psychology international

MDPI

Systematic Review

Assessing Childhood Development: Systematic Review and Meta-Analysis on the Validation of Local Assessment Tools in the Context of Developing Countries

Seep Lassi [1], Maira Niaz [1], Zoya Navid Ansari [1], Hamza Iftikar [1], Shanzay Rizvi [1], Hamna Amir [1], Zain Hasnain [1], Sidra Kaleem Jafri [2] and Jai K. Das [1,2,*]

[1] Institute for Global Health and Development, Aga Khan University, Karachi 74800, Pakistan
[2] Department of Paediatrics and Child Health, Aga Khan University, Karachi 74800, Pakistan
* Correspondence: jai.das@aku.edu

Abstract

Background: Accurate child development assessment is crucial, particularly in developing countries where access to validated tools remains limited. Many assessment tools are adapted for local contexts, but their psychometric properties require evaluation. Objective: This systematic review examines the reliability, validity, and overall psychometric properties of new and adapted child development assessment tools used in developing countries. The focus on these settings stems from the need to assess tools that are culturally appropriate, feasible, and accurate in resource-constrained environments, where early identification of developmental delays can significantly impact long-term child outcomes. Methods: Descriptive and meta-analyses were conducted to synthesize findings from eligible studies. Psychometric properties such as internal consistency, inter-rater reliability, construct validity, sensitivity and specificity were assessed. This review is registered on Open Science Framework (OSF) doi:10.17605/OSF.IO/GU28K. Results: The findings indicate that although some adapted tools demonstrate strong reliability and validity, others exhibit inconsistencies, highlighting challenges in adaptation. The meta-analysis provided pooled estimates of key psychometric properties with a net sensitivity and specificity of 0.859 and 0.805, respectively, illustrating the validity of these local tools but also variability in performance across different tools. Conclusion: The results emphasize the need for rigorous validation processes to ensure that adapted tools maintain their psychometric integrity. Future research should focus on refining these measures to improve their applicability in diverse cultural and socioeconomic settings.

Keywords: childhood development; neurodevelopmental disorders; developmental tools; low- and middle-income countries (LMICs); psychometric properties

Check for updates

Academic Editor: Marianna Mazza

Received: 13 March 2026
Revised: 18 May 2026
Accepted: 24 May 2026
Published: 5 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Children in developing countries face significant challenges as they grow up, often in environments that lack adequate support, leading to a widened gap between children in low- and middle-income countries (LMICs) and their peers in high-income countries (HICs). Such disparity puts them at a disadvantage in reaching their full potential. Over 250 million children under the age of five worldwide are estimated to be living in poverty or are stunted, putting them at risk of not reaching their full potential in terms of physical growth, cognitive abilities, and social–emotional development (Black et al., 2017). For children with

neurodevelopmental disorders (NDDs), early identification and intervention is incredibly important for managing symptoms and improving outcomes (Pallanti & Salerno, 2023). The burden of neurodevelopmental disorders (NDD) persists as a significant public health concern in low- and middle-income countries (LMICs). Neurodevelopmental disorders are complex and diverse conditions that impact brain development and can cause impairment in cognitive, motor, communication, adaptive domains, and activities of daily living (ADLs) (American Psychiatric Association, n.d.; Pallanti & Salerno, 2023). These disorders manifest in the early years of life and hinder the typical growth of a child by prominently impairing social, personal, academic, and occupational functioning.

The transition from infancy to adolescence is marked by significant psychological changes across various domains, including cognitive, behavioral, physiological, and social–emotional development. Failure to meet age-specific milestones may lead to a classification of developmental delays (I. Khan & Leventhal, 2023). Several factors contributing to developmental delay include birth complications, brain injury, medical issues, genetic abnormalities, lack of stimulation, malnutrition, anemia, chronic illnesses, and adverse environmental or familial conditions. To measure growth and learning in children, various tools have been developed; however, most of these tools have been primarily focused on HICs (Angrist et al., 2021) and developed in the English language (Beaton et al., 2000). They are typically translated into local languages to adapt to various contexts, but mere translation does not guarantee validity to a specific context. Without adequate cultural adaptation, these tools may lack the applicability and effectiveness in local contexts, hence compromise the validity of the test findings (Gjersing et al., 2010).

Despite these concerns, developing new tests with robust psychometric properties and establishing normative standards is an expensive and time-consuming process, especially in LMICs where resources are frequently limited (Geisinger, 1994). Some progress has occurred as tools have been developed and tested in LICs and LMICs, but the need for a systematic review to validate these locally adapted tools with standard measures remains. There is a significant need for evidence-based, cost-effective, and culturally appropriate screening tools for the early identification of developmental disorders. Several gaps in the availability and applicability of neurodevelopmental tools in LICs, LMICs and UMICs have been identified (Geisinger, 1994). The gaps included the lack of standardized tools, limited access to trained professionals, and cultural and linguistic barriers.

Some systematic reviews have highlighted the urgent demand for culturally appropriate screening tools for neurodevelopmental disorders, specifically evaluating their reliability and validity (Huda et al., 2024; Semrud-Clikeman et al., 2017). However, a scoping review further denoted developmental tools such as the Ages and Stages Questionnaire, third edition (ASQ-3), Bayley Scales of Infant and Toddler Development, third edition (Bayley-III) and Social Attention and Communication Surveillance-Revised (SACS-R) that may be effectively utilized across multiple cultures, as proven by their predictive or discriminative ability (Burgess et al., 2025). Various studies have been conducted to validate local tools, showing mixed patterns in their reliability, validity, and effectiveness. There is a lack of systematic reviews evaluating the net sensitivity and specificity of the various culturally adapted screening tools. Hence, this review aims to conduct a meta-analysis to assess the strengths and weaknesses of local neurodevelopmental tools in developing countries.

2. Materials and Methods

2.1. Objectives

This systematic review was conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for diagnostic test accuracy (McInnes et al., 2018). It is registered on Open Science Framework (OSF) (Lassi et al., 2026).

The primary objectives were to assess the validity of various local neurodevelopmental tools used in developing countries against the global gold standard tools and assess the strengths and weaknesses of locally developed neurodevelopmental tools.

2.2. Eligibility Criteria

2.2.1. Study Selection

Studies were included if they focused on developing, modifying, and validating a chosen instrument as a key objective. The selection criteria focused on studies that evaluated locally adapted tools against established gold standard measures and addressing neurodevelopment, intelligence (IQ) assessment, developmental screening, and the gaps and relevance of these tools within low-income countries (LICs), low- and middle-income countries (LMICs) and upper middle-income countries (UMICs) were included. The review included all relevant studies published between 1990 and 2026 indexed in the database.

2.2.2. Participants

The review included studies conducted on children and adolescents from birth up to 18 years of age, recruited from both healthy and clinical settings and residing in LICs, LMICs and UMICs. Research involving populations that were not considered representative of the general population was not included, such as studies conducted solely on premature children and children with HIV, malaria or AIDS.

2.2.3. Study Design

All validation study designs were included. This review excluded studies that used the instrument solely as a measurement tool without reporting its local adaptation or psychometric properties. Additionally, news articles, blog posts, literature reviews, case series, case reports, abstracts without full texts and special-group studies (such as the prevalence of cerebral palsy) were excluded.

2.2.4. Study Setting

Studies conducted in developing country settings, defined by the World Bank's classification of "developing countries" were included and encompassed LICs, LMICs and UMICs as classified in 2023 (World Bank, 2023).

2.2.5. Literature Search and Screening

A search strategy was developed using the Population, Intervention, Control and Outcome (PICO) framework. The search was conducted across the following databases: PsycINFO, PubMed, Cochrane and Embase, focusing on literature published from January 1990 to May 2026 (Appendix A).

Medical Subject Headings (MeSH) terms included, but were not limited to: "adolescents", "infants", "developmental disability", "cognitive dysfunction", "intellectual disability", "intelligence tests", "neuropsychological tests" and "LMICs". The search was carried out in two stages. After removing duplicates in EndNote, the records were imported into Covidence (Veritas Health Innovation, n.d.) for screening. Two authors independently reviewed titles and abstracts to determine eligibility, followed by a full-text review. Any disagreements were resolved through discussion, with a third reviewer involved when necessary. The reasons for excluding studies at full-text stage were documented.

2.2.6. Data Extraction

Two review authors independently extracted data from the included studies onto a standardized data extraction table in Excel. The table detailed important aspects of each study, including the publication year, country of origin, name of the screening tool, gold standard tool(s) used for validation, study design, setting, sample size, sampling approach, participant age range, inclusion criteria, and the sensitivity and specificity of the screening tool. For studies reporting multiple outcomes, results from more than one questionnaire, various cut-offs, or different results based on 1 and 2 standard deviations, 2 standard deviation values were consistently selected to maintain uniformity across our findings. Any discrepancies encountered during the literature selection, data extraction, or risk assessment phases were addressed and resolved through discussion among the authors to ensure consensus.

2.2.7. Data Analysis

A meta-analysis and a descriptive analysis were conducted to evaluate the validity, reliability, and diagnostic accuracy of neurodevelopmental assessment tools adapted for developing countries. For the descriptive analysis, data were extracted from all included studies, focusing on reported reliability and validity measures such as Cronbach's alpha for internal consistency, inter-rater reliability scores, and various validity metrics, including concurrent and construct validity. These findings were summarized, and trends were analyzed across studies to assess the consistency of adaptation processes and psychometric performance. The meta-analysis was conducted by entering data on Review Manager version 5.4.1 and using forest plots to present sensitivity and specificity (The Cochrane Collaboration, 2020). Although the studies provided sensitivity and specificity values, the true positives, true negatives, false positives, and false negatives were calculated for the meta-analysis. This analysis employed the Metandi command in STATA version 18.5 (StataCorp, 2023) to perform a bivariate model and generate the hierarchical Summary Receiver Operating Characteristic (HSROC) curve. A sensitivity analysis was conducted for studies reporting sub-domain outcomes of development.

