation and use of parent-completed developmental screening test in Shanghai. *Journal of Early Childhood Research*, *10*(2), 162–175. https://doi.org/10.1177/1476718X11430071

Biasini, F. J., De Jong, D., Ryan, S., Thorsten, V., Bann, C., Bellad, R., Mahantshetti, N. S., Dhaded, S. M., Pasha, O., Chomba, E., Goudar, S. S., Carlo, W. A., & McClure, E. (2015). Development of a 12 month screener based on items from the Bayley II Scales of Infant Development for use in Low Middle Income countries. *Early Human Development*, *91*(4), 253–258. https://doi.org/10.1016/j.earlhumdev.2015.02.001

Boivin, M. J., Weiss, J., Chhaya, R., Seffren, V., Awadu, J., Sikorskii, A., & Giordani, B. (2017). The feasibility of automated eye tracking with the early childhood vigilance test of attention

in younger HIV-exposed Ugandan children. *Neuropsychology*, *31*(5), 525–534. https://doi.org/10.1037/neu0000382

Boivin, M. J., Zoumenou, R., Sikorskii, A., Fievet, N., Alao, J., Davidson, L., Cot, M., Massougbodji, A., & Bodeau-Livinec, F. (2021). Neurodevelopmental assessment at one year of age predicts neuropsychological performance at six years in a cohort of West African Children. *Child Neuropsychology*, *27*(4), 548–571. https://doi.org/10.1080/09297049.2021.1876012

Bornman, J., Sevcik, R. A., Romski, M. A., & Pae, H. K. (2010). Successfully translating language and culture when adapting assessment measures. *Journal of Policy and Practice in Intellectual Disabilities*, *7*(2), 111–118. https://doi.org/10.1111/j.1741-1130.2010.00254.x

Brinkman, S. A., Kinnell, A., Maika, A., Hasan, A., Jung, H., & Pradhan, M. (2017). Validity and Reliability of the Early Development Instrument in Indonesia. *Child Indicators Research*, *10*(2), 331–352. https://doi.org/10.1007/s12187-016-9372-4

Caldera, A. ., Wickremasinghe, A. R., Muttiah, N., Godamunne, P. K. S., Jayasena, B. ., Chathurika, L. K. E., Perera, K. M. N., Mendis, M., Tilakarathne, D., Peiris, M. K. R. ., Wijesinghe, T., Senarathna, N. ., Saubhagya, W. D. L., Chandraratne, M., & Sumanasena, S. . (2023). REACh for the preschoolers; a developmental assessment tool for 2–5 year old children in Sri Lanka. *BMC Pediatrics*, *23*(1), 80. https://doi.org/10.1186/s12887-023-03895-5

Chen, C. Y., Xie, H., Filgueiras, A., Squires, J., Anunciação, L., & Landeira-Fernandez, J. (2017). Examining the Psychometric Properties of the Brazilian Ages & Stages Questionnaires-Social-Emotional: Use in Public Child Daycare Centers in Brazil. *Journal of Child and Family Studies*, *26*(9), 2412–2425. https://doi.org/10.1007/s10826-017-0770-0

Chunsuwan, I., Hansakunachai, T., & Pornsamrit, S. (2016). Parent Evaluation of Developmental Status (PEDS) in screening: The Thai experience. *Pediatrics International*, *58*(12), 1277–1283. https://doi.org/10.1111/ped.13055

Colbert, A. M., Connery, A. K., Lamb, M. M., Bauer, D., Olson, D., Paniagua-Avila, A., Martínez, M. A., Arroyave, P., Hernández, S., Mirella Calvimontes, D., Bolaños, G. A., El Sahly, H. M., Muñoz, F. M., & Asturias, E. J. (2021). Caregiver rating of early childhood development: Reliability and validity of the ASQ-3 in rural Guatemala. *Early Human Development*, *161*(November 2020). https://doi.org/10.1016/j.earlhumdev.2021.105453

Colbert, A. M., Lamb, M. M., Asturias, E. J., Muñoz, F. M., Bauer, D., Arroyave, P., Hernández, S., Martínez, M. A., Paniagua-Avila, A., Olson, D., Calvimontes, D. M., Bolaños, G. A., El Sahly, H. M., & Connery, A. K. (2020). Reliability and Validity of an Adapted and Translated Version of the Mullen Scales of Early Learning (AT-MSEL) in Rural Guatemala. *Child: Care, Health and Development*, *46*(3), 327–335. https://doi.org/10.1111/cch.12748

Dagvadorj, A., Takehara, K., Bavuusuren, B., Morisaki, N., Gochoo, S., & Mori, R. (2015). The quick and easy Mongolian Rapid Baby Scale shows good concurrent validity and sensitivity. *Acta Paediatrica, International Journal of Paediatrics*, *104*(3), e94–e99. https://doi.org/10.1111/apa.12884

Danişman, Ş., Dereli Iman, E., Demircan, Z. A., & Yaya, D. (2016). Una aproximación al enfoque Reggio Emilia: Claves para entender por qué motiva a los estudiantes. *Electronic Journal of Research in Educational Psychology*, *14*(3), 534–556.

de Albuquerque, K. A., da Cunha, A. C. B., & Berkovits, M. D. (2022). Cross-cultural adaptation of the Battelle Development Inventory, 2nd Edition for Brazil. *Estudos de Psicologia*, *39*, 1–13. https://doi.org/10.1590/1982-0275202239e190170

Dias, T. P., Freitas, L. C., Del Prette, Z. A. P., & Del Prette, A. (2011). Validação da escala de comportamentos sociais de pré-escolares para o Brasil. *Psicologia Em Estudo*, *16*(3), 447–457. https://doi.org/10.1590/S1413-73722011000300012

Dursun, O. B., Öğütlü, H., & Esin, İ. S. (2020). Psychometric properties of Turkish version of the strength and difficulties questionnaire for age 2-4. *Noropsikiyatri Arsivi*, *57*(1), 44–49. https://doi.org/10.29399/npa.23674

Familiar-Lopez, I., Sikorskii, A., Chhaya, R., Holmes, A., Arima, E. G., Caesar, O. J., Nakasujja, N., & Boivin, M. J. (2023). Predictive validation of Ugandan infant eye-tracking test for memory of human faces. *Child Neuropsychology*, *29*(3), 486–502. https://doi.org/10.1080/09297049.2022.2099537

Fernandes, M., Bassani, D., Albernaz, E., Bertoldi, A. D., Silveira, M. F., Matijsevich, A., Anselmi, L., Cruz, S., Halal, C. S., Tovo-Rodrigues, L., Cruz, G. I. N., Metgud, D., & Santos, I. S. (2022). Construction and validation of the Oxford Neurodevelopment Assessment (OX-NDA) in 1-year-old Brazilian children. *BMC Pediatrics*, *22*(1), 1–12. https://doi.org/10.1186/s12887-022-03794-1

Ferreira-Vasques, A. T., Santos, C. F., & Lamônica, D. A. C. (2019). Transcultural adaptation process of the Griffiths-III Mental Development Scale. *Child: Care, Health and Development*, *45*(3), 403–408. https://doi.org/10.1111/cch.12664

Filgueiras, A., Pires, P., Maissonette, S., & Landeira-Fernandez, J. (2013). Psychometric properties of the Brazilian-adapted version of the Ages and Stages Questionnaire in public child daycare centers. *Early Human Development*, *89*(8), 561–576. https://doi.org/10.1016/j.earlhumdev.2013.02.005

Finch, J. E., Yousafzai, A. K., Rasheed, M., & Obradović, J. (2018). Measuring and understanding social-emotional behaviors in preschoolers from rural Pakistan. *PLoS ONE*, *13*(11), 1–22. https://doi.org/10.1371/journal.pone.0207807

Flórez Romero, R., Castro Martínez, J., & Arias Velandia, N. (2013). Propiedades psicométricas de la escala de lenguaje para preescolares (PLS-3) colombianos TT - Psychometric Properties of the Pre-school Language Scale (PLS-3) with Colombian Children TT - Propriedades psicométricas da Escada de linguagem para pré-escola. *Pensamiento Psicológico*, *11*(1), 131–140.

Gladstone, M., Lancaster, G. a., Umar, E., Nyirenda, M., Kayira, E., van den Broek, N. R., & Smyth, R. L. (2010). The Malawi Developmental Assessment Tool (MDAT): The Creation, Validation, and Reliability of a Tool to Assess Child Development in Rural African Settings. *PLoS Medicine*, *7*(5), e1000273. https://doi.org/10.1371/journal.pmed.1000273

Golmohammadi, G., Zarifian, T., Qhanbari, S., Bakhshi, E., & Sakhai, F. (2022). Cross-cultural adaptation and validation of the Persian version of Children's Behavior Questionnaire in Iranian children. *Current Psychology*, *41*(6), 4041–4053. https://doi.org/10.1007/s12144-020-00918-7

Gomez, J. A., Brown, J. L., & Spier, E. (2022). Factor structure, measurement invariance, and construct validity of MELQO measure of child development and early learning in Laos. *International Journal of School and Educational Psychology*, *10*(2), 181–207. https://doi.org/10.1080/21683603.2020.1791770

Guedes, D. Z., Primi, R., & Kopelman, B. I. (2011). BINS validation - Bayley neurodevelopmental screener in Brazilian preterm children under risk conditions. *Infant Behavior and Development*, *34*(1), 126–135. https://doi.org/10.1016/j.infbeh.2010.11.001

Guo, B., Aveyard, P., & Dai, X. (2009). The Chinese Intelligence Scale for Young Children. *Educational and Psychological Measurement*, *69*(3), 459–474. https://doi.org/10.1177/0013164409332209

Hahn, K. R., Schaefer, B. A., Merino, C., & Worrell, F. C. (2009). The factor structure of preschool learning behaviors scale scores in peruvian children. *Canadian Journal of School Psychology*, *24*(4), 318–331. https://doi.org/10.1177/0829573509349015

Halpin, P. F., Wolf, S., Yoshikawa, H., Rojas, N., Kabay, S., Pisani, L., & Dowd, A. J. (2019). Measuring early learning and development across cultures: Invariance of the IDELA across five countries. *Developmental Psychology*, *55*(1), 23–37. https://doi.org/10.1037/dev0000626

Hamadani, J. D., Baker-Henningham, H., Tofail, F., Mehrin, F., Huda, S. N., & Grantham-McGregor, S. M. (2010). Validity and Reliability of Mothers' Reports of Language Development in 1-Year-Old Children in a Large-Scale Survey in Bangladesh. *Food and Nutrition Bulletin*, *31*(2_suppl2), S198–S206. https://doi.org/10.1177/15648265100312S212

Hamadani, J. D., Tofail, F., Cole, T., & Grantham-McGregor, S. (2013). The relation between age of attainment of motor milestones and future cognitive and motor development in

Bangladeshi children. *Maternal and Child Nutrition*, *9*(SUPPL. 1), 89–104. https://doi.org/10.1111/mcn.12020

Hanlon, C., Medhin, G., Worku, B., Tomlinson, M., Alem, A., Dewey, M., & Prince, M. (2016). Adapting the Bayley Scales of infant and toddler development in Ethiopia: evaluation of reliability and validity. *Child: Care, Health and Development*, *42*(5), 699–708. https://doi.org/10.1111/cch.12371

Hao, M., Shu, H., Xing, A., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods*, *40*(3), 728–733. https://doi.org/10.3758/BRM.40.3.728

Hilburn, N., Potterton, J., Stewart, A., & BeckerC, P. (2011). The development of a screening tool to evaluate gross motor function in HIV-infected infants. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV*, *23*(12), 1619–1625. https://doi.org/10.1080/09540121.2011.579942

Hsiao, C., Richter, L., Makusha, T., Matafwali, B., van Heerden, A., & Mabaso, M. (2017). Use of the ages and stages questionnaire adapted for South Africa and Zambia. *Child: Care, Health and Development*, *43*(1), 59–66. https://doi.org/10.1111/cch.12413

Hua, J., Gu, G., Meng, W., & Wu, Z. (2013). Age band 1 of the Movement Assessment Battery for Children-Second Edition: Exploring its usefulness in mainland China. *Research in Developmental Disabilities*, *34*(2), 801–808. https://doi.org/10.1016/j.ridd.2012.10.012

Jackson-Maldonado, D., Marchman, V. A., & Fernald, L. C. H. (2013). Short-form versions of the Spanish MacArthur-Bates Communicative Development Inventories. *Applied Psycholinguistics*, *34*(4), 837–868. https://doi.org/10.1017/S0142716412000045

Jianduan, Z., Huishan, W., Shuhua, S., Xiaonan, H., Guoyan, L., Guangli, L., & Junxin, S. (2009). Reliability and validity of standardized Chinese version of Urban Infant-Toddler Social and Emotional Assessment. *Early Human Development*, *85*(5), 331–336. https://doi.org/10.1016/j.earlhumdev.2008.12.012

Karabekiroglu, K., Rodopman-Arman, A., Ay, P., Ozkesen, M., Akbas, S., Tasdemir, G. N., Boke, O., & Peksen, Y. (2009). The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). *Infant Behavior and Development*, *32*(3), 291–297. https://doi.org/10.1016/j.infbeh.2009.03.003

Kariuki, S. M., Abubakar, A., Murray, E., Stein, A., & Newton, C. R. J. C. (2016). Evaluation of psychometric properties and factorial structure of the pre-school child behaviour checklist at the Kenyan Coast. *Child and Adolescent Psychiatry and Mental Health*, *10*(1), 1–10. https://doi.org/10.1186/s13034-015-0089-9

Khan, N. Z., Muslima, H., Shilpi, A. B., Begum, D., Parveen, M., Akter, N., Ferdous, S., Nahar, K., McConachie, H., & Darmstadt, G. L. (2013). Validation of rapid neurodevelopmental

assessment for 2- to 5-year-old children in bangladesh. *Pediatrics*, *131*(2). https://doi.org/10.1542/peds.2011-2421

Khan, N. Z., Muslima, H., Begum, D., Shilpi, A. B., Akhter, S., Bilkis, K., Begum, N., Parveen, M., Ferdous, S., Morshed, R., Batra, M., & Darmstadt, G. L. (2010). Validation of rapid neurodevelopmental assessment instrument for under-two-year-old children in Bangladesh. *Pediatrics*, *125*(4). https://doi.org/10.1542/peds.2008-3471

Knauer, H. A., Kariger, P., Jakiela, P., Ozier, O., & Fernald, L. C. H. (2019). Multilingual assessment of early child development: Analyses from repeated observations of children in Kenya. *Developmental Science*, *22*(5), 1–23. https://doi.org/10.1111/desc.12875

Koura, K. G., Boivin, M. J., Davidson, L. L., Ouédraogo, S., Zoumenou, R., Alao, M. J., Garcia, A., Massougbodji, A., Cot, M., & Bodeau-Livinec, F. (2013). Usefulness of child development assessments for low-resource settings in Francophone Africa. *Journal of Developmental and Behavioral Pediatrics*, *34*(7), 486–493. https://doi.org/10.1097/DBP.0b013e31829d211c

Kucuker, S., Kapci, E. G., & Uslu, R. I. (2011). Evaluation of the turkish version of the "ages and stages questionnaires: Social-emotional" in identifying children with social-emotional problems. *Infants and Young Children*, *24*(2), 207–220. https://doi.org/10.1097/IYC.0b013e31820eae26

Kumar, R., Shankar, K., Kush, V., Kumar, C., Bhave, A., & Agarwal, V. (2016). Adaptation: Vineland Adaptive Behavior Scale for 3-9 year-old Indian children. *International Journal on Disability and Human Development*, *15*(1), 49–55. https://doi.org/10.1515/ijdhd-2014-0026

Kvestad, I., Silpakar, J. S., Hysing, M., Ranjitkar, S., Strand, T. A., Schwinger, C., Shrestha, M., Chandyo, R. K., & Ulak, M. (2023). The reliability and predictive ability of the Test of Infant Motor Performance (TIMP) in a community-based study in Bhaktapur, Nepal. *Infant Behavior and Development*, *70*(January), 101809. https://doi.org/10.1016/j.infbeh.2023.101809

Kvestad, I., Taneja, S., Kumar, T., Bhandari, N., Strand, T. A., & Hysing, M. (2013). The assessment of developmental status using the Ages and Stages questionnaire-3 in nutritional research in north Indian young children. *Nutrition Journal*, *12*(1), 1–11. https://doi.org/10.1186/1475-2891-12-50

Lacunza, A. B., Castro Solano, A., & Contini, N. (2009). Habilidades sociales preescolares: una escala para niños de contextos de pobreza. *Revista de Psicología*, *27*(1), 3–28. https://doi.org/10.18800/psico.200901.001

Lara Díaz, M. F. L., Fonseca, Á. M. G., Bohórquez, D. M. G., Guechá, C. M., & Sellabona, E. S. (2011). Normativización del inventario del desarrollo comunicativo MacArthur-bates al Español, Colombia. *Revista Latinoamericana de Psicologia*, *43*(2), 241–254.

Leppänen, J. M., Butcher, J. W., Godbout, C., Stephenson, K., Taylor Hendrixson, D., Griswold, S., Rogers, B. L., Webb, P., Koroma, A. S., & Manary, M. J. (2022). Assessing infant cognition in field settings using eye-tracking: A pilot cohort trial in Sierra Leone. *BMJ Open*, *12*(2). https://doi.org/10.1136/bmjopen-2021-049783

Li, Y., Zhu, J. J., Coplan, R. J., Gao, Z. Q., Xu, P., Li, L., & Zhang, H. (2016). Assessment and Implications of Social Withdrawal Subtypes in Young Chinese Children: The Chinese Version of the Child Social Preference Scale. *Journal of Genetic Psychology*, *177*(3), 97–101. https://doi.org/10.1080/00221325.2016.1174100

Li, Y., Tang, L., Bai, Y., Zhao, S., & Shi, Y. (2020). Reliability and validity of the Caregiver Reported Early Development Instruments (CREDI) in impoverished regions of China. *BMC Pediatrics*, *20*(1). https://doi.org/10.1186/s12887-020-02367-4

Liu, J., Cheng, H., & Leung, P. W. L. (2011). The application of the preschool child behavior checklist and the Caregiver-Teacher Report Form to Mainland Chinese children: Syndrome structure, gender differences, country effects, and inter-informant agreement. *Journal of Abnormal Child Psychology*, *39*(2), 251–264. https://doi.org/10.1007/s10802-010-9452-8

Lohaus, A., Lamm, B., Keller, H., Teubert, M., Fassbender, I., Glüer, M., Borchert, S., Vöhringer, I., Teiser, J., Freitag, C., Suhrke, J., Knopf, M., & Schwarzer, G. (2014). Gross and Fine Motor Differences Between Cameroonian and German Children Aged 3 to 40 Months: Results of a Cross-Cultural Longitudinal Study. *Journal of Cross-Cultural Psychology*, *45*(8), 1328–1341. https://doi.org/10.1177/0022022114537703

Lopez Boo, F., Cubides Mateus, M., & Llonch Sabatés, A. (2020). Initial psychometric properties of the Denver II in a sample from Northeast Brazil. *Infant Behavior and Development*, *58*(August 2019), 101391. https://doi.org/10.1016/j.infbeh.2019.101391

Madaschi, V., Mecca, T. P., Macedo, E. C., & Paula, C. S. (2016). Bayley-III scales of infant and toddler development: Transcultural adaptation and psychometric properties. *Paideia*, *26*(64), 189–197. https://doi.org/10.1590/1982-43272664201606

Manandhar, S. R., Dulal, S., Manandhar, D. S., Saville, N., & Prost, A. (2016). Acceptability and Reliability of the Bayley Scales of Infant Development III Cognitive and Motor Scales among Children in Makwanpur. *Journal of Nepal Health Research Council*, *14*(32), 47–50.

McCarthy, A. M., Wehby, G. L., Barron, S., Aylward, G. P., Castilla, E. E., Javois, L. C., Goco, N., & Murray, J. C. (2012). Application of neurodevelopmental screening to a sample of South American infants: The Bayley Infant Neurodevelopmental Screener (BINS). *Infant Behavior and Development*, *35*(2), 280–294. https://doi.org/10.1016/j.infbeh.2011.12.003

McCoy, D. C., Sudfeld, C. R., Bellinger, D. C., Muhihi, A., Ashery, G., Weary, T. E., Fawzi, W., & Fink, G. (2017). Development and validation of an early childhood development scale for use in low-resourced settings. *Population Health Metrics*, *15*(1), 1–18. https://doi.org/10.1186/s12963-017-0122-8

McCoy, D. C., Waldman, M., & Fink, G. (2018). Measuring early childhood development at a global scale: Evidence from the Caregiver-Reported Early Development Instruments. *Early Childhood Research Quarterly*, *45*, 58–68. https://doi.org/10.1016/j.ecresq.2018.05.002

McCray, G., McCoy, D., Kariger, P., Janus, M., Black, M. M., Chang, S. M., Tofail, F., Eekhout, I., Waldman, M., Van Buuren, S., Khanam, R., Sazawal, S., Nizar, A., Schönbeck, Y., Zongo, A., Brentani, A., Zhang, Y., Dua, T., Cavallera, V., … Gladstone, M. (2023). The creation of the Global Scales for Early Development (GSED) for children aged 0-3 years: Combining subject matter expert judgements with big data. *BMJ Global Health*, *8*(1), 1–11. https://doi.org/10.1136/bmjgh-2022-009827

McHenry, M. S., Oyungu, E., Yang, Z., Hines, A. C., Ombitsa, A. R., Vreeman, R. C., Abubakar, A., & Monahan, P. O. (2021). Cultural adaptation of the Bayley Scales of Infant and Toddler Development, 3rd Edition for use in Kenyan children aged 18–36 months: A psychometric study. *Research in Developmental Disabilities*, *110*(December 2020), 103837. https://doi.org/10.1016/j.ridd.2020.103837

Mehrani, M. B. (2018). An elicited imitation test for measuring preschoolers' language development. *Psychological Reports*, *121*(4), 767–786. https://doi.org/10.1177/0033294117732976

Meyer, E. L., Schaefer, B. A., Soto, C. M., Simmons, C. S., Anguiano, R., Brett, J., Holman, A., Martin, J. F., Hata, H. K., Roberts, K. J., Mello, Z. R., & Worrell, F. C. (2011). Factor structure of child behavior scale scores in peruvian preschoolers. *Psychology in the Schools*, *48*(10), 931–942. https://doi.org/10.1002/pits.20596

Milosavljevic, B., Vellekoop, P., Maris, H., Halliday, D., Drammeh, S., Sanyang, L., Darboe, M. K., Elwell, C., Moore, S. E., & Lloyd-Fox, S. (2019). Adaptation of the Mullen Scales of Early Learning for use among infants aged 5- to 24-months in rural Gambia. *Developmental Science*, *22*(5), 1–17. https://doi.org/10.1111/desc.12808

Mukherjee, D., Bhavnani, S., Swaminathan, A., Verma, D., Parameshwaran, D., Divan, G., Dasgupta, J., Sharma, K., Thiagarajan, T. C., & Patel, V. (2020). Proof of Concept of a Gamified DEvelopmental Assessment on an E-Platform (DEEP) Tool to Measure Cognitive Development in Rural Indian Preschool Children. *Frontiers in Psychology*, *11*(June), 1–12. https://doi.org/10.3389/fpsyg.2020.01202

Mukherjee, S. B., Verma, S., Sharma, S., & Aneja, S. (2022). Diagnostic Accuracy of Parents' Evaluation of Developmental Status (PEDS), PEDS Developmental Milestones, and PEDS Combined in Indian Children Aged Less than 2 Years. *Indian Journal of Pediatrics*, *89*(5), 459–465. https://doi.org/10.1007/s12098-020-03651-y

Nair, M. K. C., Russell, P. S., Rekha, R. S., Lakshmi, M. A., Latha, S., & Rajee, K. (2009). Validation of developmental assessment tool for Anganwadis (DATA). *Indian Pediatrics*, *46 Suppl*(2), s27-36.

Nair, M. K. C., & Russell, P. S. (2013). Development and normative validation of Developmental Assessment Tool for Anganwadis for 3- to 4-year-old children (DATA-II). *Journal of Clinical Epidemiology*, *66*(1), 23–29. https://doi.org/10.1016/j.jclinepi.2012.06.011

Nampijja, M., Apule, B., Lule, S., Akurut, H., Muhangi, L., Elliott, A. M., & Alcock, K. J. (2010). Adaptation of western measures of cognition for assessing 5-year-old semi-urban Ugandan children. *British Journal of Educational Psychology*, *80*(1), 15–30. https://doi.org/10.1348/000709909X460600

Ngoun, C., Stoey, L. S., van't Ende, K., & Kumar, V. (2012). Creating a Cambodia-specific developmental milestone screening tool - A pilot study. *Early Human Development*, *88*(6), 379–385. https://doi.org/10.1016/j.earlhumdev.2011.09.014

Obradović, J., Finch, J. E., Portilla, X. A., Rasheed, M. A., Tirado-Strayer, N., & Yousafzai, A. K. (2019). Early executive functioning in a global context: Developmental continuity and family protective factors. *Developmental Science*, *22*(5). https://doi.org/10.1111/desc.12795

Okuda, P. M. M., Félix, E., Cogo-Moreira, H., Liu, T., Surkan, P. J., Martins, S. S., & Caetano, S. C. (2020). Construct validity of the motor development scale (MDS). *Trends in Psychiatry and Psychotherapy*, *42*(1), 82–85. https://doi.org/10.1590/2237-6089-2018-0114

Ortiz-León, S., Granados-Rojas, A., Cavazos-Olivo, J., Benito-Avendaño, L. De, Obregón-García, V. H., Duran-Avendaño, X. M., Ramírez-Pérez, R. V., Cárdenas-Medina, J. A., Vargas-Terréz, B. E., Luna-Guevara, G., Rodríguez-Machain, A. C., & Jaimes-Medrano, A. L. (2018). Internal and inter-rater reliability of the ASQ-3 in Mexican preschoolers. *Salud Mental*, *41*(2), 65–72. https://doi.org/10.17711/SM.0185-3325.2018.011

Oryadi-Zanjani, M. M. (2020). Development of the Childhood Nonverbal Communication Scale. *Journal of Autism and Developmental Disorders*, *50*(4), 1238–1248. https://doi.org/10.1007/s10803-019-04356-8

Parveen, M., Rahman, S. T., Islam, S., Zaman, S. S., Hamadani, J., & Khan, N. Z. (2014). Adaptation of items of bayley scales of infant development-II (BSID-II) suitable for Bangladeshi infants. *Dhaka University Journal of Biological Sciences*, *23*(2), 187–195. https://doi.org/10.3329/dujbs.v23i2.20099

Pathak, S., & Sovani-Kelkar, P. (2023). Development of a screening tool to identify babies at risk of language delay in India: A preliminary study. *Language Acquisition*, *30*(1), 29–49. https://doi.org/10.1080/10489223.2022.2078212

Pendergast, L. L., Schaefer, B. A., Murray-Kolb, L. E., Svensen, E., Shrestha, R., Rasheed, M. A., Scharf, R. J., Kosek, M. N., Vasquez, A. O., Maphula, A., Costa, H., Rasmussen, Z. A., Yousafzai, A., Tofail, F., Seidman, J. C., Acosta, A. M., de Burga, R. R., Chavez, C. B., Flores, J. T., … Strand, T. (2018). Assessing development across cultures: Invariance of the

Bayley-III scales across seven international MAL-ED sites. *School Psychology Quarterly*, *33*(4), 604–614. https://doi.org/10.1037/spq0000264

Phạm, B., McLeod, S., & Harrison, L. J. (2017). Validation and norming of the Intelligibility in Context Scale in Northern Viet Nam. *Clinical Linguistics and Phonetics*, *31*(7–9), 665–681. https://doi.org/10.1080/02699206.2017.1306110

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). *International Journal of Educational Research*, *91*(July), 1–15. https://doi.org/10.1016/j.ijer.2018.06.007

Pisani, L., Seiden, J., & Wolf, S. (2022). Longitudinal evidence on the predictive validity of the International Development and Early Learning Assessment (IDELA). *Educational Assessment, Evaluation and Accountability*, *34*(2), 173–194. https://doi.org/10.1007/s11092-021-09374-8

Pitchik, H. O., Tofail, F., Akter, F., Shoab, A. K. M., Sultana, J., Huda, T. M. N., Rahman, M., Winch, P. J., Luby, S. P., & Fernald, L. C. H. (2023). Concurrent validity of the Ages and Stages Questionnaire Inventory and the Bayley Scales of Infant and Toddler Development in rural Bangladesh. *BMC Pediatrics*, *23*(1), 1–11. https://doi.org/10.1186/s12887-022-03800-6

Prado, E. L., Phuka, J., Ocansey, E., Maleta, K., Ashorn, P., Ashorn, U., Adu-Afarwuah, S., Oaks, B. M., Lartey, A., & Dewey, K. G. (2018). A method to develop vocabulary checklists in new languages and their validity to assess early language development. *Journal of Health, Population, and Nutrition*, *37*(1), 13. https://doi.org/10.1186/s41043-018-0145-1

Prathanee, B., Pongjanyakul, A., & Chano, J. (2008). Thai speech and language test for children between 1 and 2 years of age. *International Journal of Language and Communication Disorders*, *43*(1), 125–140. https://doi.org/10.1080/13682820601181208

Puglisi, M. L., Blasi, H. F., & Snowling, M. J. (2020). Screening for the identification of oral language difficulties in brazilian preschoolers: A validation study. *Language, Speech, and Hearing Services in Schools*, *51*(3), 852–865. https://doi.org/10.1044/2020_LSHSS-19-00083

Ranjitkar, S., Kvestad, I., Strand, T. A., Ulak, M., Shrestha, M., Chandyo, R. K., Shrestha, L., & Hysing, M. (2018). Acceptability and reliability of the Bayley Scales of infant and Toddler Development-III Among Children in Bhaktapur, Nepal. *Frontiers in Psychology*, *9*(JUL), 1–10. https://doi.org/10.3389/fpsyg.2018.01265

Rasheed, M. A., Tabussum, G., Habib, Z. H., & Yousafzai, A. K. (2017). Adaptation of MacArthur Communicative Development Inventory in rural Pakistan – useful tool for early childhood studies. *Child: Care, Health and Development*, *43*(3), 427–434. https://doi.org/10.1111/cch.12450

Rasheed, M. A., Kvestad, I., Shaheen, F., Memon, U., & Strand, T. A. (2023). The predictive validity of Bayley Scales of Infant and Toddler Development-III at 2 years for later general abilities: Findings from a rural, disadvantaged cohort in Pakistan. *PLOS Global Public Health*, *3*(1), e0001485. https://doi.org/10.1371/journal.pgph.0001485

Rasheed, M. A., Pham, S., Memon, U., Siyal, S., Obradović, J., & Yousafzai, A. K. (2018). Adaptation of the Wechsler Preschool and Primary Scale of Intelligence-III and lessons learned for evaluating intelligence in low-income settings. *International Journal of School and Educational Psychology*, *6*(3), 197–207. https://doi.org/10.1080/21683603.2017.1302851

Razak, R. A., Madison, C. L., Siow, Y. K., & Aziz, M. A. A. (2010). Preliminary Content Validity and Reliability of a Newly Developed Malay Preschool Language Assessment Tool. *Asia Pacific Journal of Speech, Language and Hearing*, *13*(4), 217–234. https://doi.org/10.1179/136132810805334985

Rodriguez, V. J., Zegarac, M., La Barrie, D. L., Parrish, M. S., Matseke, G., Peltzer, K., & Jones, D. L. (2020). Validation of the Bayley Infant Neurodevelopmental Screener among HIV-Exposed Infants in Rural South Africa. *Journal of Acquired Immune Deficiency Syndromes*, *85*(4), 507–516. https://doi.org/10.1097/QAI.0000000000002479

Rubio-Codina, M., Araujo, M. C., Attanasio, O., Muñoz, P., & Grantham-McGregor, S. (2016). Concurrent validity and feasibility of short tests currently used to measure early childhood development in large scale studies. *PLoS ONE*, *11*(8), 1–17. https://doi.org/10.1371/journal.pone.0160962

Rubio-Codina, M., & Grantham-McGregor, S. (2020). Predictive validity in middle childhood of short tests of early childhood development used in large scale studies compared to the Bayley-III, the Family Care Indicators, height-for-age, and stunting: A longitudinal study in Bogota, Colombia. *PLoS ONE*, *15*(4), 1–20. https://doi.org/10.1371/journal.pone.0231317

Saccani, R., Valentini, N. C., & Pereira, K. R. G. (2016). New Brazilian developmental curves and reference values for the Alberta infant motor scale. *Infant Behavior and Development*, *45*, 38–46. https://doi.org/10.1016/j.infbeh.2016.09.002

Sallum, I., da Mata, F. G., Cheib, N. F., Mathias, C. W., Miranda, D. M., & Malloy-Diniz, L. F. (2017). Development of a version of the self-ordered pointing task: a working memory task for Brazilian preschoolers. *Clinical Neuropsychologist*, *31*(2), 459–470. https://doi.org/10.1080/13854046.2016.1275818

Samarakkody, D. C., Fernando, D. N., Perera, H., McClure, R. J., & De Silva, H. (2010). The Child Behaviour Assessment Instrument: Development and validation of a measure to screen for externalising child behavioural problems in community setting. *International Journal of Mental Health Systems*, *4*, 1–8. https://doi.org/10.1186/1752-4458-4-13

Sánchez-Vincitore, L. V., Schaettle, P., & Castro, A. (2019). Validation of the Malawi Developmental Assessment Tool for children in the Dominican Republic: Preliminary results. *PLoS ONE*, *14*(8), 1–11. https://doi.org/10.1371/journal.pone.0221162

Scherzer, A. L. (2009). Experience in Cambodia with the use of a culturally relevant developmental milestone chart for children in low- and middle-income countries. *Journal of Policy and Practice in Intellectual Disabilities*, *6*(4), 287–292. https://doi.org/10.1111/j.1741-1130.2009.00234.x

Shah, M. B., Schaefer, B. A., & Clark, T. P. (2013). Reliability of Bracken School Readiness Assessment, Third Edition Scores With Young Children in Mumbai, India. *International Journal of School and Educational Psychology*, *1*(1), 58–66. https://doi.org/10.1080/21683603.2013.782595

Shaoli, S. S., Islam, S., Haque, S., & Islam, A. (2019). Validating the Bangla version of the Dimensions of Mastery Questionnaire (DMQ-18) for preschoolers. *Asian Journal of Psychiatry*, *44*(May 2019), 143–149. https://doi.org/10.1016/j.ajp.2019.07.044

Shrestha, M., Strand, T. A., Ulak, M., Chandyo, R. K., Ranjitkar, S., Hysing, M., Shrestha, L., & Kvestad, I. (2019). The feasibility of the Ages and Stages Questionnaire for the assessment of child development in a community setting in Nepal. *Child: Care, Health and Development*, *45*(3), 394–402. https://doi.org/10.1111/cch.12654

Siegle, C. B. H., & dos Santos Cardoso de Sá, C. (2018). Concurrent validity between instruments of assessment of motor development in infants exposed to HIV. *Infant Behavior and Development*, *50*(February), 198–206. https://doi.org/10.1016/j.infbeh.2018.01.005

Sincovich, A., Gregory, T., Zanon, C., Santos, D. D., Lynch, J., & Brinkman, S. A. (2019). Measuring early childhood development in multiple contexts: The internal factor structure and reliability of the early Human Capability Index in seven low and middle income countries. *BMC Pediatrics*, *19*(1), 1–14. https://doi.org/10.1186/s12887-019-1852-5

Snelling, M., Dawes, A., Biersteker, L., Girdwood, E., & Tredoux, C. (2019). The development of a South African Early Learning Outcomes Measure: A South African instrument for measuring early learning program outcomes. *Child: Care, Health and Development*, *45*(2), 257–270. https://doi.org/10.1111/cch.12641

Soleimani, F., & Dadkhah, A. (2007). Validity and reliability of infant Neurological International Battery for detection of gross motor developmental delay in Iran. *Child: Care, Health and Development*, *33*(3), 262–265. https://doi.org/10.1111/j.1365-2214.2006.00704.x

Soleymani, Z., Nematzadeh, S., Tehrani, L. G., Rahgozar, M., & Schneider, P. (2016). Language sample analysis: development of a valid language assessment tool and determining the reliability of outcome measures for Farsi-speaking children. *European Journal of Developmental Psychology*, *13*(2), 275–291. https://doi.org/10.1080/17405629.2016.1152175

Solís Cordero K, Quezada Ugalde AM, González Serrano S, Zúñiga Baldi C. Costa Rican test: Evaluation of the children's integral development from 0 months to 6 years of age (EDIN II). Enfermería Actual Costa Rica 2019; 37. DOI:http://dx.doi.org/10.15517/revenf.v0ino.37.34959.

