

Portland State University

PDXScholar

Special Education Faculty Publications and
Presentations

Special Education

5-2018

Comparative Use of the Ages and Stages Questionnaires in Low- and Middle-Income Countries

Jason W. Small
Oregon Research Institute

Hollie Hix-Small
Portland State University, hixsmall@pdx.edu

Emily Vargas-Baron
RISE Institute

Kevin P. Marks
PeaceHealth Medical Group

Follow this and additional works at: https://pdxscholar.library.pdx.edu/sped_fac



Part of the [Special Education and Teaching Commons](#)

Let us know how access to this document benefits you.

Citation Details

Small, Jason W.; Hix-Small, Hollie; Vargas-Baron, Emily; and Marks, Kevin P., "Comparative Use of the Ages and Stages Questionnaires in Low- and Middle-Income Countries" (2018). *Special Education Faculty Publications and Presentations*. 19.

https://pdxscholar.library.pdx.edu/sped_fac/19

This Post-Print is brought to you for free and open access. It has been accepted for inclusion in Special Education Faculty Publications and Presentations by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.

Comparative use of the Ages and Stages Questionnaires in low- and middle-income countries

Jason W. Small^{1,3}, Hollie Hix-Small^{2,3}, Emily Vargas-Baron³, Kevin P. Marks⁴

1-Oregon Research Institute, 1776 Millrace, Eugene, OR 97403-1983, USA. Email: jasons@ori.org; Phone +1 541-484-2123.

2-Portland State University, Graduate School of Education, 1900 SW 4th Ave, #200, Portland, OR 97201, USA. Email: hollie.hixsmall@gmail.com; Phone +1 503-725-5490

3-RISE Institute, 3012 Porter Street, N.W., Washington, DC 20008, USA. Email: vargasbaron@hotmail.com; Phone +1 202-686-7044

4-PeaceHealth Medical Group, 4010 Aerial Way, Eugene, OR 97402. Email: kmarksodinson@gmail.com; Phone +1 45 51 80 43 40

Submit to Developmental Medicine and Child Neurology

Author correspondence and reprint requests to: Jason W. Small
Oregon Research Institute
1776 Millrace
Eugene, OR 97403
Ph: 541/484-2123;
E-mail: jasons@ori.org

Abstract

AIM To map the breadth of use of the Ages and Stages Questionnaires (ASQ) in low- and middle-income countries (LMICs) and across world regions and economies and examine procedures for ASQ translation, adaptation, psychometric evaluation, and administration.

METHOD We conducted a review of all original, peer-reviewed studies written in English referencing use of the ASQ in LMICs. We used a consensus rating procedure to classify each article into one of four categories: feasibility study, psychometric study, prevalence study, or research study.

RESULTS We analyzed 53 peer-reviewed articles written in English detailing use of the ASQ in LMIC. We found evidence of ASQ use in 23 LMICs distributed across all world regions.

The ASQ was translated into 16 languages. Just over half of the studies reported parent completion of the ASQ (50.9%). We identified 8 feasibility studies, 12 psychometric studies, and 9 prevalence studies. Study type varied by economy and region.

INTERPRETATION Findings suggest broad global use of the ASQ in a range of countries and cultural and linguistic contexts. There is need for further validation studies across all cited regions and countries and in countries ready to begin to design systems for providing universal developmental screening services.

What this paper adds

- The ASQ has been used in at least 23 low- and middle-income countries (LMICs)
- The ASQ has been translated into at least 16 languages in LMICs
- Over half of the studies identified reported parent completion of the ASQ

Introduction

Over the last decade, interest in developmental screening and assessment in low- and middle-income countries (LMICs) has increased dramatically. Numerous literature reviews have examined feasibility,^{1,2} explored prevalence,³ and provided broad summaries of the research evidence for developmental screening measures,^{4,5} neurodevelopmental assessment,⁶ and measures of child disability in LMICs.⁷ Additional contributions to the literature have elaborated on the complexity and challenges of creating new measures or translating and adapting existing measures to new countries and contexts^{8,9} and have reviewed developmental monitoring, screening, and assessment procedures in LMICs.¹⁰

In combination, this literature base provides researchers and practitioners an invaluable resource that documents screening, monitoring, and assessment measures created within LMICs; describes – to an extent – the breadth and depth of efforts to translate and adapt measures originally created in high income countries (HICs); and reports the psychometric properties of many of these measures. However, the current literature has limitations.