2.2.8. Risk of Bias Assessment

Two authors independently performed quality assessments for this review using the Quality Assessment for Diagnostic Accuracy Studies 2 (QUADAS-2) tool (Whiting et al., 2011). This involved evaluating the risk of bias (categorized as high, low, or unclear) across four domains: patient selection, index test, reference standard, and flow and timing. Additionally, the first three domains were evaluated for applicability concerns. If any item within a domain indicated a high risk of bias, the entire domain was classified as high risk.

3. Results

The search strategy identified 22,658 studies. After screening, 443 articles were selected for full-text evaluation. Of these, 54 validation studies were included in the review, all of which accounted for cultural sensitivity in their adaptation processes (Figure 1).

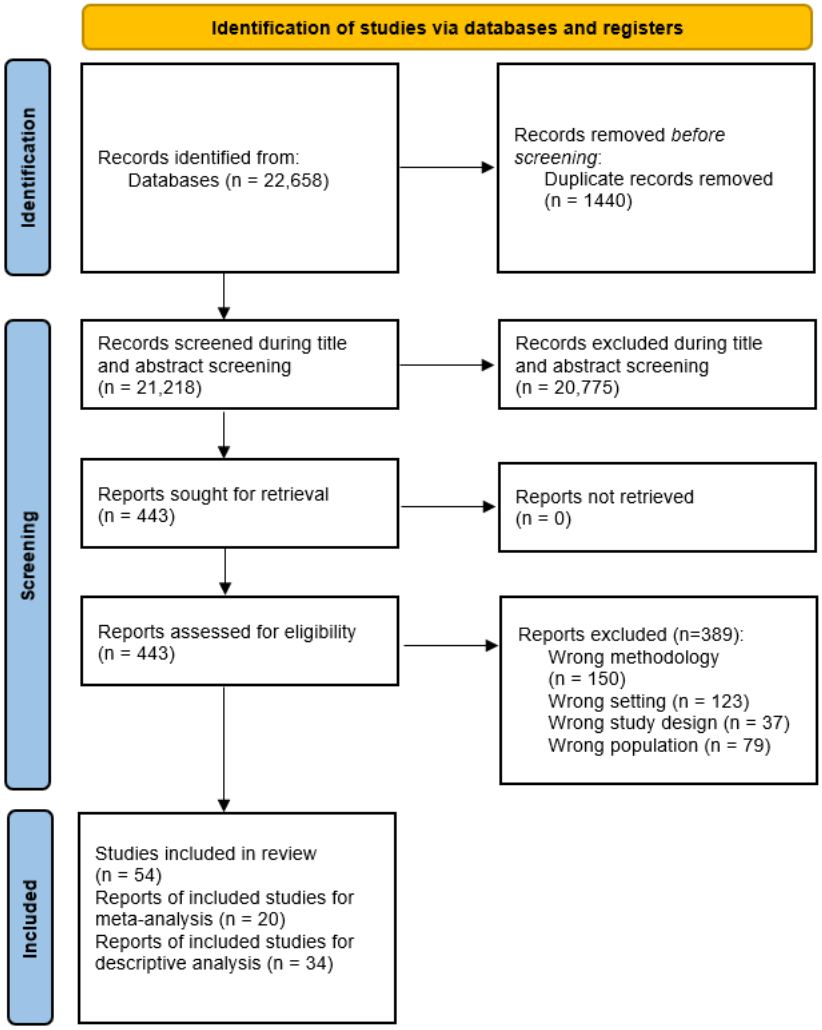


Figure 1. PRISMA Diagram.

3.1. Description of Included Studies

A total of 54 studies were included in this review, with 5 conducted in LICs, 28 in LMICs and 22 in UMICs. The studies reporting the validity and reliability of newly developed or adapted tools were used for the descriptive analysis, whereas those recording sensitivity and specificity were included in the meta-analysis. Tool development processes typically involved defining the tool's purpose, selecting or adapting items from existing measures, translating and culturally adapting content, and refining through pilot testing. Developers often incorporated expert input and feedback from target populations before finalizing the tools through reliability and validity testing. Most of these studies were conducted in a community setting (n = 34), with some carried out in clinics or health centers (n = 10) or both community and health centers (n = 9). The sample sizes ranged from 30 in Bangladesh (N. Z. Khan et al., 2010) to 45,640 in Brazil (Filgueiras et al., 2013) and involved children aged 0 to 18 years (Table 1).

Table 1. Description of Included Studies.

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
1	Aina (2001)	Nigeria	LMIC	2–30 months	128 children	Clinic and Community	DSI	Fields of development—adaptive, personal–social, language, fine motor and gross motor	BSID
2	Abubakar et al. (2008)	Kenya	LMIC	6–35 months 24–35 months	319 = rural l104 = urban	Community	KDI	Two functional areas—locomotor skills and eye–hand coordination	Kilifi Developmental Checklist
3	Abubakar et al. (2010)	Kenya	LMIC	2–10 months	95 children	Clinic and Community	Developmental Milestone Checklist (DMC)	Child functioning: motor, language and personal–social development	Kilifi Developmental Inventory (KDI)
4	Amani et al. (2018)	Iran	UMIC	5–18 years	360 students	Community	BRIEF Persian version	Executive functions—inhibit, shift, working memory, emotional control, planning, and organizing of material, initiate, and monitor	BRIEF English version
5	Anderson et al. (2021)	South Africa	UMIC	55–69 months	Study 1: 49 Study 2: 62	Community	ELOM	Gross motor development (GMD), fine motor coordination and visual motor integration (FMC & VMI), emergent numeracy and mathematics (ENM), cognition and executive functioning (CEF), and emergent literacy and language (ELL)	WPPSI-IV
6	Bhave et al. (2010)	India	LMIC	6–24 months	142 children	Clinic	LDS	27 milestones—gross motor, fine motor, language and social domains	Developmental Assessment Scale for Indian Infants
7	Burkey et al. (2018)	Nepal	LMIC	5–15 years	268 children	Community	DBIS-N	Common behavior-related problems	ECBI

Table 1. *Cont.*

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
8	Chopra et al. (1999)	India	LMIC	0–6 years	3560 children	Community	DSS	Screen for all major disabilities, physical, motor, sensory and mental retardation	NIMH Development Screening Schedule NIMH Development Assessment Schedule TQ CDQ
9	Charafeddine et al. (2013)	Lebanon	LMIC	4–60 months	733 children	Clinic and Community	ASQ-2	Aspects of development communication, gross motor, fine motor, problem solving, and personal–social skills	ASQ
10	Durkin et al. (1995)	Bangladesh Jamaica Pakistan	LMIC UMIC LMIC	2–9 years	22,125 children	Community	TQQ	Neurodevelopmental abilities—cognitive, motor and seizure	TQ
11	Dagvadorj et al. (2015)	Mongolia	UMIC	0 months 16 days—42 months 15 days	150 children	Clinic	MORBAS	Developmental domains—cognitive, gross motor, fine motor, social–emotional, expressive communication, receptive communication, adaptive behavior	BSID-III
12	Dang et al. (2017)	Vietnam	LMIC	6–16 years	208 children 1314 children	Hospitals Community	CBCL—Vietnamese version, SDQ—Vietnamese version	Internalizing and externalizing problems, social, conduct and hyperactivity	

Table 1. *Cont.*

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
13	Ertem et al. (2008)	Turkey	UMIC	0–24 months	510 children	Clinic	Guide for monitoring child development	Developmental milestones—Expressive language and communication Receptive language Fine and gross motor Relationship (social–emotional) Play (social–emotional, cognitive) Self-help skills	DDST-II Vineland Brigance Screening Test ASQ BSID-II
14	Filgueiras et al. (2013)	Brazil	UMIC	4–60 months	45,640 children	Community	ASQ-BR	Developmental delays	ASQ-3
15	Fernandes et al. (2022)	Brazil	UMIC	12 months	104 children	Community	OX-NDA	Cognition, motor, language, positive and negative behavior	BSID-III
16	M. J. Gladstone et al. (2008)	Malawi	LIC	0–6 years	1130 children	Clinic	DDST	Gross motor, fine motor, language, and social domains	Denver II
17	Goldberg et al. (2009)	Vietnam	LMIC	3–6 years	120 mothers of pre-school aged children	Community	VVABS	Communication, daily living skills, socialization and motor skills	VABS
18	M. Gladstone et al. (2010)	Malawi	LIC	0–6 years	1426 children	Community	MDAT	Domains of development—gross motor, fine motor, language, and social	Denver II
19	Gustawan et al. (2016)	Indonesia	UMIC	3–12 months	170 infants	Clinic	PEDS	Global/cognitive, speech/expressive language, receptive language, behavior, social–emotional, school, self-help, fine motor, gross motor and other	BSID-II

Table 1. *Cont.*

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
20	Holding et al. (2004)	Kenya	LMIC	5–7 years (Phase 1) 5 years 7 months–6 years 11 months (Phase 2)	>100 (Phase 1) 56 (Phase 2)	Community	K-ABC—Kilifi	Mental Processing Index and Fluid-Crystallised Index	K-ABC
21	Hanlon et al. (2016)	Ethiopia	LIC	30–42 months	*n*= 440 (30 months) *n*= 456 (42 months)	Community	Adapted version of BSID-III	Developmental functioning and delay: cognitive, expressive and receptive language, and fine and gross motor	BSID-III
22	Hsiao et al. (2017)	South Africa Zambia	UMIC LMIC	2 months–60 months	853	Clinic/community	ASQ-3	Five developmental domains: communication, gross motor, fine motor, problem solving and personal–social	
23	Jeong et al. (2025)	Kenya	LMIC	0–24 months	647 children	Community	GSED	Domains of cognitive, language, and motor development	Bayley-II, CREDI
24	N. Z. Khan et al. (2010)	Bangladesh	LMIC	Sample A: 0–3 months Sample B: 3–24 months	Sample A: 50 children Sample B: 30 children	Clinic and Community	RNDA	Functional status—primitive reflexes, gross motor, fine motor, vision, hearing, speech, cognition, behavior and seizures	BSID-II
25	Kandawasvika et al. (2012)	Zimbabwe	LMIC	6–8 years	101 children	Community	Short form MSCA	Intelligence and motor abilities	Educational psychologists' assessment
26	N. Z. Khan et al. (2013)	Bangladesh	LMIC	2–5 years	77 children	Community	RNDA	Neurodevelopmental impairments (NDIs)—gross motor, fine motor, vision, hearing, speech, cognition, behavior, and seizures	IBAS BSID-II SB5 WPPSI