Soysal, A. S., Gucuyener, K., Ergenekon, E., Turan, Ö., Koc, E., Turkylmaz, C., Önal, E., & Atalay, Y. (2014). The prediction of later neurodevelopmental status of preterm infants at ages 7 to 10 years using the bayley infant neurodevelopmental screener. *Journal of Child Neurology*, *29*(10), 1349–1355. https://doi.org/10.1177/0883073813520495

Thompson, L., Peñaloza, R. A., Stormfields, K., Kooistra, R., Valencia-Moscoso, G., Muslima, H., & Khan, N. Z. (2015). Validation and adaptation of rapid neurodevelopmental assessment instrument for infants in Guatemala. *Child: Care, Health and Development*, *41*(6), 1131–1139. https://doi.org/10.1111/cch.12279

Toor, H. K., & Hanif, R. (2022). Executive Function Touch battery: Translation and preliminary measure validation for Pakistani preschoolers. *PLoS ONE*, *17*(9 September), 1–14. https://doi.org/10.1371/journal.pone.0274431

Uyanik, Ö., & Kandir, A. (2014). Adaptation of the Kaufman Survey of Early Academic and Language Skills to Turkish Children Aged 61 to 72 Months. *Kuram ve Uygulamada Egitim Bilimleri*, *14*(2), 682–692. https://doi.org/10.12738/estp.2014.2.1682

Valentini, N. C., & Saccani, R. (2012). Brazilian Validation of the Alberta Infant Motor Scale. *Physical Therapy*, *92*(3), 440–447. https://doi.org/10.2522/ptj.20110036

Vameghi, R., Sajedi, F., Kraskian Mojembari, A., Habiollahi, A., Lornezhad, H. R., & Delavar, B. (2013). Cross-cultural adaptation, validation and standardization of Ages and Stages Questionnaire (ASQ) in Iranian children. *Iranian Journal of Public Health*, *42*(5), 522–528.

van Der Linde, J., Swanepoel, D. W., Glascoe, F. P., Louw, E. M., & Vinck, B. (2015). Developmental screening in South Africa: Comparing the national developmental checklist to a standardized tool. *African Health Sciences*, *15*(1), 188–196. https://doi.org/10.4314/ahs.v15i1.25

van Heerden, A., Hsiao, C., Matafwali, B., Louw, J., & Richter, L. (2017). Support for the feasibility of the ages and stages questionnaire as a developmental screening tool: A cross-sectional study of South African and Zambian children aged 2-60 months. *BMC Pediatrics*, *17*(1), 1–9. https://doi.org/10.1186/s12887-017-0802-3

Venancio, S. I., Bortoli, M. C., Frias, P. G., Giugliani, E. R. J., Alves, C. R. L., & Santos, M. O. (2020). Development and validation of an instrument for monitoring child development indicators. *Jornal de Pediatria (Versão Em Português)*, *96*(6), 778–789. https://doi.org/10.1016/j.jpedp.2019.10.008

Venter, A., Pienaar, A. E., & Coetzee, D. (2015). Suitability of the 'Little DCDQ'' for the identification of DCD in a selected group of 3–5-year-old South African children.' *Early Child Development and Care*, *185*(8), 1359–1371. https://doi.org/10.1080/03004430.2014.1000887

Waechter, R., Evans, R., Hanna, S., Murray, T., Mobley, C., Holmes, S., Isaac, R., Wolfe, R., Andrew, E., Landon, B., & Fernandes, M. (2022). Adaptation of the INTERGROWTH-21st neurodevelopment assessment (INTER-NDA) to the context of the English-speaking Caribbean. *BMC Pediatrics*, *22*(1), 1–10. https://doi.org/10.1186/s12887-021-03039-7

Waldman, M., McCoy, D. C., Seiden, J., Cuartas, J., Fink, G., Altafim, E., Brentani, A., Castellanos, A., Chen, A., Chomat, A. M., Fawzi, W., de Pin˜eres, C. G., Hamadani, J., Henao, N., Jervis, P., Kane, C., Measelle, J., Medrano, P., Pisani, L., … Yousafzai, A. K. (2021). Validation of motor, cognitive, language, and socio-emotional subscales using the Caregiver Reported Early Development Instruments: An application of multidimensional item factor analysis. *International Journal of Behavioral Development*, *45*(4), 368–377. https://doi.org/10.1177/01650254211005560

Washington, K. N., McDonald, M. M., McLeod, S., Crowe, K., & Devonish, H. (2017). Validation of the intelligibility in context scale for Jamaican Creole-speaking preschoolers. *American Journal of Speech-Language Pathology*, *26*(3), 750–761. https://doi.org/10.1044/2016_AJSLP-15-0103

Weber, A. M., Marchman, V. A., DIop, Y., & Fernald, A. (2018). Validity of caregiver-report measures of language skill for Wolof-learning infants and toddlers living in rural African villages. *Journal of Child Language*, *45*(4), 939–958. https://doi.org/10.1017/S0305000917000605

Wijedasa, D. (2012). Developmental screening in context: Adaptation and standardization of the Denver Developmental Screening Test-II (DDST-II) for Sri Lankan children. *Child: Care, Health and Development*, *38*(6), 889–899. https://doi.org/10.1111/j.1365-2214.2011.01332.x

Willoughby, M. T., Piper, B., Kwayumba, D., & McCune, M. (2019). Measuring executive function skills in young children in Kenya. *Child Neuropsychology*, *25*(4), 425–444. https://doi.org/10.1080/09297049.2018.1486395

Willoughby, M. T., Piper, B., Oyanga, A., & Merseth King, K. (2019). Measuring executive function skills in young children in Kenya: Associations with school readiness. *Developmental Science*, *22*(5). https://doi.org/10.1111/desc.12818

Wolf, S., Halpin, P., Yoshikawa, H., Dowd, A. J., Pisani, L., & Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Childhood Research Quarterly*, *41*(June), 21–36. https://doi.org/10.1016/j.ecresq.2017.05.001

Wu, Z., Hu, B. Y., & Fan, X. (2019). Cross-Cultural Validity of Preschool Learning Behavior Scale in Chinese Cultural Context. *Journal of Psychoeducational Assessment*, *37*(1), 125–130. https://doi.org/10.1177/0734282916651538

Wu, Z., Mak, M. C. K., Hu, B. Y., He, J., & Fan, X. (2019). A validation of the Social Skills domain of the Social Skills Improvement System-Rating Scales with Chinese preschoolers. *Psychology in the Schools*, *56*(1), 126–147. https://doi.org/10.1002/pits.22193

Xie, H., Bian, X., Chen, C. Y., Squires, J., & Lu, P. (2019). Examining the Convergent Evidence of a Parent-Completed, Social-Emotional Screening Tool in China. *Journal of Child and Family Studies*, *28*(6), 1471–1480. https://doi.org/10.1007/s10826-019-01362-3

Xie, H., Clifford, J., Squires, J., Chen, C. Y., Bian, X., & Yu, Q. (2017). Adapting and validating a developmental assessment for chinese infants and toddlers: The ages & stages questionnaires: Inventory. *Infant Behavior and Development*, *49*(June), 281–295. https://doi.org/10.1016/j.infbeh.2017.09.004

Xie, S., & Li, H. (2019). Development and Validation of the Chinese Preschool Readiness Scale. *Early Education and Development*, *30*(4), 522–539. https://doi.org/10.1080/10409289.2019.1596462

Yuan, H., Ocansey, M., Adu-Afarwuah, S., Sheridan, M., Hamoudi, A., Okronipa, H., Kumordzie, S. M., Oaks, B. M., & Prado, E. (2022). Evaluation of a tablet-based assessment tool for measuring cognition among children 4–6 years of age in Ghana. *Brain and Behavior*, *12*(10), 1–12. https://doi.org/10.1002/brb3.2749

Yue, A., Jiang, Q., Wang, B., Abbey, C., Medina, A., Shi, Y., & Rozelle, S. (2019). Concurrent validity of the Ages and Stages Questionnaire and the Bayley Scales of Infant Development III in China. *PLoS ONE*, *14*(9). https://doi.org/10.1371/journal.pone.0221675

Yunilda, E., Gunardi, H., Medise, B. E., & Oswari, H. (2023). The Indonesian version of Ages and Stages Questionnaire III accuracy compared to Bayley Scales of Infant Development III. *Infant and Child Development*, *32*(1), 1–12. https://doi.org/10.1002/icd.2387

Zakaria, S., Chua, B. S., Sombuling, A., Ahmad, S., & Hashmi, S. I. (2012). Reliability and Validity for Malay Version of Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III): Preliminary Study. *International Proceedings of Economics Development and Research*, *40*(June), 116–119.

Zhang, L., Li, Z., & He, K. (2018). Validation of the Chinese Revision of Emotional and Behavioral Screener among Preschoolers (4–7 years old). *Journal of Child and Family Studies*, *27*(3), 771–779. https://doi.org/10.1007/s10826-017-0940-0

Zhang, Y., Xu, X., Jiang, Y., Sun, W., Wang, Y., Song, Y., Dong, S., Zhu, Q., Jiang, F., & Sheng, L. (2021). Early language and communication development in Chinese children:

Adaption and validation of a parent report instrument. *International Journal of Speech-Language Pathology*, *23*(4), 372–381. https://doi.org/10.1080/17549507.2020.1817558

Zirakashvili, M., Gabunia, M., Tatishvili, N., Ediberidze, T., Lomidze, G., Chachava, T., & Hix-Small, H. (2018). Cultural Adaptation and Psychometric Validation of the Ages and Stages Questionnaires for use in Georgia. *Journal of Child and Family Studies*, *27*(3), 739–749. https://doi.org/10.1007/s10826-017-0917-z

BMJ Open

Psychometric properties of early childhood development assessment tools in low-and-middle-income countries: a systematic review

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2024-096365.R1
Article Type:	Original research
Date Submitted by the Author:	27-Mar-2025
Complete List of Authors:	Bliznashka, Lilia; International Food Policy Research Institute, Nutrition, Diets, and Health Unit; University of Edinburgh, Global Academy of Agriculture and Food Systems Hentschel, Elizabeth; Harvard T H Chan School of Public Health, Department of Global Health and Population Ali, Nazia Binte; Harvard T H Chan School of Public Health, Department of Global Health and Population Hunt, Xanthe; Stellenbosch University, Department of Global Health; Africa Health Research Institute (AHRI) Neville, Sarah ; Brown University School of Public Health, International Health Institute Olney, D.; International Food Policy Research Institute, Nutrition, Diets, and Health Unit Pitchik, Helen; University of California Berkeley, School of Public Health Roy, Aditi; Centre for Chronic Disease Control Seiden, Jonathan; Harvard Graduate School of Education Solís-Cordero, Katherine ; University of Costa Rica, School of Nursing Thapa, Aradhana; Emory University Hubert Department of Global Health Jeong, Joshua; Emory University Hubert Department of Global Health
Primary Subject Heading:	Paediatrics
Secondary Subject Heading:	Global health, Paediatrics
Keywords:	PAEDIATRICS, Child, Community child health < PAEDIATRICS, Patient Reported Outcome Measures, Psychometrics

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Psychometric properties of early childhood development assessment tools in low-and-middle-income countries: a systematic review

Lilia Bliznashka, PhD[1,2,†], Elizabeth Hentschel, PhD [3], Nazia Binte Ali, MPH[3], Xanthe Hunt, PhD [4,5], Sarah Elizabeth Neville, PhD[6], Deanna Olney, PhD[1], Helen O. Pitchik, PhD[7], Aditi Roy, PhD[8], Jonathan Seiden, EdM[9], Katherine Solís Cordero, PhD[10], Aradhana Thapa, MPH[11], Joshua Jeong, ScD[11]

[1] International Food Policy Research Institute, Washington, DC, 20001

[2] Global Academy of Agriculture and Food Systems, University of Edinburgh, UK

[3]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA

[4] Institute for Life Course Health Research, Department of Global Health, Stellenbosch University, South Africa

[5] Africa Health Research Institute (AHRI), Somkele, South Africa

[6] Brown University School of Public Health, International Health Institute, Providence, RI, USA

[7] Division of Epidemiology, School of Public Health, University of California, Berkeley, Berkeley, California, USA

[8] Centre for Chronic Disease Control, New Delhi, India

[9] Harvard Graduate School of Education, Cambridge, MA, USA

[10] School of Nursing, University of Costa Rica, San José, Costa Rica

[10] Hubert Department of Global Health, Emory University, Atlanta, GA, USA

† After the first, second, and last authors, all authors are listed in alphabetical order by last name.

Corresponding author:

Lilia Bliznashka

l.bliznashka@cgiar.org

1201 I St NW, Washington, DC 20001, US

Abstract

Objective: Valid and reliable measurement of early childhood development (ECD) is critical for monitoring and evaluating ECD-related policies and programmes. Although ECD tools developed in high-income countries may be applicable to low- and middle-income countries (LMICs), directly applying them in LMICs can be problematic without psychometric evidence for new cultures and contexts. Our objective was to systematically appraised available evidence on the psychometric properties of tools used to measure ECD in LMIC.

Design: A systematic review following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.

Data sources: MEDLINE, Embase, PubMed, PsycInfo, SciELO, and BVS were searched from inception to February 2025.

Eligibility criteria: We included studies that examined the reliability, validity, and measurement invariance of tools assessing ECD in children 0-6 years of age living in LMICs.

Data extraction and synthesis: Each study was independently screened by two researchers and data extracted by one randomly assigned researcher. Risk of bias was assessed using a checklist developed by the study team assessing bias due to training/administration, selective reporting, and missing data. Results were synthesized narratively by country, location, age group at assessment, and developmental domain.

Results: A total of 160 articles covering 117 tools met inclusion criteria. Most reported psychometric properties were internal consistency reliability (n=117, 64%), concurrent validity (n=81, 45%), convergent validity (n=74, 41%), test-retest reliability (n=73, 40%), and structural validity (n=72, 40%). Measurement invariance was least commonly reported (n=16, 9%). Most articles came from Brazil, China, India, and South Africa. Most psychometric evidence was from

urban (n=92, 51%) or urban-rural (n=41, 23%) contexts. Study samples focused on children aged 6-17.9 or 48-59.9 months. The most assessed developmental domains were language (n=111, 61%), motor (n=104, 57%) and cognitive (n=82, 45%). Bias due to missing data was most common.

Conclusions: Psychometric evidence is fragmented, limited, and heterogeneous. More rigorous psychometric analyses, especially on measurement invariance, are needed to establish the quality and accuracy of ECD tools for use in LMICs.

Strengths and limitations of this study

- This systematic review used extensive search filters, made no restrictions on developmental domains, and included both single- and multi-domain tools.
- The review team was rigorously trained and brought diverse backgrounds in ECD measurement in LMICs.
- We could not conduct full text review and extraction for articles in Chinese, Farsi, and Turkish because these languages were not spoken by the review team.
- We did not use a validated tool to assess risk of bias, which may limit the comparability of our findings.

INTRODUCTION

The Sustainable Development Goals (SDGs) recognise the importance of early childhood development (ECD),[1] which sets the foundation for children's later learning and economic outcomes.[2–4] Rigorous ECD measurement is critical for accurate comparability of children's skills across populations and time, evaluating the effectiveness of ECD interventions, and tracking ECD-related policies.

In low-and-middle-income countries (LMICs), >140 tools (60% of which originated in high-income countries (HICs)) have been used to assess ECD in children 0-8 years old.[5] Although ECD tools developed in HICs may be applicable to LMICs, directly applying ECD tools from HICs in LMICs can be problematic without psychometric evidence for new cultures and contexts. Few studies to-date have synthesised the evidence on the psychometric properties of ECD tools.

Evidence pertaining to a tool's reliability, validity, and measurement invariance in a given context is critical for selecting an ECD outcome or indicator. Other factors include the purpose of measurement, the population and age-range of interest, the developmental domain(s) of interest, and administration time and csot.[6] Evidence of reliability and validity ensures consistent and accurate ECD measurement,[7] whereas evidence of measurement invariance guarantees assessment of the same construct across countries and sub-groups.[7,8]

Prior reviews have provided guidance for selecting ECD tools for use in LMICs[5,9–11] and underscored that evidence on tool reliability and validity is fundamental.[9,10] However, prior work has often focused on individual ECD domains,[12–14] which has limited use for population level assessment or tracking of SDG-related policies, or a subset of psychometric properties,[10] which is not evidence of reliability and validity as a whole. Further, prior reviews do not disaggregate use of ECD tools by HICs vs LMICs or urban vs rural settings,[12–14] despite known and persistent

disparities in ECD by country income level and urban/rural residence[15,16] and the fact that young child populations are increasingly diverse due to migration and urbanization.[13] These are important distinctions given that psychometric properties can vary by population characteristics.

We reviewed available evidence on 10 psychometric properties of tools used to assess ECD in children 0-6 years old living in LMICs. We summarised the current landscape by ECD tool, developmental domain, country, and age group, focusing on 10 types of psychometric evidence. We sought to deepen our understanding and consistently summarise whether psychometric evidence exists for tools used to measure ECD in LMICs. Our findings can assist stakeholders in selection of ECD tools for their intended use and context, and inform what research is needed to improve how we track ECD-related SDGs, programmes, and policies in LMICs.

METHODS

Search strategy and selection criteria

We identified articles in any language through MEDLINE, Embase, PubMed, PsycInfo, SciELO, and BVS. The search strategy used medical subject headings (MeSH), keywords, and free text words along four key elements: population, construct, measurement properties, and location (**Supplemental Tables 1**). Search terms were combined using Boolean operators. Truncation wildcards were used to include variations of the search terms. The search strategy used a combination of searches through titles, abstracts, and keywords. The COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) measurement properties filter was used for the third search element.[17] The search strategy was piloted in PubMed and then adapted for the remaining databases.

Full-text, peer-reviewed articles were included if: (1) the study was conducted in a LMIC, (2) included children 0-6 years old, (3) included at least one ECD domain (cognitive, language, motor, social-emotional, attention/executive function, personal-social, and pre-academic/academic) defined in **Supplemental Table 2**, (4) developed a new tool or adapted an existing one, and provided primary evidence of at least one of 10 psychometric properties in terms of reliability, validity, or measurement invariance (**Table 1**), and (5) were published between January 1, 2007 and March 9, 2023. An updated search was conducted on February 28, 2025 and seven additional studies were identified for inclusion: five through the database searches and two from a Google Scholar alert. The 10 psychometric properties were selected based on prior systematic reviews on psychometrics properties of ECD tools,[9,10,12,18,19] classical test theory,[20] and reviews of measurement in cross-cultural psychology.[21,22] We considered all ECD tools regardless of their intended or actual use (diagnosis, screening, or surveillance). We included both articles that adapted original tools ("adaptation articles" hereafter) and articles that developed new tools ("development articles" hereafter). Articles were included if at least 50% of the study sample was within the range 0-6 years and the average age in the sample was <6 years or unspecified. If information on the child age range was not available or clear, the article was included. Multi-country articles were included if ≥50% of the countries were LMICs. If an article reported that measurement properties were reported elsewhere (e.g., referencing another study and thus providing secondary evidence), the article was ineligible. We included prospective, retrospective, cross-sectional, and longitudinal quantitative study designs. Cited references were reviewed for potential inclusion.

We excluded articles where a tool was used to assess an outcome measure (e.g., trials reporting impacts on ECD outcomes), but the article did not include measurement objectives.[23]

Among articles with an explicit measurement focus, we excluded those reporting only on convergent validity (e.g., correlations with socio-demographic variables) which alone provides limited psychometric evidence.[23] We excluded articles studying children with developmental disabilities and disorders (e.g., autism spectrum disorder, cerebral palsy), and children with physical disabilities that impair performance on ECD measures (e.g., deafness, blindness). Lastly, we excluded the following study designs: animal studies, simulation studies, case studies, opinions, letters, preprints, protocols, conference abstracts, ecological studies, dissertations/theses, reviews or systematic reviews, and meta-analyses.

Data Extraction

Search results were imported into Covidence, where duplicates were automatically removed. Two reviewers independently screened titles and abstracts for inclusion and reviewed full texts of retained articles. Disagreements in screening or full text review were resolved through discussion with a third reviewer.

One reviewer developed the data extraction sheet, and two reviewers piloted it. Revisions were made through discussion between the two reviewers. Seven reviewers were trained on using the data extraction sheet through pilot extractions of included articles. Further revisions were made to the data extraction sheet based on discussions of the piloting process. Data extraction included information on publication details, study meta-data, characteristics of the ECD tool, type of administration, ECD domains assessed, and psychometric properties with respect to reliability, validity, and measurement invariance. After piloting, each article was extracted by one reviewer. For quality assurance, 20% of articles were randomly selected for independent extraction by a second reviewer. Any discrepancies regarding data extraction were resolved by a third reviewer. Authors of included articles were not contacted when information was missing or unclear.

Risk of Bias Assessment

Since we aimed to assess the quality of the underlying studies, rather than the quality of the psychometric properties, we could not use a validated risk of bias tool and instead developed a new one. One reviewer created a risk of bias checklist by adapting items from the COSMIN[23] and Cochrane's ROBINS-E[24] risk of bias tools. The checklist was then refined via discussions with two other reviewers and a psychometrician, and after piloting by four reviewers. The final checklist contained three categories of bias due to: (1) training/administration (not assessed for tools relying on self-assessment), (2) selective reporting (only assessed for studies reporting on convergent or predictive validity), and (3) missing data. Each article was rated separately on each category using the ROBINS-E risk of bias ratings: low risk, some concerns, high risk, and very high risk. We also assessed indirectness of populations [25] by assessing whether the sample was limited to a specific setting, the tool covered the entire age range it was intended for, subgroups were generalisable, and results were generalisable. Where insufficient information was provided in the article, we rated the study as "unable to assess". Seven reviewers were trained and conducted the risk of bias assessment at the article level. Disagreements for articles assessed by two reviewers (20%) were resolved through discussion with a third reviewer.

Data Synthesis

Data analysis was conducted at the article-tool level because some articles reported on multiple tools. We created binary indicators for whether evidence on each one of the 10 psychometric properties was reported. In longitudinal studies, we considered the psychometric properties reported at any time point. We then summarised the evidence by country, location, age group at assessment, type of article (development vs adaptation), developmental domain, and ECD tool. For all tools, child age at assessment was converted to months based on the information

provided in the article. For longitudinal studies, age at first assessment was used. For multi-site/multi-country studies, the full age range at assessment was used across the sites/countries. Because the objective the paper was to take stock of the available evidence rather than to report the adequacy and relevance of specific psychometric properties, we did not summarize evidence on the psychometric properties themselves. Results were synthesised narratively.

This review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[26] It was preregistered on PROSPERO: CRD42022372305.

Patient and Public Involvement

Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

Results

Study characteristics

Our search strategy identified 6,430 records from six databases and two other sources (**Figure 1**). After removing duplicates, 5,388 records were excluded during title and abstract screening. After full text review of the remaining 250 records, 98 records were excluded. After updating the search in February 2025, we included seven more articles (five identified from the six databases and two identified from other sources). We included 160 articles, covering 117 tools. Four articles reported on multiple tools, resulting in 182 article-tool combinations (referred to as articles for brevity).

Psychometric evidence available

Most articles were adaptation article (n=145, 80%) (**Figure 2, Supplemental Table 3**). The most often evaluated psychometric properties were internal consistency reliability (n=117, 64%), concurrent validity (n=81, 45%), convergent validity (n=74, 41%), test-retest reliability

(n=73, 40%), and structural validity (n=72, 40%). Measurement invariance was the least frequently reported psychometric property (n=16, 9%) and was primarily evaluated over countries and child sex. The number of articles increased over time (**Supplemental Figure 1**). Between 2007-2010, most included articles were development articles; since 2011, most included articles have been adaptation articles.

Psychometric evidence available by country

Psychometric evidence came from 55 countries (**Figure 3**); one-third of articles came from Brazil (n=22, 12%), China (n=18, 10%), India (n=14, 8%), and South Africa (n=10, 5%). For 40% of countries represented, there was only one article reporting psychometric evidence (**Supplemental Table 4**). Psychometric evidence for each ECD tool included was generally limited to 1-2 countries with a few notable exceptions: the Global Scales for Early Development (GSED, 32 countries), the Caregiver Reported Early Development Instrument (CREDI, 18 countries), the International Development and Early Learning Assessment (IDELA, 17 countries), the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-III, 12 countries), and the Ages and Stages Questionnaire-3 (ASQ-3, 11 countries) (**Supplemental Table 5**).

Psychometric evidence available by location

Most psychometric evidence came from urban (n=92, 51%) or urban-rural (n=41, 23%) settings. In South Asia and Sub-Saharan Africa, a similar number of articles originated from urban, rural, and urban-rural settings (**Supplemental Table 6**).

Psychometric evidence available by ECD tool

For development articles, a single article reported psychometric properties for all tools except for the CREDI covered in two articles (**Supplemental Table** 7). For adaptation articles, ASQ-3 and BSID-III were most often studied (14 and 12 articles, respectively). For 75% of tools

only a single article provided psychometric evidence (**Supplemental Table 8**). Other frequently studied tools included the Mullen Scales of Early Learning (n=5 articles), the Alberta Infant Motor Scale (n=4 articles), the Bayley Infant Neurodevelopmental Screener (n=4 articles), Denver Developmental Screening Test (Denver-II, n=4 articles), and International Development and Early Learning Assessment (IDELA, n=4 articles) (Supplemental Table 5).

Psychometric evidence available by age group

Although included tools targeted 0-71.9-month-old children, most studies focused on children 12–17.9 month (n=86, 47%), 6-11.9 months (n=82, 45%), or 54-59.9 months (n=76, 42%) old. Articles largely did not cover the full age groups defined here, or the full age range the ECD tool can be used for. Some articles included age groups as narrow as one month (**Supplemental Figures 2-3**).

Psychometric evidence available by developmental domain

Most articles reported on language (n=111, 61%), motor (n=104, 57%), and cognitive (n82, 45%) development (**Supplemental Table 9**). Academic/pre-academic was the least studied domain (n=10, 5%). Within domains, internal consistency reliability was most frequently reported (**Figure 4**).

Risk of bias assessment

Bias due to training/administration was assessed for 59 articles (32%). Of these, 80% had low risk of bias (n=47) and 20% some concerns (n=12) (**Supplemental Tables 10-11**). Risk of bias due to selective reporting was assessed for 95 articles (52%). Of these, most articles (n=86, 91%) had low risk of bias. Risk of bias due to missing data was assessed for 137 articles (75%). Of these, 48% had some concerns (n=52) and 5% high risk (n=7). For indirectness, 20 articles (11%) were rated as generalisable and 82 (45%) as probably generalisable.

DISCUSSION

Based on 160 articles, available evidence on 10 psychometric properties of 117 ECD tools for children 0-6 years old is fragmented, limited, and heterogeneous. The most frequently provided evidence was on internal consistency reliability, test-retest reliability and/or concurrent, convergent, and structural validity. Psychometric evidence on measurement invariance was the least commonly available. Although evidence came from 55 LMICs, four countries were most represented. Most evidence came from urban or urban-rural settings. The most studied tools were ASQ-3 and BSID-III.

Our findings support ECD measurement trends found in other work:[27] much of the work on psychometric properties is recent, with ECD tools being developed and adapted concurrently. Psychometric efforts remain limited to a few ECD tools,[28,29] individual ECD domains,[12–14] and few psychometric properties.[10,12] This fragmentation is evidenced by included articles focusing on individual countries, limited age ranges, and single developmental domains. In addition, included studies focused on individual reliability or validity properties (e.g., internal consistency reliability and concurrent reliability, respectively), thus providing a limited picture of reliability and validity as a whole. The resulting heterogeneous psychometric evidence can hinder comparability and large-scale monitoring of ECD policies and programmes within and across LMICs, which is crucial for identifying and implementing effective approaches to support ECD.

Despite efforts to consolidate ECD measurement through tools like CREDI, GSED, and IDELA, such tools do not fully meet research, programmatic, and policy needs as evidenced by the increase of adaptation and development articles since 2015. This is not surprising given that no ECD tool is suitable for all populations.[18] Our disaggregation by development and adaptation

articles permitted better understanding of the psychometric evidence available and highlighted evidence gaps for both existing and newly developed tools.

This review highlights four important limitations of existing psychometric evidence for ECD tools in LMICs. First, although most tools are designed for a wide age range, the psychometric evidence behind most tools pertained to narrower age ranges and in some cases as narrow as one month. This may have limited applicability to diverse age ranges (given that there is a natural variability in child development in the early years[13]) in these specific contexts. Relatedly, most psychometric evidence pertained to urban contexts. Given existing urban-rural disparities in ECD[15,16] and increasingly diverse young child populations in urban settings,[13] ECD tools whose psychometric properties were examined only in urban settings might be inadequate for rural settings. Those developing or adapting ECD tools should consider establishing psychometric properties across the full intended child age range and across both urban and rural settings. Although this implies longer, more expensive, and logistically difficult studies, it would ensure the tool can be used broadly and in more diverse populations.

A second limitation is the very little evidence on measurement invariance. While several studies reported on samples drawn from multiple countries, few conducted statistical analysis to test for equivalence across countries, thus providing no evidence of measurement invariance.[8] Cross-country invariance, which guarantees assessment of the same construct across countries, is key for tracking global SDG goals. More work in this domain is needed particularly for tools widely used for policy making and programme evaluations. Measurement work should be considered relative to competing and more urgent ECD priorities in LMICs.