Some articles have reinterpreted or even misrepresented definitions of screening and monitoring. Ertrem,¹⁰ for example, promotes developmental monitoring as a holistic, family-centered, child-focused process that improves child and family well-being over time. She then recasts developmental screening in LMICs as an approach “in which groups of children are screened to ascertain whether they have developmental delay, by testers who do not necessarily have a continuous relationship with the families or access to health or social information other than that provided by the screening instrument” (p. 41). As will be seen below, Ertrem’s reinterpretation of developmental screening as something done to groups of children by “testers” who have limited knowledge of child development or family history does not align with the

literature. In a recent article describing the Guide for Monitoring Child Development (GMCD),¹¹ Ertrem further argues that the words ‘surveillance,’ ‘screening,’ and ‘monitoring’ represent different underlying philosophies. She alleges that surveillance and screening are “associated with security and policing” and keeping out “something undesirable” (p. 83). Monitoring, in contrast, “means ‘to watch, keep track of, or check, usually for a special purpose’ and implies an accepting, humble, positive and hopeful stance” (p. 84).

Other articles from this growing literature on developmental screening in LMICs inadvertently omit or overlook evidence in support of some measures. For example, Semrud-Clikeman et al.’s recent review of neurodevelopmental assessments used in LMICs,⁶ includes the Ages and Stages Questionnaires (ASQ) but inaccurately states there are no normative data on the measure in low- and middle-income contexts. As discussed herein, normative ASQ data have been collected in multiple LMICs including China, Georgia, Iran, and Turkey. Still other articles emphasize the cost and complexity of the translation and adaptation process but primarily apply it to measures created in HICs and not to measures transported from one low or middle-income country to another. As an example, Fischer, Morris, and Martines,⁴ who examine the feasibility of implementing developmental screening in primary healthcare centers (PHCs) in low- and middle-income settings, promote the use of tools developed in LMICs over those developed in HICs. The authors argue that “in addition to their adaptation needs, their costs, training requirements, and time for application”¹ make existing measures from HICs “less suitable for use” than measures developed in LMICs (p. 323). Although the time and expense associated with

¹ It is worth noting Fischer et al (2014) state they are examining the nine feasibility characteristics most valued by a convenience sample of 23 international experts on child development (p. 316); however, four of the nine criteria they highlight do not correspond to the nine top-ranked items reported in Table 1 of their paper. The three items referred to here (cost, training requirements, and time for application) are ranked 14th, 11th, and 12th, respectively, among a list of 19 feasibility items.

measurement translation, adaption, and training are important considerations, they are equally applicable to measures developed in LMICs.

Finally, some of the articles within this literature base discuss reliability and validity as though they were unchanging characteristics of the measures themselves. Fischer et al.⁴ recommend use of the Disability Screening Test (DST), GMCD, and Malawian Developmental Assessment Tool (MDAT) citing only sensitivity and specificity data from the country in which each measure was developed. As we discuss below, reliability and validity data contribute evidence to a broader validity argument that must be continuously updated and reconsidered as measures are transported to new contexts and countries. Measures are not inherently valid and reliable and a single source of psychometric data should not be used to endorse the broad use of a measure.

Given the aforementioned definitional shifts and omissions of evidence, there is a need to compile and evaluate across studies the evidence for a given measure. Such a process enables understanding of whether these shifting definitions of developmental screening are warranted when implementing them in LMICs and facilitates examination of the breadth and depth of use across a range of settings and contexts. In this manuscript, we use the ASQ as an example of this process. Recent articles on the ASQ have offered general descriptions of global use¹² and summarized previous research on the tool's psychometric properties for some age intervals.¹³ However, to our knowledge, no previous article has documented the range of ASQ studies conducted in LMIC and examined ASQ use with respect to study type (i.e., feasibility, psychometric, prevalence, and research).

To begin, we discuss widely accepted definitions of developmental monitoring and screening which contrast sharply with those described above. We then report findings from a

literature review of published, peer-reviewed articles on the ASQ in LMICs to better understand current ASQ use; describe the extent to which these identified studies adhere to recommended procedures for translation, adaptation, and psychometric evaluation; and discuss ASQ screening procedures (e.g., setting, form of administration, parent involvement) in low- and middle-income contexts. Finally, we identify gaps in currently available evidence and make recommendations for broadening and increasing support for rigorous translation, adaptation and evaluation procedures.

Developmental surveillance/monitoring and screening

Developmental surveillance (often labeled “monitoring” by non-medical professionals) is “a flexible, longitudinal, continuous and cumulative process whereby knowledgeable health care professionals identify children who may have developmental problems.”¹⁴ There are five fundamental components to developmental surveillance according to the American Academy of Pediatrics though others have recommended up to 12 components.¹⁵ During surveillance a health professional – at a minimum – should (a) elicit and attend to parent concerns about their child’s development; (b) document and maintain a developmental history; (c) make accurate observations of the child; (d) identify the presence of risk and protective factors; and (e) maintain an accurate record of the process and findings.¹⁴ Surveillance can be unstructured (i.e., guided by subjective impression or casual observation) or structured. Structured surveillance incorporates an evidence-based screening measure such as the ASQ or Parents’ Evaluation of Development Status (PEDS) into the decision-making process, which increases the likelihood children will be identified and linked with an appropriate agency for further assessment.¹⁶