Table 1. *Cont.*

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
27	Koura et al. (2013)	Benin	LMIC	12 months	357 children	Clinic and Community	MSEL French Translation TQQ	Childhood development—gross motor, fine motor, visual reception, receptive language, and expressive language	MSEL
28	Kakooza-Mwesige et al. (2014)	Uganda	LIC	2–9 years	1169 children	Community	TQQ	Neurodevelopmental abilities	
29	N. Z. Khan et al. (2014)	Bangladesh	LMIC	5–9 years	121 children	Community	RNDA	Neurodevelopmental impairments (NDIs)—gross motor, fine motor, vision, hearing, speech, cognition, behavior, and seizures	IBAS (Gold Standard I) IQ tests (WPPSI and WISC, or Gold Standard II)
30	Kvestad et al. (2023)	Nepal	LMIC	8–12 weeks and 6 months	705 infants	Community	TIMP	Linear growth and neurodevelopment	BSID-III
31	Munir et al. (1999)	Bangladesh	LMIC	2–9 years	1404 children	Community	IBAS	Adaptive behavior skills—motor skills, socialization, communication and daily living skills	
32	Mung'ala-Odera et al. (2004)	Kenya	LMIC	6–9 years	10,218 children	Community	TQQ	Impairment in cognitive, motor, epilepsy, hearing and vision domains	
33	Malda et al. (2010)	India	LMIC	6–10 years	598 children	Community	KABC-II	Psychological and visual spatial memory, speed of processing, number processing and spatial and non-verbal abilities	
34	Mammen et al. (2013)	India	LMIC	4–18 years	124 children	Clinic	BIDS	Intellectual disability	BKT GDS VSMS

Table 1. *Cont.*

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
35	Maleka et al. (2016)	South Africa	UMIC	6–18 years	207	Clinic	PEDS—m-Health version PEDS:DM—m-Health version	Developmental domains: expressive language, receptive language, fine motor, gross motor, social–emotional, self-help, and academics	PEDS tool operated by professionals
36	Muslima et al. (2016)	Bangladesh	LMIC	10–16 years	47 young adolescents	Community	RNDA	Neurodevelopmental domains: gross motor, fine motor, vision, hearing, expressive language, cognition, behavior, self-care and unprovoked seizures	WISC, R
37	McCoy et al. (2017)	Tanzania	LMIC	18–36 months	2481 children	Community	ECDS	Motor, cognitive and social–emotional domains	BSID-III
38	Mashhadi et al. (2021)	Iran	UMIC	6–18 years	Parents of 2295 children and adolescents	Clinic and Community	BDEFS-CA	Deficit in executive functioning	CHEXI
39	Morse et al. (2026)	South Africa	UMIC	4–5 years	59 children	Community	MSEL KABC-II	Fine motor, visual reception, receptive language, and expressive language	
40	Nair et al. (2009)	India	LMIC	1.6–3 years	429	Community	DATA	Six domains of gross motor, fine motor, cognitive, personal–social, expressive language, and receptive language	

Table 1. *Cont.*

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
41	Pisani et al. (2018)	Bangladesh Bhutan Egypt Ethiopia Indonesia Malawi Mali Mozambique Pakistan Rwanda Zambia	LIC LMIC UMIC	3.5–6 years	138 children	Community	IDELA	Four developmental domains: physical development, language/literacy, numeracy/cognitive development and social–emotional development	ASQ
42	Rubio-Codina et al. (2016)	Colombia	UMIC	6–42 months	1311 children	Clinic and Community	ASQ-3 DDST-II BDI-2 SFI and SFII WHO-Motor	Early childhood development—cognitive, receptive and expressive language, and fine and gross motor development	BSID-II
43	Richard's et al. (2020)	Argentina	UMIC	9–12 years	103	Community	TAC	Perceptual inhibition tasks, WM tasks, and cognitive flexibility tasks	
44	Singhi et al. (2007)	India	LMIC	2–9 years	1763	Community	TQS–translated to Hindi and Punjabi	Detection of common disabilities (physical, mental, speech, hearing, visual and epilepsy)	Pre-structured Medical and Neurodevelop-ment Assessment Form Malin's adaptation of WPPSI
45	Soleimani and Dadkhah (2007)	Iran	UMIC	4–18 months	6150 children	Clinic	INFANIB	Gross motor developmental delay	CDI PEDS BINS ASQ
46	Thorburn et al. (1992)	Jamaica	UMIC	2–9 years	5478 children	Community	TQQ	Childhood disabilities (motor, hearing, visual, speech, cognitive and fits)	ADLQ DDST

Table 1. *Cont.*

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
47	Vameghi et al. (2013)	Iran	UMIC	1–66 months	11,740	Community	ASQ-3—Persian	Five developmental domains—communication, gross motor skills, fine motor skills, problem solving, and personal–social (30 items)	
48	Van Der Linde et al. (2015)	South Africa	UMIC	6–12 months	201	Clinic	RTHB developmental checklist	Developmental domains, including sensory functioning such as sight and hearing, communication and gross motor and fine motor development	PEDS
49	Wantanakorn et al. (2016)	Thailand	UMIC	18–30 months	137 children	Clinic	PEDS	Developmental delays—fine motor, gross motor, self-help, receptive language, expressive language, social–emotional and academic	MSEL
50	Waechter et al. (2022)	Grenada	UMIC	22–30 months	Inter-rater and test–retest reliability (n = 21) Internal consistency (n = 145)	Community	INTER-NDA	Cognitive, language, motor and behavioral outcomes	
51	Xie et al. (2017)	China	UMIC	1–25 months	812	Clinic	ASQ-3—translated to Chinese (ASQ-C)	Five developmental domains: communication, gross motor, fine motor, problem solving, and personal–social	ASQ
52	Yue et al. (2019)	China	UMIC	5–24 months	1831	Community	ASQ-3	Domains—problem solving, communication, fine motor, gross motor, and personal–social skills	BSID-III

Table 1. *Cont.*

	Author/Year	Country	Country Classification	Population	Sample Size	Setting	Development Tool	Tool Measures	Gold Standard
53	Yuan et al. (2022)	Ghana	UMIC	4–6 years	966	Community	RACER	Assessing inhibitory control (e.g., slower responses on inhibition trials), declarative memory (e.g., higher accuracy on previously seen items), and procedural memory (e.g., faster responses on sequence blocks)	
54	Zaman et al. (1990)	Bangladesh	LMIC	2–9 years	2576 children	Community	TQQ	Disabilities—gross motor, fine motor, vision, hearing, seizures, cognition, speech, nutritional status, psychiatric status, and other	

ADLQ (Activities of Daily Living Questionnaire); ASQ (Ages and Stages Questionnaire); (ASQ-2) Ages and Stages Questionnaire 2nd Edition; (ASQ-3) Ages and Stages Questionnaire, 3rd Edition; ASQ-BR (Ages and Stages Questionnaires—Brazilian translation); BDI-2 (Battelle Developmental Inventory Screener, 2nd Edition); BDEFS-CA (Barkley Deficits in Executive Functioning Scale—Children and Adolescents); BINS (Bayley Infant Neuro-developmental Screener); BSID (Bayley Scales of Infant Development); BSID-II (Bayley Scales of Infant Development 2nd Edition); BSID-III (Bayley Scales of Infant Development 3rd Edition); BRIEF (Behavioral Rating Inventory of Executive Functions); BKT (Binet–Kamat Test of Intelligence); BIDS (Brief Intellectual Disability Scale); CBCL (Child Behavior Checklist); CDI (Child Development Inventory); CDQ (Child Disability Questionnaire); CHEXI (Childhood Executive Functioning Inventory); DDST (Denver Developmental Screening Test); DDST-II (Denver Developmental Screening Test, 2nd Edition); DATA (Developmental Assessment Tool for Anganwadis); DSI (Developmental Screening Inventory); DSS (Disability Screening Schedule); DBIS-N (Disruptive Behavior International Scale—Nepal version); ECDS (Early Childhood Development Scale); ECBI (Eyberg Child Behavior Inventory); ELOM (Early Learning Outcomes Measure); GDS (Gesell's Developmental Schedule); GSED (Global Scales of Early Development); IBAS (Independent Behavior Assessment Scale); IDELA (International Development and Early Learning Assessment); INFANIB (Infant Neurological International Battery); INTER-NDA (INTERGROWTH—21st Neurodevelopment Assessment); K-ABC (Kaufman Assessment Battery for Children); KABC-II (Kaufman Assessment Battery for Children, 2nd edition); KDI (Kilifi Developmental Inventory); LDS (Lucknow Development Screen); MDAT (Malawi Developmental Assessment Tool); MSCA (McCarthy Scales of Children's Abilities); MORBAS (Mongolian Rapid Baby Scale); MSEL (Mullen Scales of Early Learning); OX-NDA (Oxford Neurodevelopment Assessment); PEDS (Parents Evaluation Developmental Status); PEDS:DM (Parents Evaluation Developmental Status: Developmental Milestones); RACER (Rapid Assessment of Cognitive and Emotional Regulation); RNDA (Rapid Neurodevelopmental Assessment); RTHB (Road to Health Booklet); SB5 (Stanford Binet Intelligence Scale); SDQ (Strengths and Difficulties Questionnaire); TIMP (Test of Infant Motor Performance); TQQ (Ten Questions Questionnaire); TQS (Ten Questions Screen); TAC (Cognitive Self-Regulation Test Battery); VABS (Vineland Adaptive Behavior Scale); VSMS (Vineland Social Maturity Scale); VVABS (Vietnamese Vineland Adaptive Behavior Scale); WHO-Motor (World Health Organization Gross Motor Milestones); WISC (Wechsler Intelligence Scales for Children); WISC, R (Wechsler Intelligence Scale for Children, Revised); WPPSI (Wechsler Preschool and Primary Scales of Intelligence); WPPSI-IV (Wechsler Preschool and Primary Scale of Intelligence Fourth Edition).