Third, consistent with existing literature, we found limited psychometric evidence on tools measuring socio-emotional and personal-social development.[13,19] This is surprising given that

these two domains are among the most culturally specific,[13] implying they require more comprehensive and rigorous adaptation. In addition, psychometric evidence on tools to assess attention/executive function and academic/preacademic development was very limited. Without additional work to establish a psychometric base, this poses major challenges for those seeking to monitor these domains in early life.

Lastly, in contrast to prior reviews which assessed the quality of psychometric properties using COSMIN guidelines,[11,12] we assessed the risk of bias and quality of the underlying studies themselves. We observed a common lack of reporting and transparency in training of assessors and data management. Risk of bias could not be assessed for many included studies or was considered high. Better reporting standards and guidelines for psychometric studies can help strengthen the field and ensure more critical evaluation of the evidence is possible.

Nevertheless, some ECD tools had multiple forms of psychometric properties assessed. Using IDELA as an example: one article examined three types of reliability and one type of validity, drawing on pilots from 12 countries,[27] followed by another article examining measurement invariance across countries.[30] Subsequent articles have examined additional psychometric properties, albeit in individual countries.[31,32] Although such examples of consecutively examining psychometric properties should be the norm, they are often financially and logistically infeasible and the timing does not always align with programmatic and policy agendas. Likewise, ample psychometric evidence was available for ECD tools that have been implemented for longer duration (e.g., ASQ and BSID), or have had more available funding (e.g., GSED). As a result, the quantity of psychometric evidence available should not be the criterion used to determine the psychometric quality or useability of an ECD tool. When prior psychometric evidence is unavailable and psychometric studies not possible, statistical analysis should be

conducted to verify psychometric properties and at a minimum should include reporting internal consistency reliability and structural validity (where relevant). This can help build evidence across multiple settings and populations and confirm the usefulness of ECD tools with diverse populations. In addition, since our review focused on whether psychometric evidence exists, our findings on the availability of psychometric evidence do not inform our understanding of the underlying quality, strength, or rigor of the psychometric evidence. An important next step in this line of work is to fully unpack the utility of existing psychometric evidence. The results of psychometric analyses along with other characteristics of an ECD tool (e.g., domain assessed, age range, and administration time and cost among others) should be used to determine the most relevant ECD tool for the given context.

Several strengths of our review should be noted. We used extensive search filters, made no restrictions on domains, included both single- and multi-domain tools, and had a rigorously trained team with diverse backgrounds in ECD measurement in LMICs. Nevertheless, we acknowledge some limitations. We excluded studies published prior to 2007, the search start year when the inaugural *Lancet* Series on Child Development in Developing Countries was published which fundamentally changed the breadth and scale of ECD research in LMICs. Therefore, the search period captured the most important years for the evolution of psychometric research on ECD tools in LMICs. Further, we did not use a validated risk of bias tool, which may limit the comparability of our findings. Lastly, although we applied no language restrictions, we were limited to the languages spoken by the review team. We were unable to conduct full text review and extraction for articles in Chinese, Farsi, and Turkish. Consequently, results from China, Iran, and Turkey may be underrepresented.

CONCLUSION

Psychometric evidence on ECD tools used in LMICs is fragmented, limited, and heterogeneous. More research is warranted to establish the applicability of existing tools in diverse populations, including urban and rural settings, and on establishing measurement invariance over countries. Nevertheless, the results by country and ECD tool presented here can serve stakeholders by providing a database of available psychometric evidence for ECD tools in LMICs. To improve monitoring, evaluation, and accountability for ECD globally, psychometric evidence should be a key consideration when selecting ECD tools together with other important consideration including the purpose of measurement, available resources for training and administration, and the population and developmental domain of interest. As psychometric properties can vary by geography, population, and age, among other characteristics, greater psychometric validation can help facilitate ECD tool selection across diverse contexts in LMICs. Improved reporting for psychometric studies can help ensure transparency, replication, and adequate ability to assess the quality of evidence.

Acknowledgements

Contributors: LB conceptualised the review, designed the methodology, screened articles, extracted and analysed data, drafted and revised the manuscript. EH extracted data and analysed data, drafted and revised the manuscript. NBA and AR extracted data, and revised the manuscript. XH and KSC screened articles, extracted data, and revised the manuscript. SEN and JS extracted data and revised the manuscript. DO revised the manuscript. HOP screened articles and revised the manuscript. AT curated the data and revised the manuscript. JJ conceptualised the review,

designed the methodology, screened articles, and revised the manuscript. All authors read and approved the final manuscript. LB is the guarantor.

Funding: This work was supported by the CGIAR Initiative on Resilient Cities. We would like to thank all funders who supported this research through their contributions to the CGIAR Trust Fund: https://www.cgiar.org/funders. SEN's time was supported by the National Institute of Mental Health (grant number T32 MH078788).

Competing interests: JS has contributed and continues to contribute to the development of the CREDI, GSED, IDELA and AIM-ECD. The remaining authors declare no conflicts of interest.

Patient consent for publication: Not applicable.

Ethical considerations: Not applicable.

Provenance and peer review: Not commissioned; externally peer reviewed.

Data availability statement: All data are included in the manuscript and appendix.

References

1. United Nations. Transforming our world: the 2030 Agenda for Sustainable Development. Published online 2015. https://sdgs.un.org/sites/default/files/publications/21252030 Agenda for Sustainable Development web.pdf
2. Walker SP, Chang SM, Vera-Hernandez M, Grantham-McGregor S. Early Childhood Stimulation Benefits Adult Competence and Reduces Violent Behavior. *Pediatrics*. 2011;127(5):849-857. doi:10.1542/peds.2010-2231
3. Gertler P, Heckman J, Pinto R, et al. Labor market returns to an early childhood stimulation intervention in Jamaica. *Science (80-)*. 2014;344(6187):998-1001. doi:10.1126/science.1251178
4. Campbell F, Conti G, Heckman JJ, et al. Early Childhood Investments Substantially Boost Adult Health. *Science (80-)*. 2014;343(6178):1478-1485. doi:10.1126/science.1248429
5. Fernald LCH, Prado EL, Kariger PK, Raikes A. A Toolkit for Measuring Early Childhood Development in Low- and Middle-Income Countries. Published online 2017.
6. Pushparatnam A, Seiden JM, Luna Bazaldua DA. Guiding Questions for Choosing the Right Tools to Measure Early Childhood Outcomes: Why, What, Who, and How. Published online 2022. https://openknowledge.worldbank.org/handle/10986/37030
7. American Psychological Association. APA Dictionary of Psychology. Accessed February 26, 2023. https://dictionary.apa.org/
8. Putnick DL, Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Dev Rev*. 2016;41:71-90. doi:10.1016/j.dr.2016.06.004
9. Semrud-Clikeman M, Romero RAA, Prado EL, Shapiro EG, Bangirana P, John CC. Selecting measures for the neurodevelopmental assessment of children in low- and middle-income countries. *Child Neuropsychol*. Published online September 9, 2016:1-42. doi:10.1080/09297049.2016.1216536
10. Munoz-Chereau B, Ang L, Dockrell J, Outhwaite L, Heffernan C. Measuring early child development across low and middle-income countries: A systematic review. *J Early Child Res*. 2021;19(4):443-470. doi:10.1177/1476718X211020031
11. Boggs D, Milner KM, Chandna J, et al. Rating early child development outcome measurement tools for routine health programme use. *Arch Dis Child*. 2019;104(Suppl 1):S22-S33. doi:10.1136/archdischild-2018-315431
12. Griffiths A, Toovey R, Morgan PE, Spittle AJ. Psychometric properties of gross motor assessment tools for children: a systematic review. *BMJ Open*. 2018;8(10):e021734. doi:10.1136/bmjopen-2018-021734
13. Halle TG, Darling-Churchill KE. Review of measures of social and emotional development. *J Appl Dev Psychol*. 2016;45:8-18. doi:10.1016/j.appdev.2016.02.003
14. Humphrey N, Kalambouka A, Wigelsworth M, Lendrum A, Deighton J, Wolpert M. Measures of Social and Emotional Skills for Children and Young People. *Educ Psychol Meas*. 2011;71(4):617-637. doi:10.1177/0013164410382896
15. Lu C, Cuartas J, Fink G, et al. Inequalities in early childhood care and development in low/middle-income countries: 2010–2018. *BMJ Glob Heal*. 2020;5(2):e002314. doi:10.1136/bmjgh-2020-002314
16. McCoy DC, Peet ED, Ezzati M, et al. Early Childhood Developmental Status in Low- and Middle-Income Countries: National, Regional, and Global Prevalence Estimates Using Predictive Modeling. *PLoS Med*. 2016;13(6):1-18. doi:10.1371/journal.pmed.1002034

17. Terwee CB, Jansma EP, Riphagen II, de Vet HCW. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res*. 2009;18(8):1115-1123. doi:10.1007/s11136-009-9528-5
18. Sabanathan S, Wills B, Gladstone M. Child development assessment tools in low-income and middle-income countries: how can we use them more appropriately? *Arch Dis Child*. 2015;100(5):482-488. doi:10.1136/archdischild-2014-308114
19. Gridley N, Blower S, Dunn A, Bywater T, Bryant M. Psychometric Properties of Child (0–5 Years) Outcome Measures as used in Randomized Controlled Trials of Parent Programs: A Systematic Review. *Clin Child Fam Psychol Rev*. 2019;22(3):388-405. doi:10.1007/s10567-019-00277-1
20. Crocker LM, Algina J. *Introduction to Classical and Modern Test Theory*. First Edit. Holt, Rinehart, and Winston; 1986.
21. Hui CH, Triandis HC. Measurement in Cross-Cultural Psychology. *J Cross Cult Psychol*. 1985;16(2):131-152. doi:10.1177/0022002185016002001
22. Peña ED. Lost in Translation: Methodological Considerations in Cross-Cultural Research. *Child Dev*. 2007;78(4):1255-1264. doi:10.1111/j.1467-8624.2007.01064.x
23. Mokkink LB, Prinsen CA, Patrick DL, et al. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs). Published online 2018:1-78. https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf
24. ROBINS-E Development Group. Risk Of Bias In Non-randomized Studies - of Exposure (ROBINS-E). Launch version, 1 June 2022. Published online 2022. https://www.riskofbias.info/welcome/robins-e-tool
25. GRADE Working Group. *Handbook for Grading the Quality of Evidence and the Strength of Recommendations Using the GRADE Approach. Updated October 2013.* (Schünemann H, Brożek J, Guyatt G, Oxman A, eds.).; 2013. https://gdt.gradepro.org/app/handbook/handbook.html
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLoS Med*. 2009;6(7):e1000097. doi:10.1371/journal.pmed.1000097
27. Pisani L, Borisova I, Dowd AJ. Developing and validating the International Development and Early Learning Assessment (IDELA). *Int J Educ Res*. 2018;91(July):1-15. doi:10.1016/j.ijer.2018.06.007
28. Kersten P, Czuba K, McPherson K, et al. A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *Int J Behav Dev*. 2016;40(1):64-75. doi:10.1177/0165025415570647
29. Velikonja T, Edbrooke-Childs J, Calderon A, Sleed M, Brown A, Deighton J. The psychometric properties of the Ages & Stages Questionnaires for ages 2-2.5: a systematic review. *Child Care Health Dev*. 2017;43(1):1-17. doi:10.1111/cch.12397
30. Halpin PF, Wolf S, Yoshikawa H, et al. Measuring early learning and development across cultures: Invariance of the IDELA across five countries. *Dev Psychol*. 2019;55(1):23-37. doi:10.1037/dev0000626
31. Wolf S, Halpin P, Yoshikawa H, Dowd AJ, Pisani L, Borisova I. Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Child Res Q*. 2017;41(June):21-36. doi:10.1016/j.ecresq.2017.05.001

32. Pisani L, Seiden J, Wolf S. Longitudinal evidence on the predictive validity of the International Development and Early Learning Assessment (IDELA). *Educ Assessment, Eval Account*. 2022;34(2):173-194. doi:10.1007/s11092-021-09374-8

Table 1 Definitions of the psychometric properties used in the selection criteria for included articles

Domain	Measurement property	Example test statistic	Definition
Reliability			The consistency of a test or measurement, i.e., how consistently a measure produces similar results with repeated measures over a short period of time or across assessors at the same time point. This can also be thought of as the correlation between observed scores across replications.
	Test-retest reliability	correlation coefficient	Correlation between scores from the same test from assessments conducted over a short time interval.
	Inter-rater reliability	kappa, Bland-Altman test	The extent to which independent assessors produce similar ratings in judging the same abilities or characteristics in the same target person at the same time.
	Internal consistency reliability	Cronbach's alpha, alpha	Degree of interrelatedness among items on the same tool, i.e., how well the items work together to provide information on an underlying construct.
Validity			The degree to which the tool measures what it is supposed to measure, i.e., the degree to which the tool reflects the underlying construct.
	Content/face validity		The degree to which the content of the tool is adequate for the construct being measured, i.e., assessing the extent to which a tool appears to reflect the underlying construct.
	Concurrent/criterion validity	correlation coefficient; regression estimate	The degree to which scores on one measurement tool are related to scores obtained at about the same point in time from another tool considered the gold standard.
	Convergent validity	correlation coefficient; regression estimate	Evidence that scores on a test or measurement are associated with theoretically related measures or variables.

Domain	Measurement property	Example test statistic	Definition
	Predictive	correlation coefficient; regression estimate	Evidence that a score correlates with a variable that can only be assessed at some point after the test has been administered or the measurement made, e.g., evidence that scores now are correlated with scores at a future timepoint.
	Structural validity (dimensionality)	Exploratory factor analysis: number of factors, eigen values Confirmatory factor analysis: model fit statistics such as Comparative Fit Index (CF), Root mean square error of approximation (RMSEA)	The degree to which the scores of an assessment are an adequate reflection of the dimensionality of the construct to be measured.
Invariance			The property when a scale or construct provides the same results across different samples, populations, settings, or characteristics.
	Measurement invariance over countries	Likelihood ratio chi squared statistic and p-value from freeing parameters across groups	The degree to which an assessment or construct provides the same results across separate samples in different countries.
	Measurement invariance over other groups	Likelihood ratio chi squared statistic and p-value from freeing parameters across groups	The degree to which an assessment or construct provides the same results across different groups.

Note: All definitions based on the APA Dictionary of Psychology.[7]

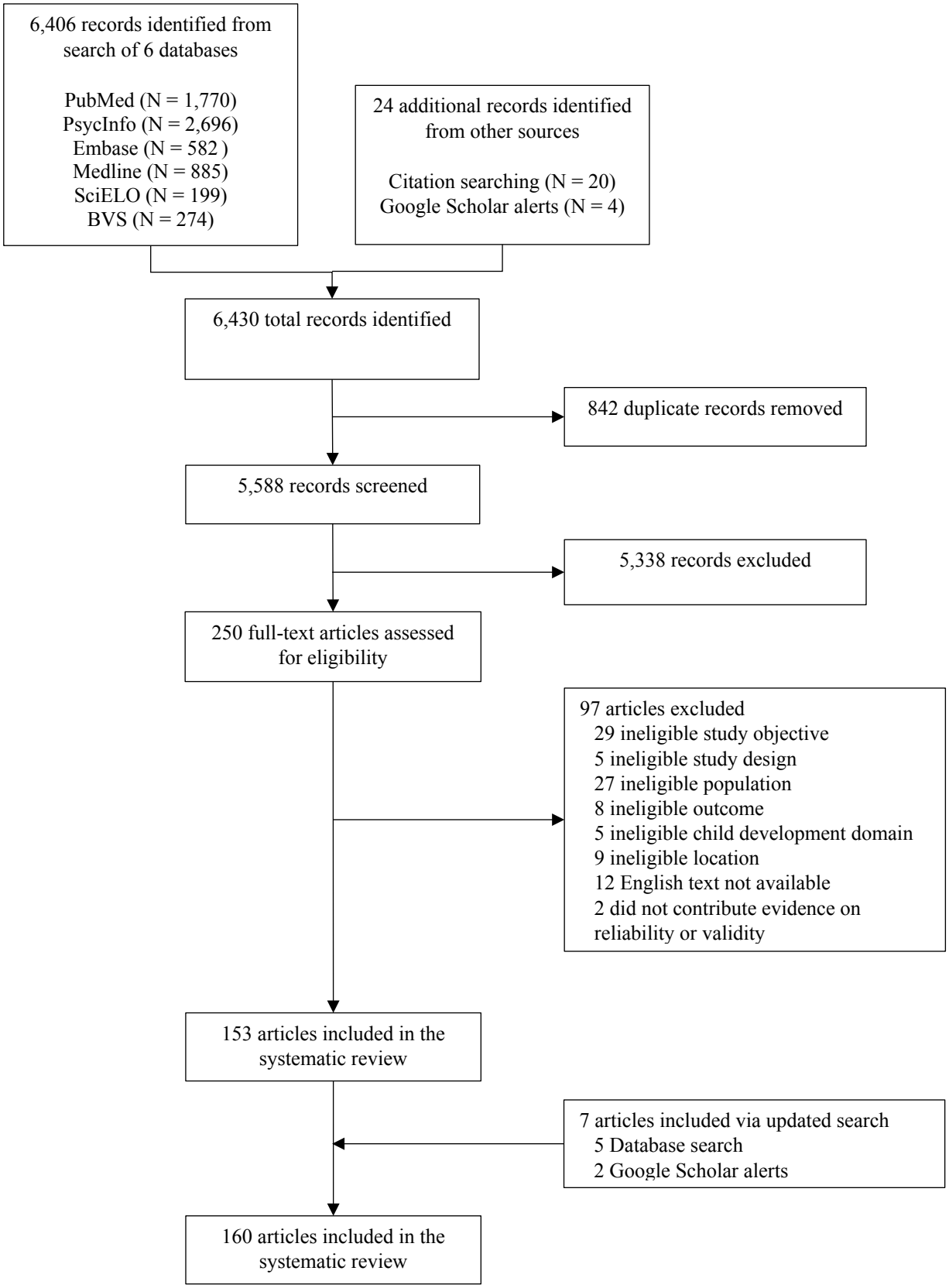
Figure Captions

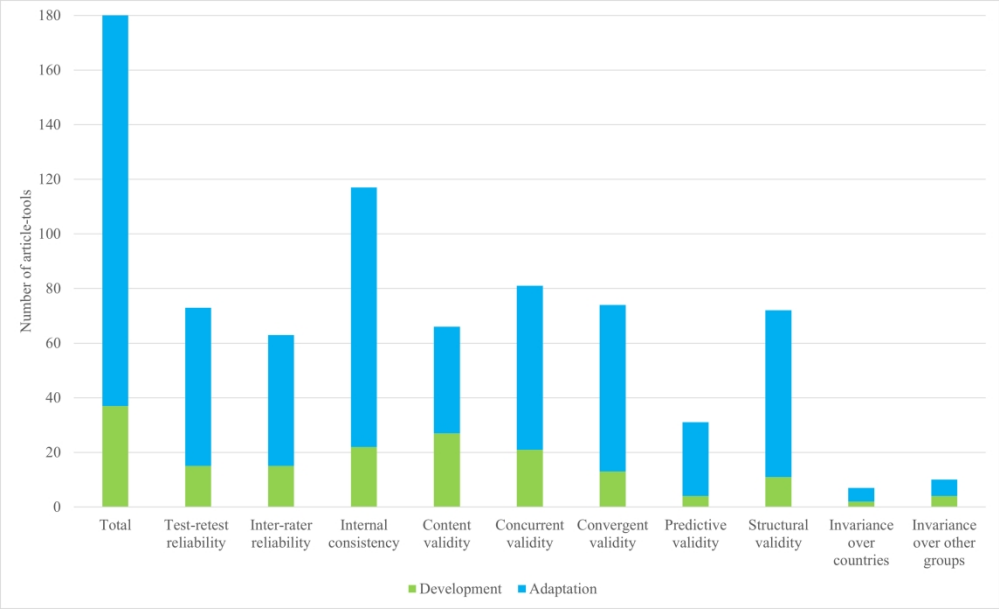
Figure 1 Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) flow diagram of search results and included articles

Figure 2 Number of included article-tools and type of psychometric evidence provided

Figure 3 Countries where studies providing evidence on at least one psychometric property were conducted

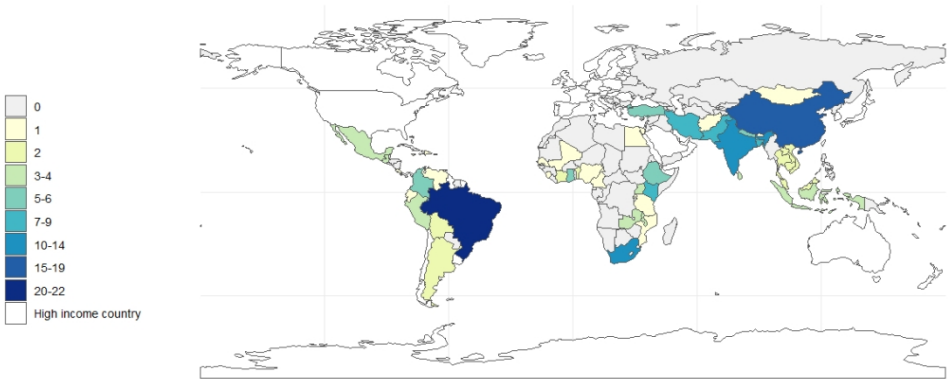
Figure 4 Number of articles providing psychometric evidence by article type and early childhood development domain





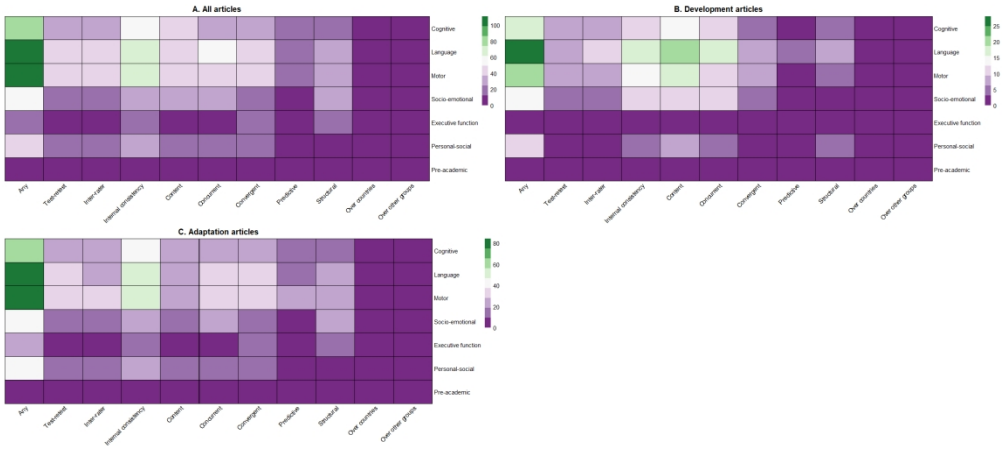
Number of included article-tools and type of psychometric evidence provided

276x168mm (300 x 300 DPI)



Countries where studies providing evidence on at least one psychometric property were conducted

265x116mm (115 x 115 DPI)



Number of articles providing psychometric evidence by article type and early childhood development domain

593x264mm (108 x 108 DPI)

Supplemental appendix

Table of Contents

Supplemental Table 1 Search strategy as executed in each database

PubMed
(“Child”[MeSH] OR “Child, Preschool”[MeSH] OR “child*”[tiab] OR “Infant”[MeSH] OR “infant*”[tiab] OR “paediatric”[tiab] OR “pediatric”[tiab] OR “toddler*”[tiab]) AND ("Child Development"[MeSH] OR “child development*”[tiab] OR “early child development”[tiab] OR “early childhood development”[tiab] OR “ECD”[tiab] OR "Child Behavior"[MeSH] OR “child behavior”[tiab] OR “child behaviour”[tiab] OR "Cognition"[MeSH] OR "Executive Function"[MeSH] OR “executive function*”[tiab] OR "Emotional Intelligence"[MeSH] OR “attachment”[tiab] OR “attention”[tiab] OR “cognition”[tiab] OR “cognitive development”[tiab] OR “cognitive skills”[tiab] OR “neurocognitive”[tiab] OR “communication”[tiab] OR “language”[tiab] OR “language skills”[tiab] OR “language development”[tiab] OR “receptive language”[tiab] OR “expressive language”[tiab] OR “speech development”[tiab] OR “communication skills”[tiab] OR "Motor Skills"[MeSH] OR “motor skills”[tiab] OR “fine motor”[tiab] OR “gross motor”[tiab] OR “motor”[tiab] OR “prosocial”[tiab] OR “socialisation”[tiab] OR “socialization”[tiab] OR “socioemotion*”[tiab] OR “socio-emotion*”[tiab] OR “neurodevelopment”[tiab] OR “neurobehavior”[tiab] OR “neurobehaviour”[tiab] OR “neurobehavioral”[tiab] OR “neurobehavioural”[tiab] OR “Intelligence”[MeSH] OR “intelligence”[tiab] OR “IQ”[tiab] OR “psychomotor”[tiab] OR “sensorimotor”[tiab] OR “attention”[tiab] OR “inhibition”[tiab] OR “impulse control”[tiab] OR “working memory”[tiab] OR “problem solving”[tiab] OR “self-regulation”[tiab] OR “prosocial”[tiab] OR “psychosocial development”[tiab] OR “internali*”[tiab] OR “externali*”[tiab] OR “empathy”[tiab] OR “neuropsychology*”[tiab] OR “school readiness”[tiab] or “numeracy”[tiab] or “literacy”[tiab] OR “pre-academic”[tiab] OR “early learning”[tiab] or “personal-social”[tiab] OR “personal social”[tiab]) AND ((“developing countr*”[tiab] OR “under developed countr*”[tiab] OR lmic*[tiab] OR “less developed”[tiab] OR “low income”[tiab] OR “lower income”[tiab] OR “low- and middle-income”[tiab] OR “low middle income”[tiab] OR “resource poor”[tiab] OR “resource constrained”[tiab] OR “low resource”[tiab] OR “limited resource*”[tiab] OR “resource limited”[tiab]) AND (Afghan*[tiab] OR Africa[tiab] OR African[tiab] OR Algeria*[tiab] OR “American Samoa*”[tiab] OR Angola*[tiab] OR Argentin*[tiab] OR Bangladesh*[tiab] OR Barbad*[tiab] OR Belorussian[tiab] OR Beliz*[tiab] OR Benin*[tiab] OR Bhutan*[tiab] OR Bolivia*[tiab] OR Botswan*[tiab] OR Brazil*[tiab] OR "Burkina Faso"[tiab] OR Burkinabe[tiab] OR Burund*[tiab] OR Cambodia*[tiab] OR Cameroon*[tiab] OR "Cape Verde"[tiab] OR "Cape Verdean"[tiab] OR "Central African Republic"[tiab] OR Chad*[tiab] OR Chile*[tiab] OR China[tiab] OR Chinese[tiab] OR Colombia*[tiab] OR Comoros[tiab] OR Comorian[tiab] OR Congo[tiab] OR Congolese[tiab] OR Costa Rica*[tiab] OR "Côte

d'Ivoire"[tiab] OR "Ivory Coast"[tiab] OR Ivorian[tiab] OR Croatia*[tiab] OR Croat[tiab] OR Cuba*[tiab] OR Djibouti*[tiab] OR Dominica*[tiab] OR "Dominican Republic"[tiab] OR Ecuador*[tiab] OR Egypt*[tiab] OR "El Salvador"[tiab] OR Salvadorian[tiab] OR "Equatorial Guinea"[tiab] OR Guinean[tiab] OR Eritrea*[tiab] OR Ethiopia*[tiab] OR Fiji*[tiab] OR Gabon*[tiab] OR Gambia*[tiab] OR Gaza[tiab] OR Gazan[tiab] OR Ghana[tiab] OR Ghanaian[tiab] OR Grenad*[tiab] OR Guatemala*[tiab] OR Guinea[tiab] OR Guyan*[tiab] OR Haiti*[tiab] OR Hondura*[tiab] OR Hungar*[tiab] OR India[tiab] OR Indian[tiab] OR Indonesia*[tiab] OR Iran*[tiab] OR Iraq*[tiab] OR Jamaica*[tiab] OR Jordan*[tiab] OR Kenya[tiab] OR Kenyan[tiab] OR Kiribati[tiab] OR Korea*[tiab] OR Kyrgy*[tiab] OR Laos[tiab] OR Laotian*[tiab] OR Lebanon[tiab] OR Lebanese[tiab] OR Lesotho[tiab] OR Liberia*[tiab] OR Libya*[tiab] OR Macedonia*[tiab] OR Madagasca*[tiab] OR Malawi*[tiab] OR Malaysia*[tiab] OR Maldives[tiab] OR Maldivian[tiab] OR Mali[tiab] OR Malian*[tiab] OR "Marshall Islands"[tiab] OR Mauritania*[tiab] OR Mauritius[tiab] OR Mauritian[tiab] OR Mayotte[tiab] OR Mexic*[tiab] OR Micronesia*[tiab] OR Moldov*[tiab] OR Mongolia*[tiab] OR Morocc*[tiab] OR Mozambique[tiab] OR Mozambican[tiab] OR Myanmar[tiab] OR Namibia*[tiab] OR Nepal*[tiab] OR Nevis[tiab] OR Nicaragua*[tiab] OR Niger*[tiab] OR "Northern Mariana Islands"[tiab] OR Oman*[tiab] OR Pakistan*[tiab] OR Palau*[tiab] OR Panama*[tiab] OR "Papua New Guinea"[tiab] OR Paraguay*[tiab] OR Peru*[tiab] OR Philippine*[tiab] OR Filipino*[tiab] OR Poland[tiab] OR Polish[tiab] OR Rwanda*[tiab] OR Samoa*[tiab] OR Sao Tome*[tiab] OR Principe[tiab] OR Senegal*[tiab] OR Seychell*[tiab] OR Sierra Leon*[tiab] OR Solomon Island*[tiab] OR Somali*[tiab] OR South Africa*[tiab] OR Sri Lanka*[tiab] OR "Saint Kitts"[tiab] OR "St Kitts"[tiab] OR "Saint Lucia"[tiab] OR "St Lucia"[tiab] OR "Saint Vincent" [tiab] OR "St Vincent"[tiab] OR Sudan*[tiab] OR Suriname*[tiab] OR Swaziland Or Swazi[tiab] OR Syria*[tiab] OR Tajik*[tiab] OR Tanzania*[tiab] OR Thailand[tiab] OR Thai[tiab] OR "Timor Leste"[tiab] OR Togo*[tiab] OR Tonga*[tiab] OR Trinidad[tiab] OR Trinidadian[tiab] OR Tobago[tiab] OR Tobagonian[tiab] OR Tunisia*[tiab] OR Turky[tiab] OR Turkish[tiab] OR Uganda*[tiab] OR Uruguay*[tiab] OR Vanuat*[tiab] OR Venezuela*[tiab] OR Vietnam*[tiab] OR "West Bank"[tiab] OR Yemen*[tiab] OR Zambia*[tiab] OR Zimbabwe*))

AND

(instrumentation[sh] OR methods[sh] OR "Validation Studies"[pt] OR "Comparative Study"[pt] OR "psychometrics"[MeSH] OR psychometr*[tiab] OR clinimetr*[tw] OR clinometr*[tw] OR "outcome assessment (health care)"[MeSH] OR "outcome assessment"[tiab] OR "outcome measure*"[tw] OR "observer variation"[MeSH] OR "observer variation"[tiab] OR "Health Status Indicators"[Mesh] OR "reproducibility of results"[MeSH] OR reproducib*[tiab] OR "discriminant analysis"[MeSH] OR reliab*[tiab] OR unreliab*[tiab] OR valid*[tiab] OR "coefficient of variation"[tiab] OR coefficient[tiab] OR homogeneity[tiab] OR homogeneous[tiab] OR "internal consistency"[tiab] OR (cronbach*[tiab] AND (alpha[tiab] OR alphas[tiab])) OR (item[tiab] AND (correlation*[tiab] OR selection*[tiab] OR reduction*[tiab])) OR agreement[tw] OR precision[tw] OR imprecision[tw] OR "precise values"[tw] OR test-retest[tiab] OR (test[tiab] AND retest[tiab]) OR (reliab*[tiab] AND