Use of a standardized broadband developmental screening tool is central to an on-going, structured surveillance process. The AAP recommends developmental screening at 9, 18 and 24-

30 months.^{14,15} A parent, early intervention specialist, preschool teacher, community outreach worker, public health nurse, medical professional or paraprofessional may administer a screening tool. Developmental screening is not limited to healthcare and public health settings. They may be completed in the child's home, childcare center, or any other community settings of the parents' choosing. If the child has a potential delay based on a screener's cutoff criteria, a practitioner (with parental consent) or even a parent can make a referral to an Early Childhood Intervention (ECI) program or another similar program for an in-depth developmental assessment, decision regarding eligibility, and the development of an individualized family service plan, after which individualized ECI services are provided.

Several broadband developmental screening tools with evidence of reliability and validity are currently available, including the ASQ-3; PEDS; Parents' Evaluation of Developmental Status – Developmental Milestones (PEDS-DM); Brigance Screens; and Developmental Assessment of Young Children (DAYC-2). Examples of narrowband tools include the Ages and Stages Questionnaires: Social Emotional (ASQ: SE-2); Communication and Symbolic Behavior Scales Developmental Profile (CSBS DP); and the Modified Checklist for Autism in Toddlers (M-CHAT-R/F).

The Ages and Stages Questionnaires

The ASQ is a series of 21 simply worded developmental screening questionnaires that can be used with children between 1 and 66 months (5 ½ years) of age who are not known to have an established delay or disability. Now in its third edition, the ASQ is a flexible, low cost, first level, broadband developmental screening tool designed for universal application in order to answer the question, "Does this child require further evaluation?" Since it is designed to be parent completed, ASQ items reflect skills that are easily observed or elicited by parents; are

likely to occur in home and early childhood settings; and reflect key developmental milestones. The developers encourage parents to try each item with their child when their child is healthy, rested and fed and to make completing the questionnaire a fun activity.

The ASQ questionnaires have a uniform format across all age intervals. Each interval assesses five developmental domains (communication, gross motor, fine motor, problem-solving, and personal social) via six scored items hierarchically arranged from easier to harder. Gross motor items look at the child's large muscle movement and coordination. Fine motor items ask about finger and hand movements and hand-eye coordination. Problem-solving items are assessed by observing how the child plays with toys and solves problems. The personal social items address the child's self-help skills and interaction with others.

Parents select "Yes", "Sometimes", and "Not Yet" for each of the 30 scored items. A parent selects "Sometimes" for emerging behaviors the child is just beginning to demonstrate. Some items contain graphics to clarify item intent. The ASQ also has a section with open-ended items that are not scored. These items look at the quality of skills to detect parent concerns, including early language and behavioral concerns. ASQ authors recommend follow-up for any parent concerns because they are highly predictive of developmental needs.^{17,18,19}

The developers recommend parents independently complete the measure with support as needed from trained program personnel who subsequently score the form. Support can range from limited to intensive program personnel involvement depending on the setting and population. For example, limited support would entail personnel providing parents with an explanation of the screening process, providing information on the questionnaire's item and domain structure, answering parent questions, allaying any parent fears about the process, and scoring and interpreting the results for the parents. Intensive supports might include reading

items to parents or helping parents elicit and interpret their child's response to an item.²⁰ Other family members, teachers, and individuals who know the child well or who can easily observe and elicit the child's developmental skills and behaviors may complete the ASQ.

Evidence of reliability and validity

A measure is neither inherently valid nor invalid.²¹ Instead, measurement developers collect evidence to support the reliability and validity of a measure and then utilize this evidence to construct an argument to support its use for a particular purpose.²¹ To date, the majority of research on the adaptation and validation of developmental screening tools in LMICs describe findings in terms of validity types (e.g., face, criterion, content, construct) rather than utilizing more recent unified frameworks for validity evidence or argument-based approaches. In contrast to outdated discussions of validity types, unified and argument-based validity models focus on the targeted accumulation of evidence. Within these models, validity is not an inherent quality of a measure; instead, it is context specific and dependent on the decisions and consequences associated with its implementation.²¹ Furthermore, validation is an iterative, on-going process through which a body of evidence is assembled to either support or refute the use and interpretation of a measure.²²