3.2. Methodological Quality of Included Studies

The risk of bias assessment revealed that out of the 54 studies evaluated, 70% exhibited a high risk of patient selection bias due to their convenience sampling or case–control design. In contrast, 26% of studies reported a low risk of bias, whereas 4% were classified as having an unclear risk. The index test bias indicated that 56% had an unclear risk of bias, and 35% indicated a low risk of bias, whereas 9% of studies were found to have a high risk of bias. The main contributor of the unclear risk was the unavailability of a pre-specified threshold for tools in the studies. Most of the studies (i.e., 78%) had an unclear risk of reference standard bias, 20% reported low risk and 2% reported high risk. On the flow and timing domain, 22% of the studies with a high risk of bias were due to not all patients receiving a reference standard or the same one, or not all patients being included in the analysis, with 44% having a low risk and 37% indicating unclear risk.

For applicability concerns, three domains were assessed: patient selection, index test and reference standard. Patient selection concerns, according to the inclusion criteria, found that 80% of studies had a low risk of bias, whereas 20% had high risk. In the index test domain, 87% had a low risk of bias, and 13% had a high risk. In the reference standard domain, 66% of studies had a low risk of bias, 30% had unclear risk, and 4% had a high risk. The notable proportion of studies with high or unclear risk of bias highlights methodological limitations, emphasizing the need for more rigorous validation and standardization of neurodevelopmental tools, especially in developing countries. Details of the risk of bias and applicability concerns are highlighted in Figure 2.

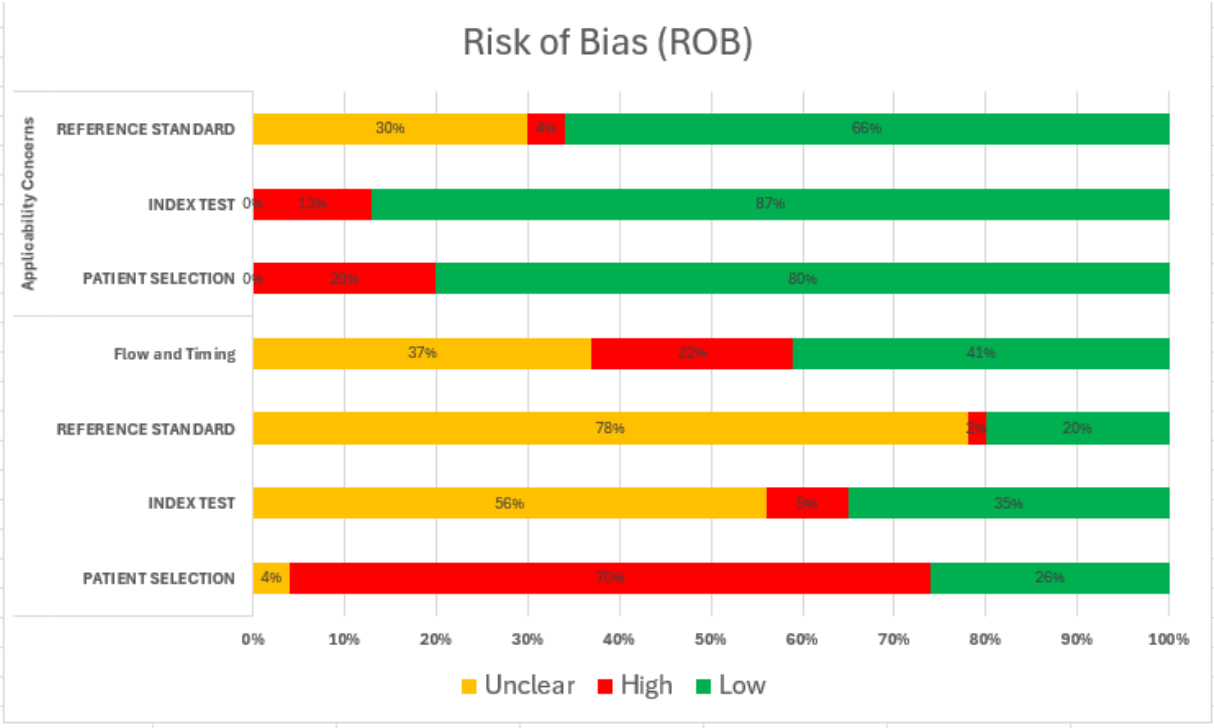


Figure 2. Risk of Bias and Applicability Bar Graph.

3.3. Meta-Analysis

The 20 studies (Bhave et al., 2010; Chopra et al., 1999; Dagvadorj et al., 2015; Ertem et al., 2008; Fernandes et al., 2022; M. Gladstone et al., 2010; Gustawan et al., 2016; Kakooza-Mwesige et al., 2014; Kandawasvika et al., 2012; N. Z. Khan et al., 2013, 2014; Mammen et al., 2013; Mung'ala-Odera et al., 2004; Singhi et al., 2007; Soleimani & Dadkhah, 2007; Thorburn et al., 1992; Van Der Linde et al., 2015; Wantanakorn et al., 2016; Yue et al., 2019; Zaman et al., 1990) included in the meta-analysis reported the sensitivity and speci-

ficity of the adapted tools and were validated against gold standard measures. Of the 20 studies, nine studies (45%) were conducted in UMICs, whereas nine studies (45%) were from LMICs and two studies (10) were from LICs. The sample size of these studies ranged from 77 in Bangladesh (N. Z. Khan et al., 2013) to 10,218 children in Kenya (Mung'ala-Odera et al., 2004).

A variety of developmental assessment tools were used across the studies. Some adapted tools were not validated against gold standard tools, such as the Disability Screening Schedule, and three studies that adapted the Ten Questions. The remaining tools were validated against gold standards, with the most used being the Bayley Scales of Infant and Toddler Development, Third Edition (BSID-3). Other commonly used gold standard tools included the Second Edition of the Bayley Scales (BSID-II), the Denver Developmental Screening Test (Denver II), the Ages and Stages Questionnaire (ASQ), the Vineland Social Maturity Scale (VSMS), and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI). These tools were selected for their established validity and wide use in assessing multiple developmental domains in children.

The adapted tools were administered by a range of assessors, reflecting diverse implementation settings. One of the key strengths of the locally adapted tools is their cultural relevance. Many were specifically designed or modified to align with local developmental expectations, caregiving practices, and language, which likely contributed to their strong sensitivity in detecting developmental delays. For example, several tools incorporated input from local caregivers and experts to adapt language, scoring, and milestone benchmarks, steps that may enhance ecological validity and ensure assessments are grounded in the child's actual environment.

All studies assessed developmental delays or milestones, covering domains such as cognitive, motor, epilepsy, visual, and hearing. Importantly, all tools focused on the development of typically developing children up to 18 years of age. Additionally, all studies demonstrate cultural sensitivity by adapting the tools to align with local languages, norms, and contextual factors.

Certain tools, such as the Malawi Developmental Assessment Tool (MDAT), have demonstrated strong psychometric properties, achieving 97% sensitivity and 81% specificity when validated against the Denver Developmental Screening Test and the Griffiths instrument (M. Gladstone et al., 2010). Similarly, the Lucknow Developmental Screen, validated against Malin's Developmental Assessment Scale for Indian Infants and the Vineland Social Maturity Scale, reported high sensitivity at 95.7% but relatively lower specificity at 72% (Bhave et al., 2010). These findings highlight a recurring challenge where higher sensitivity often comes at the cost of lower specificity, potentially leading to overestimations of developmental delays. The sensitivity of the 20 studies ranged from 25% (Van Der Linde et al., 2015) to 100% (Zaman et al., 1990), whereas the specificity range was from 25% (Thorburn et al., 1992) to 98% (Chopra et al., 1999).

The overall findings suggest that the local tools effectively screened children with developmental issues when compared to gold standard measures. The net sensitivity across all domains was 0.859 (95% CI: 0.78–0.91), indicating a strong ability to correctly detect children with developmental concerns. The net specificity was also reasonably high at 0.805 (95% CI: 0.71–0.87) (Figure 3), reflecting a good ability to correctly identify children without developmental concerns. Between-study variance was 1.30 (95% CI: 0.58–2.88) for sensitivity and 1.35 (95% CI: 0.70–2.59) for specificity, indicating moderate heterogeneity (τ^2 equivalents). The ROC analysis (Figure 4) provides a comprehensive overview of each study's findings, detailing the psychometric properties of these tools across developmental domains. HSROC variance components were $s^2\alpha = 2.33$ (95% CI: 1.13–4.81), which reflects between-study variability in test accuracy, and $s^2\theta = 0.74$ (95% CI:

0.37–1.48), noting between-study variability in threshold effects. A sensitivity analysis was conducted by removing three studies reporting a sub-domain analysis, with no significant differences in the overall results (Kandawasvika et al., 2012; Mammen et al., 2013; Soleimani & Dadkhah, 2007) (Appendix C).