(test[tiab] OR retest[tiab])) OR stability[tiab] OR interrater[tiab] OR inter-rater[tiab] OR intrarater[tiab] OR intra-rater[tiab] OR intertester[tiab] OR inter-tester[tiab] OR intratester[tiab] OR intra-tester[tiab] OR interobserver[tiab] OR inter-observer[tiab] OR intraobserver[tiab] OR intra-observer[tiab] OR intertechnician[tiab] OR inter-technician[tiab] OR intratechnician[tiab] OR intra-technician[tiab] OR interexaminer[tiab] OR inter-examiner[tiab] OR intraexaminer[tiab] OR intra-examiner[tiab] OR interassay[tiab] OR inter-assay[tiab] OR intraassay[tiab] OR intra-assay[tiab] OR interindividual[tiab] OR inter-individual[tiab] OR intraindividual[tiab] OR intra-individual[tiab] OR interparticipant[tiab] OR inter-participant[tiab] OR intraparticipant[tiab] OR intra-participant[tiab] OR kappa[tiab] OR kappa's[tiab] OR kappas[tiab] OR repeatab*[tw] OR ((replicab*[tw] OR repeated[tw]) AND (measure[tw] OR measures[tw] OR findings[tw] OR result[tw] OR results[tw] OR test[tw] OR tests[tw])) OR generaliza*[tiab] OR generalisa*[tiab] OR concordance[tiab] OR (intraclass[tiab] AND correlation*[tiab]) OR discriminative[tiab] OR "known group"[tiab] OR "factor analysis"[tiab] OR "factor analyses"[tiab] OR "factor structure"[tiab] OR "factor structures"[tiab] OR dimension*[tiab] OR subscale*[tiab] OR (multitrait[tiab] AND scaling[tiab] AND (analysis[tiab] OR analyses[tiab])) OR "item discriminant"[tiab] OR "interscale correlation*"[tiab] OR error[tiab] OR errors[tiab] OR "individual variability"[tiab] OR "interval variability"[tiab] OR "rate variability"[tiab] OR (variability[tiab] AND (analysis[tiab] OR values[tiab])) OR (uncertainty[tiab] AND (measurement[tiab] OR measuring[tiab])) OR "standard error of measurement"[tiab] OR sensitiv*[tiab] OR responsive*[tiab] OR (limit[tiab] AND detection[tiab]) OR "minimal detectable concentration"[tiab] OR interpretab*[tiab] OR ((minimal[tiab] OR minimally[tiab] OR clinical[tiab] OR clinically[tiab]) AND (important[tiab] OR significant[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR (small*[tiab] AND (real[tiab] OR detectable[tiab]) AND (change[tiab] OR difference[tiab])) OR "meaningful change"[tiab] OR "ceiling effect"[tiab] OR "floor effect"[tiab] OR "Item response model"[tiab] OR IRT[tiab] OR Rasch[tiab] OR "Differential item functioning"[tiab] OR DIF[tiab] OR "computer adaptive testing"[tiab] OR "item bank"[tiab] OR "cross-cultural equivalence"[tiab] OR "measurement invariance"[tiab] OR "invariance"[tiab] OR assessment[tiab] OR adapt*[tiab])

NOT

("editorial"[Publication Type] OR "comment"[Publication Type] OR "case reports"[Publication Type] OR "congress"[Publication Type] OR "consensus development conference"[Publication Type] OR "interview"[Publication Type] OR "lecture"[Publication Type] OR "legal case"[Publication Type] OR "legislation"[Publication Type] OR "letter"[Publication Type] OR "news"[Publication Type] OR "newspaper article"[Publication Type] OR "Observational Study, Veterinary"[Publication Type] OR "Patient Education Handout"[Publication Type] OR "practice guideline"[Publication Type] OR "preprint"[Publication Type])

MEDLINE
(Child OR Child, Preschool OR child OR Infant OR infant OR paediatric OR pediatric OR toddler) AND (Child Development OR child development OR early child development OR early childhood development OR ECD OR Child Behavior OR child behavior OR child behaviour OR Cognition OR Executive Function OR executive function OR Emotional Intelligence OR attachment OR attention OR cognition OR cognitive development OR cognitive skills OR neurocognitive OR communication OR language OR language skills OR language development OR receptive language OR expressive language OR speech development OR communication skills OR Motor Skills OR motor skills OR fine motor OR gross motor OR prosocial OR socialisation OR socialization OR socioemotion OR socio-emotion OR neurodevelopment OR neurobehavior OR neurobehaviour OR neurobehavioral OR neurobehavioural OR Intelligence OR intelligence OR IQ OR psychomotor OR sensorimotor OR attention OR inhibition OR impulse control OR working memory OR problem solving OR self-regulation OR prosocial OR psychosocial development OR internali OR externali OR empathy) AND ((developing countr OR under developed countr OR lmic OR less developed OR low income OR lower income OR low- and middle-income OR low middle income OR resource poor OR resource constrained OR low resource OR limited resource OR resource limited) AND (Afghan OR Africa OR African OR Algeria OR American Samoa OR Angola OR Argentin OR Bangladesh OR Barbad OR Belorussian OR Beliz OR Benin OR Bhutan OR Bolivia OR Botswan OR Brazil OR Burkina Faso OR Burkinabe OR Burund OR Cambodia OR Cameroon OR Cape Verde OR Cape Verdean OR Central African Republic OR Chad OR Chile OR China OR Chinese OR Colombia OR Comoros OR Comorian OR Congo OR Congolese OR Costa Rica OR Cote d Ivoire OR Ivory Coast OR Ivorian OR Croatia OR Croat OR Cuba OR Djibouti OR Dominica OR Dominican Republic OR Ecuador OR Egypt OR El Salvador OR Salvadorian OR Equatorial Guinea OR Guinean OR Eritrea OR Ethiopia OR Fiji OR Gabon OR Gambia OR Gaza OR Gazan OR Ghana OR Ghanaian OR Grenad OR Guatemala OR Guinea OR Guyan OR Haiti OR Hondura OR Hungar OR India OR Indian OR Indonesia OR Iran OR Iraq OR Jamaica OR Jordan OR Kenya OR Kenyan OR Kiribati OR Korea OR Kyrgy OR Laos OR Laotian OR Lebanon OR Lebanese OR Lesotho OR Liberia OR Libya OR Macedonia OR Madagasca OR Malawi OR Malaysia OR Maldives OR Maldivian OR Mali OR Malian OR Marshall Islands OR Mauritania OR Mauritius OR Mauritian OR Mayotte OR Mexic OR Micronesia OR Moldov OR Mongolia OR Morocc OR Mozambique OR Mozambican OR Myanmar OR Namibia OR Nepal OR Nevis OR Nicaragua OR Niger OR Northern Mariana Islands OR Oman OR Pakistan OR Palau OR Panama OR Papua New Guinea OR Paraguay OR Peru OR Philippine OR Filipino OR Poland OR Polish OR Rwanda OR Samoa OR Sao Tome OR Principe OR Senegal OR Seychell OR Sierra Leon OR Solomon Island OR Somali OR

South Africa OR Sri Lanka OR Saint Kitts OR St Kitts OR Saint Lucia OR St Lucia OR Saint Vincent OR St Vincent OR Sudan OR Suriname OR Swaziland Or Swazi OR Syria OR Tajik OR Tanzania OR Thailand OR Thai OR Timor Leste OR Togo OR Tonga OR Trinidad OR Trinidadian OR Tobago OR Tobagonian OR Tunisia OR Turky OR Turkish OR Uganda OR Uruguay OR Vanuat OR Venezuela OR Vietnam OR West Bank OR Yemen OR Zambia OR Zimbabwe))

AND

(instrumentation OR methods OR Validation Studies OR Comparative Study OR psychometrics OR psychometr OR clinimetr OR clinometr OR outcome assessment OR outcome measure OR observer variation OR observer variation OR Health Status Indicators OR reproducibility of results OR reproducib OR discriminant analysis OR reliab OR unreliab OR valid OR coefficient of variation OR coefficient OR homogeneity OR homogeneous OR internal consistency OR (cronbach AND (alpha OR alphas)) OR (item AND (correlation OR selection OR reduction)) OR agreement OR precision OR imprecision OR precise values OR test-retest OR (test AND retest) OR (reliab AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappas OR repeatab OR ((replicab OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza OR generalisa OR concordance OR (intraclass AND correlation) OR discriminative OR known group OR factor analysis OR factor analyses OR factor structure OR factor structures OR dimension OR subscale OR (multitrait AND scaling AND (analysis OR analyses)) OR item discriminant OR interscale correlation OR error OR errors OR individual variability OR interval variability OR rate variability OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR standard error of measurement OR sensitiv OR responsive OR (limit AND detection) OR minimal detectable concentration OR interpretab OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small AND (real OR detectable) AND (change OR difference)) OR meaningful change OR ceiling effect OR floor effect OR Item response model OR IRT OR Rasch OR Differential item functioning OR DIF OR computer adaptive testing OR item bank OR cross-cultural equivalence)

Embase
(Child OR Child, Preschool OR child OR Infant OR infant OR paediatric OR pediatric OR toddler) AND (Child Development OR child development OR early child development OR early childhood development OR ECD OR Child Behavior OR child behavior OR child behaviour OR Cognition OR Executive Function OR executive function OR Emotional Intelligence OR attachment OR attention OR cognition OR cognitive development OR cognitive skills OR neurocognitive OR communication OR language OR language skills OR language development OR receptive language OR expressive language OR speech development OR communication skills OR Motor Skills OR motor skills OR fine motor OR gross motor OR prosocial OR socialisation OR socialization OR socioemotion OR socio-emotion OR neurodevelopment OR neurobehavior OR neurobehaviour OR neurobehavioral OR neurobehavioural OR Intelligence OR intelligence OR IQ OR psychomotor OR sensorimotor OR attention OR inhibition OR impulse control OR working memory OR problem solving OR self-regulation OR prosocial OR psychosocial development OR internali OR externali OR empathy) AND ((developing countr OR under developed countr OR lmic OR less developed OR low income OR lower income OR low- and middle-income OR low middle income OR resource poor OR resource constrained OR low resource OR limited resource OR resource limited) AND (Afghan OR Africa OR African OR Algeria OR American Samoa OR Angola OR Argentin OR Bangladesh OR Barbad OR Belorussian OR Beliz OR Benin OR Bhutan OR Bolivia OR Botswan OR Brazil OR Burkina Faso OR Burkinabe OR Burund OR Cambodia OR Cameroon OR Cape Verde OR Cape Verdean OR Central African Republic OR Chad OR Chile OR China OR Chinese OR Colombia OR Comoros OR Comorian OR Congo OR Congolese OR Costa Rica OR Cote d Ivoire OR Ivory Coast OR Ivorian OR Croatia OR Croat OR Cuba OR Djibouti OR Dominica OR Dominican Republic OR Ecuador OR Egypt OR El Salvador OR Salvadorian OR Equatorial Guinea OR Guinean OR Eritrea OR Ethiopia OR Fiji OR Gabon OR Gambia OR Gaza OR Gazan OR Ghana OR Ghanaian OR Grenad OR Guatemala OR Guinea OR Guyan OR Haiti OR Hondura OR Hungar OR India OR Indian OR Indonesia OR Iran OR Iraq OR Jamaica OR Jordan OR Kenya OR Kenyan OR Kiribati OR Korea OR Kyrgy OR Laos OR Laotian OR Lebanon OR Lebanese OR Lesotho OR Liberia OR Libya OR Macedonia OR Madagasca OR Malawi OR Malaysia OR Maldives OR Maldivian OR Mali OR Malian OR Marshall Islands OR Mauritania OR Mauritius OR Mauritian OR Mayotte OR Mexic OR Micronesia OR Moldov OR Mongolia OR Morocc OR Mozambique OR Mozambican OR Myanmar OR Namibia OR Nepal OR Nevis OR Nicaragua OR Niger OR Northern Mariana Islands OR Oman OR Pakistan OR Palau OR Panama OR Papua New Guinea OR Paraguay OR Peru OR Philippine OR Filipino OR Poland OR Polish OR Rwanda OR Samoa OR Sao Tome OR Principe OR Senegal OR Seychell OR Sierra Leon OR Solomon Island OR Somali OR

South Africa OR Sri Lanka OR Saint Kitts OR St Kitts OR Saint Lucia OR St Lucia OR Saint Vincent OR St Vincent OR Sudan OR Suriname OR Swaziland Or Swazi OR Syria OR Tajik OR Tanzania OR Thailand OR Thai OR Timor Leste OR Togo OR Tonga OR Trinidad OR Trinidadian OR Tobago OR Tobagonian OR Tunisia OR Turky OR Turkish OR Uganda OR Uruguay OR Vanuat OR Venezuela OR Vietnam OR West Bank OR Yemen OR Zambia OR Zimbabwe))

AND

(instrumentation OR methods OR Validation Studies OR Comparative Study OR psychometrics OR psychometr OR clinimetr OR clinometr OR outcome assessment OR outcome measure OR observer variation OR observer variation OR Health Status Indicators OR reproducibility of results OR reproducib OR discriminant analysis OR reliab OR unreliab OR valid OR coefficient of variation OR coefficient OR homogeneity OR homogeneous OR internal consistency OR (cronbach AND (alpha OR alphas)) OR (item AND (correlation OR selection OR reduction)) OR agreement OR precision OR imprecision OR precise values OR test-retest OR (test AND retest) OR (reliab AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappas OR repeatab OR ((replicab OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza OR generalisa OR concordance OR (intraclass AND correlation) OR discriminative OR known group OR factor analysis OR factor analyses OR factor structure OR factor structures OR dimension OR subscale OR (multitrait AND scaling AND (analysis OR analyses)) OR item discriminant OR interscale correlation OR error OR errors OR individual variability OR interval variability OR rate variability OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR standard error of measurement OR sensitiv OR responsive OR (limit AND detection) OR minimal detectable concentration OR interpretab OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small AND (real OR detectable) AND (change OR difference)) OR meaningful change OR ceiling effect OR floor effect OR Item response model OR IRT OR Rasch OR Differential item functioning OR DIF OR computer adaptive testing OR item bank OR cross-cultural equivalence)

PsycInfo
(Child OR Child, Preschool OR child OR Infant OR infant OR paediatric OR pediatric OR toddler) AND (Child Development OR child development OR early child development OR early childhood development OR ECD OR Child Behavior OR child behavior OR child behaviour OR Cognition OR Executive Function OR executive function OR Emotional Intelligence OR attachment OR attention OR cognition OR cognitive development OR cognitive skills OR neurocognitive OR communication OR language OR language skills OR language development OR receptive language OR expressive language OR speech development OR communication skills OR Motor Skills OR motor skills OR fine motor OR gross motor OR prosocial OR socialisation OR socialization OR socioemotion OR socio-emotion OR neurodevelopment OR neurobehavior OR neurobehaviour OR neurobehavioral OR neurobehavioural OR Intelligence OR intelligence OR IQ OR psychomotor OR sensorimotor OR attention OR inhibition OR impulse control OR working memory OR problem solving OR self-regulation OR prosocial OR psychosocial development OR internali OR externali OR empathy) AND ((developing countr OR under developed countr OR lmic OR less developed OR low income OR lower income OR low- and middle-income OR low middle income OR resource poor OR resource constrained OR low resource OR limited resource OR resource limited) AND (Afghan OR Africa OR African OR Algeria OR American Samoa OR Angola OR Argentin OR Bangladesh OR Barbad OR Belorussian OR Beliz OR Benin OR Bhutan OR Bolivia OR Botswan OR Brazil OR Burkina Faso OR Burkinabe OR Burund OR Cambodia OR Cameroon OR Cape Verde OR Cape Verdean OR Central African Republic OR Chad OR Chile OR China OR Chinese OR Colombia OR Comoros OR Comorian OR Congo OR Congolese OR Costa Rica OR Cote d Ivoire OR Ivory Coast OR Ivorian OR Croatia OR Croat OR Cuba OR Djibouti OR Dominica OR Dominican Republic OR Ecuador OR Egypt OR El Salvador OR Salvadorian OR Equatorial Guinea OR Guinean OR Eritrea OR Ethiopia OR Fiji OR Gabon OR Gambia OR Gaza OR Gazan OR Ghana OR Ghanaian OR Grenad OR Guatemala OR Guinea OR Guyan OR Haiti OR Hondura OR Hungar OR India OR Indian OR Indonesia OR Iran OR Iraq OR Jamaica OR Jordan OR Kenya OR Kenyan OR Kiribati OR Korea OR Kyrgy OR Laos OR Laotian OR Lebanon OR Lebanese OR Lesotho OR Liberia OR Libya OR Macedonia OR Madagasca OR Malawi OR Malaysia OR Maldives OR Maldivian OR Mali OR Malian OR Marshall Islands OR Mauritania OR Mauritius OR Mauritian OR Mayotte OR Mexic OR Micronesia OR Moldov OR Mongolia OR Morocc OR Mozambique OR Mozambican OR Myanmar OR Namibia OR Nepal OR Nevis OR Nicaragua OR Niger OR Northern Mariana Islands OR Oman OR Pakistan OR Palau OR Panama OR Papua New Guinea OR Paraguay OR Peru OR Philippine OR Filipino OR Poland OR Polish OR Rwanda OR Samoa OR Sao Tome OR Principe OR Senegal OR Seychell OR Sierra Leon OR Solomon Island OR Somali OR

South Africa OR Sri Lanka OR Saint Kitts OR St Kitts OR Saint Lucia OR St Lucia OR Saint Vincent OR St Vincent OR Sudan OR Suriname OR Swaziland Or Swazi OR Syria OR Tajik OR Tanzania OR Thailand OR Thai OR Timor Leste OR Togo OR Tonga OR Trinidad OR Trinidadian OR Tobago OR Tobagonian OR Tunisia OR Turky OR Turkish OR Uganda OR Uruguay OR Vanuat OR Venezuela OR Vietnam OR West Bank OR Yemen OR Zambia OR Zimbabwe))

AND

(instrumentation OR methods OR Validation Studies OR Comparative Study OR psychometrics OR psychometr OR clinimetr OR clinometr OR outcome assessment OR outcome measure OR observer variation OR observer variation OR Health Status Indicators OR reproducibility of results OR reproducib OR discriminant analysis OR reliab OR unreliab OR valid OR coefficient of variation OR coefficient OR homogeneity OR homogeneous OR internal consistency OR (cronbach AND (alpha OR alphas)) OR (item AND (correlation OR selection OR reduction)) OR agreement OR precision OR imprecision OR precise values OR test-retest OR (test AND retest) OR (reliab AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappas OR repeatab OR ((replicab OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza OR generalisa OR concordance OR (intraclass AND correlation) OR discriminative OR known group OR factor analysis OR factor analyses OR factor structure OR factor structures OR dimension OR subscale OR (multitrait AND scaling AND (analysis OR analyses)) OR item discriminant OR interscale correlation OR error OR errors OR individual variability OR interval variability OR rate variability OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR standard error of measurement OR sensitiv OR responsive OR (limit AND detection) OR minimal detectable concentration OR interpretab OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small AND (real OR detectable) AND (change OR difference)) OR meaningful change OR ceiling effect OR floor effect OR Item response model OR IRT OR Rasch OR Differential item functioning OR DIF OR computer adaptive testing OR item bank OR cross-cultural equivalence)

SciELO
(Child OR Child, Preschool OR child OR Infant OR infant OR paediatric OR pediatric OR toddler) AND (Child Development OR child development OR early child development OR early childhood development OR ECD OR Child Behavior OR child behavior OR child behaviour OR Cognition OR Executive Function OR executive function OR Emotional Intelligence OR attachment OR attention OR cognition OR cognitive development OR cognitive skills OR neurocognitive OR communication OR language OR language skills OR language development OR receptive language OR expressive language OR speech development OR communication skills OR Motor Skills OR motor skills OR fine motor OR gross motor OR prosocial OR socialisation OR socialization OR socioemotion OR socio-emotion OR neurodevelopment OR neurobehavior OR neurobehaviour OR neurobehavioral OR neurobehavioural OR Intelligence OR intelligence OR IQ OR psychomotor OR sensorimotor OR attention OR inhibition OR impulse control OR working memory OR problem solving OR self-regulation OR prosocial OR psychosocial development OR internali OR externali OR empathy) AND (instrumentation OR methods OR Validation Studies OR Comparative Study OR psychometrics OR psychometr OR clinimetr OR clinometr OR outcome assessment OR outcome measure OR observer variation OR observer variation OR Health Status Indicators OR reproducibility of results OR reproducib OR discriminant analysis OR reliab OR unreliab OR valid OR coefficient of variation OR coefficient OR homogeneity OR homogeneous OR internal consistency OR (cronbach AND (alpha OR alphas)) OR (item AND (correlation OR selection OR reduction)) OR agreement OR precision OR imprecision OR precise values OR test-retest OR (test AND retest) OR (reliab AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappas OR repeatab OR ((replicab OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza OR generalisa OR concordance OR (intraclass AND correlation) OR discriminative OR known group OR factor analysis OR factor analyses OR “factor structure” OR “factor structures” OR dimension OR subscale OR (multitrait AND scaling AND (analysis OR analyses)) OR “item discriminant” OR “interscale correlation” OR error OR errors OR “individual variability” OR "interval variability" OR "rate variability" OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR “sensitive” OR “responsive” OR (limit AND detection) OR "minimal

detectable concentration" OR interpretab OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR ceiling effect OR floor effect OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR item bank OR "cross-cultural equivalence")

BVS
(Child OR Child, Preschool OR child OR Infant OR infant OR paediatric OR pediatric OR toddler) AND (Child Development OR child development OR early child development OR early childhood development OR ECD OR Child Behavior OR child behavior OR child behaviour OR Cognition OR Executive Function OR executive function OR Emotional Intelligence OR attachment OR attention OR cognition OR cognitive development OR cognitive skills OR neurocognitive OR communication OR language OR language skills OR language development OR receptive language OR expressive language OR speech development OR communication skills OR Motor Skills OR motor skills OR fine motor OR gross motor OR prosocial OR socialisation OR socialization OR socioemotion OR socio-emotion OR neurodevelopment OR neurobehavior OR neurobehaviour OR neurobehavioral OR neurobehavioural OR Intelligence OR intelligence OR IQ OR psychomotor OR sensorimotor OR attention OR inhibition OR impulse control OR working memory OR problem solving OR self-regulation OR prosocial OR psychosocial development OR internali OR externali OR empathy) AND (instrumentation OR methods OR Validation Studies OR Comparative Study OR psychometrics OR psychometr OR clinimetr OR clinometr OR outcome assessment OR outcome measure OR observer variation OR observer variation OR Health Status Indicators OR reproducibility of results OR reproducib OR discriminant analysis OR reliab OR unreliab OR valid OR coefficient of variation OR coefficient OR homogeneity OR homogeneous OR internal consistency OR (cronbach AND (alpha OR alphas)) OR (item AND (correlation OR selection OR reduction)) OR agreement OR precision OR imprecision OR precise values OR test-retest OR (test AND retest) OR (reliab AND (test OR retest)) OR stability OR interrater OR inter-rater OR intrarater OR intra-rater OR intertester OR inter-tester OR intratester OR intra-tester OR interobserver OR inter-observer OR intraobserver OR intra-observer OR intertechnician OR inter-technician OR intratechnician OR intra-technician OR interexaminer OR inter-examiner OR intraexaminer OR intra-examiner OR interassay OR inter-assay OR intraassay OR intra-assay OR interindividual OR inter-individual OR intraindividual OR intra-individual OR interparticipant OR inter-participant OR intraparticipant OR intra-participant OR kappa OR kappas OR repeatab OR ((replicab OR repeated) AND (measure OR measures OR findings OR result OR results OR test OR tests)) OR generaliza OR generalisa OR concordance OR (intraclass AND correlation) OR discriminative OR known group OR factor analysis OR factor analyses OR “factor structure” OR “factor structures” OR dimension OR subscale OR (multitrait AND scaling AND (analysis OR analyses)) OR “item discriminant” OR “interscale correlation” OR error OR errors OR “individual variability” OR "interval variability" OR "rate variability" OR (variability AND (analysis OR values)) OR (uncertainty AND (measurement OR measuring)) OR "standard error of measurement" OR “sensitive” OR “responsive” OR (limit AND detection) OR "minimal

detectable concentration" OR interpretab OR ((minimal OR minimally OR clinical OR clinically) AND (important OR significant OR detectable) AND (change OR difference)) OR (small AND (real OR detectable) AND (change OR difference)) OR "meaningful change" OR ceiling effect OR floor effect OR "Item response model" OR IRT OR Rasch OR "Differential item functioning" OR DIF OR "computer adaptive testing" OR item bank OR "cross-cultural equivalence")

Supplemental Table 2 Definitions of the seven early childhood domains of interest

Domain	Definition
Cognitive	Cognitive development includes abilities such as general intellectual ability, problem solving, reasoning, visual-spatial ability, memory, learning, etc.
Language	Language development includes the ability to understand and express verbal communication, and vocabulary development and acquisition.
Motor	Moto development includes gross motor development (the ability to control and coordinate gross movements of the legs and arms (e.g., jumping, throwing)) and fine motor development (the ability to control and coordinate fine movements of the fingers and toes.
Socio-emotional/temperament	Socio-emotional development is the ability to regulate emotional responses and social interactions. It includes behaviour problems, social competency, emotional competency, and self-regulation. Temperament includes biological influences on the experience and expression of emotion: extraversion (positive affect, activity level, impulsivity, risk-taking), negative affectivity (fear, anger, sadness, discomfort) and effortful control (attention shifting and focusing, perceptual sensitivity, inhibitory and activational control). Socio-emotional development and temperament are often overlapping constructs, particularly in the early years.
Attention/Executive function	Executive function, including attention, entails international control over behaviour and cognition. It covers inhibitory control, cognitive flexibility, planning, and working memory.
Personal-social/adaptive	Personal-social and adaptive behaviour comprises the ability to perform daily-life and self-help skills, such as feeding, dressing, toilet training, recognising others, interacting with others, and adjusting to new situations.
Academic/pre-academic	Preacademic skills are skills needed to learn reading and math, such as counting and letters.

Note: Definitions based on (Boggs et al., 2019; Fernald et al., 2017).

Supplemental Table 3 Number of included articles by type of psychometric evidence provided

Type of article	Total number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance	
		Test-retest	Inter-rater	Internal consistency	Content	Concurrent	Convergent	Predictive	Structural	Invariance over countries	Invariance over other groups
All types of articles	182	73	63	117	66	81	74	31	72	7	10
Development articles	37	15	15	22	27	21	13	4	11	2	4
Adaptation articles	145	58	48	95	39	60	61	27	61	5	6

Supplemental Table 4 List of countries and assessment tools used in included articles

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Afghanistan	International Development and Early Learning Assessment (IDELA)	1	0	0	0	0	0	0	0	1	0	0	Halpin 2019
Argentina	Bayley Infant Neurodevelopmental Screener (BINS)	1	1	1	0	0	0	0	0	0	0	0	McCarthy 2012
Argentina	Escala de Habildades Sociales (Scale of Social Abilities)	1	0	0	1	0	0	0	0	1	0	0	Lacunza 2009
Bangladesh	Ages & Stages Questionnaire Inventory (ASQ:I)	1	0	1	1	0	1	1	0	0	0	0	Pitchik 2023
Bangladesh	Bayley Scales of Infant Development, Second Edition (BSID-II)	1	0	0	0	1	0	0	0	0	0	0	Parveen 2014
Bangladesh	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	1	0	0	0	1	0	0	Pendergast 2018
Bangladesh	Dimensions of Mastery Questionnaire (DMQ-18)	1	1	1	1	1	1	0	0	1	0	0	Shaoli 2019
Bangladesh	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	Pisani 2018

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Bangladesh	Rapid Neurodevelopmental Assessment (RNDA)	2	0	2	0	1	2	0	1	0	0	0	Khan 2010; Khan 2013
Bangladesh	The language inventory	1	1	0	0	1	1	1	1	0	0	0	Hamadani 2010
Bangladesh	WHO Gross Motor Milestones	1	0	1	0	0	1	1	1	0	0	0	Hamadani 2013
Benin	Mullen Scales of Early Learning (MSEL)	2	0	1	0	0	1	1	1	0	0	0	Koura 2013; Boivin 2021
Bhutan	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	Pisani 2018
Bolivia	Bayley Infant Neurodevelopmental Screener (BINS)	1	1	1	0	0	0	0	0	0	0	0	McCarthy 2012
Bolivia	International Development and Early Learning Assessment (IDELA)	1	0	0	0	0	0	0	0	1	0	0	Halpin 2019
Brazil	Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	1	0	0	0	0	0	0	0	1	0	0	Anunciação 2019
Brazil	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	0	1	0	0	0	0	1	0	1	Filgueiras 2013
Brazil	Ages & Stages Questionnaires: Social Emotional (ASQ:SE-BR)	1	0	0	0	0	0	0	0	1	0	0	Chen 2017

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Brazil	Alberta Infant Motor Scale (AIMS)	4	3	3	1	1	3	0	1	0	0	0	Saccani 2016; Siegle 2018; de Albuquerque 2018; Valentini 2012
Brazil	Battelle Developmental Inventory Second Edition, BDI-2	1	0	0	0	1	0	0	0	0	0	0	de Albuquerque 2022
Brazil	Bayley Infant Neurodevelopment Screener (BINS)	2	1	1	1	1	1	0	1	1	0	0	Guedes 2011; McCarthy 2012
Brazil	Bayley Scales of Infant Development, Third Edition (BSID-III)	2	1	0	1	1	1	0	0	2	0	0	Pendergast 2018; Madaschi 2016
Brazil	Caregiver Reported Early Development Instruments (CREDI), short-form	1	1	0	1	1	1	1	0	0	0	0	Altafim 2018
Brazil	Denver Developmental Screening Test (Denver-II)	1	0	1	1	1	0	1	0	1	0	0	Lopez Boo 2020
Brazil	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	1	Sincovich 2019
Brazil	Motor Development Scale	1	0	0	1	0	0	0	0	1	0	0	Okuda 2020
Brazil	Oxford Neurodevelopment Assessment (OX-NDA)	1	1	1	1	0	1	0	0	0	0	0	Fernandes 2022

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Brazil	Preschool and Kindergarten Behavior Scale (PKBS)	1	0	0	1	0	0	0	0	1	0	0	Dias 2011
Brazil	Questionário para Avaliação do Desenvolvimento Infantil (QAD-PIPAS)	1	0	0	0	1	0	0	0	0	0	0	Venancio 2020
Brazil	Screening for Identification of Oral Language Difficulties by Preschool Teachers (SIOLD)	1	0	0	0	0	0	0	0	1	0	0	Puglisi 2020
Brazil	Self-ordered Pointing Task (SOPT)	1	0	0	0	0	1	1	0	0	0	0	Sallum 2017
Brazil	Test of Gross Motor Development (TGMD-2)	1	0	1	1	0	0	1	0	1	0	0	Bandeira 2020
Cambodia	Angkor Hospital for Children Developmental Milestone Assessment Tool (AHC DMAT)	1	0	0	0	1	0	0	0	0	0	0	Ngoun 2012
Cambodia	Developmental Milestone Chart (DMC)	1	0	0	0	1	0	0	0	0	0	0	Scherzer 2009
Cameroon	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	0	0	0	1	0	1	0	Lohaus 2014
China	Ages & Stages Questionnaire Inventory (ASQ:I)	1	1	0	1	1	1	0	0	0	0	0	Xie 2017

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
China	Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	1	1	1	1	1	0	0	0	1	0	0	Bian 2017
China	Ages & Stages Questionnaire: Social-Emotional, First Edition (ASQ:SE-1)	1	1	1	1	0	1	0	0	0	0	0	Xie 2019 (Journal of Child & Dev Studies)
China	Ages & Stages Questionnaire-3 (ASQ-3)	2	1	1	1	0	2	1	0	1	0	0	Yue 2019; Bian 2012
China	Caregiver Reported Early Development Instruments (CREDI), long-form	1	0	0	1	0	1	1	0	0	0	0	Li 2020
China	Child Behavior Checklist/1.5-5 (CBCL/1.5-5) and the Caregiver-Teacher Report Form (C-TRF)	1	0	0	0	0	0	0	0	1	0	0	Liu 2011
China	Child Social Preference Scale	1	0	0	1	0	0	0	1	1	0	1	Li 2016
China	Chinese Preschool Readiness Scale (CPRS)	1	0	0	1	1	1	0	0	1	0	0	Xie 2019 (Early Educ & Dev)
China	Chinese version of Urban Infant-Toddler Social and Emotional Assessment	1	1	1	1	1	1	0	0	1	0	0	Jianduan 2009
China	Chinese Vocabulary Checklist	1	0	0	0	0	0	1	0	0	0	0	Hao 2008

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
China	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	1	Sincovich 2019
China	Emotional and Behavioral Screener (EBS)	1	0	1	1	0	1	0	0	1	0	0	Zhang 2018
China	Movement Assessment Battery for Children- second edition (MBAC-2)	1	1	1	1	1	1	0	0	1	0	0	Hua 2013
China	Preschool Learning Behavior Scale (PLBS)	1	0	0	1	0	1	0	0	1	0	0	Wu 2019 (JPA)
China	Social Skills domain of System-Rating Scales (SSIS-RS)	1	0	0	1	0	0	1	0	1	0	0	Wu 2019 (PS)
China	The Chinese Early Language and Communication Questionnaire	1	0	0	1	0	1	0	0	0	0	0	Zhang 2021
China	The Chinese Intelligence Scale for Young Children	1	0	0	0	0	0	0	0	1	0	1	Guo 2009
Colombia	Ages & Stages Questionnaire-3 (ASQ-3)	2	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	Battelle Developmental Inventory Second Edition, BDI-2	2	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	Bayley Infant Neurodevelopmental Screener (BINS)		1	1	0	0	0	0	0	0	0	0	McCarthy 2012