A unified theory of validity situates all forms within construct validity.²³ Tool developers then collect five types of evidence to construct an argument in support of the reliability and validity of a measure. The five types of evidence collected pertain to a measure's (a) content, (b) response process, (c) internal structure, (d) relationship to other variables, and (e) consequences. Broadly speaking, the first three forms (content, response process, and internal structure) provide evidence to support the use of a measure; whereas the latter two forms (relationship to other variables and consequences) attend to decision-making and implications associated with a given

measure. Researchers often obtain evidence in support of content and response process during tool development²⁴ or – in the context of this paper – during the translation and adaptation process. For example, providing documentation of how a measure was developed and the extent to which the process accords with accepted guidelines for cross-cultural adaptation^{25,26} would provide strong evidence in support of a measure's *content*. This evidence might include documentation of how and by whom the measure was translated; the qualification of those involved in the translation, adaptation, and review process; and whether content experts possessing country, context, and substantive knowledge conducted and reviewed the adaptations. *Response processes* pertain to how a respondent interacts with a given assessment or measure. Evidence of response processes might be obtained through observations, focus groups, interviews, think-aloud sessions, or measurement of eye tracking and response time.²⁴ Evidence in support of a measure's *internal structure* and *relationship to other variables* align closely with standard conceptions of validity. Evidence of internal structure include statistical methods assessing dimensionality, measurement invariance, and reliability;²⁷ whereas evidence of a measure's relationship to other variables include traditional forms of criterion-related validity such as convergent, divergent, and concurrent.²³ Finally, evidence of a measure's *consequences* include not only statistical evidence such as Receiver Operating Characteristic (ROC) Curve analysis and the sensitivity and specificity of cutoffs but also consideration of the potential positive and negative consequences of administering a tool or identifying a child using the cutoffs.²⁸

Method

Inclusion criteria

Articles eligible for inclusion in this study were all original, peer-reviewed studies written in English referencing use of the ASQ in a low- or middle-income country. Since we wanted to capture the broad use of the ASQ, we included articles using all versions of the ASQ (e.g., ASQ-2 and ASQ-3) as well as studies reporting use of modified versions of the measure (e.g., combining age intervals or using altered administration procedures).

We utilized World Bank (WB) region and income categories to group the studies. For the 2018 fiscal year, WB classifies a country as low-income if it had a Gross National Income (GNI) per capita of US \$1,005 or less in 2016. If a country had a GNI per capita between US \$1,006 and US \$3,995, WB classifies the country as lower middle-income. Countries with a GNI per capita between US \$3,956 and US \$12,235 are classified as upper middle-income.²⁹ Based on these criteria, studies on the ASQ from up to 31 low-income countries and 109 lower middle-income and upper middle-income countries were eligible for inclusion. Additionally, studies from all world regions, with the exception of North America, were eligible. We excluded North America because all countries in the region (e.g., Bermuda, Canada, and the United States) have high-income economies (e.g., $GNI \geq US \$12,336$).

Search strategy

We compiled the articles included in this review between September 2016 and September 2017. We used a multi-stage process for study identification. First, we conducted a search of the following databases: ERIC, Google Scholar, Medline, and PsycInfo. For each database search, we included the key terms “Ages and Stages Questionnaire,” “ASQ,” “ASQ-2,” and “ASQ-3” along with additional terms to limit the scope of each search to low- and middle-income countries. These terms included the following: “low-income countries,” “middle -income countries,” “low and middle-income countries,” “developing countries,” and “developing

nations.” Although we acknowledge that some of these terms are outdated descriptors (e.g., developing countries), we sought to narrow our search from including high-income countries while simultaneously using broad search terms to identify studies in low- and middle-income contexts. After completing the initial search as described, we conducted a second search using various search terms for the ASQ and the country name for each low- and middle-income country. After completing the database searches, we reviewed the references of each article we identified during our search to find additional articles. Finally, we contacted the lead author of the ASQ, Dr. Jane Squires, and obtained a list of published articles on the ASQ against which we cross checked our article list.

Article categorization

To establish agreement on study categorization, the first and second authors of this article used a consensus rating procedure. First, each author independently reviewed and categorized the articles using the criteria described in Table 1. After independent review, we agreed on 49 of 55 articles. We then reviewed and discussed the six discrepant cases until reaching consensus on all articles. Inter-rater reliability was acceptable ($\kappa = .83$). We classified each article into one of four categories: feasibility study, psychometric study, prevalence study, or research study. These categories represent a continuum with respect to measurement development, moving from initial development and testing through systematic, large-scale validation to use in research or practice. Table 1 briefly details the criteria for each category. Feasibility studies assess “the extent to which a new treatment, or innovation, can be successfully used or carried out within a given agency or setting.”³⁰ In general, feasibility studies involved smaller samples as compared to psychometric studies; involved examination of fewer ASQ age intervals; utilized convenience samples; and reported preliminary reliability or prevalence data. The distinguishing characteristic

of psychometric studies pertained to the types of analyses conducted. We categorized an article as a psychometric study if it included evidence in support of at least one form of validity.³¹ Typically, psychometric studies involved larger samples; analysis of all ASQ age intervals; and reported data on a representative sample of participants. Psychometric studies also included information on the translation and adaptation process and scale reliability or referenced an article or manual containing evidence detailing this information. We grouped articles in the prevalence category if the authors described a clearly defined target population, identified the sample via probability sampling, and reported ASQ prevalence rates using valid, country-specific cutoffs.³² Finally, we categorized any article in which the authors used the ASQ to assess child development as a research study.