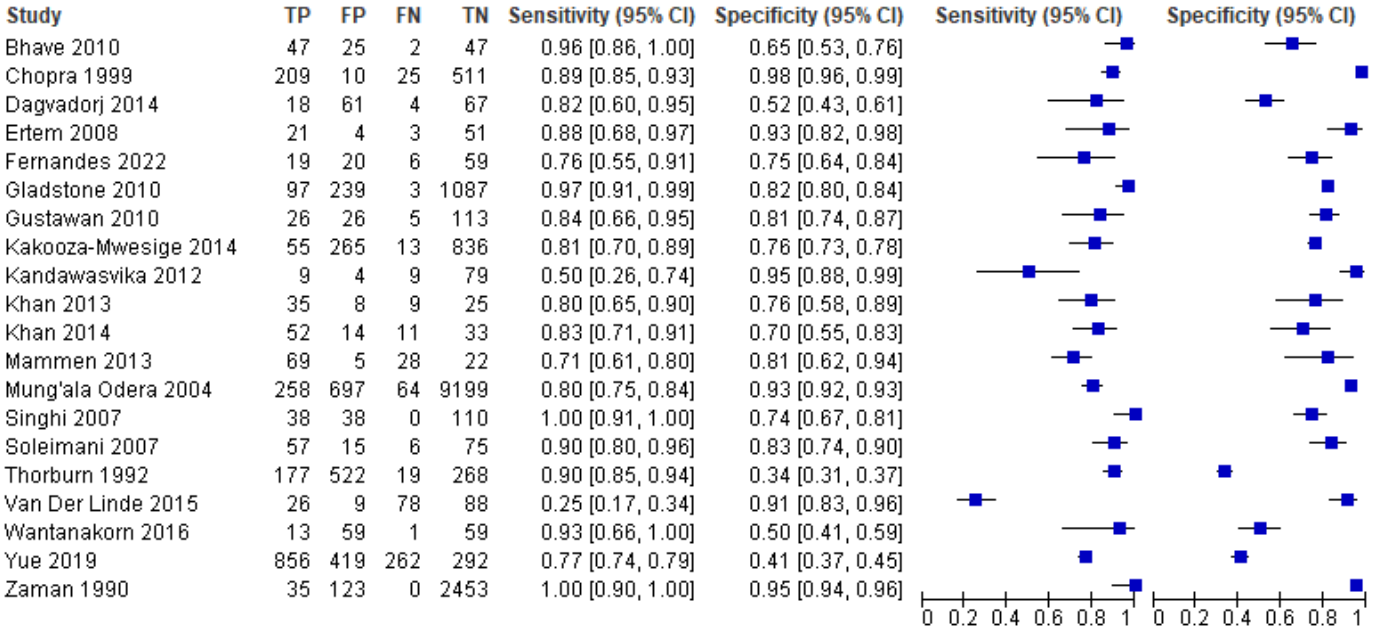


Figure 3. Forest Plot for Developed Tools Compared to Gold Standards for Neurodevelopmental Screening in Children.

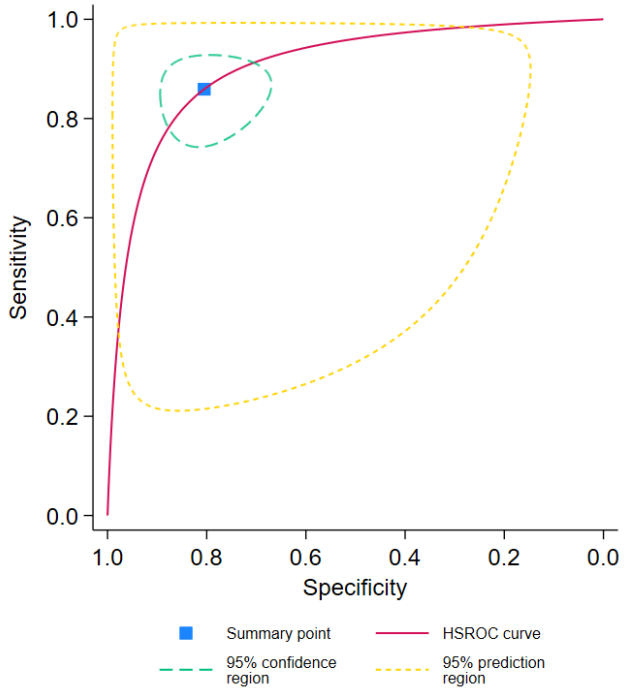


Figure 4. ROC Analysis.

3.4. Descriptive Analysis

There were 34 studies (Abubakar et al., 2008, 2010; Aina, 2001; Amani et al., 2018; Anderson et al., 2021; Burkey et al., 2018; Charafeddine et al., 2013; Dang et al., 2017; Durkin et al., 1995; Filgueiras et al., 2013; M. J. Gladstone et al., 2008; Goldberg et al., 2009; Hanlon et al., 2016; Holding et al., 2004; Hsiao et al., 2017; Jeong et al., 2025; N. Z. Khan

et al., 2010; Koura et al., 2013; Kvestad et al., 2023; Malda et al., 2010; Maleka et al., 2016; Mashhadi et al., 2021; McCoy et al., 2017; Morse et al., 2026; Munir et al., 1999; Muslima et al., 2016; Nair et al., 2009; Pisani et al., 2018; Richard's et al., 2020; Rubio-Codina et al., 2016; Vameghi et al., 2013; Waechter et al., 2022; Xie et al., 2017; Yuan et al., 2022) utilized for the descriptive analysis based in diverse settings as follows: LIC (n = 3), LMIC (n = 20) and UMIC (n = 15). Additionally, three studies were carried out within clinical settings, twenty-two were community-based, and nine were overlapping.

3.4.1. Low-Income Countries

Three studies conducted in LICs (Ethiopia, Malawi, Mali, Mozambique and Rwanda) evaluated the following screening tools: Denver Developmental Screening Test (DDST), an adapted version of the Bayley Scales of Infant Development, third version (BSID-III) and International Development and Early Learning Assessment (IDELA) (M. J. Gladstone et al., 2008; Hanlon et al., 2016; Pisani et al., 2018). These scales measured motor, language, social, physical and cognitive development. Evaluation of these tools showed a range of moderate to high inter-rater reliability and good content validity. DDST was found to have good respondent and face validity, whereas the adapted version of BSID-III indicated varying convergent validity in children of different age groups. IDELA reported good test–retest reliability and convergent validity with ASQ. Translated versions of the tools developed in high-income countries fared well in the cultural norms of these LICs.

3.4.2. Low- and Middle-Income Countries

A sum of 20 studies investigated neurodevelopmental tools in LMICs such as Kenya, Bangladesh, and India, amongst others. These studies focused on overall neurodevelopment (n = 13), motor development (n = 2), cognitive development (n = 2) and behavior development (n = 2).

Overall Neurodevelopmental Skills

Some of the scales utilized in these studies were the Ages and Stages Questionnaire (ASQ), Independent Behavior Scale (IBAS) and Rapid Neurodevelopmental Assessment (RNDA), which were compared to gold standard tools such as the Bayley Scales of Infant Development, second edition (BSID-II) and the Wechsler Intelligence Scale for Children, revised (WISC-R) (Charafeddine et al., 2013; N. Z. Khan et al., 2010; Munir et al., 1999; Muslima et al., 2016). These measured the following developmental domains: communication, problem solving, gross and fine motor, personal–social, vision, hearing, cognition, behavior and seizures. Translation, back-translation and field testing were utilized when adapting these tools to specific cultures. The ASQ indicated a range of internal consistency from 0.70 to 0.76 and high test–retest reliability. Similarly, IBAS denoted high test–retest reliability (0.96) and inter-rater reliability (0.95), as well as high construct validity of 0.92 to 0.97. Two studies utilizing RNDA found that there was very good to excellent agreement between testers, with some domains scoring 0.76 as the lowest value and 0.93 as the highest value. Face validity was good, and the tools were tested based off concurrent validity as well.

Other tests included the Developmental Milestone Checklist (Abubakar et al., 2010), the Vietnamese Vineland Adaptive Behavior Scale (VVABS) (Goldberg et al., 2009), the Developmental Assessment Tool for Anganwadis (DATA) (Nair et al., 2009) and the Global Scales of Early Development (GSED) (Jeong et al., 2025). These measures focused on cognitive, language, motor skills and personal–social development. The Developmental Milestone Checklist and DATA had good internal consistency of 0.85 and 0.86, respectively, whereas VVABs had a satisfactory internal consistency of 0.27.

Motor Skills

Two studies assessed scales of motor development: Kilifi Developmental Inventory (KDI) and the Test of Infant Motor Performance (TIMP) in Kenya and Nepal, respectively (Abubakar et al., 2008; Kvestad et al., 2023). These measured locomotor skills, eye–hand coordination and overall linear growth and neurodevelopment. Both tools, with TIMP assessed against BSID-III and no gold standard for KDI, found acceptable levels of reliability, with excellent inter-rater agreement (ICCs > 0.93) in TIMP. Although validity was not tested for TIMP, there was an acceptable range of validity within the KDI.

Cognitive Skills

These studies were conducted in Kenya and India and evaluated cognitive development using the first and second version of the Kaufman Assessment Battery for Children (K-ABC) (Holding et al., 2004; Malda et al., 2010). The domains focused on measuring the Mental Processing Index and the Fluid–Crystallised Index, psychological and visual spatial memory, speed of processing, number processing, and spatial and non-verbal abilities. Internal consistency was acceptable, ranging from 0.61 to 0.96, in different sub-scale areas. K-ABC was not evaluated against any gold standard tools but was adapted for the cultural context of LMICs and was found to be valid in such norms.

Behavioral Skills

These studies were conducted in Nepal and Vietnam and assessed behavioral development utilizing the following scales: Disruptive Behavior International Scale—Nepal version (DBIS-N), Child Behavior Checklist (CBCL)—Vietnamese version and Strengths and Difficulties Questionnaire (SDQ)—Vietnamese version (Burkey et al., 2018; Dang et al., 2017). The measured outcomes included common behavior-related problems, internalizing and externalizing problems, social and conduct problems and hyperactivity. The two scales had good internal consistency, ranging from 0.76 to 0.96, in different sub-scale areas, as well as acceptable validity. CBCL and SDQ Vietnamese versions were not validated against any gold standard tools but were adapted for the cultural context of LMICs and found to be reliable and valid in the relevant context.