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Colombia	Bayley Scales of Infant Development, Third Edition (BSID-III)	2	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	Denver Developmental Screening Test (Denver-II)	2	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	MacArthur-Bates Communicative Development Inventories CDI	1	0	0	1	1	0	1	0	0	0	0	Lara Díaz 2011
Colombia	MacArthur-Bates Communicative Development Inventories Short-Forms I vocabulary checklist	2	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	MacArthur-Bates Communicative Development Inventories Short-Forms II vocabulary checklist	2	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Colombia	Preschool Language Scale-3 (PLS- 3)	1	0	0	1	0	0	0	0	1	0	0	Romero 2013
Colombia	WHO Gross Motor Milestones	2	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
Costa Rica	EDIN II (Evaluación del Desarrollo Integral de la niña y el niño)	1	0	0	0	1	0	1	0	0	0	0	Cordero 2019

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Côte d'Ivoire	Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	1	0	0	1	0	0	1	0	1	0	0	Barthel 2021
Côte d'Ivoire	Kilifi Developmental Inventory	1	0	1	1	0	0	0	1	1	0	0	Barthel 2022
Dominican Republic	Malawi Developmental Assessment Tool (MDAT), renamed Dominican Child Development Screening Tool	1	0	1	1	0	0	1	0	0	0	0	Sánchez-Vincitore 2019
Ecuador	Bayley Infant Neurodevelopmental Screener (BINS)	1	1	1	0	0	0	0	0	0	0	0	McCarthy 2012
Egypt	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	Pisani 2018
Ethiopia	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	1	1	0	0	1	1	1	0	0	Hanlon 2016
Ethiopia	Denver Developmental Screening Test (Denver-II)	1	1	1	0	1	0	0	0	0	0	0	Abessa 2016
Ethiopia	Global Scales for Early Development	1	0	1	0	1	0	0	0	0	0	0	Yibeltal 2025

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Ethiopia	International Development and Early Learning Assessment (IDELA)	3	1	1	1	1	2	2	0	2	1	0	Halpin 2019; Pisani 2018; Wolf 2017
Georgia	Ages & Stages Questionnaire-3 (ASQ-3)	1	1	0	1	1	0	1	0	0	0	0	Zirakashvili 2018
Ghana	Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	1	0	0	1	0	0	1	0	1	0	0	Barthel 2021
Ghana	International Development and Early Learning Assessment (IDELA)	1	0	0	1	0	0	0	1	1	0	0	Pisani 2022
Ghana	Kilifi Developmental Inventory	1	0	1	1	0	0	0	1	1	0	0	Barthel 2022
Ghana	Rapid Assessment of Cognitive and Emotional Regulation (RACER)	1	1	1	0	0	1	1	0	0	0	0	Yuan 2022
Ghana	Vocabulary Checklists	1	0	0	0	0	1	0	1	0	0	0	Prado 2018
Grenada	Grenada Learning and Memory Scale	1	1	1	1	1	1	1	0	1	0	0	Blackmon 2024
Grenada	INTER- GROWTH-21st Project Neurodevelopment Assessment (INTER-NDA)	1	1	1	1	0	0	0	0	0	0	0	Waechter 2022
Guatemala	Ages & Stages Questionnaire-3 (ASQ-3)	1	1	0	1	0	1	0	1	0	0	0	Colbert 2021

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Guatemala	Mullen Scales of Early Learning (MSEL)	1	1	1	1	0	0	1	1	0	0	0	Colbert 2020
Guatemala	Rapid Neurodevelopmental Assessment (RNDA)	1	0	1	0	1	1	1	0	0	0	0	Thompson 2015
India	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	1	1	1	0	0	0	0	0	0	Kvestad 2013
India	Bayley Scales of Infant Development, Second Edition (BSID-II)	1	0	0	0	0	1	0	0	0	0	0	Biasini 2015
India	Bayley Scales of Infant Development, Third Edition (BSID-III)	2	0	0	0	1	0	0	1	1	0	0	Pendergast 2018; Koshy 2024
India	Caregiver Reported Early Development Instruments (CREDI), long-form	1	0	0	1	0	1	1	0	0	0	0	Alderman 2021
India	DEvelopmental assessment on an E-Platform (DEEP)	2	0	0	0	1	1	0	0	0	0	0	Bhavnani 2019; Mukherjee 2020
India	Developmental Assessment Tool for Anganwadis (DATA)	1	0	0	1	1	0	0	0	1	0	0	Nair 2009
India	Developmental Assessment Tool for Anganwadis (DATA-II)	1	0	0	1	1	0	0	0	1	0	0	Nair 2012
India	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	Pisani 2018

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
India	Parents' Evaluation of Developmental Status (PEDS)	1	0	0	0	0	1	0	0	0		0	Mukherjee 2022
India	Screening Test of Early Language Development-Test version (STELD-T)	1	0	0	1	0	1	0	0	0		0	Pathak 2023
India	Bracken School Readiness Assessment, Third Edition (BSRA-3)	1	0	0	1	1	0	0	0	0		0	Shah 2013
India	Vineland Adaptive Behaviour Scale (VABS II)	1	1	1	0	0	0	1	0	0		0	Kumar 2016
Indonesia	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	1	1	1	1	0	0	0		0	Yunilda 2022
Indonesia	Early Development Instrument	1	0	1	1	1	1	0	1	0		0	Brinkman 2017
Indonesia	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0		0	Pisani 2018
Iran	Ages & Stages Questionnaire (ASQ)	1	0	1	1	1	0	0	0	1		0	Vameghi 2013
Iran	Bayley Scales of Infant and Toddler Development (BSID)	1	1	1	1	1	0	1	0	1		0	Azari 2017
Iran	Childhood Nonverbal Communication Scale (CNCS)	1	0	0	1	1	1	1	0	1	0	0	Oryadi-Zanjani 2020
Iran	Children's Behavior Questionnaire (CBQ)	1	1	0	1	1	1	1	0	1	0	0	Golmohammadi 2022

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Iran	Elicited Imitation Test (EIT)	1	0	1	0	0	1	0	0	0		1	Mehrani 2018
Iran	Farsi Narrative Norms Instrument	1	1	1	0	1	0	0	0	0		0	Soleymani 2016
Iran	Infant Neurological International Battery	1	0	1	0	0	1	0	0	0		0	Soleimani 2007
Jamaica	Intelligibility in Context Scale (ICS) / ICS-Jamaican Creole	1	1	1	1	0	1	1	0	1		0	Washington 2017
Kenya	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	1	0	1	1	0	1	0	1		0	McHenry 2021
Kenya	Child behaviour checklist (CBCL)	1	1	0	1	0	0	1	0	1		0	Kariuki 2016
Kenya	Communicative Development Inventory (CDI)	1	1	0	1	0	1	0	0	0		0	Alcock 2015
Kenya	Developmental Milestones Checklist (DMC)	1	1	0	1	1	1	1	0	1		0	Abubakar 2010
Kenya	Executive Function Touch (EF Touch)	2	0	1	1	1	1	2	0	1		0	Willoughby 2019 (Child Neuropsych); Willoughby 2019 (Dev Science)
Kenya	Global Scales of Early Development	1	0	0	0	1	1	1	0	0	0	0	Jeong 2025
Kenya	Kilifi Developmental Inventory	1	1	1	1	1	1	1	0	0	0	0	Abubakar 2008

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Kenya	Receptive and expressive vocabulary test battery (English, Luo and Swahili)	1	0	1	1	0	1	1	1	0	0	0	Knauer 2019
Kiribati	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	1	Sincovich 2019
Laos	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	1	Sincovich 2019
Laos	MELQO Measure of Development and Learning Outcomes	1	0	0	0	0	0	1	0	1	0	1	Gomez 2022
Malawi	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	Pisani 2018
Malawi	Malawi Developmental Assessment Tool (MDAT)	1	1	1	0	1	0	1	0	0	0	0	Gladstone 2010
Malawi	Vocabulary Checklists	1	0	0	0	0	1	0	1	0	0	0	Prado 2018
Malaysia	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	1	0	0	0	0	1	0	0	Zakaria 2012
Malaysia	Malay Preschool Language Assessment Tool (MPLAT)	1	1	1	1	1	0	0	0	0	0	0	Razak 2010

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Mali	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0		0	Pisani 2018
Mexico	Ages & Stages Questionnaire-3 (ASQ-3)	1	0	1	1	0	0	0	0	1		0	Ortiz-Leon 2018
Mexico	Early Childhood Development Index 2030	1	0	0	1	1	1	0	0	1		1	Halpin 2024
Mexico	Spanish-language MacArthur–Bates Communicative Development Inventories short forms (S-CDI SFI and SFII)	1	1	0	0	0	0	1	0	0		0	Jackson-Maldonado 2013
Mongolia	Mongolian Rapid Baby Scale (MORBAS)	1	0	0	0	1	1	0	0	0		0	Dagvadorj 2015
Mozambique,	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0		0	Pisani 2018
Nepal	Ages & Stages Questionnaire-3 (ASQ-3)	2	0	2	1	1	0	1	1	0		0	Shrestha 2019; Shrestha 2024
Nepal	Bayley Scales of Infant Development, Third Edition (BSID-III)	3	0	2	1	1	0	0	0	1	1	0	Pendergast 2018; Ranjitkar 2018; Manandhar 2016

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Nigeria	Early Childhood Development Index (ECDI)	1	0	0	1	1	0	0	0	1	0	0	Enelamah 2024
Nepal	The Test of Infant Motor Performance (TIMP)	1	0	1	1	0	1	0	1	0	0	0	Kvestad 2023
Pakistan	Bayley Scales of Infant Development, Second Edition (BSID-II)	1	0	0	0	0	1	0	0	0	0	0	Biasini 2015
Pakistan	Bayley Scales of Infant Development, Third Edition (BSID-III)	2	0	0	0	1	1	0	1	1	0	0	Pendergast 2018; Rasheed 2023
Pakistan	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	Pisani 2018
Pakistan	Executive function battery	1	0	0	1	1	0	1	1	1	0	0	Obradovic 2019
Pakistan	Executive Function Touch (EF Touch)	1	0	0	0	0	0	1	0	1	0	0	Toor 2022
Pakistan	MacArthur Communicative Development Inventory Level II	1	1	1	1	1	1	0	0	0	0	0	Rasheed 2017
Pakistan	Strengths and Difficulties Questionnaire (SDQ)	1	1	1	1	1	1	1	0	1	0	0	Finch 2018
Pakistan	Wechsler Preschool and Primary Scale of Intelligence, thid edition (WPPSI-III)	1	0	1	1	1	0	1	0	1	0	0	Rasheed 2018

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Palestine	Early Childhood Development Index 2030	1	0	0	1	1	1	0	0	1	1	1	Halpin 2024
Peru	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	1	0	0	0	1	0	0	Pendergast 2018
Peru	Child Behavior Scale (CBS)	1	0	0	1	0	0	0	0	1	0	0	Meyer 2011
Peru	Escala de Conductas de Aprendizaje Preescolar (ECAP), a Spanish translation of the Preschool Learning Behaviors Scale (PLBS)	1	0	0	0	0	0	0	0	1	0	0	Hahn 2009
Rwanda	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	Pisani 2018
Samoa	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	1	Sincovich 2019
Senegal	Milestones Checklist; Vocabulatory Inventory	1	0	0	1	1	1	1	0	0	0	0	Weber 2018
Sierra Leone	Saccadic reaction time (SRTs)	1	1	0	0	0	0	1	0	0	0	0	Leppänen 2022
South Africa	Ages & Stages Questionnaire (ASQ)	1	0	0	0	1	0	0	0	0	0	0	Bornman 2010
South Africa	Ages & Stages Questionnaire-3 (ASQ-3)	2	0	0	2	0	0	1	0	1	0	0	van Heerden 2017; Hsiao 2016

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
South Africa	Bayley Infant Neurodevelopmental Screener (BINS)	1	0	0	1	0	0	1	0	1	0	0	Rodriguez 2020
South Africa	Bayley Scales of Infant Development, Third Edition (BSID-III)	1	0	0	0	1	0	0	0	1	0	0	Pendergast 2018
South Africa	Early Learning Outcomes Measure (ELOM)	1	0	1	1	0	0	0	0	1	0	1	Snelling 2019
South Africa	Little Developmental Coordination Disorder Questionnaire (Little DCDQ)	1	0	0	1	0	1	0	0	0	0	0	Venter 2015
South Africa	Mullen Scales of Early Learning (MSEL)	1	0	0	0	1	0	0	0	0	0	0	Bornman 2010
South Africa	Parents' Evaluation of Developmental status (PEDS)	1	0	0	0	0	1	0	0	0	0	0	Abdoola 2021
South Africa	The infant gross motor screening test (IGMST)	1	1	1	0	1	1	0	0	0	0	0	Hilburn 2011
South Africa	The Road to Health Booklet (RTHB)	1	0	0	0	0	1	0	0	0	0	0	van der Linde 2015
Sri Lanka	Child Behaviour Assessment Instrument (CBAI)	1	1	0	1	1	1	0	0	0	0	0	Samarakkody 2010

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Sri Lanka	Denver Developmental Screening Test for Sri Lankan Children (DDST-SL)	1	0	1	0	0	0	0	0	0	0	0	Wijedasa 2012
Sri Lanka	Ragama Early Assessment for Children (REACh)	1	0	1	1	1	1	0	0	0	0	0	Caldera 2023
Tanzania	Caregiver Reported Early Development Instruments (CREDI)	1	1	1	1	1	1	1	0	0	0	0	McCoy 2017
Thailand	Parents' Evaluation of Developmental Status (PEDS)	1	0	0	0	0	1	0	0	0	0	0	Chunsuwan 2016
Thailand	Thai Speech and Language Test (TSLT)	1	0	1	1	1	0	0	0	0	0	0	Prathanee 2008
The Gambia	Mullen Scales of Early Learning (MSEL)	1	0	0	0	0	0	1	1	0	0	0	Milosavljevic 2019
Tonga	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	1	Sincovich 2019
Turkey	Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	1	1	1	1	0	1	0	0	0	0	0	Kucuker 2011
Turkey	Bayley Infant Neurodevelopmental Screener (BINS)	1	0	0	0	0	0	0	1	0	0	0	Soysal 2014
Turkey	Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	1	1	1	1	0	1	0	0	0	0	0	Karabekiroglu 2009

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Turkey	Emotional Regulation Checklist	1	1	0	1	1	1	0	0	1	0	0	Danişman 2016
Turkey	Kaufman Survey of Early Academic and Language Skills (K-SEALS)	1	1	1	1	1	0	0	0	1	0	0	Uyanik 2014
Turkey	Strengths and Difficulties Questionnaire (SDQ)	1	0	0	1	0	1	1	0	0	0	0	Dursun 2020
Tuvalu	Early Human Capability Index (eHCI)	1	0	0	1	0	0	0	0	1	0	1	Sincovich 2019
Uganda	Bayley Scales of Infant Development, bead threding task	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	British Ability Scales – 3rd ed., block design task	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	International Development and Early Learning Assessment (IDELA)	1	0	0	0	0	0	0	0	1	0	0	Halpin 2019
Uganda	Kilifi Developmental Inventory, coin box task	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	Kilifi Picture Vocabulary Test	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	Modified Fagan Test for Infant Intelligence (FTII)	1	0	0	0	0	0	0	1	0	0	0	Familiar-Lopez 2022

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Uganda	Movement Assessment Battery for Children, balancing on one leg	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	NEPSY verbal fluence task	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	NEPSY, knock tap game	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	NEPSY, sentence repetition task	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	Sky Search in the Tests of Everyday Attention for Children, Sky Search Task	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	Tap Once Tap Twice Task	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Uganda	Tobii Professional Studio eye tracking programming	1	0	0	0	0	1	0	0	0	0	0	Boivin 2017
Uganda	Wisconsin Card Sort Test	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Venezuela	Bayley Infant Neurodevelopmental Screener (BINS)	1	1	1	0	0	0	0	0	0	0	0	McCarthy 2012
Viet Nam	Vietnamese version of the Intelligibility in Context Scale (ICS-VN).	1	0	0	1	0	1	1	0	0	0	0	Phạm 2017

Country	Name of assessment tool	Number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consistency	Content	Concur-rent	Conver-gent	Predictive	Structu-ral	Over countries	Over other groups	First author & year
Vietnam	International Development and Early Learning Assessment (IDELA)	1	0	0	0	0	0	0	0	1	0	0	Halpin 2019
Zambia	Ages & Stages Questionnaire-3 (ASQ-3)	2	0	0	2	0	0	1	0	1	0	0	van Heerden 2017; Hsiao 2016
Zambia	Bayley Scales of Infant Development, Second Edition (BSID-II)	1	0	0	0	0	1	0	0	0	0	0	Biasini 2015
Zambia	International Development and Early Learning Assessment (IDELA)	1	1	1	1	1	1	1	0	0	0	0	Pisani 2018

Supplemental Table 5 List of child development assessment tools and countries where they were applied

Name assessment tool	Number of countries where the tool was applied	List of countries where the tool was applied	Number of articles where tool was applied	Age range at assessment (in months)	Type of administration	References (first author & year)
Ages & Stages Questionnaire (ASQ)	2	Iran; South Africa	1	4-60	Indirect/reported	Vameghi 2013; Bornman 2010
Ages & Stages Questionnaire Inventory (ASQ:I)	2	Bangladesh; China	1	1-27	Indirect/reported	Pitchik 2023; Xie 2017
Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	3	Brazil; China; Turkey	3	4-72	Indirect/reported	Anunciação 2019; Bian 2017; Kucuker 2011
Ages & Stages Questionnaire: Social-Emotional, First Edition (ASQ:SE-1)	1	China	1	15-63	Indirect/reported	Xie 2019 (Journal of Child & Dev Studies)
Ages & Stages Questionnaire-3 (ASQ-3)	11	Brazil; China; Colombia; Georgia; Guatemala; India; Indonesia; Mexico; Nepal; South Africa; Zambia	14	1-66	Indirect/reported	Filgueiras 2013; Yue 2019; Rubio-Codina 2016; Rubio-Codina 2020; Zirakashvili 2018; Colbert 2021; Kvestad 2013; Yunilda 2022; Ortiz-Leon 2018; Shrestha 2019; van Heerden 2017; Hsiao 2016; Bian 2012; Shrestha 2024
Ages & Stages Questionnaires: Social Emotional (ASQ:SE-BR)	1	Brazil	1	21-66	Indirect/reported	Chen 2017
Alberta Infant Motor Scale (AIMS)	1	Brazil	4	0-19	Direct assessment	Saccani 2016; Siegle 2018; de Albuquerque 2018; Valentini 2012
Angkor Hospital for Children Developmental Milestone Assessment Tool (AHC DMAT)	1	Cambodia	1	1-72	Direct assessment, indirect/reported	Ngoun 2012
Battelle Developmental Inventory Second Edition, BDI-2	2	Brazil; Colombia	3	0-95	Direct assessment, indirect/reported	de Albuquerque 2022; Rubio-Codina 2016; Rubio-Codina 2020
Bayley Infant Neurodevelopmental Screener (BINS)	8	Argentina, Brazil, Ecuador, Bolivia, Colombia, Venezuela; Turkey; South Africa	4	3-24	Direct assessment	McCarthy 2012; Soysal 2014; Rodriguez 2020; Guedes 2011
Bayley Scales of Infant and Toddler Development (BSID)	1	Iran	1	1-42	Not reported	Azari 2017

Name assessment tool	Number of countries where the tool was applied	List of countries where the tool was applied	Number of articles where tool was applied	Age range at assessment (in months)	Type of administration	References (first author & year)
Bayley Scales of Infant Development, bead threding task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Bayley Scales of Infant Development, Second Edition (BSID-II)	4	Bangladesh, India, Pakistan, Zambia	2	1-42	Direct assessment	Parveen 2014
Bayley Scales of Infant Development, Third Edition (BSID-III)	12	Bangladesh, Brazil, India, Nepal, Pakistan, Peru, South Africa; Colombia; Ethiopia; Cameroon; Kenya; Malaysia	12	6-42	Direct assessment	Pendergast 2018; Madaschi 2016; Rubio-Codina 2016; Rubio-Codina 2020; Hanlon 2016; Lohaus 2014; McHenry 2021; Zakaria 2012; Ranjitkar 2018; Manandhar 2016; Rasheed 2023; Koshy 2024
Bracken School Readiness Assessment, Third Edition (BSRA-3)	1	India	1	36-60	Direct assessment	Shah 2013
Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	3	Côte d'Ivoire; Ghana; Turkey	2	23-32	Indirect/reported	Barthel 2021; Karabekiroglu 2009
British Ability Scales – 3rd ed., block design task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Caregiver Reported Early Development Instruments (CREDI)	18	17 countries not listed, Tanzania	2	0-35	Indirect/reported	Waldman 2021; McCoy 2017
Caregiver Reported Early Development Instruments (CREDI), long-form	2	China; India	2	5-36	Indirect/reported	Li 2020; Alderman 2021
Caregiver Reported Early Development Instruments (CREDI), short-form	18	17 countries not listed; Brazil	2	0-35	Indirect/reported	McCoy 2018; Altafim 2018
Child Behavior Checklist/1.5-5 (CBCL/1.5-5) and the Caregiver-Teacher Report Form (C-TRF)	1	China	1	50-71	Indirect/reported	Liu 2011
Child Behavior Scale (CBS)	1	Peru	1	24-72	Indirect/reported	Meyer 2011
Child Behaviour Assessment Instrument (CBAI)	1	Sri Lanka	1	48-72	Indirect/reported	Samarakkody 2010
Child behaviour checklist (CBCL)	1	Kenya	1	12-72	Indirect/reported	Kariuki 2016
Child Social Preference Scale	1	China	1	36-60	Indirect/reported	Li 2016
Childhood Nonverbal Communication Scale (CNCS)	1	Iran	1	3-18	Indirect/reported	Oryadi-Zanjani 2020
Children's Behavior Questionnaire (CBQ)	1	Iran	1	48-72	Indirect/reported	Golmohammadi 2022
Chinese Preschool Readiness Scale (CPRS)	1	China	1	31-62	Indirect/reported	Xie 2019 (Early Educ & Dev)

Name assessment tool	Number of countries where the tool was applied	List of countries where the tool was applied	Number of articles where tool was applied	Age range at assessment (in months)	Type of administration	References (first author & year)
Chinese version of Urban Infant-Toddler Social and Emotional Assessment	1	China	1	12-36	Indirect/reported	Jianduan 2009
Chinese Vocabulary Checklist	1	China	1	12-30	Indirect/reported	Hao 2008
Communicative Development Inventory (CDI)	1	Kenya	1	6-36	Indirect/reported	Alcock 2015
Denver Developmental Screening Test for Sri Lankan Children (DDST-SL)	1	Sri Lanka	1	0-80	Direct assessment	Wijedasa 2012
Denver Developmental Screening Test (Denver-II)	3	Brazil; Colombia; Ethiopia	4	0-70.6	Direct assessment; indirect/reported	Lopez Boo 2020; Rubio-Codina 2016; Rubio-Codina 2020; Abessa 2016
DEvelopmental assessment on an E-Platform (DEEP)	1	India	2	34-40	Direct assessment	Bhavnani 2019; Mukherjee 2020
Developmental Assessment Tool for Anganwadis (DATA)	1	India	1	19-36	Direct assessment	Nair 2009
Developmental Assessment Tool for Anganwadis (DATA-II)	1	India	1	36-48	Direct assessment	Nair 2012
Developmental Milestone Chart (DMC)	1	Cambodia	1	1-96	Direct assessment; indirect/reported	Scherzer 2009
Developmental Milestones Checklist (DMC)	1	Kenya	1	2.53–12.06	Indirect/reported	Abubakar 2010
Dimensions of Mastery Questionnaire (DMQ-18)	1	Bangladesh	1	36-72	Indirect/reported	Shaoli 2019
Early Childhood Development Index (ECDI)	1	Nigeria	1	36-48	Indirect/reported	Enelamah 2024
Early Childhood Development Index 2030 (ECDI2030)	2	Mexio, Palestine	1	24-59	Indirect/reported	Halpin 2024
Early Development Instrument	1	Indonesia	1	48-96	Indirect/reported	Brinkman 2017
Early Human Capability Index (eHCI)	7	Brazil, Kiribati, China, Laos, Samoa, Tonga, Tuvalu	1	24-72	Indirect/reported	Sincovich 2019
Early Learning Outcomes Measure (ELOM)	1	South Africa	1	50-69	Direct assessment	Snelling 2019
EDIN II (Evaluación del Desarrollo Integral de la niña y el niño)	1	Costa Rica	1	0-72	Direct assessment	Cordero, 2019
Elicited Imitation Test (EIT)	1	Iran	1	36-83	Direct assessment	Mehrani 2018
Emotional and Behavioral Screener (EBS)	1	China	1	48-84	Indirect/reported	Zhang 2018
Emotional Regulation Checklist	1	Turkey	1	48-60	Indirect/reported	Danişman 2016

Name assessment tool	Number of countries where the tool was applied	List of countries where the tool was applied	Number of articles where tool was applied	Age range at assessment (in months)	Type of administration	References (first author & year)
Escala de Conductas de Aprendizaje Preescolar (ECAP), a Spanish translation of the Preschool Learning Behaviors Scale (PLBS)	1	Peru	1	24-72	Indirect/reported	Hahn 2009
Escala de Habildades Sociales (Scale of Social Abilities)	1	Argentina	1	36-60	Indirect/reported	Lacunza 2009
Executive function battery	1	Pakistan	1		Direct assessment	Obradovic 2019
Executive Function Touch (EF Touch)	2	Kenya	3	36-72	Self-administrated	Willoughby 2019 (Child Neuropsych); Willoughby 2019 (Dev Science); Toor 2022
Farsi Narrative Norms Instrument	1	Iran	1	60-72	Direct assessment	Soleymani 2016
Global Scales for Early Development	32	Not listed Ethiopia, Kenya	3	0-36	Direct assessment, indirect/reported	McCray 2023; Jeong 2025; Yibeltal 2025
Grenada Learning and Memory Scale	1	Grenada	1	36-72	Direct assessment	Blackmon 2024
Infant Neurological International Battery	1	Iran	1	4-18	Direct assessment	Soleimani 2007
Intelligibility in Context Scale (ICS) / ICS-Jamaican Creole	1	Jamaica	1	39-75	Indirect/reported	Washington 2017
INTER- GROWTH-21st Project Neurodevelopment Assessment (INTER-NDA)	1	Grenada	1	22-30	Direct assessment, indirect/reported	Waechter 2022
International Development and Early Learning Assessment (IDELA)	17	Afghanistan, Ethiopia, Uganda, Vietnam, Bolivia, Bangladesh, Bhutan, Egypt, India, Indonesia, Mali, Malawi, Mozambique, Pakistan, Rwanda, Zambia, Ghana	4	36-96	Direct assessment	Halpin 2019; Pisani 2018; Wolf 2017; Pisani 2022
Kaufman Survey of Early Academic and Language Skills (K-SEALS)	1	Turkey		61-72	Direct assessment	Uyanik 2014
Kilifi Developmental Inventory	2	Ghana, Côte d'Ivoire, Kenya	2	6-35	Direct assessment	Barthel 2022; Abubakar 2008

Name assessment tool	Number of countries where the tool was applied	List of countries where the tool was applied	Number of articles where tool was applied	Age range at assessment (in months)	Type of administration	References (first author & year)
Kilifi Developmental Inventory, coin box task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Kilifi Picture Vocabulary Test	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Little Developmental Coordination Disorder Questionnaire (Little DCDQ)	1	South Africa	1	36-49	Indirect/reported	Venter 2015
MacArthur Communicative Development Inventory Level II	1	Pakistan	1		Indirect/reported	Rasheed 2017
MacArthur-Bates Communicative Development Inventories CDI	1	Colombia	1	8-30	Indirect/reported	Lara Díaz 2011
MacArthur-Bates Communicative Development Inventories Short-Forms I vocabulary checklist	1	Colombia	2	6-42	Indirect/reported	Rubio-Codina 2016; Rubio-Codina 2020
MacArthur-Bates Communicative Development Inventories Short-Forms II vocabulary checklist	1	Colombia	2	6-42	Indirect/reported	Rubio-Codina 2016; Rubio-Codina 2020
Malawi Developmental Assessment Tool (MDAT)	1	Malawi	1	0-84	Direct assessment, indirect/reported	Gladstone 2010
Malawi Developmental Assessment Tool (MDAT), renamed Dominican Child Development Screening Tool	1	Dominican Republic	1	2-24	Direct assessment, indirect/reported	Sánchez-Vincitore 2019
Malay Preschool Language Assessment Tool (MPLAT)	1	Malaysia	1	48-83	Direct assessment	Razak 2010
MELQO Measure of Development and Learning Outcomes	1	Laos	1	48-49	Direct assessment	Gomez 2022
Milestones Checklist; Vocabulatory Inventory	1	Senegal	1	4-30	Indirect/reported	Weber 2018
Modified Fagan Test for Infant Intelligence (FTII)	1	Uganda	1	6-12	Direct assessment	Familiar-Lopez 2022
Mongolian Rapid Baby Scale (MORBAS)	1	Mongolia	1	0-42	Indirect/reported	Dagvadorj 2015
Motor Development Scale	1	Brazil	1	48-72	Not reported	Okuda 2020
Movement Assessment Battery for Children-second edition (MBAC-2)	1	China	1	36-72	Direct assessment	Hua 2013
Movement Assessment Battery for Children, balancing on one leg	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Mullen Scales of Early Learning (MSEL)	4	Guatemala; South Africa; The Gambia	5	0-68	Direct assessment	Koura 2013; Boivin 2021; Colbert 2020; Bornman 2010; Milosavljevic 2019

Name assessment tool	Number of countries where the tool was applied	List of countries where the tool was applied	Number of articles where tool was applied	Age range at assessment (in months)	Type of administration	References (first author & year)
NEPSY verbal fluence task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
NEPSY, knock tap game	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
NEPSY, sentence repetition task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Oxford Neurodevelopment Assessment (OX-NDA)	1	Brazil	1		Direct assessment, indirect/reported	Fernandes 2022
Parents' Evaluation of Developmental Status (PEDS)	3	India; South Africa; Thailand	3	0-30	Indirect/reported	Mukherjee 2022; Abdoola 2021; Chunsuwan 2016
Preschool and Kindergarten Behavior Scale (PKBS)	1	Brazil	1	36-72	Indirect/reported	Dias 2011
Preschool Language Scale-3 (PLS- 3)	1	Colombia	1	48-84	Direct assessment	Romero 2013
Preschool Learning Behavior Scale (PLBS)	1	China	1		Indirect/reported	Wu 2019 (JPA)
Questionário para Avaliação do Desenvolvimento Infantil (QAD-PIPAS)	1	Brazil	1	0-59	Indirect/reported	Venancio 2020
Ragama Early Assessment for Children (REACh)	1	Sri Lanka	1	24-60	Indirect/reported	Caldera 2023
Rapid Assessment of Cognitive and Emotional Regulation (RACER)	1	Ghana	1	48-72	Self-administration	Yuan 2022
Rapid Neurodevelopmental Assessment (RNDA)	2	Bangladesh; Guatemala	3	0-60	Direct assessment	Khan 2010; Khan 2013; Thompson 2015
Receptive and expressive vocabulary test battery (English, Luo and Swahili)	1	Kenya	1	24-83	Direct assessment	Knauer 2019
Saccadic reaction time (SRTs)	1	Sierra Leone	1	7-11.9	Self-administration	Leppänen 2022
Screening for Identification of Oral Language Difficulties by Preschool Teachers (SIOLD)	1	Brazil	1	51-65	Indirect/reported	Puglisi 2020
Screening Test of Early Language Development-Test version (STELD-T)	1	India	1	6-21	Indirect/reported	Pathak 2023
Self-ordered Pointing Task (SOPT)	1	Brazil	1	42-60	Self-administration	Sallum 2017
Sky Search in the Tests of Everyday Attention for Children, Sky Search Task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010