Results

We identified 55 peer-reviewed articles written in English detailing use of the ASQ in LMIC. All articles were published between 2007 and 2017, with the majority (93%) published between 2010 and 2017. We identified 51 of 55 articles via database searches. Four additional articles were identified through a cross check of our article list with a list from the lead author of the ASQ. Prior to analysis, we removed two research articles that referenced the ASQ but did not utilize the tool in their research. Thus, our final analysis included 53 articles.

Regions and countries

Figure 1 provides a flowchart of the articles by world region and country. Although only 53 articles are included in this review, four articles referenced two countries,^{33,34,35,36} one article referenced three countries,³⁷ and one article referenced four countries.³⁸ Thus, by country, the total number of articles referenced is 62 rather than 53. In total, we identified ASQ use in 23 of 140 LMIC (16%), including use in 3 of 31 (10%) low-income countries, 8 of 53 (15%) lower

middle-income countries, and 12 of 56 (21%) upper middle-income countries. As can be seen in Table 1, there were between 1 and 13 studies conducted in the 23 countries.

The 23 countries are distributed across all world regions. Specifically, we identified articles detailing ASQ use in 4 of 24 (16.7%) LMICs in East Asia and the Pacific (EAP; 9 articles); 3 of 21 (14.3%) LMICs in Europe and Central Asia (ECA; 4 articles); 5 of 26 (19.2%) LMICs in Latin America and the Caribbean (LAC; 11 articles); 2 of 13 (15.4%) LMICs in the Middle East and North Africa (MENA; 15 articles); 4 of 8 (50%) LMICs in South Asia (13 articles); and 5 of 47 (10.6%) LMICs in Sub-Saharan Africa (SSA; 12 articles). Within world region, articles are not distributed equally across the represented countries. In four regions – ECA, MENA, South Asia, and SSA – 50% or more of the articles pertain to work in one country. For example, 2 of 4 (50%) articles in ECA are from Turkey; 13 of 14 (92.9%) articles in MENA are from Iran; 10 of 13 (76.9%) articles in South Asia pertain to work in India; and 6 of 12 (50%) articles in SSA represent various studies conducted in South Africa. In EAP and LAC, the articles were distributed more evenly across countries.

Administration language

For 41 of the 53 (77.4%) articles the authors specified the language in which they administered the ASQ. The ASQ was translated into 16 languages: Afrikaans (n = 2), Arabic (n = 1), Bengali (n = 1), Brazilian Portuguese (n = 2), Chinese (n = 2), Georgian (n = 1), Hindi (n = 6), Nepalese (n = 1), Nyanja (n = 2), Persian (n = 13), Quechua (n = 1), Sesotho (n = 1), Spanish (n = 7), Thai (n = 1), Turkish (n = 2), and Zulu (n = 2).² For the remaining 13 studies (24%), the authors did not explicitly state the language of administration. For four studies, trained assessors translated the ASQ as needed during administration with the parent and child. For the remaining

² Four studies administered the ASQ in two different languages. Thus, the total sums to 46 rather than 42.

eight studies, the authors did not describe whether they used a translated version of the ASQ or the English version.

Administration procedures

Just over half of the studies reported parent completion of the ASQ (50.9%). Of the 27 studies reporting parent completion, 13 reported the parent completed the questionnaire independently and 14 reported completion via parent interview. For the remaining 26 in which the parent did not complete the ASQ, 17 reported a trained assessor conducted a direct assessment of the child, 2 reported teacher completion, 1 reported physician completion, and the other six studies did not provide information on who completed the questionnaire. Parent completion was highest among feasibility (75.0%) and prevalence studies (77.8%). For psychometric studies, 40% reported parent completion of the ASQ, which was comparable to the percent reported for research studies (38.5%). In East Asia and the Pacific all five studies reported parent completion of the ASQ. Parent completion rates were comparable across studies conducted in Europe and Central Asia (75%), the Middle East and North Africa (71.4%), and Sub-Saharan Africa (70%). In Latin America and the Caribbean and South Asia, only 10% of the studies in each region utilized parent completion. Parent completion rates were higher in upper middle-income countries (61.1%) as compared to lower middle-income (28.6%) and low-income countries (33.3%).

Study Type

As detailed in Table 2, study type varied considerably by economy. In low-income and lower middle-income contexts, the majority of identified articles reported findings from research studies (100% and 70%, respectively). The percentage of feasibility and psychometric studies

identified were comparable across lower middle-income and upper middle-income contexts.

Prevalence studies were unique to upper middle-income countries.