3.4.3. Upper–Middle-Income Countries

There were 15 studies conducted in UMICs, including Iran, South Africa, Brazil, and China, with twelve studies focusing on overall neurodevelopment and three studies evaluating cognitive development.

Overall Neurodevelopmental Skills

These studies utilized scales focusing on the overall neurodevelopment of children such as the Ages and Stages Questionnaire (ASQ) (Filgueiras et al., 2013; Hsiao et al., 2017; Rubio-Codina et al., 2016; Vameghi et al., 2013; Xie et al., 2017), and Behavioral Rating Inventory of Executive Functions (BRIEF) (Amani et al., 2018), Early Learning Outcomes Measure (ELOM) (Anderson et al., 2021), Ten Questions Questionnaire (TQ) (Durkin et al., 1995), Mullen Scales of Early Learning (MSEL) and Kauffman Assessment Battery for Children (second edition) (KABC-II) (Morse et al., 2026). The gold standard tools utilized for comparison were the ASQ and BSID-III. The majority of these studies (n = 5) evaluated the ASQ, which assessed five developmental domains: communication, gross motor skills, fine motor skills, problem solving, and personal–social skills. The tools were translated into respective languages and were cross-culturally adapted. Overall, the internal consistency of the ASQ was found to be low, with Cronbach's alpha falling within the range of 0.31 to 0.96, differing based on skill domains and age intervals. There was low concurrent validity

for children below 19 months of age in some domains, whereas this was higher in others, such as gross motor scales, when compared to BSID-III.

Other scales such as BRIEF, ELOM and TQ measured executive functions, language and motor skills, and disabilities regarding cognitive, motor and seizures (Amani et al., 2018; Anderson et al., 2021; Durkin et al., 1995). BRIEF had a high internal consistency of 0.98, whereas TQ showed an internal consistency of 0.60. Both BRIEF and ELOM had high test–retest reliability, 0.81 and 0.80, respectively. Findings also denoted that there was good convergent validity between BRIEF and Wechsler's Intelligence scale and ELOM and WPPSI-IV.

Cognitive Skills

A diverse range of tools were utilized for measuring cognitive development in children, which included Barkley Deficits in Executive Functioning Scale—Children and Adolescents (BDEFS-CA) (Mashhadi et al., 2021), Cognitive Self-Regulation Test Battery (TAC) (Richard's et al., 2020), and Rapid Assessment of Cognitive and Emotional Regulation (RACER) (Yuan et al., 2022). These tools measured executive functioning, control, memory, and self-regulation and were translated and back-translated. An acceptable internal consistency of 0.91 was found in BDEFS-CA, with test–retest ranging from 0.66 to 0.83 for its subscales. It also had good convergent and divergent validity when compared with CHEXI. TAC showed a moderate construct validity of 0.53. Overall, BDEFS-CA, TAC and RACER were reliable and valid tools, with good feasibility and temporal efficiency in varying cultural contexts.

4. Discussion

This systematic review provides the first comprehensive assessment of the diagnostic accuracy of developmental assessment tools in developing countries, comparing them against gold standard measures such as BSID-III, Denver II, and WPPSI. Tools such as the Malawi Developmental Assessment Tool (MDAT) were found to have demonstrated strong psychometric properties (i.e., 97% sensitivity and 81% specificity), whereas the Lucknow Developmental Screen indicated high sensitivity at 95.7%, but lower specificity at 72%. Similarly, the Ten Questions Questionnaire showed a high sensitivity but low specificity, 100% and 25%, respectively. These findings denoted a recurring challenge: higher sensitivity was accompanied by lower specificity, contributing to overestimations of developmental delay. High rates of false positive cases may lead to the incorrect identification of children as developmentally delayed (Semrud-Clikeman et al., 2017); hence, requiring better evaluation of these tools. Consequently, the results highlighted both encouraging progress and persistent challenges in ensuring these tools' validity and reliability within diverse cultural and economic contexts.

The adaptation and validation processes varied significantly across studies, directly impacting the applicability of these tools. Although some scales effectively captured culturally specific developmental milestones, inconsistencies in validation methodologies and limited reporting hinder their generalizability. Other systematic reviews support these claims, as psychometric properties range from poor to excellent for culturally adapted tools (Huda et al., 2024). Moreover, many tools originally developed in Western settings required extensive modifications to ensure cultural and linguistic appropriateness, highlighting the central role of local context in tool development. Some of the adapted tools designed specifically with cultural nuances in mind may offer more contextually accurate and meaningful assessments of child development than traditional gold standards. These scales were not only tailored to local practices and environments but may also reflect a more nuanced understanding of development across diverse settings.

The findings indicate that the adapted tools exhibit reasonable diagnostic accuracy, with pooled sensitivity and specificity values of 0.859 and 0.805, respectively. This demonstrates their effectiveness in identifying developmental delays, though refinement is still needed, particularly to improve specificity and minimize false positives. Comprehensive analysis across multiple tools and populations strengthens our conclusions and offers a nuanced understanding of how well these tools perform across developing countries' contexts.

Nonetheless, several limitations must be acknowledged. Twenty-nine studies relied solely on validity or reliability assessments and were excluded from the meta-analysis. The QUADAS tool revealed significant methodological limitations in many studies, particularly in the documentation of participant demographics, validation steps, and data collection procedures. Additionally, considerable heterogeneity in study design, statistical approaches, sample characteristics, and geographic focus made direct comparisons challenging. For instance, the Ox-NDA (Fernandes et al., 2022), designed as a rapid, low-cost tool for LMICs, was validated solely in the city of Pelotas, Brazil, limiting its applicability across broader contexts. Similarly, the evaluation of PEDS (Gustawan et al., 2016) included a narrow age range (3–12 months), which restricted the developmental information that could be reliably gathered. These examples underscore the need for more inclusive and representative validation studies that test tools across various regions, age groups, and populations.

This review emphasizes the critical need for rigorous and standardized validation frameworks that are culturally grounded, and future research should adopt participatory, culturally sensitive methodologies that engage local stakeholders in tool development and adaptation. Additionally, longitudinal validation strategies are essential to assess the consistency and predictive power of these tools over time, and a growing interest in digital tools presents a promising avenue for improving accessibility and standardization, though further research is needed to evaluate their feasibility and accuracy in low-resource settings.

5. Conclusions

This systematic review evaluated the psychometric properties and assessed the strengths and weaknesses of various locally adapted or developed neurodevelopmental tools used in developing countries against global gold standard tools. The findings indicate that the adapted tools exhibited reasonable diagnostic accuracy, with pooled sensitivity and specificity values of 0.859 and 0.805, respectively. Although this demonstrates their effectiveness in identifying developmental delays, further assessment is required to improve specificity and minimize false positives. The Malawi Developmental Assessment Tool (MDAT) demonstrated strong psychometric properties, whereas the Lucknow Developmental Screen and the Ten Questions Questionnaire showed high sensitivities accompanied by relatively lower specificities. This comprehensive assessment highlights the progress and persistent challenges for wider use and acceptability, as more robust studies would be needed for general recommendations for diverse cultural and economic contexts.

Author Contributions: Conceptualization, J.K.D.; screening, search and extraction, M.N., S.L., Z.N.A., H.I., S.R., H.A. and Z.H., methodology, M.N., S.L., Z.N.A.; software, S.L., Z.N.A. and M.N.; analysis, S.L., Z.N.A. and J.K.D.; writing—original draft preparation, S.L. and Z.N.A.; writing—review and editing, Z.N.A., S.L., S.K.J. and J.K.D.; supervision, J.K.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data presented in this systematic review was derived from published papers and which are available in the public domain and we used aggregate estimates.

Acknowledgments: We would like to acknowledge Arjumand Rizvi who helped with the analysis and the support from Aga Khan University.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A

Search Strategy

((Child[MeSH Terms]) OR (Young Adult[MeSH Terms])) OR (Adolescent[MeSH Terms])) OR (Infant[MeSH Terms])) OR (early life)) OR (child, preschool[MeSH Terms])) OR (early childhood)) OR (early childhood education)) OR (early childhood education and care)) AND (pervasive child development disorders[MeSH Terms])) OR (developmental disabilities[MeSH Terms])) OR (impairment)) OR (cognitive dysfunction[MeSH Terms])) OR (cognition disorders[MeSH Terms])) OR (cognitive impairment)) OR (cognitive disabilit)) OR (motor skills disorders[MeSH Terms])) OR (motor disorders[MeSH Terms])) OR (motor impairment)) OR (motor disabilit)) OR (learning disabilities[MeSH Terms])) OR (disorders, language development[MeSH Terms])) OR (intellectual disabilit[MeSH Terms])) OR (intellectual impairment)) OR (autistic disorder[MeSH Terms])) OR (autism spectrum disorder[MeSH Terms])) OR (asperger syndrome[MeSH Terms])) OR (cerebral palsy[MeSH Terms])) OR (disorders, social behavior[MeSH Terms])) OR (attention deficit disorder with hyperactivity[MeSH Terms])) OR (neurodisabilit)) OR (developmental delay)) OR (communication disorders[MeSH Terms])) OR (specific language disorder[MeSH Terms])) OR (language development disorders[MeSH Terms])) OR (speech disorders[MeSH Terms])) OR (hearing disorders[MeSH Terms])) OR (down syndrome[MeSH Terms])) OR (dyslexia[MeSH Terms])) OR (childhood onset fluency disorder[MeSH Terms])) OR (neurocognitive deficit)) OR (social emotional functioning)) OR (poor productivity)) OR (intellectual disability[MeSH Terms])) OR (child development[MeSH Terms])) OR (infant development[MeSH Terms])) OR (cognition[MeSH Terms])) OR (cognitive development)) OR (cognitive outcome)) OR (language development)) OR (speech development)) OR (motor skills[MeSH Terms])) OR (motor performance)) OR (psychomotor performance[MeSH Terms])) OR (motor outcome)) OR (motor development)) OR (social development)) OR (emotional development)) OR (behavioral outcome)) OR (developmental outcome)) OR (social skills[MeSH Terms])) OR (adaptation, psychological[MeSH Terms])) AND (milestone)) OR (surveillance)) OR (screening tool)) OR (screening measure)) OR (screening assessment)) OR (development tool)) OR (assessment)) OR (checklist)) OR (tool)) OR (scale)) OR (measure)) OR (test)) OR (screening)) OR (inventory)) OR (early identification)) OR (physical examination[MeSH Terms])) OR (growth assessment)) OR (intelligence tests[MeSH Terms])) OR (neuropsychological tests[MeSH Terms])) OR (child development[MeSH Terms])) OR (surveys and questionnaires[MeSH Terms])) OR (developmental assessment)) OR (psychometrics[MeSH Terms])) OR (psychological tests[MeSH Terms])) AND (LMIC)) OR (low-resourced setting)) OR (global south)) OR (low-income countries)) OR (upper-income countries)) OR (UMIC)) OR (low resource setting)) OR (developing countries[MeSH Terms]))