Name assessment tool	Number of countries where the tool was applied	List of countries where the tool was applied	Number of articles where tool was applied	Age range at assessment (in months)	Type of administration	References (first author & year)
Social Skills domain of System-Rating Scales (SSIS-RS)	1	China	1	36-60	Indirect/reported	Wu 2019 (PS)
Spanish-language MacArthur–Bates Communicative Development Inventories short forms (S-CDI SFI and SFII)	1	Mexico	1	8-30	Indirect/reported	Jackson-Maldonado 2013
Strengths and Difficulties Questionnaire (SDQ)	2	Pakistan; Turkey	2	20-51	Indirect/reported	Finch 2018; Dursun 2020
Tap Once Tap Twice Task	1	Uganda	1	55-67	Direct assessment	Nampijja 2010
Test of Gross Motor Development (TGMD-2)	1	Brazil	1	36-60	Direct assessment	Bandeira 2020
Thai Speech and Language Test (TSLT)	1	Thailand	1	3-24	Direct assessment	Prathanee 2008
The Chinese Early Language and Communication Questionnaire	1	China	1	2-48	Indirect/reported	Zhang 2021
The Chinese Intelligence Scale for Young Children	1	China	1	36-84	Direct assessment	Guo 2009
The infant gross motor screening test (IGMST)	1	South Africa	1	6-18	Direct assessment	Hilburn 2011
The language inventory	1	Bangladesh	1	12-18	Indirect/reported	Hamadani 2010
The Road to Health Booklet (RTHB)	1	South Africa	1	6-12	Indirect/reported	van der Linde 2015
The Test of Infant Motor Performance (TIMP)	1	Nepal	1	2-3	Direct assessment	Kvestad 2023
Tobii Professional Studio eye tracking programming	1	Uganda	1	44-65	Direct assessment	Boivin 2017
Vietnamese version of the Intelligibility in Context Scale (ICS-VN).	1	Viet Nam	1	24-71	Indirect/reported	Phạm 2017
Vineland Adaptive Behaviour Scale (VABS II)	1	India	1	36-108	Indirect/reported	Kumar 2016
Vocabulary Checklists	2	Malawi, Ghana	1	12-17; 48-72	Indirect/reported	Prado 2018
Wechsler Preschool and Primary Scale of Intelligence, thid edition (WPPSI-III)	1	Pakistan	1		Direct assessment	Rasheed 2018
WHO Gross Motor Milestones	2	Bangladesh; Colombia	3	3-42	Direct assessment, indirect/reported	Hamadani 2013; Rubio-Codina 2016; Rubio-Codina 2020
Wisconsin Card Sort Test	1	Uganda	1	55-67	Direct assessment	Nampijja 2010

Supplemental Table 6 Number of articles providing psychometric evidence by world region and urban/rural location

Region	Urban	Rural	Urban-rural
Caribbean	2	0	0
Central America	1	3	0
East Asia	10	1	8
Oceania	0	0	1
North Africa and the Middle East	0	0	2
North America	1	0	1
South America	18	0	7
South Asia	13	13	9
Southeast Asia	5	1	3
Sub-Saharan Africa	11	16	14
West Asia	8	0	2

Supplemental Table 7 Number of articles providing psychometric evidence by child development assessment tools (development articles)

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu -ral	Over countries	Over other groups	First author & year
Angkor Hospital for Children Developmental Milestone Assessment Tool (AHC DMAT)	1	1	0	0	0	1	0	0	0	0	0	0	Ngoun 2012
Caregiver Reported Early Development Instruments (CREDI)	2	18	2	1	2	2	2	2	0	1	0	0	McCoy 2017; Waldman 2021
Caregiver Reported Early Development Instruments (CREDI), short-form	1	17	1	0	1	1	1	1	0	0	0	1	McCoy 2018
Child Behaviour Assessment Instrument (CBAI)	1	1	1	0	1	1	1	0	0	0	0	0	Samarakkody 2010
Childhood Nonverbal Communication Scale (CNCS)	1	1	0	0	1	1	1	1	0	1	0	0	Oryadi-Zanjani 2020
Chinese Preschool Readiness Scale (CPRS)	1	1	0	0	1	1	1	0	0	1	0	0	Xie 2019 (Early Educ & Dev)
Chinese Vocabulary Checklist	1	1	0	0	0	0	0	1	0	0	0	0	Hao 2008
DEvelopmental assessment on an E-Platform (DEEP)	1	1	0	0	0	1	0	0	0	0	0	0	Bhavnani 2019
Developmental Assessment Tool for Anganwadis (DATA)	1	1	0	0	1	1	0	0	0	1	0	0	Nair 2009
Developmental Assessment Tool for Anganwadis (DATA-II)	1	1	0	0	1	1	0	0	0	1	0	0	Nair 2012
Developmental Milestone Chart (DMC)	1	1	0	0	0	1	0	0	0	0	0	0	Scherzer 2009
Developmental Milestones Checklist (DMC)	1	1	1	0	1	1	1	1	0	1	0	0	Abubakar 2010
Early Childhood Development Index 2030 (ECDI203)	1	2	0	0	1	1	1	0	0	1	0	1	Halpin 2024
Early Learning Outcomes Measure (ELOM)	1	1	0	1	1	0	0	0	0	1	0	1	Snelling 2019
Elicited Imitation Test (EIT)	1	1	0	1	0	0	1	0	0	0	0	1	Mehrani 2018

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence of measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over other countries	Over other groups	First author & year
Escala de Habildades Sociales (Scale of Social Abilities)	1	1	0	0	1	0	0	0	0	1	0	0	Lacunza 2009
Farsi Narrative Norms Instrument	1	1	1	1	0	1	0	0	0	0	0	0	Soleymani 2016
Global Scales for Early Development	1	32	0	0	1	1	0	0	0	0	0	0	McCray 2023
Grenada Learning and Memory Scalea	1	1	1	1	1	1	1	1	0	1	0	0	Blackmon 2024
International Development and Early Learning Assessment (IDELA)	1	12	1	1	1	1	1	1	0	0	0	0	Pisani 2018
Kilifi Developmental Inventory	1	1	1	1	1	1	1	1	0	0	0	0	Abubakar 2008
Malawi Developmental Assessment Tool (MDAT)	1	1	1	1	0	1	0	1	0	0	0	0	Gladstone 2010
Malay Preschool Language Assessment Tool (MPLAT)	1	1	1	1	1	1	0	0	0	0	0	0	Razak 2010
Mongolian Rapid Baby Scale (MORBAS)	1	1	0	0	0	1	1	0	0	0	0	0	Dagvadorj 2015
Oxford Neurodevelopment Assessment (OX-NDA)	1	1	1	1	1	0	1	0	0	0	0	0	Fernandes 2022
Questionário para Avaliação do Desenvolvimento Infantil (QAD-PIPAS)	1	1	0	0	0	1	0	0	0	0	0	0	Venancio 2020
Ragama Early Assessment for Children (REACh)	1	1	0	1	1	1	1	0	0	0	0	0	Caldera 2023
Rapid Neurodevelopmental Assessment (RNDA)	1	1	0	1	0	1	1	0	1	0	0	0	Khan 2010
Receptive and expressive vocabulary test battery (English, Luo and Swahili)	1	1	0	1	1	0	1	1	1	0	0	0	Knauer 2019
Saccadic reaction time (SRTs)	1	1	1	0	0	0	0	1	0	0	0	0	Leppänen 2022
Screening for Identification of Oral Language Difficulties by Preschool Teachers (SIOLD)	1	1	0	0	0	0	0	0	0	1	0	0	Puglisi 2020
Screening Test of Early Language Development-Test version (STELD-T)	1	1	0	0	1	0	1	0	0	0	0	0	Pathak 2023

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence of measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu -ral	Over countries	Over other groups	First author & year
Thai Speech and Language Test (TSLT)	1	1	0	1	1	1	0	0	0	0	0	0	Prathanee 2008
The infant gross motor screening test (IGMST)	1	1	1	1	0	1	1	0	0	0	0	0	Hilburn 2011
The language inventory	1	1	1	0	0	1	1	1	1	0	0	0	Hamadani 2010
Vocabulary Checklists	1	1	0	0	0	0	1	0	1	0	0	0	Prado 2018

Supplemental Table 8 Number of articles providing psychometric evidence by child development assessment tools (adaptation articles)

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over other groups	First author & year
Ages & Stages Questionnaire (ASQ)	2	2	0	1	1	2	0	0	0	1	0	0	Vameghi 2013; Bornman 2010
Ages & Stages Questionnaire Inventory (ASQ:I)	2	2	1	1	2	1	2	1	0	0	0	0	Xie 2017; Pitchik 2023
Ages & Stages Questionnaire: Social-Emotional (ASQ-SE)	3	3	2	2	2	1	1	0	0	2	0	0	Kucuker 2011; Anunciação 2019; Bian 2017
Ages & Stages Questionnaire: Social-Emotional, First Edition (ASQ:SE-1)	1	1	1	1	1	0	1	0	0	0	0	0	Xie 2019 (Journal of Child & Dev Studies)
Ages & Stages Questionnaire-3 (ASQ-3)	14	11	5	6	12	4	5	5	3	4	0	0	Zirakashvili 2018; Colbert 2021; Rubio-Codina 2016; Rubio-Codina 2020; Shrestha 2019; Hsiao 2016; Filgueiras 2013; van Heerden 2017; Ortiz-Leon 2018; Kvestad 2013; Yue 2019; Yunilda 2022; Bian 2012; Shrestha 2024
Ages & Stages Questionnaires: Social Emotional (ASQ:SE-BR)	1	1	0	0	0	0	0	0	0	1	0	0	Chen 2017
Alberta Infant Motor Scale (AIMS)	4	1	3	3	2	1	3	1	1	0	0	0	Saccani 2016; Siegle 2018; de Albuquerque 2018; Valentini 2012
Battelle Developmental Inventory Second Edition, BDI-2	3	2	2	0	2	1	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020; de Albuquerque 2022
Bayley Infant Neurodevelopment Screener (BINS)	4	8	1	1	2	1	1	1	2	2	0	0	Guedes 2011; McCarthy 2012; Soysal 2014; Rodriguez 2020
Bayley Scales of Infant and Toddler Development (BSID)	1	1	1	1	1	1	0	1	0	1	0	0	Azari 2017

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over other groups	First author & year
Bayley Scales of Infant Development, bead threding task	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Bayley Scales of Infant Development, Second Edition (BSID-II)	2	4	0	0	0	1	1	0	0	0	0	0	Biasini 2015; Parveen 2014
Bayley Scales of Infant Development, Third Edition (BSID-III)	12	12	4	3	7	2	3	3	5	5	2	1	Rubio-Codina 2016; Rubio-Codina 2020; Hanlon 2016; Ranjitkar 2018; Madaschi 2016; Pendergast 2018; Manandhar 2016; Zakaria 2012; Lohaus 2014; McHenry 2021; Rasheed 2023; Koshy 2024
Bracken School Readiness Assessment, Third Edition (BSRA-3)	1	1	0	0	1	1	0	0	0	0	0	0	Shah 2013
Brief Infant-Toddler Social and Emotional Assessment (BITSEA)	2	3	1	1	2	0	1	1	0	1	0	0	Karabekiroglu 2009; Barthel 2021
British Ability Scales – 3rd ed., block design task	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Caregiver Reported Early Development Instruments (CREDI), long-form	2	2	0	0	2	0	2	2	0	0	0	0	Alderman 2021; Li 2020
Caregiver Reported Early Development Instruments (CREDI), short-form	1	1	1	0	1	1	1	1	0	0	0	0	Altafim 2018
Child Behavior Checklist/1.5-5 (CBCL/1.5-5) and the Caregiver-Teacher Report Form (C-TRF)	1	1	0	0	0	0	0	0	0	1	1	0	Liu 2011
Child Behavior Scale (CBS)	1	1	0	0	1	0	0	0	0	1	0	0	Meyer 2011
Child behaviour checklist (CBCL)	1	1	1	0	1	0	0	1	0	1	0	0	Kariuki 2016
Child Social Preference Scale	1	1	0	0	1	0	0	0	1	1	0	1	Li 2016
Children's Behavior Questionnaire (CBQ)	1	1	1	0	1	1	1	1	0	1	0	0	Golmohammadi 2022

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over other groups	First author & year
Chinese version of Urban Infant-Toddler Social and Emotional Assessment	1	1	1	1	1	1	1	0	0	1	0	0	Jianduan 2009
Communicative Development Inventory (CDI)	1	1	1	0	1	0	1	0	0	0	0	0	Alcock 2015
Denver Developmental Screening Test for Sri Lankan Children (DDST-SL)	1	1	0	1	0	0	0	0	0	0	0	0	Wijedasa 2012
Denver Developmental Screening Test (Denver-II)	4	3	3	2	3	2	1	2	1	1	0	0	Abessa 2016; Rubio-Codina 2016; Rubio-Codina 2020; Lopez Boo 2020
DEvelopmental assessment on an E-Platform (DEEP)	1	1	0	0	0	0	1	0	0	0	0	0	Mukherjee 2020
Dimensions of Mastery Questionnaire (DMQ-18)	1	1	1	1	1	1	1	0	0	1	0	0	Shaoli 2019
Early Childhood Development Index (ECDI)	1	1	0	0	1	1	0	0	0	1	0	0	Enelamah 2024
Early Development Instrument	1	1	0	1	1	1	1	0	1	0	0	0	Brinkman 2017
Early Human Capability Index (eHCI)	1	7	0	0	1	0	0	0	0	1	0	1	Sincovich 2019
EDIN II (Evaluación del Desarrollo Integral de la niña y el niño)	1	1	0	0	0	1	0	1	0	0	0	0	Cordero, 2019
Emotional and Behavioral Screener (EBS)	1	1	0	1	1	0	1	0	0	1	0	0	Zhang 2018
Emotional Regulation Checklist	1	1	1	0	1	1	1	0	0	1	0	0	Danişman 2016
Escala de Conductas de Aprendizaje Preescolar (ECAP), a Spanish translation of the Preschool Learning Behaviors Scale (PLBS)	1	1	0	0	0	0	0	0	0	1	0	0	Hahn 2009
Executive function battery	1	1	0	0	1	1	0	1	1	1	0	0	Obradovic 2019
Executive Function Touch (EF Touch)	3	2	0	1	1	1	1	3	0	2	0	0	Willoughby 2019 (Child Neuropsych); Willoughby 2019 (Dev Science); Toor 2022

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over other groups	First author & year
Global Scales of Early Development	2	2	0	1	0	2	1	1	0	0	0	0	Yibeltal 2025; Jeong 2025
Infant Neurological International Battery	1	1	0	1	0	0	1	0	0	0	0	0	Soleimani 2007
Intelligibility in Context Scale (ICS) / ICS-Jamaican Creole	1	1	1	1	1	0	1	1	0	1	0	0	Washington 2017
INTER- GROWTH-21st Project Neurodevelopment Assessment (INTER-NDA)	1	1	1	1	1	0	0	0	0	0	0	0	Waechter 2022
International Development and Early Learning Assessment (IDELA)	3	17	1	1	1	0	1	1	1	2	1	0	Halpin 2019; Pisani 2022; Wolf 2017
Kaufman Survey of Early Academic and Language Skills (K-SEALS)	1	1	1	1	1	1	0	0	0	1	0	0	Uyanik 2014
Kilifi Developmental Inventory	1	2	0	1	1	0	0	0	1	1	0	0	Barthel 2022
Kilifi Developmental Inventory, coin box task	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Kilifi Picture Vocabulary Test	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Little Developmental Coordination Disorder Questionnaire (Little DCDQ)	1	1	0	0	1	0	1	0	0	0	0	0	Venter 2015
MacArthur Communicative Development Inventory Level II	1	1	1	1	1	1	1	0	0	0	0	0	Rasheed 2017
MacArthur-Bates Communicative Development Inventories CDI	1	1	0	0	1	1	0	1	0	0	0	0	Lara Díaz 2011
MacArthur-Bates Communicative Development Inventories Short-Forms I vocabulary checklist	2	1	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020
MacArthur-Bates Communicative Development Inventories Short-Forms II vocabulary checklist	2	1	2	0	2	0	1	1	1	0	0	0	Rubio-Codina 2016; Rubio-Codina 2020

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over other groups	First author & year
Malawi Developmental Assessment Tool (MDAT), renamed Dominican Child Development Screening Tool	1	1	0	1	1	0	0	1	0	0	0	0	Sánchez-Vincitore 2019
MELQO Measure of Development and Learning Outcomes	1	1	0	0	0	0	0	1	0	1	1	0	Gomez 2022
Milestones Checklist; Vocabulatory Inventory	1	1	0	0	1	1	1	1	0	0	0	0	Weber 2018
Modified Fagan Test for Infant Intelligence (FTII)	1	1	0	0	0	0	0	0	1	0	0	0	Familiar-Lopez 2022
Motor Development Scale	1	1	0	0	1	0	0	0	0	1	0	0	Okuda 2020
Movement Assessment Battery for Children- second edition (MBAC-2)	1	1	1	1	1	1	1	0	0	1	0	0	Hua 2013
Movement Assessment Battery for Children, balancing on one leg	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Mullen Scales of Early Learning (MSEL)	5	4	1	2	1	1	1	3	3	0	0	0	Colbert 2020; Koura 2013; Milosavljevic 2019; Bornman 2010; Boivin 2021
NEPSY verbal fluence task	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
NEPSY, knock tap game	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
NEPSY, sentence repetition task	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Parents' Evaluation of Developmental status (PEDS)	3	3	0	0	0	0	3	0	0	0	0	0	Abdoola 2021; Chunsuwan 2016; Mukherjee 2022
Preschool and Kindergarten Behavior Scale (PKBS)	1	1	0	0	1	0	0	0	0	1	0	0	Dias 2011
Preschool Language Scale-3 (PLS- 3)	1	1	0	0	1	0	0	0	0	1	0	0	Romero 2013
Preschool Learning Behavior Scale (PLBS)	1	1	0	0	1	0	1	0	0	1	0	0	Wu 2019 (JPA)
Rapid Assessment of Cognitive and Emotional Regulation (RACER)	1	1	1	1	0	0	1	1	0	0	0	0	Yuan 2022

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over other groups	First author & year
Rapid Neurodevelopmental Assessment (RNDA)	2	2	0	2	0	1	2	1	0	0	0	0	Thompson 2015; Khan 2013
Self-ordered Pointing Task (SOPT)	1		0	0	0	0	1	1	0	0	0		Sallum 2017
Sky Search in the Tests of Everyday Attention for Children, Sky Search Task	1	1	1	0	1	0	0	1	0	1	0		Nampijja 2010
Social Skills domain of System-Rating Scales (SSIS-RS)	1	1	0	0	1	0	0	1	0	1	0		Wu 2019 (PS)
Spanish-language MacArthur–Bates Communicative Development Inventories short forms (S-CDI SFI and SFII)	1	1	1	0	0	0	0	1	0	0	0		Jackson-Maldonado 2013
Strengths and Difficulties Questionnaire (SDQ)	2	2	1	1	2	1	2	2	0	1	0		Finch 2018; Dursun 2020
Tap Once Tap Twice Task	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010
Test of Gross Motor Development (TGMD-2)	1	1	0	1	1	0	0	1	0	1	0	0	Bandeira 2020
The Chinese Early Language and Communication Questionnaire	1	1	0	0	1	0	1	0	0	0	0	0	Zhang 2021
The Chinese Intelligence Scale for Young Children	1	1	0	0	0	0	0	0	0	1	0	1	Guo 2009
The Road to Health Booklet (RTHB)	1	1	0	0	0	0	1	0	0	0	0	0	van der Linde 2015
The Test of Infant Motor Performance (TIMP)	1	1	0	1	1	0	1	0	1	0	0	0	Kvestad 2023
Tobii Professional Studio eye tracking programming	1	1	0	0	0	0	1	0	0	0	0	0	Boivin 2017
Vietnamese version of the Intelligibility in Context Scale (ICS-VN).	1	1	0	0	1	0	1	1	0	0	0	0	Phạm 2017
Vineland Adaptive Behaviour Scale (VABS II)	1	1	1	1	0	0	0	1	0	0	0	0	Kumar 2016
Wechsler Preschool and Primary Scale of Intelligence, thid edition (WPPSI-III)	1	1	0	1	1	1	0	1	0	1	0	0	Rasheed 2018

Name of assessment tool	Number of articles providing psycho-metric evidence	Number of countries where the tool was applied	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance		References
			Test-retest	Inter-rater	Internal consis-tency	Content	Concur-rent	Conver-gent	Predic-tive	Structu-ral	Over countries	Over other groups	First author & year
WHO Gross Motor Milestones	3	2	2	1	2	0	2	2	2	0	0	0	Hamadani 2013; Rubio-Codina 2016; Rubio-Codina 2020
Wisconsin Card Sort Test	1	1	1	0	1	0	0	1	0	1	0	0	Nampijja 2010

Supplemental Table 9 Number of articles providing psychometric evidence by article type and early childhood development domain

Child development domain	Total number of articles	Number of articles providing evidence on reliability			Number of articles providing evidence on validity					Number of articles providing evidence on measurement invariance	
		Test-retest	Inter-rater	Internal consistency	Content/ face	Concurrent /Criterion	Convergent	Predictive	Structural	Over countries	Over other groups
All articles (N=182 article-tool dyads)											
Cognitive	82	28	31	53	39	35	29	18	3	3	
Language	111	39	38	72	49	51	41	21	4	7	
Motor	104	38	41	62	41	46	38	21	5	5	
Socioemotional/ temperament	52	19	17	37	25	30	16	4	4	5	
Attention/ executive function	22	10	4	14	4	7	16	1	1	4	
Personal-social/ adaptive	49	19	18	32	20	22	16	6	0	1	
Academic/ preacademic	10	2	4	6	3	4	3	1	2	5	
Development articles (N=37 article-tool dyads)											
Cognitive	17	7	7	12	14	10	6	1	2	2	
Language	28	9	10	16	21	16	9	4	2	3	
Motor	20	9	9	13	18	12	7	1	2	3	
Socioemotional/ temperament	13	6	4	10	12	10	4	0	2	2	
Attention/ executive function	3	1	2	3	2	2	1	0	0	2	
Personal-social/ adaptive	10	2	3	6	9	4	2	1	0	0	
Academic/ preacademic	3	1	2	1	2	2	1	0	0	2	
Adaptation articles (N=145 article-tool dyads)											
Cognitive	65	21	24	41	25	25	23	17	1	1	
Language	83	30	28	56	28	35	32	17	2	4	
Motor	84	29	32	49	23	34	31	20	3	2	
Socioemotional/ temperament	39	13	13	27	13	20	12	4	2	3	
Attention/ executive function	19	9	2	11	2	5	15	1	1	2	
Personal-social/ adaptive	39	17	15	26	11	18	14	5	0	1	
Academic/ preacademic	7	1	1	3	1	2	2	1	2	3	

Supplemental Table 10 Number of articles by category of risk of bias

	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data
Low risk	47	86	78
Some concerns	12	8	52
High risk	0	1	7
Unable to assess	123	9	44
Not applicable	0	78	1

	Indirectness (are results generalisable?)
Yes	20
Probably yes	82
Probably no	53
No	24
No information	3

Supplemental Table 11 Risk of bias assessment by included article

First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Abdoola 2021	unable to assess	N/A	unable to assess	probably no
Abessa 2016	low risk	N/A	unable to assess	probably yes
Abubakar 2008	unable to assess	Low risk	low risk	probably yes
Abubakar 2010	unable to assess	Some concerns	High risk	No
Alcock 2015	unable to assess	N/A	low risk	Probably yes
Alderman 2021	Low risk	unable to assess	Low risk	Yes
Altafim 2018	low risk	unable to assess	some concerns	probably yes
Anunciação 2019	unable to assess	n/a	some concerns	Yes
Azari 2017	unable to assess	high risk	unable to assess	probably no
Bandeira 2020	unable to assess	low risk	Low risk	probably yes
Barthel 2021	unable to assess	low risk	low risk	probably yes
Barthel 2022	Some concerns	Low risk	Low risk	Probably no
Bhavnani 2019	unable to assess	N/A	low risk	Yes
Bian 2012	unable to assess	N/A	Low risk	no
Bian 2017	unable to assess	n/a	unable to assess	Yes
Biasini 2015	unable to assess	N/A	Some concerns	Probably yes
Blackmon 2024	Some concerns	Low risk	Some concerns	Probably no
Boivin 2017	unable to assess	Low risk	Low risk	No
Boivin 2021	unable to assess	low risk	low risk	probably yes
Bornman 2010	unable to assess	N/A	Low risk	no
Bornman 2010	unable to assess	N/A	low risk	no
Brinkman 2017	some concerns	Low risk	unable to assess	probably yes
Caldera 2023	unable to assess	N/A	unable to assess	Yes
Chen 2017	unable to assess	N/A	some concerns	yes
Chunsuwan 2016	unable to assess	N/A	unable to assess	probably no
Colbert 2020	low risk	low risk	unable to assess	probably yes
Colbert 2021	unable to assess	low risk	Low risk	No information
Cordero, 2019	unable to assess	N/A	unable to assess	Yes
Dagvadorj 2015	unable to assess	N/A	Low risk	Probably no
Danişman 2016	unable to assess	N/A	high risk	no
de Albuquerque 2018	unable to assess	N/A	Low risk	probably yes
de Albuquerque 2022	unable to assess	N/A	Low risk	probably yes
Dias 2011	unable to assess	N/A	low risk	Probably yes
Dursun 2020	unable to assess	low risk	Some concerns	probably no
Enelamah 2024	Unable to assess	N/A	Low risk	Yes
Familiar-Lopez 2022	unable to assess	Some concerns	Some concerns	No
Fernandes 2022	unable to assess	low risk	low risk	probably yes
Filgueiras 2013	unable to assess	N/A	unable to assess	probably yes

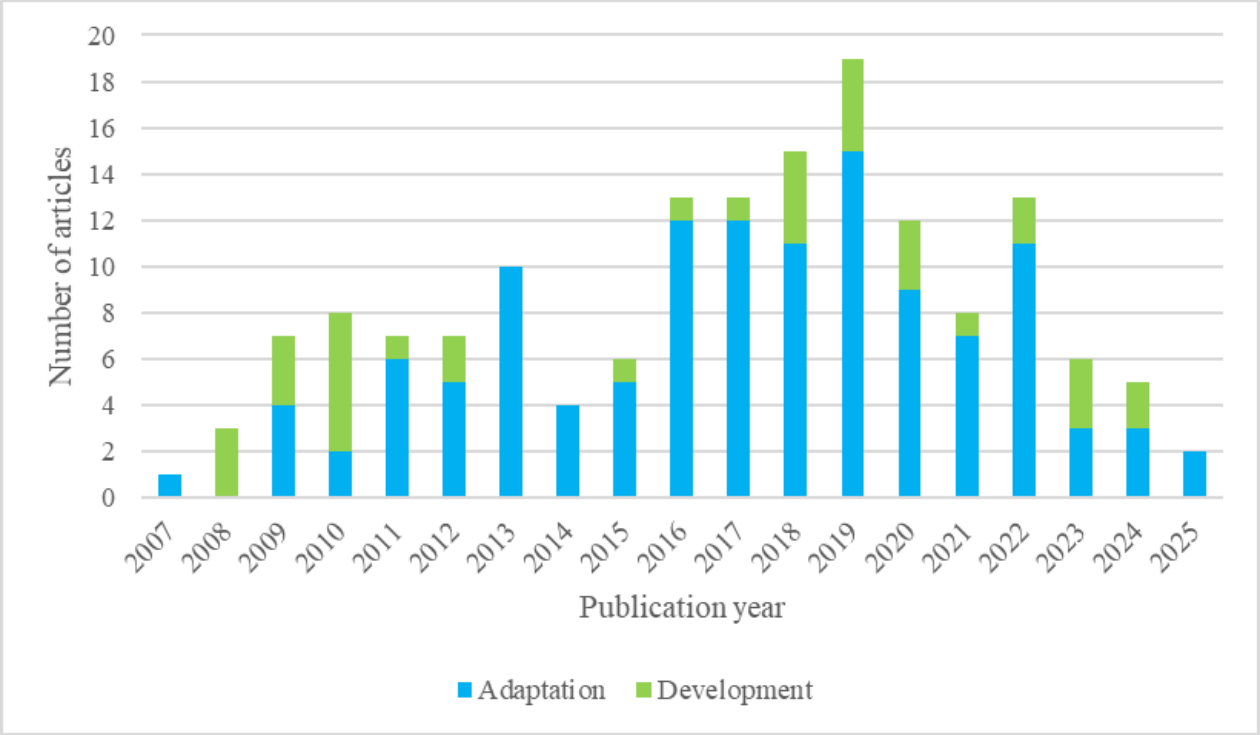
First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Finch 2018	low risk	Low risk	low risk	probably no
Gladstone 2010	unable to assess	low risk	some concerns	Probably yes
Golmohammadi 2022	unable to assess	Low risk	low risk	probably no
Gomez 2022	unable to assess	Low risk	Low risk	Probably yes
Guedes 2011	unable to assess	low risk	unable to assess	probably no
Guo 2009	unable to assess	N/A	unable to assess	Probably yes
Hahn 2009	unable to assess	N/A	high risk	probably yes
Halpin 2019	unable to assess	N/A	unable to assess	yes
Halpin 2024	Low risk	Low risk	Some concerns	Yes
Hamadani 2010	low risk	low risk	high risk	probably yes
Hamadani 2013	unable to assess	low risk	low risk	Probably yes
Hanlon 2016	Some concerns	Some concerns	Low risk	Probably no
Hao 2008	unable to assess	low risk	unable to assess	no
Hilburn 2011	low risk	N/A	unable to assess	Probably yes
Hsiao 2016	some concerns	low risk	low risk	probably yes
Hua 2013	unable to assess	N/A	Some concerns	probably yes
Jackson-Maldonado 2013	unable to assess	low risk	Low risk	Yes
Jeong 2025	Low risk	Low risk	Low risk	Probably yes
Jianduan 2009	unable to assess	N/A	unable to assess	probably yes
Karabekiroglu 2009	unable to assess	N/A	low risk	no
Kariuki 2016	Some concerns	Low risk	Some concerns	Probably yes
Khan 2010	unable to assess	Some concerns	low risk	Probably no
Khan 2013	unable to assess	N/A	unable to assess	probably yes
Knauer 2019	unable to assess	low risk	some concerns	probably yes
Koshy 2024	Low risk	Low risk	Low risk	Probably no
Koura 2013	low risk	unable to assess	unable to assess	probably no
Kucuker 2011	unable to assess	N/A	some concerns	Probably yes
Kumar 2016	unable to assess	some concerns	low risk	probably no
Kvestad 2013	Low risk	N/A	some concerns	Probably yes
Kvestad 2023	low risk	low risk	some concerns	probably no
Lacunza 2009	unable to assess	N/A	unable to assess	probably no
Lara Díaz 2011	unable to assess	low risk	Low risk	Probably yes
Leppänen 2022	unable to assess	low risk	high risk	probably no
Li 2016	unable to assess	Low risk	Some concerns	No
Li 2020	unable to assess	unable to assess	low risk	probably yes
Liu 2011	unable to assess	N/A	Some concerns	probably yes
Lohaus 2014	unable to assess	Low risk	Some concerns	No
Lopez Boo 2020	Some concerns	low risk	some concerns	probably yes
Madaschi 2016	unable to assess	N/A	Some concerns	probably yes
Manandhar 2016	low risk	N/A	unable to assess	probably no
McCarthy 2012	low risk	N/A	unable to assess	probably yes
McCoy 2017	Low risk	low risk	low risk	yes

First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
McCoy 2018	unable to assess	low risk	low risk	yes
McCray 2023	unable to assess	N/A	Low risk	Probably yes
McHenry 2021	low risk	low risk	some concerns	Probably no
Mehrani 2018	unable to assess	N/A	Low risk	Probably no
Meyer 2011	unable to assess	N/A	unable to assess	Probably no
Milosavljevic 2019	low risk	low risk	some concerns	no
Mukherjee 2020	unable to assess	N/A	low risk	No
Mukherjee 2022	unable to assess	N/A	low risk	probably yes
Nair 2009	unable to assess	N/A	unable to assess	probably yes
Nair 2012	unable to assess	N/A	unable to assess	probably yes
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Nampijja 2010	unable to assess	low risk	low risk	probably no
Ngoun 2012	unable to assess	N/A	unable to assess	yes
Obradovic 2019	unable to assess	Low risk	low risk	probably yes
Okuda 2020	unable to assess	N/A	unable to assess	probably yes
Ortiz-Leon 2018	unable to assess	N/A	some concerns	probably no
Oryadi-Zanjani 2020	some concerns	N/A	low risk	probably yes
Parveen 2014	unable to assess	N/A	N/A	no
Pathak 2023	unable to assess	N/A	low risk	probably no
Pendergast 2018	low risk	N/A	Some concerns	yes
Phạm 2017	unable to assess	low risk	low risk	probably no
Pisani 2018	unable to assess	low risk	unable to assess	yes
Pisani 2022	unable to assess	low risk	some concerns	probably yes
Pitchik 2023	low risk	low risk	low risk	probably yes
Prado 2018	Low risk	Low risk	low risk	No
Prathanee 2008	low risk	N/A	low risk	probably yes
Puglisi 2020	Some concerns	N/A	low risk	Probably yes
Ranjitkar 2018	low risk	low risk	low risk	probably yes
Rasheed 2017	some concerns	N/A	unable to assess	probably no
Rasheed 2018	low risk	Low risk	unable to assess	probably yes
Rasheed 2023	unable to assess	Low risk	unable to assess	no information
Razak 2010	unable to assess	N/A	low risk	probably no
Rodriguez 2020	low risk	unable to assess	unable to assess	probably yes