The distribution of article type also varied by region. Findings from a psychometric study were reported for at least one country in each world region. Every region also had at least one feasibility study with the exception of Europe and Central Asia. In three regions the majority of articles were research focused: East Asia and the Pacific (5 of 8 articles; 62.5%); Latin America and the Caribbean (7 of 11 articles; 63.6%); and South Asia (11 of 13 articles; 84.6%). Only three regions reported data from prevalence studies: East Asia and the Pacific (1 article); Europe and Central Asia (1 article); and the MENA (7 articles from Iran). Two regions had a relatively equal distribution of articles across three of four types. Articles on countries in Sub-Saharan Africa were distributed equally (e.g., four articles each) across feasibility, psychometric, and research studies. In Europe and Central Asia, articles were distributed across psychometric (2 articles), prevalence (1 article), and research (1 article) studies. Below, we discuss pertinent characteristics of the articles classified as feasibility, psychometric, and prevalence. A detailed discussion of identified research studies is outside the scope of this paper.

Feasibility studies

We identified eight feasibility studies, two conducted in lower middle-income countries (i.e., Ghana and India) and six in upper middle-income countries: Lebanon ($n = 1$), Peru ($n = 1$), South Africa ($n = 3$), and Thailand ($n = 1$). Reflecting their exploratory nature, most of the feasibility studies involved a small number of participating children (range = 15 – 733) identified predominately via convenience sampling (87.5%). As described in Table 3, three of eight studies examined all ASQ intervals (e.g., 2 – 60 or 4 – 60 months, depending on version), with the remainder translating and adapting a smaller range of age intervals. Across the eight studies,

participants resided in rural (n = 2), urban (n = 3), and suburban (n = 1) settings.³ As an example, the three feasibility studies conducted in South Africa targeted participants from rural,⁴⁵ urban,⁴⁴ and suburban⁴³ settings. Additionally, South African researchers examined the feasibility of translating the ASQ to three of the country's 11 languages (e.g., Afrikaans, Hindi, and Sesotho) for use with “middle-class” households^{43,44} as well as families facing socioeconomic inequalities.⁴⁵

A high-quality translation and adaptation process includes (a) forward-back translation, (b) cultural adaptation, (c) expert review, and (d) pilot testing.^{25,47} The feasibility studies identified in this review varied in terms of detail and quality. Two studies reported on all four steps^{41,44} and two other studies reported completing all steps with the exception of pilot testing.^{40,46} Two studies reported conducting a forward-back translation process prior to commencing research but neither study culturally-adapted the ASQ prior to using it. In one study, the researchers reported they found it necessary to make cultural adaptations after completing three interviews.⁴⁵ In the second,⁴³ the authors suggested follow-up studies to address potential cultural bias. Overall, six studies (75%) noted that some adaptation of ASQ items was necessary when using the measure in the new context.

Measures of reliability include inter-rater, test-retest, and coefficient alpha. Only one of eight feasibility studies (12.5%) provided information on all three forms of reliability. Two additional studies reported on two of three forms of reliability. Thus, across the eight studies, two (25%) reported inter-rater reliability, two (25%) reported test-retest reliability, and three (37.5%) reported coefficient alpha. Additionally, three studies provided valuable information on the

³ Two studies did not provide sufficient detail to ascertain setting type.

appropriateness, utility, and functionality of the ASQ via direct report from parents who had completed the screening process.

The identified feasibility studies did not systematically examine other feasibility factors⁴ such as acceptability, utility, training procedures, or administration time. Four studies^{41,43,45,46} reported information on parent acceptability, two studies discussed parent-reported utility of screening procedures,^{41,46} and one study⁴⁶ detailed information on early childhood staff's acceptability and utility of the ASQ. One study⁴⁰ described the education level and training required to administer the ASQ "home procedure." Three studies^{40,43,45} reported an estimated window of time for administration (e.g., 20-30 minutes). Finally, only one study⁴⁰ discussed cost, noting that the ASQ could be "conducted at a low cost."

Psychometric studies

In total, we identified 12 psychometric studies.⁴ Table 4 provides a description of the sample, language, ASQ age intervals, and study purpose. All 12 studies were completed in middle- income countries, with the majority (66.7%) completed in upper middle-income countries. The four studies conducted in lower middle-income (LMI) countries were completed in Georgia (n = 1), India (n = 1), and Zambia (n = 2). The seven studies completed in UMI countries were completed in Brazil (n = 2), China (n = 1), Iran (n = 1), Peru (n = 1), South Africa (n = 2), and Turkey (n = 1).

Reporting psychometric data for each study is beyond the scope of this review. Instead, we will describe the extent to which the studies (a) detail information on the adaptation and translation process, (b) report evidence of reliability (e.g., inter-rater, test-retest, and internal consistency), and (c) present information in support of the ASQ's internal structure, relationship

⁴ Two studies reported findings on two countries. Thus, the 12 studies discussed refer to 10 journal articles.

to other variables, and consequences in the new country and context. Since two studies reported findings on two countries,^{33,36} we will not treat these separately. Thus, the information presented below pertains to 10 rather than 12 psychometric studies.