Appendix B

PRISMA Checklist

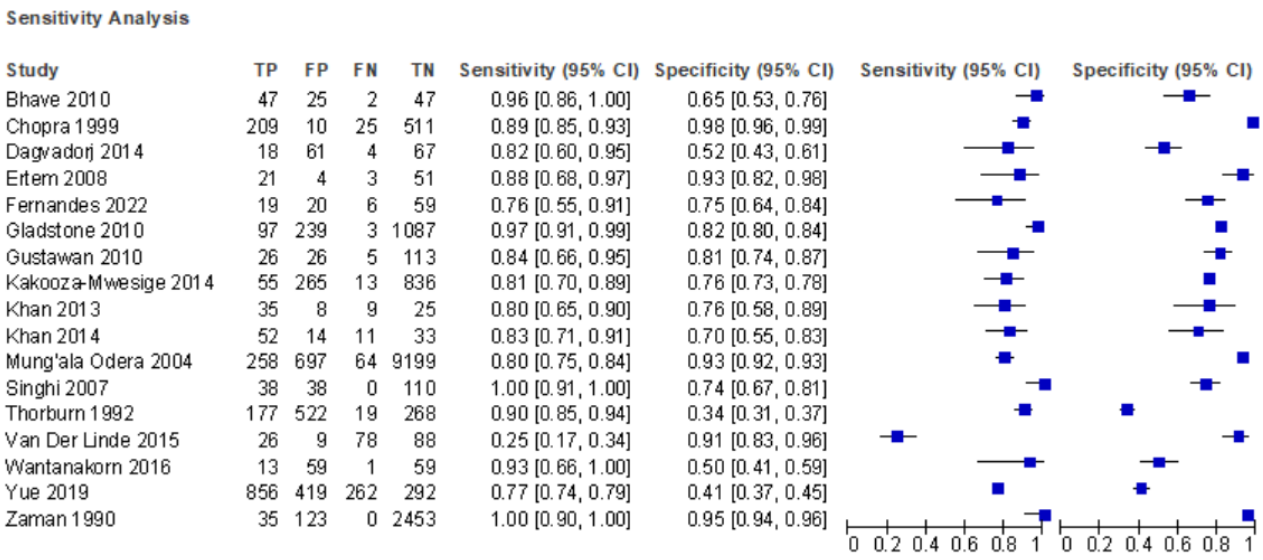
Section and Topic	Item #	Checklist Item	Location Where Item Is Reported
TITLE			
Title	1	Identify the report as a systematic review.	Title Page
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Objectives
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Study Selection
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Study Selection
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Literature Search and Screening
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Data Extraction
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Data Extraction
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g., for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Study Selection
	10b	List and define all other variables for which data were sought (e.g., participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Study Selection
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Risk of Bias Assessment
Effect measures	12	Specify, for each outcome, the effect measure(s) (e.g., risk ratio, mean difference) used in the synthesis or presentation of results.	Data Analysis

Section and Topic	Item #	Checklist Item	Location Where Item Is Reported
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g., tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Data Extraction
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Data Extraction
	13c	Describe any methods used to tabulate or visually display the results of individual studies and syntheses.	Meta-analysis
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If a meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Meta-analysis
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g., subgroup analysis, meta-regression).	Meta-analysis
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Meta-analysis
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	N/A
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	N/A
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 1: PRISMA diagram
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	N/A
Study characteristics	17	Cite each included study and present its characteristics.	Table 1: Description of Included Studies
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Risk of bias assessment
Results of individual studies	19	For all outcomes, present for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g., confidence/credible interval), ideally using structured tables or plots.	Table 1: Full table attached
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Risk of Bias Assessment
	20b	Present results of all statistical syntheses conducted. If a meta-analysis was conducted, present for each the summary estimate and its precision (e.g., confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Meta-analysis
	20c	Present results of all investigations of possible causes of heterogeneity among the study results.	Meta-analysis
	20d	Present the results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Meta-analysis, Appendix B

Section and Topic	Item #	Checklist Item	Location Where Item Is Reported
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Risk of Bias Assessment
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	N/A
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Discussion Paragraph 3
	23b	Discuss any limitations of the evidence included in the review.	Discussion paragraph 4
	23c	Discuss any limitations of the review processes used.	Discussion paragraph 4
	23d	Discuss implications of the results for practice, policy, and future research.	Conclusion
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Methodology
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Methodology
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	N/A
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Patents
Competing interests	26	Declare any competing interests of review authors.	Patents
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Data extracted from included studies

Appendix C

Sensitivity Analysis



References

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. [CrossRef] [PubMed]

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*, *99*(2), 291–297. [CrossRef]

Aina, O. F. (2001). The validation of developmental screening inventory (DSI) on Nigerian children. *Journal of Tropical Pediatrics*, *47*(6), 323–328. [CrossRef]

Amani, M., Asady Gandomani, R., & Nesayan, A. (2018). The reliability and validity of behavior rating inventory of executive functions tool teacher's form among Iranian primary school students. *Iranian Rehabilitation Journal*, *16*(1), 25–34. [CrossRef]

American Psychiatric Association. (n.d.). *Neurodevelopmental disorders*. American Psychiatric Association. Available online: https://www.psychiatry.org/patients-families/neurodevelopment-disorders (accessed on 22 February 2025).

Anderson, K. J., Henning, T. J., Moonsamy, J. R., Scott, M., Du Plooy, C., & Dawes, A. R. L. (2021). Test–retest reliability and concurrent validity of the South African early learning outcomes measure (ELOM). *South African Journal of Childhood Education*, *11*(1), a881. [CrossRef]

Angrist, N., Djankov, S., Goldberg, P. K., & Patrinos, H. A. (2021). Measuring human capital using global learning data. *Nature*, *592*(7854), 403–408. [CrossRef]

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186–3191. [CrossRef]

Bhave, A., Bhargava, R., & Kumar, R. (2010). Development and validation of a new Lucknow development screen for Indian children aged 6 months to 2 years. *Journal of Child Neurology*, *25*(1), 57–60. [CrossRef]

Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., McCoy, D. C., Fink, G., Shawar, Y. R., Shiffman, J., Devercelli, A. E., Wodon, Q. T., Vargas-Barón, E., & Grantham-McGregor, S. (2017). Early childhood development coming of age: Science through the life course. *The Lancet*, *389*(10064), 77–90. [CrossRef] [PubMed]

Burgess, A., Luke, C., Jackman, M., Wotherspoon, J., Whittingham, K., Benfer, K., Goodman, S., Caesar, R., Nesakumar, T., Bora, S., Honeyman, D., Copplin, D., Reedman, S., Cairney, J., Reid, N., Sakzewski, L., & Boyd, R. N. (2025). Clinical utility and psychometric properties of tools for early detection of developmental concerns and disability in young children: A scoping review. *Developmental Medicine and Child Neurology*, *67*(3), 286–306. [CrossRef]

Burkey, M. D., Adhikari, R. P., Ghimire, L., Kohrt, B. A., Wissow, L. S., Luitel, N. P., Haroz, E. E., & Jordans, M. J. D. (2018). Validation of a cross-cultural instrument for child behavior problems: The disruptive behavior international scale—Nepal version. *BMC Psychology*, *6*(1), 51. [CrossRef]

Charafeddine, L., Sinno, D., Ammous, F., Yassin, W., Al-Shaar, L., & Mikati, M. A. (2013). Ages and stages questionnaires: Adaptation to an Arabic speaking population and cultural sensitivity. *European Journal of Paediatric Neurology*, *17*(5), 471–478. [CrossRef]

Chopra, G., Verma, I. C., & Seetharaman, P. (1999). Development and assessment of a screening test for detecting childhood disabilities. *The Indian Journal of Pediatrics*, *66*(3), 331–335. [CrossRef] [PubMed]

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*, *104*(3), e94–e99. [CrossRef] [PubMed]

Dang, H.-M., Nguyen, H., & Weiss, B. (2017). Incremental validity of the child behavior checklist (CBCL) and the strengths and difficulties questionnaire (SDQ) in Vietnam. *Asian Journal of Psychiatry*, *29*, 96–100. [CrossRef]

Durkin, M. S., Wang, W., Shrout, P. E., Zaman, S. S., Hasan, Z. M., Desai, P., & Davidson, L. L. (1995). Evaluating a ten questions screen for childhood disability: Reliability and internal structure in different cultures. *Journal of Clinical Epidemiology*, *48*(5), 657–666. [CrossRef]

Ertem, I. O., Dogan, D. G., Gok, C. G., Kizilates, S. U., Caliskan, A., Atay, G., Vatandas, N., Karaaslan, T., Baskan, S. G., & Cicchetti, D. V. (2008). A guide for monitoring child development in low- and middle-income countries. *Pediatrics*, *121*(3), e581–e589. [CrossRef]

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), 733. [CrossRef]

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. [CrossRef]

Geisinger, K. F. (1994). Cross-cultural normative assessment: Translation and adaptation issues influencing the normative interpretation of assessment instruments. *Psychological Assessment*, *6*(4), 304–312. [CrossRef]