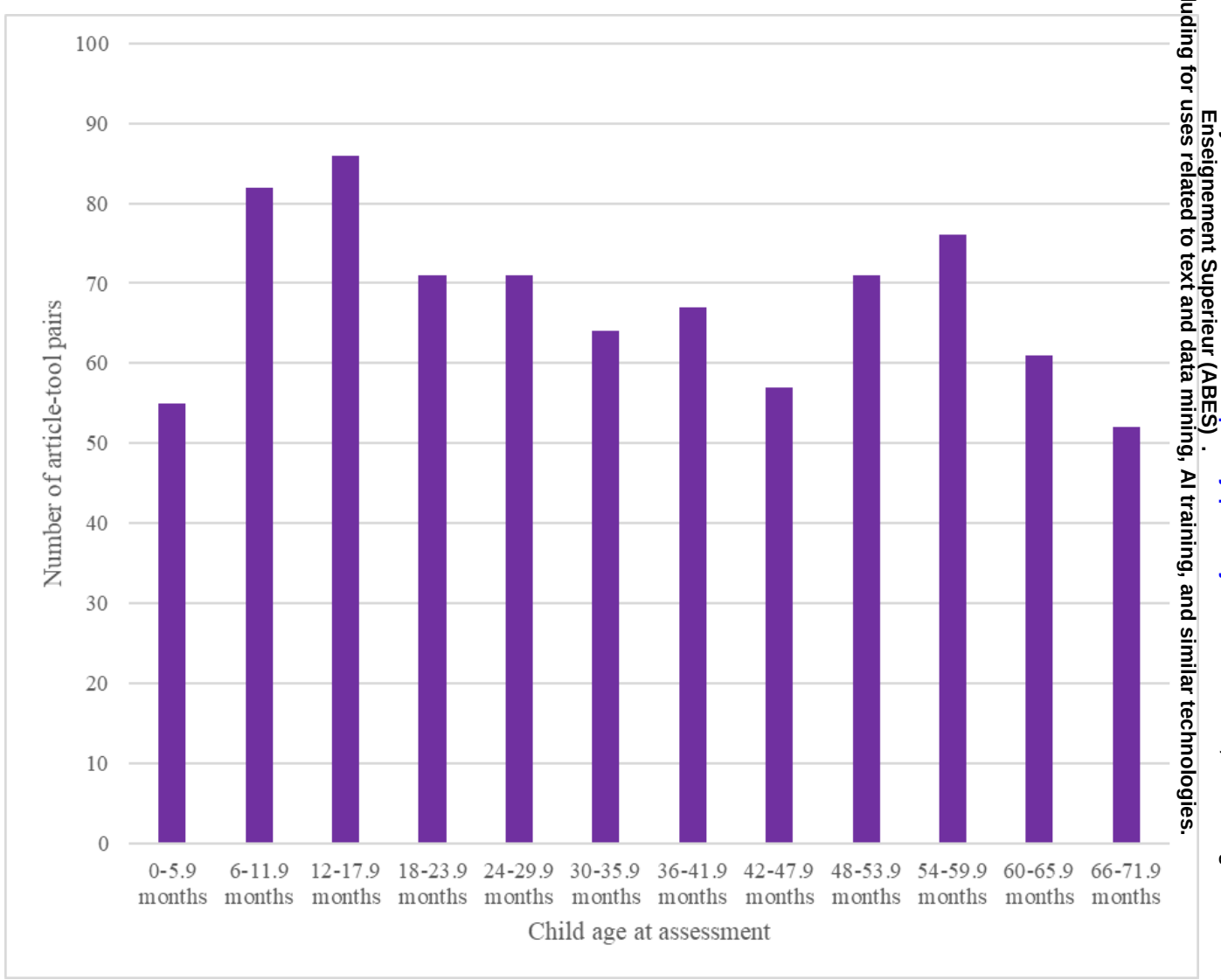
First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Romero 2013	unable to assess	N/A	unable to assess	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	low risk	Probably yes
Rubio-Codina 2016	low risk	Low risk	Some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	Some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	unable to assess	Low risk	some concerns	Probably yes
Rubio-Codina 2020	low risk	Low risk	some concerns	Probably yes
Saccani 2016	low risk	N/A	unable to assess	yes
Sallum 2017	unable to assess	low risk	some concerns	probably yes
Samarakkody 2010	unable to assess	N/A	unable to assess	yes
Sánchez-Vincitore 2019	low risk	Low risk	low risk	probably no
Scherzer 2009	unable to assess	N/A	unable to assess	probably yes
Shah 2013	unable to assess	n/a	unable to assess	no
Shaoli 2019	unable to assess	n/a	low risk	probably yes
Shrestha 2019	some concerns	low risk	some concerns	probably no
Shrestha 2024	Low risk	Low risk	Some concerns	Probably no
Siegle 2018	unable to assess	N/A	some concerns	probably no
Sincovich 2019	unable to assess	N/A	some concerns	probably no
Snelling 2019	unable to assess	N/A	some concerns	probably yes
Soleimani 2007	low risk	N/A	Some concerns	no information
Soleymani 2016	unable to assess	N/A	Low risk	No
Soysal 2014	unable to assess	some concerns	some concerns	no
Thompson 2015	low risk	unable to assess	unable to assess	probably no
Toor 2022	unable to assess	Low risk	some concerns	No
Uyanik 2014	unable to assess	N/A	Some concerns	Probably no
Valentini 2012	unable to assess	Low risk	unable to assess	Probably no
Vameghi 2013	unable to assess	N/A	high risk	probably no
van der Linde 2015	unable to assess	low risk	low risk	probably no
van Heerden 2017	low risk	N/A	unable to assess	no
Venancio 2020	unable to assess	N/A	Some concerns	Probably yes
Venter 2015	unable to assess	N/A	high risk	probably no
Waechter 2022	low risk	N/A	unable to assess	probably yes
Waldman 2021	unable to assess	Some concerns	Low risk	Probably yes

First author & year	Risk of bias due to training or administration	Risk of bias due to selective reporting	Risk of bias due to missing data	Are results generalizable?
Washington 2017	unable to assess	Low risk	Some concerns	No
Weber 2018	unable to assess	low risk	some concerns	probably yes
Wijedasa 2012	low risk	N/A	low risk	yes
Willoughby 2019 (Child Neuropsych)	low risk	low risk	low risk	probably yes
Willoughby 2019 (Dev Science)	Low risk	Low risk	Some concerns	Probably yes
Wolf 2017	Some concerns	low risk	unable to assess	no
Wu 2019 (JPA)	unable to assess	n/a	unable to assess	Probably yes
Wu 2019 (PS)	unable to assess	Low risk	low risk	probably yes
Xie 2017	low risk	N/A	some concerns	no
Xie 2019 (Early Educ & Dev)	unable to assess	N/A	Some concerns	Probably no
Xie 2019 (Journal of Child & Dev Studies)	unable to assess	N/A	some concerns	probably yes
Yibeltal 2025	Low risk	Low risk	Low risk	Probably no
Yuan 2022	low risk	low risk	low risk	probably yes
Yue 2019	low risk	unable to assess	low risk	yes
Yunilda 2022	unable to assess	unable to assess	low risk	probably no
Zakaria 2012	unable to assess	N/A	unable to assess	no
Zhang 2018	unable to assess	N/A	some concerns	Probably no
Zhang 2021	unable to assess	n/a	unable to assess	Probably yes
Zirakashvili 2018	unable to assess	Some concerns	Some concerns	Probably yes

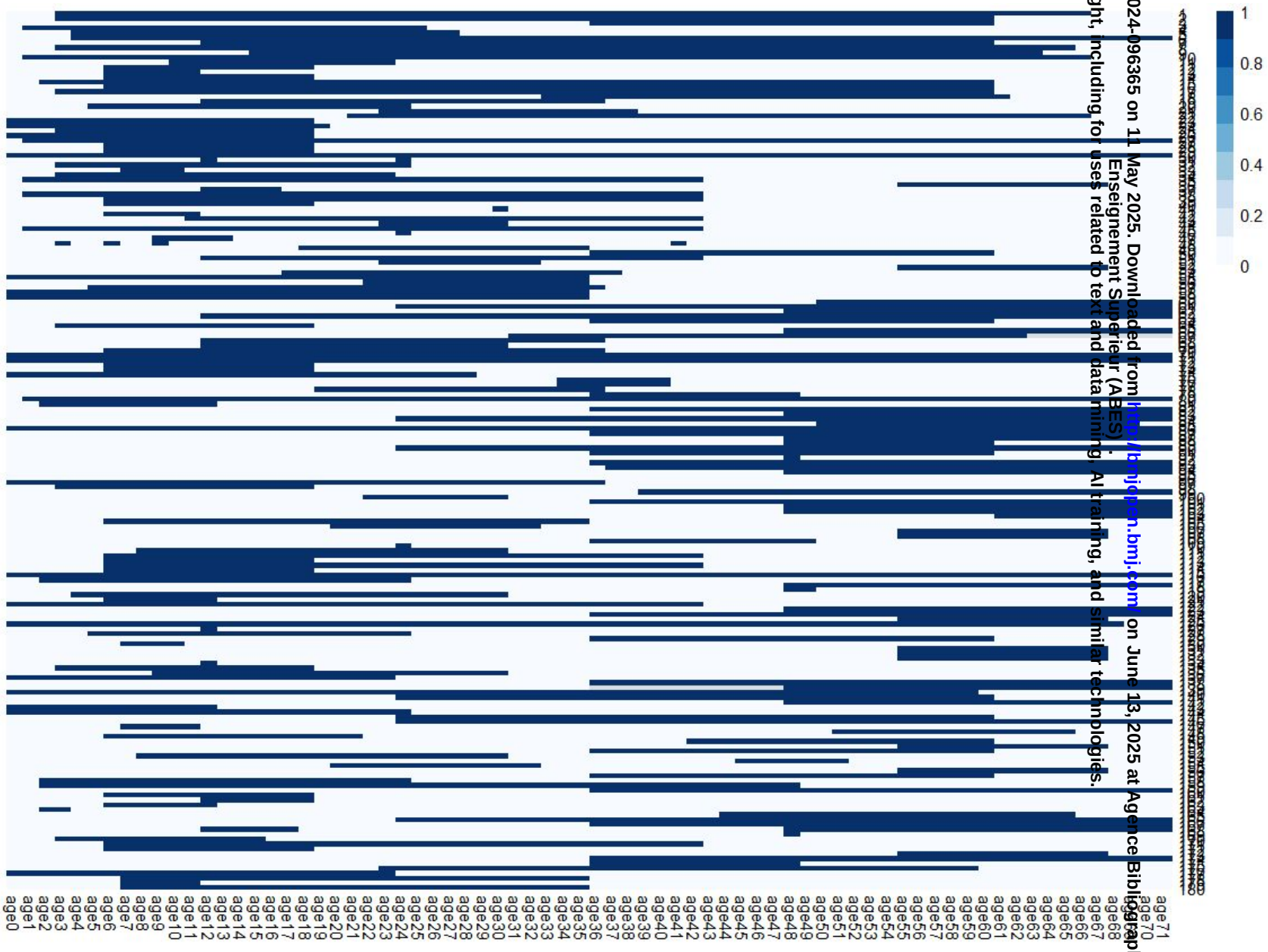
Supplemental Figure 1 Number of articles providing evidence on psychometrics by type of article



Supplemental Figure 2 Number of articles and tools providing psychometric evidence by child age group



Supplemental Figure 3 Number of articles providing evidence on psychometrics by child age at assessment



Note: Y-axis lists each article-tool combination included in the systematic review ordered by tool in alphabetical order.

References of all included articles

Abdoola, S., Swanepoel, D. W., Van Der Linde, J., & Glascoe, F. P. (2021). Detecting developmental delays in infants from a low-income South African community: comparing the BSID-III and PEDS tools. *Early Child Development and Care*, *191*(4), 545–554. https://doi.org/10.1080/03004430.2019.1628027

Abessa, T. G., Worku, B. N., Kibebew, M. W., Valy, J., Lemmens, J., Thijs, H., Yimer, W. K., Kolsteren, P., & Granitzer, M. (2016). Adaptation and standardization of a Western tool for assessing child development in non-Western low-income context. *BMC Public Health*, *16*(1), 1–13. https://doi.org/10.1186/s12889-016-3288-2

Abubakar, A., Holding, P., Van De Vijver, F., Bomu, G., & Van Baar, A. (2010). Developmental monitoring using caregiver reports in a resource-limited setting: The case of Kilifi, Kenya. *Acta Paediatrica, International Journal of Paediatrics*, *99*(2), 291–297. https://doi.org/10.1111/j.1651-2227.2009.01561.x

Abubakar, A., Holding, P., van Baar, A., Newton, C. R. J. C., & van de Vijver, F. J. R. (2008). Monitoring psychomotor development in a resource-limited setting: An evaluation of the Kilifi Developmental Inventory. *Annals of Tropical Paediatrics*, *28*(3), 217–226. https://doi.org/10.1179/146532808X335679

Albuquerque, P. L. de, Guerra, M. Q. de F., Lima, M. de C., & Eickmann, S. H. (2018). Concurrent validity of the Alberta Infant Motor Scale to detect delayed gross motor development in preterm infants: A comparative study with the Bayley III. *Developmental Neurorehabilitation*, *21*(6), 408–414. https://doi.org/10.1080/17518423.2017.1323974

Alcock, K. J., Rimba, K., Holding, P., Kitsao-Wekulo, P., Abubakar, A., & Newton, C. R. J. C. (2015). Developmental inventories using illiterate parents as informants: Communicative Development Inventory (CDI) adaptation for two Kenyan languages. *Journal of Child Language*, *42*(4), 763–785. https://doi.org/10.1017/S0305000914000403

Alderman, H., Friedman, J., Ganga, P., Kak, M., & Rubio-Codina, M. (2021). Assessing the performance of the Caregiver Reported Early Development Instruments (CREDI) in rural India. *Annals of the New York Academy of Sciences*, *1492*(1), 58–72. https://doi.org/10.1111/nyas.14543

Altafim, E. R. P., McCoy, D. C., Brentani, A., Escobar, A. M. de U., Grisi, S. J. F. E., & Fink, G. (2020). Measuring early childhood development in Brazil: validation of the Caregiver Reported Early Development Instruments (CREDI). *Jornal de Pediatria*, *96*(1), 66–75. https://doi.org/10.1016/j.jped.2018.07.008

Anunciação, L., Squires, J., Clifford, J., & Landeira-Fernandez, J. (2019). Confirmatory analysis and normative tables for the Brazilian Ages and Stages Questionnaires: Social–Emotional. *Child: Care, Health and Development*, *45*(3), 387–393. https://doi.org/10.1111/cch.12649

Azari, N., Soleimani, F., Vameghi, R., Sajedi, F., Shahshahani, S., Karimi, H., Kraskian, A., Shahrokhi, A., Teymouri, R., & Gharib, M. (2017). A psychometric study of the Bayley scales of infant and toddler development in Persian language children. *Iranian Journal of Child Neurology*, *11*(1), 50–56.

Bandeira, P. F. R., Duncan, M., Pessoa, M. L., Soares, Í., Da Silva, L., Mota, J., & Martins, C. (2020). Tgmd-2 short version: Evidence of validity and associations with sex, age, and bmi in preschool children. *Journal of Motor Learning and Development*, *8*(3), 528–543. https://doi.org/10.1123/JMLD.2019-0040

Barthel, D., Kriston, L., Guo, N., Yao, E. D., Armel, K. E. J., Fordjour, D., Mohammed, Y., Kotchi, C. E. B., Eberhardt, K. A., Hinz, R., Ehrhardt, S., & Bindt, C. (2022). Psychometric properties of the Brief Infant-Toddler Social and Emotional Assessment (BITSEA) in two-year-old Ivorian and Ghanaian children. *Applied Developmental Science*, *26*(3), 567–577. https://doi.org/10.1080/10888691.2021.1931216

Barthel, D., Kriston, L., Yao, E. D., Fordjour, D., Armel, K. E. J., Eberhardt, K. A., Hinz, R., Ehrhardt, S., & Bindt, C. (2022). Psychomotor development in two-year-old Ivorian and Ghanaian children – Psychometric properties of the Kilifi Developmental Inventory. *British Journal of Developmental Psychology*, *40*(4), 471–486. https://doi.org/10.1111/bjdp.12428

Bhavnani, S., Mukherjee, D., Dasgupta, J., Verma, D., Parameshwaran, D., Divan, G., Sharma, K. K., Thiagarajan, T., & Patel, V. (2019). Development, feasibility and acceptability of a gamified cognitive DEvelopmental assessment on an E-Platform (DEEP) in rural Indian pre-schoolers–a pilot study. *Global Health Action*, *12*(1). https://doi.org/10.1080/16549716.2018.1548005

Bian, X., Xie, H., Squires, J., & Chen, C. Y. (2017). Adapting a Parent-Completed, Socioemotional Questionnaire in China: the Ages & Stages Questionnaires: Social-Emotional. *Infant Mental Health Journal*, *38*(2), 258–266. https://doi.org/10.1002/imhj.21626

Bian, X., Yao, G., Squires, J., Hoselton, R., Chen, C. I., Murphy, K., Wei, M., & Fang, B. (2012). Translation and use of parent-completed developmental screening test in Shanghai. *Journal of Early Childhood Research*, *10*(2), 162–175. https://doi.org/10.1177/1476718X11430071

Biasini, F. J., De Jong, D., Ryan, S., Thorsten, V., Bann, C., Bellad, R., Mahantshetti, N. S., Dhaded, S. M., Pasha, O., Chomba, E., Goudar, S. S., Carlo, W. A., & McClure, E. (2015). Development of a 12 month screener based on items from the Bayley II Scales of Infant Development for use in Low Middle Income countries. *Early Human Development*, *91*(4), 253–258. https://doi.org/10.1016/j.earlhumdev.2015.02.001

Blackmon, K., Evans, R., Mohammed, L., Burgen, K. S., Ingraham, E., Punch, B., Isaac, R., Murray, T., Noel, J., Belmar-Roberts, C., Waechter, R., & Landon, B. (2024). The Grenada Learning and Memory Scale: Psychometric features and normative data in Caribbean

preschool children. Journal of the International Neuropsychological Society, 2024, 856–866. https://doi.org/10.1017/S1355617724000481

Boivin, M. J., Weiss, J., Chhaya, R., Seffren, V., Awadu, J., Sikorskii, A., & Giordani, B. (2017). The feasibility of automated eye tracking with the early childhood vigilance test of attention in younger HIV-exposed Ugandan children. *Neuropsychology*, *31*(5), 525–534. https://doi.org/10.1037/neu0000382

Boivin, M. J., Zoumenou, R., Sikorskii, A., Fievet, N., Alao, J., Davidson, L., Cot, M., Massougbodji, A., & Bodeau-Livinec, F. (2021). Neurodevelopmental assessment at one year of age predicts neuropsychological performance at six years in a cohort of West African Children. *Child Neuropsychology*, *27*(4), 548–571. https://doi.org/10.1080/09297049.2021.1876012

Bornman, J., Sevcik, R. A., Romski, M. A., & Pae, H. K. (2010). Successfully translating language and culture when adapting assessment measures. *Journal of Policy and Practice in Intellectual Disabilities*, *7*(2), 111–118. https://doi.org/10.1111/j.1741-1130.2010.00254.x

Brinkman, S. A., Kinnell, A., Maika, A., Hasan, A., Jung, H., & Pradhan, M. (2017). Validity and Reliability of the Early Development Instrument in Indonesia. *Child Indicators Research*, *10*(2), 331–352. https://doi.org/10.1007/s12187-016-9372-4

Caldera, A. ., Wickremasinghe, A. R., Muttiah, N., Godamunne, P. K. S., Jayasena, B. ., Chathurika, L. K. E., Perera, K. M. N., Mendis, M., Tilakarathne, D., Peiris, M. K. R. ., Wijesinghe, T., Senarathna, N. ., Saubhagya, W. D. L., Chandraratne, M., & Sumanasena, S. . (2023). REACh for the preschoolers; a developmental assessment tool for 2–5 year old children in Sri Lanka. *BMC Pediatrics*, *23*(1), 80. https://doi.org/10.1186/s12887-023-03895-5

Chen, C. Y., Xie, H., Filgueiras, A., Squires, J., Anunciação, L., & Landeira-Fernandez, J. (2017). Examining the Psychometric Properties of the Brazilian Ages & Stages Questionnaires-Social-Emotional: Use in Public Child Daycare Centers in Brazil. *Journal of Child and Family Studies*, *26*(9), 2412–2425. https://doi.org/10.1007/s10826-017-0770-0

Chunsuwan, I., Hansakunachai, T., & Pornsamrit, S. (2016). Parent Evaluation of Developmental Status (PEDS) in screening: The Thai experience. *Pediatrics International*, *58*(12), 1277–1283. https://doi.org/10.1111/ped.13055

Colbert, A. M., Connery, A. K., Lamb, M. M., Bauer, D., Olson, D., Paniagua-Avila, A., Martínez, M. A., Arroyave, P., Hernández, S., Mirella Calvimontes, D., Bolaños, G. A., El Sahly, H. M., Muñoz, F. M., & Asturias, E. J. (2021). Caregiver rating of early childhood development: Reliability and validity of the ASQ-3 in rural Guatemala. *Early Human Development*, *161*(November 2020). https://doi.org/10.1016/j.earlhumdev.2021.105453

Colbert, A. M., Lamb, M. M., Asturias, E. J., Muñoz, F. M., Bauer, D., Arroyave, P., Hernández, S., Martínez, M. A., Paniagua-Avila, A., Olson, D., Calvimontes, D. M., Bolaños, G. A., El

Sahly, H. M., & Connery, A. K. (2020). Reliability and Validity of an Adapted and Translated Version of the Mullen Scales of Early Learning (AT-MSEL) in Rural Guatemala. *Child: Care, Health and Development*, *46*(3), 327–335. https://doi.org/10.1111/cch.12748

Dagvadorj, A., Takehara, K., Bavuusuren, B., Morisaki, N., Gochoo, S., & Mori, R. (2015). The quick and easy Mongolian Rapid Baby Scale shows good concurrent validity and sensitivity. *Acta Paediatrica, International Journal of Paediatrics*, *104*(3), e94–e99. https://doi.org/10.1111/apa.12884

Danişman, Ş., Dereli Iman, E., Demircan, Z. A., & Yaya, D. (2016). Una aproximación al enfoque Reggio Emilia: Claves para entender por qué motiva a los estudiantes. *Electronic Journal of Research in Educational Psychology*, *14*(3), 534–556.

de Albuquerque, K. A., da Cunha, A. C. B., & Berkovits, M. D. (2022). Cross-cultural adaptation of the Battelle Development Inventory, 2nd Edition for Brazil. *Estudos de Psicologia*, *39*, 1–13. https://doi.org/10.1590/1982-0275202239e190170

Dias, T. P., Freitas, L. C., Del Prette, Z. A. P., & Del Prette, A. (2011). Validação da escala de comportamentos sociais de pré-escolares para o Brasil. *Psicologia Em Estudo*, *16*(3), 447–457. https://doi.org/10.1590/S1413-73722011000300012

Dursun, O. B., Öğütlü, H., & Esin, İ. S. (2020). Psychometric properties of Turkish version of the strength and difficulties questionnaire for age 2-4. *Noropsikiyatri Arsivi*, *57*(1), 44–49. https://doi.org/10.29399/npa.23674

Enelamah, N. V., Rao, S., Lombe, M., Yu, M., Newransky, C., Villodas, M. L., Foell, A., & Nebbitt, V. (2024). Examining the Utility of the Early Childhood Development Index (ECDI) among Children in the Nigeria Context. Children, 11(3). https://doi.org/10.3390/children11030361

Familiar-Lopez, I., Sikorskii, A., Chhaya, R., Holmes, A., Arima, E. G., Caesar, O. J., Nakasujja, N., & Boivin, M. J. (2023). Predictive validation of Ugandan infant eye-tracking test for memory of human faces. *Child Neuropsychology*, *29*(3), 486–502. https://doi.org/10.1080/09297049.2022.2099537

Fernandes, M., Bassani, D., Albernaz, E., Bertoldi, A. D., Silveira, M. F., Matijsevich, A., Anselmi, L., Cruz, S., Halal, C. S., Tovo-Rodrigues, L., Cruz, G. I. N., Metgud, D., & Santos, I. S. (2022). Construction and validation of the Oxford Neurodevelopment Assessment (OX-NDA) in 1-year-old Brazilian children. *BMC Pediatrics*, *22*(1), 1–12. https://doi.org/10.1186/s12887-022-03794-1

Ferreira-Vasques, A. T., Santos, C. F., & Lamônica, D. A. C. (2019). Transcultural adaptation process of the Griffiths-III Mental Development Scale. *Child: Care, Health and Development*, *45*(3), 403–408. https://doi.org/10.1111/cch.12664

Filgueiras, A., Pires, P., Maissonette, S., & Landeira-Fernandez, J. (2013). Psychometric properties of the Brazilian-adapted version of the Ages and Stages Questionnaire in public child daycare centers. *Early Human Development*, *89*(8), 561–576. https://doi.org/10.1016/j.earlhumdev.2013.02.005

Finch, J. E., Yousafzai, A. K., Rasheed, M., & Obradović, J. (2018). Measuring and understanding social-emotional behaviors in preschoolers from rural Pakistan. *PLoS ONE*, *13*(11), 1–22. https://doi.org/10.1371/journal.pone.0207807

Flórez Romero, R., Castro Martínez, J., & Arias Velandia, N. (2013). Propiedades psicométricas de la escala de lenguaje para preescolares (PLS-3) colombianos TT - Psychometric Properties of the Pre-school Language Scale (PLS-3) with Colombian Children TT - Propriedades psicométricas da Escada de linguagem para pré-escola. *Pensamiento Psicológico*, *11*(1), 131–140.

Gladstone, M., Lancaster, G. a., Umar, E., Nyirenda, M., Kayira, E., van den Broek, N. R., & Smyth, R. L. (2010). The Malawi Developmental Assessment Tool (MDAT): The Creation, Validation, and Reliability of a Tool to Assess Child Development in Rural African Settings. *PLoS Medicine*, *7*(5), e1000273. https://doi.org/10.1371/journal.pmed.1000273

Golmohammadi, G., Zarifian, T., Qhanbari, S., Bakhshi, E., & Sakhai, F. (2022). Cross-cultural adaptation and validation of the Persian version of Children's Behavior Questionnaire in Iranian children. *Current Psychology*, *41*(6), 4041–4053. https://doi.org/10.1007/s12144-020-00918-7

Gomez, J. A., Brown, J. L., & Spier, E. (2022). Factor structure, measurement invariance, and construct validity of MELQO measure of child development and early learning in Laos. *International Journal of School and Educational Psychology*, *10*(2), 181–207. https://doi.org/10.1080/21683603.2020.1791770

Guedes, D. Z., Primi, R., & Kopelman, B. I. (2011). BINS validation - Bayley neurodevelopmental screener in Brazilian preterm children under risk conditions. *Infant Behavior and Development*, *34*(1), 126–135. https://doi.org/10.1016/j.infbeh.2010.11.001

Guo, B., Aveyard, P., & Dai, X. (2009). The Chinese Intelligence Scale for Young Children. *Educational and Psychological Measurement*, *69*(3), 459–474. https://doi.org/10.1177/0013164409332209

Hahn, K. R., Schaefer, B. A., Merino, C., & Worrell, F. C. (2009). The factor structure of preschool learning behaviors scale scores in peruvian children. *Canadian Journal of School Psychology*, *24*(4), 318–331. https://doi.org/10.1177/0829573509349015

Halpin, P. F., de Castro, E. F., Petrowski, N., & Cappa, C. (2024). Monitoring early childhood development at the population level: The ECDI2030. Early Childhood Research Quarterly, 67(November 2023), 1–12. https://doi.org/10.1016/j.ecresq.2023.11.004

Halpin, P. F., Wolf, S., Yoshikawa, H., Rojas, N., Kabay, S., Pisani, L., & Dowd, A. J. (2019). Measuring early learning and development across cultures: Invariance of the IDELA across five countries. *Developmental Psychology*, *55*(1), 23–37. https://doi.org/10.1037/dev0000626

Hamadani, J. D., Baker-Henningham, H., Tofail, F., Mehrin, F., Huda, S. N., & Grantham-McGregor, S. M. (2010). Validity and Reliability of Mothers' Reports of Language Development in 1-Year-Old Children in a Large-Scale Survey in Bangladesh. *Food and Nutrition Bulletin*, *31*(2_suppl2), S198–S206. https://doi.org/10.1177/15648265100312S212

Hamadani, J. D., Tofail, F., Cole, T., & Grantham-McGregor, S. (2013). The relation between age of attainment of motor milestones and future cognitive and motor development in Bangladeshi children. *Maternal and Child Nutrition*, *9*(SUPPL. 1), 89–104. https://doi.org/10.1111/mcn.12020

Hanlon, C., Medhin, G., Worku, B., Tomlinson, M., Alem, A., Dewey, M., & Prince, M. (2016). Adapting the Bayley Scales of infant and toddler development in Ethiopia: evaluation of reliability and validity. *Child: Care, Health and Development*, *42*(5), 699–708. https://doi.org/10.1111/cch.12371

Hao, M., Shu, H., Xing, A., & Li, P. (2008). Early vocabulary inventory for Mandarin Chinese. *Behavior Research Methods*, *40*(3), 728–733. https://doi.org/10.3758/BRM.40.3.728

Hilburn, N., Potterton, J., Stewart, A., & BeckerC, P. (2011). The development of a screening tool to evaluate gross motor function in HIV-infected infants. *AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV*, *23*(12), 1619–1625. https://doi.org/10.1080/09540121.2011.579942

Hsiao, C., Richter, L., Makusha, T., Matafwali, B., van Heerden, A., & Mabaso, M. (2017). Use of the ages and stages questionnaire adapted for South Africa and Zambia. *Child: Care, Health and Development*, *43*(1), 59–66. https://doi.org/10.1111/cch.12413

Hua, J., Gu, G., Meng, W., & Wu, Z. (2013). Age band 1 of the Movement Assessment Battery for Children-Second Edition: Exploring its usefulness in mainland China. *Research in Developmental Disabilities*, *34*(2), 801–808. https://doi.org/10.1016/j.ridd.2012.10.012

Jackson-Maldonado, D., Marchman, V. A., & Fernald, L. C. H. (2013). Short-form versions of the Spanish MacArthur-Bates Communicative Development Inventories. *Applied Psycholinguistics*, *34*(4), 837–868. https://doi.org/10.1017/S0142716412000045

Jeong, J., McCann, J. K., Onyango, S., & Ochieng, M. (2025). Validation of the Global Scales of Early Development (GSED) tool in rural Western Kenya. BMC Public Health, 25(1). https://doi.org/10.1186/s12889-025-21801-9

Jianduan, Z., Huishan, W., Shuhua, S., Xiaonan, H., Guoyan, L., Guangli, L., & Junxin, S. (2009). Reliability and validity of standardized Chinese version of Urban Infant-Toddler Social and Emotional Assessment. *Early Human Development*, *85*(5), 331–336. https://doi.org/10.1016/j.earlhumdev.2008.12.012

Karabekiroglu, K., Rodopman-Arman, A., Ay, P., Ozkesen, M., Akbas, S., Tasdemir, G. N., Boke, O., & Peksen, Y. (2009). The reliability and validity of the Turkish version of the brief infant-toddler social emotional assessment (BITSEA). *Infant Behavior and Development*, *32*(3), 291–297. https://doi.org/10.1016/j.infbeh.2009.03.003

Kariuki, S. M., Abubakar, A., Murray, E., Stein, A., & Newton, C. R. J. C. (2016). Evaluation of psychometric properties and factorial structure of the pre-school child behaviour checklist at the Kenyan Coast. *Child and Adolescent Psychiatry and Mental Health*, *10*(1), 1–10. https://doi.org/10.1186/s13034-015-0089-9

Khan, N. Z., Muslima, H., Shilpi, A. B., Begum, D., Parveen, M., Akter, N., Ferdous, S., Nahar, K., McConachie, H., & Darmstadt, G. L. (2013). Validation of rapid neurodevelopmental assessment for 2- to 5-year-old children in bangladesh. *Pediatrics*, *131*(2). https://doi.org/10.1542/peds.2011-2421

Khan, N. Z., Muslima, H., Begum, D., Shilpi, A. B., Akhter, S., Bilkis, K., Begum, N., Parveen, M., Ferdous, S., Morshed, R., Batra, M., & Darmstadt, G. L. (2010). Validation of rapid neurodevelopmental assessment instrument for under-two-year-old children in Bangladesh. *Pediatrics*, *125*(4). https://doi.org/10.1542/peds.2008-3471

Knauer, H. A., Kariger, P., Jakiela, P., Ozier, O., & Fernald, L. C. H. (2019). Multilingual assessment of early child development: Analyses from repeated observations of children in Kenya. *Developmental Science*, *22*(5), 1–23. https://doi.org/10.1111/desc.12875

Koshy, B., Srinivasan, M., Scharf, R., Strand, T. A., Mohan, V. R., Beulah, R., John, S., Muliyil, J., & Kang, G. (2024). Stability and predictability of Bayley Scales of Infant and Toddler Development: evidence from a south Indian birth cohort prospective study. BMJ Open, 14(11), e082624. https://doi.org/10.1136/bmjopen-2023-082624

Koura, K. G., Boivin, M. J., Davidson, L. L., Ouédraogo, S., Zoumenou, R., Alao, M. J., Garcia, A., Massougbodji, A., Cot, M., & Bodeau-Livinec, F. (2013). Usefulness of child development assessments for low-resource settings in Francophone Africa. *Journal of Developmental and Behavioral Pediatrics*, *34*(7), 486–493. https://doi.org/10.1097/DBP.0b013e31829d211c

Kucuker, S., Kapci, E. G., & Uslu, R. I. (2011). Evaluation of the turkish version of the "ages and stages questionnaires: Social-emotional" in identifying children with social-emotional problems. *Infants and Young Children*, *24*(2), 207–220. https://doi.org/10.1097/IYC.0b013e31820eae26

Kumar, R., Shankar, K., Kush, V., Kumar, C., Bhave, A., & Agarwal, V. (2016). Adaptation: Vineland Adaptive Behavior Scale for 3-9 year-old Indian children. *International Journal on Disability and Human Development*, *15*(1), 49–55. https://doi.org/10.1515/ijdhd-2014-0026

Kvestad, I., Silpakar, J. S., Hysing, M., Ranjitkar, S., Strand, T. A., Schwinger, C., Shrestha, M., Chandyo, R. K., & Ulak, M. (2023). The reliability and predictive ability of the Test of Infant Motor Performance (TIMP) in a community-based study in Bhaktapur, Nepal. *Infant Behavior and Development*, *70*(January), 101809. https://doi.org/10.1016/j.infbeh.2023.101809

Kvestad, I., Taneja, S., Kumar, T., Bhandari, N., Strand, T. A., & Hysing, M. (2013). The assessment of developmental status using the Ages and Stages questionnaire-3 in nutritional research in north Indian young children. *Nutrition Journal*, *12*(1), 1–11. https://doi.org/10.1186/1475-2891-12-50

Lacunza, A. B., Castro Solano, A., & Contini, N. (2009). Habilidades sociales preescolares: una escala para niños de contextos de pobreza. *Revista de Psicología*, *27*(1), 3–28. https://doi.org/10.18800/psico.200901.001

Lara Díaz, M. F. L., Fonseca, Á. M. G., Bohórquez, D. M. G., Guechá, C. M., & Sellabona, E. S. (2011). Normativización del inventario del desarrollo comunicativo MacArthur-bates al Español, Colombia. *Revista Latinoamericana de Psicologia*, *43*(2), 241–254.