Overall, the majority of studies provided strong support for the adaptation and translation process. Four of ten studies discussed all four steps (e.g., forward-back translation, adaptation, expert review, pilot) and an additional four studies conducted and reported on the first three steps but did not pilot the measure before use. Eight of ten studies provided information on internal consistency but only three studies reported test-retest reliability and only two reported inter-rater reliability statistics. Eight of ten studies reported on only one type of evidence to support the validity of using the ASQ in the new country and context. Four studies^{33,48,49,53} provided support for the internal structure of the ASQ (e.g., factor analysis, item analysis, etc.). Two studies^{52,55} reported on the ASQ's relation to other measures via correlations with a criterion measure. Finally, six studies^{36,50,51,52,54,55} reported on the consequences of ASQ use. Three reported on sensitivity and specificity; two examined the reasonableness of ASQ cutoffs in comparison to normative cutoffs in other countries; and one study addressed clinical implications of false screening results through examination of within-interval age variation.

Prevalence studies.

We identified nine prevalence studies conducted in three upper middle-income (UMI) countries, China (n = 1), Iran (n = 7) and Turkey (n = 1). All nine prevalence studies were conducted in countries that had completed large-scale psychometric studies of a translated and adapted version of the ASQ. In turn, all nine studies cited previous research and none reported new information on reliability or validity.

Discussion

The purpose of this review was to map current use of the ASQ in LMICs as reported in peer-reviewed published literature on the measure. We found ASQ use in all world regions; though rates of use – based on the percent of articles found – were higher in middle-income countries than in low-income countries. Within region, the percent of countries reporting ASQ use in peer-reviewed journals was comparable across most world regions with the exception of South Asia. These findings suggest broad global use of the ASQ with narrow, clustered use within regions. For example, no world region reported use in more than five countries and for three world regions – the Middle East and North Africa, South Asia, and Sub-Saharan Africa – more than 50% of the published articles were from studies conducted in a single country (Iran, India, and South Africa, respectively). Given that nearly half of the articles were research studies, these variations and the clustering of studies within specific countries may reflect regional or country-level variation in research priorities, capacity, or funding.

Study type varied by world region and economy as well. In low-income and lower middle-income countries, most if not all studies utilized the ASQ within a research study to examine child development. For lower middle-income and upper-middle-income countries, the percent of feasibility studies were comparable (10% and 15.4%, respectively) as were the percent of psychometric studies (20% and 20.5%, respectively). Published prevalence studies, however, were only conducted in upper middle-income countries. Measurement development and use does not necessarily follow a linear path from initial development and testing through systematic, large-scale validation, to use in research or practice. In fact, our findings suggest at least two “paths” exist. The first “path” involves a more systematic development process during which researchers (a) translated and adapted the ASQ following accepted guidelines; (b) generated evidence in support of reliability and validity; (c) conducted prevalence studies; and then (d)

utilized the ASQ in research studies. Implementation of these steps varied by country with some countries still in the earlier stages of the process (e.g., feasibility or psychometric studies) and other countries engaging in multiple steps concurrently. As well, it is worth noting, the accumulation of evidence can occur across multiple research groups working independently or in collaboration. The second translation and adaptation “path” involved utilization of the ASQ solely for research purposes. This path typically entailed abbreviated translation and adaptation of the measure, direct translation during the parent interview, or use of the English version of the measure.

Feasibility studies provide an important first step toward broad use of a screening measure in a new context or culture. These studies provide valuable preliminary information on the appropriateness and acceptability of a measure and, depending on the design of the study, may help inform development of screening systems and procedures including identification of enablers and barriers to implementation and scale-up. Three unrelated and temporally dispersed studies conducted in South Africa^{43,44,45} provide an impromptu exemplar of the value of examining feasibility as well as the complexity of translating and adapting assessments to new contexts. Each shares a similar purpose – examining the appropriateness of the ASQ in a new language and culture – and yet, each study adapts the measure to a different language for use with a unique population within the same country. Specifically, Abdoolah et al.⁴³ translated the 60-month interval of the ASQ to Hindi, and based on their findings, described it as “an appropriate tool for use with the middle socioeconomic class Hindi (Indian) language and culture.” Bornman et al.⁴⁴ adapted the 36 – 60 month ASQ intervals to Afrikaans, reporting them as “reliable assessments of urban, middle-class Afrikaans-speaking children’s development across the preschool years.” Finally, Visser et al.⁴⁵ translated and adapted the 24 – 48 month

intervals to Afrikaans and Sesotho for caregivers in a rural area of the country. Each of these studies contributes pertinent information on the appropriateness and acceptability of the translated and adapted intervals of the ASQ used as well as the screening procedures utilized to obtain the data. These studies also demonstrate the complexity of creating or adapting appropriate measures to LMICs where a range of languages and socioeconomic contexts may affect item understanding and interpretation and the acceptability of screening procedures. Many countries may require multiple language versions of a measure and even if a translated version already exists, such as the Hindi version of the ASQ used in India, use of the measure in a new country still requires translation, adaptation, and examination of the feasibility of implementing in a new context.