Gjersing, L., Caplehorn, J. R., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. *BMC Medical Research Methodology*, *10*(1), 13. [CrossRef]

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. [CrossRef] [PubMed]

Gladstone, M. J., Lancaster, G. A., Jones, A. P., Maleta, K., Mtitimila, E., Ashorn, P., & Smyth, R. L. (2008). Can Western developmental screening tools be modified for use in a rural Malawian setting? *Archives of Disease in Childhood*, *93*(1), 23–29. [CrossRef]

Goldberg, M. R., Dill, C. A., Shin, J. Y., & Nhan, N. V. (2009). Reliability and validity of the Vietnamese vineland adaptive behavior scales with preschool-age children. *Research in Developmental Disabilities*, *30*(3), 592–602. [CrossRef] [PubMed]

Gustawan, I. W., Soetjiningsih, S., & Machfudz, S. (2016). Validity of parents' evaluation of developmental status (PEDS) in detecting developmental disorders in 3–12 month old infants. *Paediatrica Indonesiana*, *50*(1), 6. [CrossRef]

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. [CrossRef]

Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., Mbonani, L., & Stevenson, J. (2004). Assessing cognitive outcomes in a rural African population: Development of a neuropsychological battery in Kilifi District, Kenya. *Journal of the International Neuropsychological Society*, *10*(2), 246–260. [CrossRef]

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. [CrossRef]

Huda, E., Hawker, P., Cibralic, S., John, J. R., Hussain, A., Diaz, A. M., & Eapen, V. (2024). Screening tools for autism in culturally and linguistically diverse paediatric populations: A systematic review. *BMC Pediatrics*, *24*(1), 610. [CrossRef]

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), 535. [CrossRef]

Kakooza-Mwesige, A., Ssebyala, K., Karamagi, C., Kiguli, S., Smith, K., Anderson, M. C., Croen, L. A., Trevathan, E., Hansen, R., Smith, D., & Grether, J. K. (2014). Adaptation of the "ten questions" to screen for autism and other neurodevelopmental disorders in Uganda. *Autism*, *18*(4), 447–457. [CrossRef]

Kandawasvika, G. Q., Mapingure, P. M., Nhembe, M., Mtereredzi, R., & Stray-Pedersen, B. (2012). Validation of a culturally modified short form of the McCarthy scales of children's abilities in 6 to 8 year old Zimbabwean school children: A cross section study. *BMC Neurology*, *12*(1), 147. [CrossRef]

Khan, I., & Leventhal, B. L. (2023). Developmental delay. In *StatPearls*. StatPearls Publishing. Available online: http://www.ncbi.nlm.nih.gov/books/NBK562231/ (accessed on 15 March 2025).

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), e755–e762. [CrossRef] [PubMed]

Khan, N. Z., Muslima, H., El Arifeen, S., McConachie, H., Shilpi, A. B., Ferdous, S., & Darmstadt, G. L. (2014). Validation of a rapid neurodevelopmental assessment tool for 5 to 9 year-old children in Bangladesh. *The Journal of Pediatrics*, *164*(5), 1165–1170.e6. [CrossRef]

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), e486–e494. [CrossRef]

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 & Behavioral Pediatrics*, *34*(7), 486–493. [CrossRef]

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*, 101809. [CrossRef] [PubMed]

Lassi, S., Niaz, M., Ansari, Z. N., Iftikar, H., Rizvi, S., Amir, H., Hasnain, Z., Jafri, S., & Das, J. K. (2026). *Assessing childhood development: Systematic review and meta-analysis on validation of local assessment tools in the context of developing countries*. OSF Registries. [CrossRef]

Malda, M., Van De Vijver, F. J. R., Srinivasan, K., Transler, C., & Sukumar, P. (2010). Traveling with cognitive tests: Testing the validity of a KABC-II adaptation in India. *Assessment*, *17*(1), 107–115. [CrossRef] [PubMed]

Maleka, B. K., Van Der Linde, J., Glascoe, F. P., & Swanepoel, D. W. (2016). Developmental screening—Evaluation of an m-health version of the parents evaluation developmental status tools. *Telemedicine and E-Health*, *22*(12), 1013–1018. [CrossRef]

Mammen, P., Russell, P. S. S., Nair, M. K. C., Russell, S., Kishore, C., & Shankar, S. (2013). Development and psychometric validation of the brief intellectual disability scale for use in low–health resource, high-burden countries. *Journal of Clinical Epidemiology*, *66*(1), 30–35. [CrossRef] [PubMed]

Mashhadi, A., Maleki, Z. H., Hasani, J., & Rasoolzadeh Tabatabaei, S. K. (2021). Psychometric properties of Persian version of the Barkley deficits in executive functioning scale–children and adolescents. *Applied Neuropsychology: Child*, *10*(4), 369–376. [CrossRef]

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), 3. [CrossRef] [PubMed]

McInnes, M. D. F., Moher, D., Thombs, B. D., McGrath, T. A., Bossuyt, P. M., PRISMA-DTA Group, Clifford, T., Cohen, J. F., Deeks, J. J., Gatsonis, C., Hooft, L., Hunt, H. A., Hyde, C. J., Korevaar, D. A., Leeflang, M. M. G., Macaskill, P., Reitsma, J. B., Rodin, R., Rutjes, A. W. S., … Willis, B. H. (2018). Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: The PRISMA-DTA statement. *JAMA*, *319*(4), 388. [CrossRef]

Morse, K., Tatham, C., Saliwe, B., Gwampi, B., Sidloyi, L., Sherr, L., & Toska, E. (2026). Assessing cognitive development in a diverse age child cohort using the Mullen scales of early learning and the Kaufman assessment battery for children II: A correlational study among children of adolescent mothers in South Africa. *Child Neuropsychology*, *32*(2), 178–193. [CrossRef]

Mung'ala-Odera, V., Meehan, R., Njuguna, P., Mturi, N., Alcock, K., Carter, J. A., & Newton, C. R. J. C. (2004). Validity and reliability of the 'ten questions' questionnaire for detecting moderate to severe neurological impairment in children aged 6–9 years in rural Kenya. *Neuroepidemiology*, *23*(1–2), 67–72. [CrossRef]

Munir, S. Z., Zaman, S., & McConachie, H. (1999). Development of an independent behaviour assessment scale for Bangladesh. *Journal of Applied Research in Intellectual Disabilities*, *12*(3), 241–252. [CrossRef]

Muslima, H., Khan, N. Z., Shilpi, A. B., Begum, D., Parveen, M., McConachie, H., & Darmstadt, G. L. (2016). Validation of a rapid neurodevelopmental assessment tool for 10- to 16-year-old young adolescents in Bangladesh. *Child: Care, Health and Development*, *42*(5), 658–665. [CrossRef]

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*, s27–s36.

Pallanti, S., & Salerno, L. (2023). Neurodevelopmental disorders (NDDs): Beyond the clinical definition and translational approach. *Children*, *10*(1), 99. [CrossRef] [PubMed]

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*, 1–15. [CrossRef]

Richard's, M. M., Vernucci, S., Stelzer, F., Introzzi, I., & Guàrdia-Olmos, J. (2020). Exploratory data analysis of executive functions in children: A new assessment battery. *Current Psychology*, *39*(5), 1610–1617. [CrossRef]

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), e0160962. [CrossRef]

Semrud-Clikeman, M., Romero, R. A. A., Prado, E. L., Shapiro, E. G., Bangirana, P., & John, C. C. (2017). Selecting measures for the neurodevelopmental assessment of children in low- and middle-income countries. *Child Neuropsychology*, *23*(7), 761–802. [CrossRef] [PubMed]

Singhi, P., Kumar, M., Malhi, P., & Kumar, R. (2007). Utility of the WHO ten questions screen for disability detection in a rural community the north Indian experience. *Journal of Tropical Pediatrics*, *53*(6), 383–387. [CrossRef]

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. [CrossRef]

StataCorp. (2023). *Stata statistical software* (version 18.5) [Computer software]. StataCorp LLC.

The Cochrane Collaboration. (2020). *Review manager* (version 5.4.1) [Computer software]. The Cochrane Collaboration. Available online: https://revman.cochrane.org (accessed on 3 February 2025).

Thorburn, M., Desai, P., Paul, T. J., Malcolm, L., Durkin, M., & Davidson, L. (1992). Identification of childhood disability in Jamaica: The ten question screen. *International Journal of Rehabilitation Research*, *15*(2), 115–128. [CrossRef]

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., Glascoe, F., Louw, E., & Vinck, B. (2015). Developmental screening in South Africa: Comparing the national developmental checklist to a standardized tool. *African Health Sciences*, *15*(1), 188. [CrossRef]

Veritas Health Innovation. (n.d.). *Covidence systematic review software* [Computer software]. Veritas Health Innovation Ltd. Available online: www.covidence.org (accessed on 25 August 2024).

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), 21. [CrossRef] [PubMed]

Wantanakorn, P., Sawangworachart, K., Roongpraiwan, R., & Chuthapisith, J. (2016). Parents' evaluation of developmental status (PEDS) in screening for developmental delay in Thai children aged 18–30 months. *Indian Pediatrics*, *53*(12), 1110–1112. [PubMed]

Whiting, P. F., Rutjes, A. W. S., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M. G., Sterne, J. A. C., Bossuyt, P. M. M., & QUADAS-2 Group. (2011). QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, *155*(8), 529–536. [CrossRef]

World Bank. (2023). *World Bank country and lending groups*. Available online: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (accessed on 25 August 2024).

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*, 281–295. [CrossRef]

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), e2749. [CrossRef] [PubMed]

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), e0221675. [CrossRef] [PubMed]

Zaman, S. S., Khan, N. Z., Islam, S., Banu, S., Dixit, S., Shrout, P., & Durkin, M. (1990). Validity of the 'ten questions' for screening serious childhood disability: Results from urban Bangladesh. *International Journal of Epidemiology*, *19*(3), 613–620. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.