Leppänen, J. M., Butcher, J. W., Godbout, C., Stephenson, K., Taylor Hendrixson, D., Griswold, S., Rogers, B. L., Webb, P., Koroma, A. S., & Manary, M. J. (2022). Assessing infant cognition in field settings using eye-tracking: A pilot cohort trial in Sierra Leone. *BMJ Open*, *12*(2). https://doi.org/10.1136/bmjopen-2021-049783

Li, Y., Zhu, J. J., Coplan, R. J., Gao, Z. Q., Xu, P., Li, L., & Zhang, H. (2016). Assessment and Implications of Social Withdrawal Subtypes in Young Chinese Children: The Chinese Version of the Child Social Preference Scale. *Journal of Genetic Psychology*, *177*(3), 97–101. https://doi.org/10.1080/00221325.2016.1174100

Li, Y., Tang, L., Bai, Y., Zhao, S., & Shi, Y. (2020). Reliability and validity of the Caregiver Reported Early Development Instruments (CREDI) in impoverished regions of China. *BMC Pediatrics*, *20*(1). https://doi.org/10.1186/s12887-020-02367-4

Liu, J., Cheng, H., & Leung, P. W. L. (2011). The application of the preschool child behavior checklist and the Caregiver-Teacher Report Form to Mainland Chinese children: Syndrome structure, gender differences, country effects, and inter-informant agreement. *Journal of Abnormal Child Psychology*, *39*(2), 251–264. https://doi.org/10.1007/s10802-010-9452-8

Lohaus, A., Lamm, B., Keller, H., Teubert, M., Fassbender, I., Glüer, M., Borchert, S., Vöhringer, I., Teiser, J., Freitag, C., Suhrke, J., Knopf, M., & Schwarzer, G. (2014). Gross and Fine Motor Differences Between Cameroonian and German Children Aged 3 to 40

Months: Results of a Cross-Cultural Longitudinal Study. *Journal of Cross-Cultural Psychology*, *45*(8), 1328–1341. https://doi.org/10.1177/0022022114537703

Lopez Boo, F., Cubides Mateus, M., & Llonch Sabatés, A. (2020). Initial psychometric properties of the Denver II in a sample from Northeast Brazil. *Infant Behavior and Development*, *58*(August 2019), 101391. https://doi.org/10.1016/j.infbeh.2019.101391

Madaschi, V., Mecca, T. P., Macedo, E. C., & Paula, C. S. (2016). Bayley-III scales of infant and toddler development: Transcultural adaptation and psychometric properties. *Paideia*, *26*(64), 189–197. https://doi.org/10.1590/1982-43272664201606

Manandhar, S. R., Dulal, S., Manandhar, D. S., Saville, N., & Prost, A. (2016). Acceptability and Reliability of the Bayley Scales of Infant Development III Cognitive and Motor Scales among Children in Makwanpur. *Journal of Nepal Health Research Council*, *14*(32), 47–50.

McCarthy, A. M., Wehby, G. L., Barron, S., Aylward, G. P., Castilla, E. E., Javois, L. C., Goco, N., & Murray, J. C. (2012). Application of neurodevelopmental screening to a sample of South American infants: The Bayley Infant Neurodevelopmental Screener (BINS). *Infant Behavior and Development*, *35*(2), 280–294. https://doi.org/10.1016/j.infbeh.2011.12.003

McCoy, D. C., Sudfeld, C. R., Bellinger, D. C., Muhihi, A., Ashery, G., Weary, T. E., Fawzi, W., & Fink, G. (2017). Development and validation of an early childhood development scale for use in low-resourced settings. *Population Health Metrics*, *15*(1), 1–18. https://doi.org/10.1186/s12963-017-0122-8

McCoy, D. C., Waldman, M., & Fink, G. (2018). Measuring early childhood development at a global scale: Evidence from the Caregiver-Reported Early Development Instruments. *Early Childhood Research Quarterly*, *45*, 58–68. https://doi.org/10.1016/j.ecresq.2018.05.002

McCray, G., McCoy, D., Kariger, P., Janus, M., Black, M. M., Chang, S. M., Tofail, F., Eekhout, I., Waldman, M., Van Buuren, S., Khanam, R., Sazawal, S., Nizar, A., Schönbeck, Y., Zongo, A., Brentani, A., Zhang, Y., Dua, T., Cavallera, V., … Gladstone, M. (2023). The creation of the Global Scales for Early Development (GSED) for children aged 0-3 years: Combining subject matter expert judgements with big data. *BMJ Global Health*, *8*(1), 1–11. https://doi.org/10.1136/bmjgh-2022-009827

McHenry, M. S., Oyungu, E., Yang, Z., Hines, A. C., Ombitsa, A. R., Vreeman, R. C., Abubakar, A., & Monahan, P. O. (2021). Cultural adaptation of the Bayley Scales of Infant and Toddler Development, 3rd Edition for use in Kenyan children aged 18–36 months: A psychometric study. *Research in Developmental Disabilities*, *110*(December 2020), 103837. https://doi.org/10.1016/j.ridd.2020.103837

Mehrani, M. B. (2018). An elicited imitation test for measuring preschoolers' language development. *Psychological Reports*, *121*(4), 767–786. https://doi.org/10.1177/0033294117732976

Meyer, E. L., Schaefer, B. A., Soto, C. M., Simmons, C. S., Anguiano, R., Brett, J., Holman, A., Martin, J. F., Hata, H. K., Roberts, K. J., Mello, Z. R., & Worrell, F. C. (2011). Factor structure of child behavior scale scores in peruvian preschoolers. *Psychology in the Schools*, *48*(10), 931–942. https://doi.org/10.1002/pits.20596

Milosavljevic, B., Vellekoop, P., Maris, H., Halliday, D., Drammeh, S., Sanyang, L., Darboe, M. K., Elwell, C., Moore, S. E., & Lloyd-Fox, S. (2019). Adaptation of the Mullen Scales of Early Learning for use among infants aged 5- to 24-months in rural Gambia. *Developmental Science*, *22*(5), 1–17. https://doi.org/10.1111/desc.12808

Mukherjee, D., Bhavnani, S., Swaminathan, A., Verma, D., Parameshwaran, D., Divan, G., Dasgupta, J., Sharma, K., Thiagarajan, T. C., & Patel, V. (2020). Proof of Concept of a Gamified DEvelopmental Assessment on an E-Platform (DEEP) Tool to Measure Cognitive Development in Rural Indian Preschool Children. *Frontiers in Psychology*, *11*(June), 1–12. https://doi.org/10.3389/fpsyg.2020.01202

Mukherjee, S. B., Verma, S., Sharma, S., & Aneja, S. (2022). Diagnostic Accuracy of Parents' Evaluation of Developmental Status (PEDS), PEDS Developmental Milestones, and PEDS Combined in Indian Children Aged Less than 2 Years. *Indian Journal of Pediatrics*, *89*(5), 459–465. https://doi.org/10.1007/s12098-020-03651-y

Nair, M. K. C., Russell, P. S., Rekha, R. S., Lakshmi, M. A., Latha, S., & Rajee, K. (2009). Validation of developmental assessment tool for Anganwadis (DATA). *Indian Pediatrics*, *46 Suppl*(2), s27-36.

Nair, M. K. C., & Russell, P. S. (2013). Development and normative validation of Developmental Assessment Tool for Anganwadis for 3- to 4-year-old children (DATA-II). *Journal of Clinical Epidemiology*, *66*(1), 23–29. https://doi.org/10.1016/j.jclinepi.2012.06.011

Nampijja, M., Apule, B., Lule, S., Akurut, H., Muhangi, L., Elliott, A. M., & Alcock, K. J. (2010). Adaptation of western measures of cognition for assessing 5-year-old semi-urban Ugandan children. *British Journal of Educational Psychology*, *80*(1), 15–30. https://doi.org/10.1348/000709909X460600

Ngoun, C., Stoey, L. S., van't Ende, K., & Kumar, V. (2012). Creating a Cambodia-specific developmental milestone screening tool - A pilot study. *Early Human Development*, *88*(6), 379–385. https://doi.org/10.1016/j.earlhumdev.2011.09.014

Obradović, J., Finch, J. E., Portilla, X. A., Rasheed, M. A., Tirado-Strayer, N., & Yousafzai, A. K. (2019). Early executive functioning in a global context: Developmental continuity and family protective factors. *Developmental Science*, *22*(5). https://doi.org/10.1111/desc.12795

Okuda, P. M. M., Félix, E., Cogo-Moreira, H., Liu, T., Surkan, P. J., Martins, S. S., & Caetano, S. C. (2020). Construct validity of the motor development scale (MDS). *Trends in Psychiatry and Psychotherapy*, *42*(1), 82–85. https://doi.org/10.1590/2237-6089-2018-0114

Ortiz-León, S., Granados-Rojas, A., Cavazos-Olivo, J., Benito-Avendaño, L. De, Obregón-García, V. H., Duran-Avendaño, X. M., Ramírez-Pérez, R. V., Cárdenas-Medina, J. A., Vargas-Terréz, B. E., Luna-Guevara, G., Rodríguez-Machain, A. C., & Jaimes-Medrano, A. L. (2018). Internal and inter-rater reliability of the ASQ-3 in Mexican preschoolers. *Salud Mental*, *41*(2), 65–72. https://doi.org/10.17711/SM.0185-3325.2018.011

Oryadi-Zanjani, M. M. (2020). Development of the Childhood Nonverbal Communication Scale. *Journal of Autism and Developmental Disorders*, *50*(4), 1238–1248. https://doi.org/10.1007/s10803-019-04356-8

Parveen, M., Rahman, S. T., Islam, S., Zaman, S. S., Hamadani, J., & Khan, N. Z. (2014). Adaptation of items of bayley scales of infant development-II (BSID-II) suitable for Bangladeshi infants. *Dhaka University Journal of Biological Sciences*, *23*(2), 187–195. https://doi.org/10.3329/dujbs.v23i2.20099

Pathak, S., & Sovani-Kelkar, P. (2023). Development of a screening tool to identify babies at risk of language delay in India: A preliminary study. *Language Acquisition*, *30*(1), 29–49. https://doi.org/10.1080/10489223.2022.2078212

Pendergast, L. L., Schaefer, B. A., Murray-Kolb, L. E., Svensen, E., Shrestha, R., Rasheed, M. A., Scharf, R. J., Kosek, M. N., Vasquez, A. O., Maphula, A., Costa, H., Rasmussen, Z. A., Yousafzai, A., Tofail, F., Seidman, J. C., Acosta, A. M., de Burga, R. R., Chavez, C. B., Flores, J. T., … Strand, T. (2018). Assessing development across cultures: Invariance of the Bayley-III scales across seven international MAL-ED sites. *School Psychology Quarterly*, *33*(4), 604–614. https://doi.org/10.1037/spq0000264

Phạm, B., McLeod, S., & Harrison, L. J. (2017). Validation and norming of the Intelligibility in Context Scale in Northern Viet Nam. *Clinical Linguistics and Phonetics*, *31*(7–9), 665–681. https://doi.org/10.1080/02699206.2017.1306110

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). *International Journal of Educational Research*, *91*(July), 1–15. https://doi.org/10.1016/j.ijer.2018.06.007

Pisani, L., Seiden, J., & Wolf, S. (2022). Longitudinal evidence on the predictive validity of the International Development and Early Learning Assessment (IDELA). *Educational Assessment, Evaluation and Accountability*, *34*(2), 173–194. https://doi.org/10.1007/s11092-021-09374-8

Pitchik, H. O., Tofail, F., Akter, F., Shoab, A. K. M., Sultana, J., Huda, T. M. N., Rahman, M., Winch, P. J., Luby, S. P., & Fernald, L. C. H. (2023). Concurrent validity of the Ages and Stages Questionnaire Inventory and the Bayley Scales of Infant and Toddler Development in rural Bangladesh. *BMC Pediatrics*, *23*(1), 1–11. https://doi.org/10.1186/s12887-022-03800-6

Prado, E. L., Phuka, J., Ocansey, E., Maleta, K., Ashorn, P., Ashorn, U., Adu-Afarwuah, S., Oaks, B. M., Lartey, A., & Dewey, K. G. (2018). A method to develop vocabulary checklists in new languages and their validity to assess early language development. *Journal of Health, Population, and Nutrition*, *37*(1), 13. https://doi.org/10.1186/s41043-018-0145-1

Prathanee, B., Pongjanyakul, A., & Chano, J. (2008). Thai speech and language test for children between 1 and 2 years of age. *International Journal of Language and Communication Disorders*, *43*(1), 125–140. https://doi.org/10.1080/13682820601181208

Puglisi, M. L., Blasi, H. F., & Snowling, M. J. (2020). Screening for the identification of oral language difficulties in brazilian preschoolers: A validation study. *Language, Speech, and Hearing Services in Schools*, *51*(3), 852–865. https://doi.org/10.1044/2020_LSHSS-19-00083

Ranjitkar, S., Kvestad, I., Strand, T. A., Ulak, M., Shrestha, M., Chandyo, R. K., Shrestha, L., & Hysing, M. (2018). Acceptability and reliability of the Bayley Scales of infant and Toddler Development-III Among Children in Bhaktapur, Nepal. *Frontiers in Psychology*, *9*(JUL), 1–10. https://doi.org/10.3389/fpsyg.2018.01265

Rasheed, M. A., Tabussum, G., Habib, Z. H., & Yousafzai, A. K. (2017). Adaptation of MacArthur Communicative Development Inventory in rural Pakistan – useful tool for early childhood studies. *Child: Care, Health and Development*, *43*(3), 427–434. https://doi.org/10.1111/cch.12450

Rasheed, M. A., Kvestad, I., Shaheen, F., Memon, U., & Strand, T. A. (2023). The predictive validity of Bayley Scales of Infant and Toddler Development-III at 2 years for later general abilities: Findings from a rural, disadvantaged cohort in Pakistan. *PLOS Global Public Health*, *3*(1), e0001485. https://doi.org/10.1371/journal.pgph.0001485

Rasheed, M. A., Pham, S., Memon, U., Siyal, S., Obradović, J., & Yousafzai, A. K. (2018). Adaptation of the Wechsler Preschool and Primary Scale of Intelligence-III and lessons learned for evaluating intelligence in low-income settings. *International Journal of School and Educational Psychology*, *6*(3), 197–207. https://doi.org/10.1080/21683603.2017.1302851

Razak, R. A., Madison, C. L., Siow, Y. K., & Aziz, M. A. A. (2010). Preliminary Content Validity and Reliability of a Newly Developed Malay Preschool Language Assessment Tool. *Asia Pacific Journal of Speech, Language and Hearing*, *13*(4), 217–234. https://doi.org/10.1179/136132810805334985

Rodriguez, V. J., Zegarac, M., La Barrie, D. L., Parrish, M. S., Matseke, G., Peltzer, K., & Jones, D. L. (2020). Validation of the Bayley Infant Neurodevelopmental Screener among HIV-Exposed Infants in Rural South Africa. *Journal of Acquired Immune Deficiency Syndromes*, *85*(4), 507–516. https://doi.org/10.1097/QAI.0000000000002479

Rubio-Codina, M., Araujo, M. C., Attanasio, O., Muñoz, P., & Grantham-McGregor, S. (2016). Concurrent validity and feasibility of short tests currently used to measure early childhood development in large scale studies. *PLoS ONE*, *11*(8), 1–17. https://doi.org/10.1371/journal.pone.0160962

Rubio-Codina, M., & Grantham-McGregor, S. (2020). Predictive validity in middle childhood of short tests of early childhood development used in large scale studies compared to the Bayley-III, the Family Care Indicators, height-for-age, and stunting: A longitudinal study in Bogota, Colombia. *PLoS ONE*, *15*(4), 1–20. https://doi.org/10.1371/journal.pone.0231317

Saccani, R., Valentini, N. C., & Pereira, K. R. G. (2016). New Brazilian developmental curves and reference values for the Alberta infant motor scale. *Infant Behavior and Development*, *45*, 38–46. https://doi.org/10.1016/j.infbeh.2016.09.002

Sallum, I., da Mata, F. G., Cheib, N. F., Mathias, C. W., Miranda, D. M., & Malloy-Diniz, L. F. (2017). Development of a version of the self-ordered pointing task: a working memory task for Brazilian preschoolers. *Clinical Neuropsychologist*, *31*(2), 459–470. https://doi.org/10.1080/13854046.2016.1275818

Samarakkody, D. C., Fernando, D. N., Perera, H., McClure, R. J., & De Silva, H. (2010). The Child Behaviour Assessment Instrument: Development and validation of a measure to screen for externalising child behavioural problems in community setting. *International Journal of Mental Health Systems*, *4*, 1–8. https://doi.org/10.1186/1752-4458-4-13

Sánchez-Vincitore, L. V., Schaettle, P., & Castro, A. (2019). Validation of the Malawi Developmental Assessment Tool for children in the Dominican Republic: Preliminary results. *PLoS ONE*, *14*(8), 1–11. https://doi.org/10.1371/journal.pone.0221162

Scherzer, A. L. (2009). Experience in Cambodia with the use of a culturally relevant developmental milestone chart for children in low- and middle-income countries. *Journal of Policy and Practice in Intellectual Disabilities*, *6*(4), 287–292. https://doi.org/10.1111/j.1741-1130.2009.00234.x

Shah, M. B., Schaefer, B. A., & Clark, T. P. (2013). Reliability of Bracken School Readiness Assessment, Third Edition Scores With Young Children in Mumbai, India. *International Journal of School and Educational Psychology*, *1*(1), 58–66. https://doi.org/10.1080/21683603.2013.782595

Shaoli, S. S., Islam, S., Haque, S., & Islam, A. (2019). Validating the Bangla version of the Dimensions of Mastery Questionnaire (DMQ-18) for preschoolers. *Asian Journal of Psychiatry*, *44*(May 2019), 143–149. https://doi.org/10.1016/j.ajp.2019.07.044

Shrestha, M., Kvestad, I., Hysing, M., Ranjitkar, S., Ulak, M., Chandyo, R. K., & Strand, T. A. (2024). The relationship between the ages and stages questionnaire, 3rd edition scores in early childhood and future cognitive abilities in young Nepalese children. BMC Pediatrics, 24(1), 642. https://doi.org/10.1186/s12887-024-05112-3

Shrestha, M., Strand, T. A., Ulak, M., Chandyo, R. K., Ranjitkar, S., Hysing, M., Shrestha, L., & Kvestad, I. (2019). The feasibility of the Ages and Stages Questionnaire for the assessment of child development in a community setting in Nepal. *Child: Care, Health and Development*, *45*(3), 394–402. https://doi.org/10.1111/cch.12654

Siegle, C. B. H., & dos Santos Cardoso de Sá, C. (2018). Concurrent validity between instruments of assessment of motor development in infants exposed to HIV. *Infant Behavior and Development*, *50*(February), 198–206. https://doi.org/10.1016/j.infbeh.2018.01.005

Sincovich, A., Gregory, T., Zanon, C., Santos, D. D., Lynch, J., & Brinkman, S. A. (2019). Measuring early childhood development in multiple contexts: The internal factor structure and reliability of the early Human Capability Index in seven low and middle income countries. *BMC Pediatrics*, *19*(1), 1–14. https://doi.org/10.1186/s12887-019-1852-5

Snelling, M., Dawes, A., Biersteker, L., Girdwood, E., & Tredoux, C. (2019). The development of a South African Early Learning Outcomes Measure: A South African instrument for measuring early learning program outcomes. *Child: Care, Health and Development*, *45*(2), 257–270. https://doi.org/10.1111/cch.12641

Soleimani, F., & Dadkhah, A. (2007). Validity and reliability of infant Neurological International Battery for detection of gross motor developmental delay in Iran. *Child: Care, Health and Development*, *33*(3), 262–265. https://doi.org/10.1111/j.1365-2214.2006.00704.x

Soleymani, Z., Nematzadeh, S., Tehrani, L. G., Rahgozar, M., & Schneider, P. (2016). Language sample analysis: development of a valid language assessment tool and determining the reliability of outcome measures for Farsi-speaking children. *European Journal of Developmental Psychology*, *13*(2), 275–291. https://doi.org/10.1080/17405629.2016.1152175

Solís Cordero K, Quezada Ugalde AM, González Serrano S, Zúñiga Baldi C. Costa Rican test: Evaluation of the children's integral development from 0 months to 6 years of age (EDIN II). Enfermería Actual Costa Rica 2019; 37. DOI:http://dx.doi.org/10.15517/revenf.v0ino.37.34959.

Soysal, A. S., Gucuyener, K., Ergenekon, E., Turan, Ö., Koc, E., Turkylmaz, C., Önal, E., & Atalay, Y. (2014). The prediction of later neurodevelopmental status of preterm infants at ages 7 to 10 years using the bayley infant neurodevelopmental screener. *Journal of Child Neurology*, *29*(10), 1349–1355. https://doi.org/10.1177/0883073813520495

Thompson, L., Peñaloza, R. A., Stormfields, K., Kooistra, R., Valencia-Moscoso, G., Muslima, H., & Khan, N. Z. (2015). Validation and adaptation of rapid neurodevelopmental assessment instrument for infants in Guatemala. *Child: Care, Health and Development*, *41*(6), 1131–1139. https://doi.org/10.1111/cch.12279

Toor, H. K., & Hanif, R. (2022). Executive Function Touch battery: Translation and preliminary measure validation for Pakistani preschoolers. *PLoS ONE*, *17*(9 September), 1–14. https://doi.org/10.1371/journal.pone.0274431

Uyanik, Ö., & Kandir, A. (2014). Adaptation of the Kaufman Survey of Early Academic and Language Skills to Turkish Children Aged 61 to 72 Months. *Kuram ve Uygulamada Egitim Bilimleri*, *14*(2), 682–692. https://doi.org/10.12738/estp.2014.2.1682

Valentini, N. C., & Saccani, R. (2012). Brazilian Validation of the Alberta Infant Motor Scale. *Physical Therapy*, *92*(3), 440–447. https://doi.org/10.2522/ptj.20110036

Vameghi, R., Sajedi, F., Kraskian Mojembari, A., Habiollahi, A., Lornezhad, H. R., & Delavar, B. (2013). Cross-cultural adaptation, validation and standardization of Ages and Stages Questionnaire (ASQ) in Iranian children. *Iranian Journal of Public Health*, *42*(5), 522–528.

van Der Linde, J., Swanepoel, D. W., Glascoe, F. P., Louw, E. M., & Vinck, B. (2015). Developmental screening in South Africa: Comparing the national developmental checklist to a standardized tool. *African Health Sciences*, *15*(1), 188–196. https://doi.org/10.4314/ahs.v15i1.25

van Heerden, A., Hsiao, C., Matafwali, B., Louw, J., & Richter, L. (2017). Support for the feasibility of the ages and stages questionnaire as a developmental screening tool: A cross-sectional study of South African and Zambian children aged 2-60 months. *BMC Pediatrics*, *17*(1), 1–9. https://doi.org/10.1186/s12887-017-0802-3

Venancio, S. I., Bortoli, M. C., Frias, P. G., Giugliani, E. R. J., Alves, C. R. L., & Santos, M. O. (2020). Development and validation of an instrument for monitoring child development indicators. *Jornal de Pediatria (Versão Em Português)*, *96*(6), 778–789. https://doi.org/10.1016/j.jpedp.2019.10.008

Venter, A., Pienaar, A. E., & Coetzee, D. (2015). Suitability of the 'Little DCDQ'' for the identification of DCD in a selected group of 3–5-year-old South African children.' *Early Child Development and Care*, *185*(8), 1359–1371. https://doi.org/10.1080/03004430.2014.1000887

Waechter, R., Evans, R., Hanna, S., Murray, T., Mobley, C., Holmes, S., Isaac, R., Wolfe, R., Andrew, E., Landon, B., & Fernandes, M. (2022). Adaptation of the INTERGROWTH-21st neurodevelopment assessment (INTER-NDA) to the context of the English-speaking Caribbean. *BMC Pediatrics*, *22*(1), 1–10. https://doi.org/10.1186/s12887-021-03039-7

Waldman, M., McCoy, D. C., Seiden, J., Cuartas, J., Fink, G., Altafim, E., Brentani, A., Castellanos, A., Chen, A., Chomat, A. M., Fawzi, W., de Pin˜eres, C. G., Hamadani, J., Henao, N., Jervis, P., Kane, C., Measelle, J., Medrano, P., Pisani, L., … Yousafzai, A. K. (2021). Validation of motor, cognitive, language, and socio-emotional subscales using the Caregiver Reported Early Development Instruments: An application of multidimensional

item factor analysis. *International Journal of Behavioral Development*, *45*(4), 368–377. https://doi.org/10.1177/01650254211005560

Washington, K. N., McDonald, M. M., McLeod, S., Crowe, K., & Devonish, H. (2017). Validation of the intelligibility in context scale for Jamaican Creole-speaking preschoolers. *American Journal of Speech-Language Pathology*, *26*(3), 750–761. https://doi.org/10.1044/2016_AJSLP-15-0103

Weber, A. M., Marchman, V. A., DIop, Y., & Fernald, A. (2018). Validity of caregiver-report measures of language skill for Wolof-learning infants and toddlers living in rural African villages. *Journal of Child Language*, *45*(4), 939–958. https://doi.org/10.1017/S0305000917000605

Wijedasa, D. (2012). Developmental screening in context: Adaptation and standardization of the Denver Developmental Screening Test-II (DDST-II) for Sri Lankan children. *Child: Care, Health and Development*, *38*(6), 889–899. https://doi.org/10.1111/j.1365-2214.2011.01332.x

Willoughby, M. T., Piper, B., Kwayumba, D., & McCune, M. (2019). Measuring executive function skills in young children in Kenya. *Child Neuropsychology*, *25*(4), 425–444. https://doi.org/10.1080/09297049.2018.1486395

Willoughby, M. T., Piper, B., Oyanga, A., & Merseth King, K. (2019). Measuring executive function skills in young children in Kenya: Associations with school readiness. *Developmental Science*, *22*(5). https://doi.org/10.1111/desc.12818

Wolf, S., Halpin, P., Yoshikawa, H., Dowd, A. J., Pisani, L., & Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia. *Early Childhood Research Quarterly*, *41*(June), 21–36. https://doi.org/10.1016/j.ecresq.2017.05.001

Wu, Z., Hu, B. Y., & Fan, X. (2019). Cross-Cultural Validity of Preschool Learning Behavior Scale in Chinese Cultural Context. *Journal of Psychoeducational Assessment*, *37*(1), 125–130. https://doi.org/10.1177/0734282916651538

Wu, Z., Mak, M. C. K., Hu, B. Y., He, J., & Fan, X. (2019). A validation of the Social Skills domain of the Social Skills Improvement System-Rating Scales with Chinese preschoolers. *Psychology in the Schools*, *56*(1), 126–147. https://doi.org/10.1002/pits.22193

Xie, H., Bian, X., Chen, C. Y., Squires, J., & Lu, P. (2019). Examining the Convergent Evidence of a Parent-Completed, Social-Emotional Screening Tool in China. *Journal of Child and Family Studies*, *28*(6), 1471–1480. https://doi.org/10.1007/s10826-019-01362-3

Xie, H., Clifford, J., Squires, J., Chen, C. Y., Bian, X., & Yu, Q. (2017). Adapting and validating a developmental assessment for chinese infants and toddlers: The ages & stages

questionnaires: Inventory. *Infant Behavior and Development*, *49*(June), 281–295. https://doi.org/10.1016/j.infbeh.2017.09.004

Xie, S., & Li, H. (2019). Development and Validation of the Chinese Preschool Readiness Scale. *Early Education and Development*, *30*(4), 522–539. https://doi.org/10.1080/10409289.2019.1596462

Yibeltal, K., Teklehaimanot, A., Workneh, F., Fasil, N., Jensen, S. K. G., Chin, T. I., North, K., Haymanot, B., Worku, A., Lee, A. C., & Berhane, Y. (2025). Adaptation and Implementation of the Global Scales for Early Development (GSED) Tool in Ethiopia. Children, 12(3), 299. https://doi.org/10.3390/children12030299

Yuan, H., Ocansey, M., Adu-Afarwuah, S., Sheridan, M., Hamoudi, A., Okronipa, H., Kumordzie, S. M., Oaks, B. M., & Prado, E. (2022). Evaluation of a tablet-based assessment tool for measuring cognition among children 4–6 years of age in Ghana. *Brain and Behavior*, *12*(10), 1–12. https://doi.org/10.1002/brb3.2749

Yue, A., Jiang, Q., Wang, B., Abbey, C., Medina, A., Shi, Y., & Rozelle, S. (2019). Concurrent validity of the Ages and Stages Questionnaire and the Bayley Scales of Infant Development III in China. *PLoS ONE*, *14*(9). https://doi.org/10.1371/journal.pone.0221675

Yunilda, E., Gunardi, H., Medise, B. E., & Oswari, H. (2023). The Indonesian version of Ages and Stages Questionnaire III accuracy compared to Bayley Scales of Infant Development III. *Infant and Child Development*, *32*(1), 1–12. https://doi.org/10.1002/icd.2387

Zakaria, S., Chua, B. S., Sombuling, A., Ahmad, S., & Hashmi, S. I. (2012). Reliability and Validity for Malay Version of Bayley Scales of Infant and Toddler Development-Third Edition (Bayley-III): Preliminary Study. *International Proceedings of Economics Development and Research*, *40*(June), 116–119.

Zhang, L., Li, Z., & He, K. (2018). Validation of the Chinese Revision of Emotional and Behavioral Screener among Preschoolers (4–7 years old). *Journal of Child and Family Studies*, *27*(3), 771–779. https://doi.org/10.1007/s10826-017-0940-0

Zhang, Y., Xu, X., Jiang, Y., Sun, W., Wang, Y., Song, Y., Dong, S., Zhu, Q., Jiang, F., & Sheng, L. (2021). Early language and communication development in Chinese children: Adaption and validation of a parent report instrument. *International Journal of Speech-Language Pathology*, *23*(4), 372–381. https://doi.org/10.1080/17549507.2020.1817558

Zirakashvili, M., Gabunia, M., Tatishvili, N., Ediberidze, T., Lomidze, G., Chachava, T., & Hix-Small, H. (2018). Cultural Adaptation and Psychometric Validation of the Ages and Stages Questionnaires for use in Georgia. *Journal of Child and Family Studies*, *27*(3), 739–749. https://doi.org/10.1007/s10826-017-0917-z