As Sireci and Padilla assert, “a single validation study cannot provide a compelling body of evidence to support the use of a test for a particular purpose.”²² Instead, developing a compelling validity argument is an iterative and on-going process. The majority of psychometric studies in this review reported limited information on reliability and only 20% of the psychometric studies we identified reported on more than one form of validity evidence. Additionally, only two studies provided findings on the ASQ’s relationship to other variables, limiting the evidence available to determine the extent to which ASQ domains converge with other similar measures or how well the ASQ predicts *actual* developmental delay in LMICs. The limited availability of criterion measures with strong evidence of reliability and validity for use in LMICs may partially explain the absence of these data, as noted in one study.⁵³ These findings, however, suggest the need for on-going research to fill evidence gaps and bolster validity arguments. Each study provides important evidence that contributes to an overall validity argument but a single study does not provide sufficient evidence. Furthermore, all of the

psychometric studies we identified were conducted in middle-income countries, with two-thirds of the studies from upper middle-income countries. Here again, multiple factors including capacity, funding, and priorities may explain these variations.

The prevalence studies conducted in China, Iran, and Turkey provide evidence in support of a systematic development process. All nine prevalence studies we identified were conducted in these three countries and each study cited previous large-scale psychometric studies of the ASQ to justify examination of prevalence rates. Interestingly, parents completed the ASQ directly or via parent interview for all nine of these studies, in accordance with the administration procedures utilized as part of the large-scale psychometric studies conducted in each country. Although anecdotal, this finding provides preliminary evidence to suggest that procedural decisions made during psychometric testing may “set the tone” for future use and may influence how researchers’ subsequently administer the tool.

With the exception of the feasibility studies, none of the studies in this review addressed real-world implementation of the ASQ or, in particular, the ways in which practitioners can utilize the tool to inform structured developmental surveillance or inform referral procedures for in-depth developmental assessment and possible access to ECI services. Five psychometric studies detailed information on sensitivity and specificity or the reasonableness of ASQ cutoffs; however, more evidence examining the consequences of use and the development, implementation, and scale-up of developmental screening systems in LMICs would be valuable to researchers and practitioners alike. For example, there was a dearth of information in the reviewed articles, which could be used to answer important questions about where, when, and how often the ASQ should be used in LMICs.

Limitations

There are a few notable limitations to this study. First, we limited our search to articles published in English in peer-reviewed journals. This approach may have resulted in the exclusion of articles published as grey literature. Additionally, there may be peer-reviewed articles published in other languages not captured as part of this review. For example, the second author of this paper is familiar with translations of the ASQ into Somali, Swahili, Tagalog, Thai, Urdu, and Vietnamese and is contributing to on-going work translating and adapting the ASQ for use in Ukraine and Myanmar. Thus, the information we report may underestimate the extent to which researchers and practitioners are actually using the ASQ and the number of languages into which researchers have translated the tool for use in LMICs. Furthermore, published articles may not reflect the full range of psychometric information compiled in support of a translated version of the ASQ. The Turkish version of the ASQ, for example, has an accompanying manual that may contain further evidence not described in the psychometric and prevalence studies discussed in this paper. Finally, as part of this study we did not distinguish between studies that utilized the ASQ as a population-based measure and those that utilized it as an individual-level measure of developmental delay. At least one study⁴⁵ piloted the ASQ as a population-based measure, reporting it has “a potentially high level of usefulness for large-scale, population-based surveillance purposes.” Future research should examine use of the ASQ for population-based surveillance and examine, in particular, consequential evidence pertaining to its use in this manner.

Conclusion

Returning briefly to the definitional shifts and omissions of evidence in the literature base noted at the outset of this manuscript, a number of important lessons emerge from a detailed evaluation of the literature on ASQ use in LMICs. First, contrary to Ertrem’s assertion,

developmental screening is not something done to groups of children by “testers” with limited knowledge. Instead, researchers and practitioners tend to administer the ASQ in accordance with the developer’s recommendations. They either allow parents to complete the measure themselves or facilitate parent completion through provision of additional supports. Second, when considered as a whole, the evidence base for the ASQ in LMICs is vast and promising. As we note herein, individual psychometric studies often report information on only one form of reliability or validity evidence. In turn, it is imperative researchers and practitioners interested in assessing the merits of a measure compile evidence across studies and draw their own conclusions. Using the ASQ as an example, Semrud-Clikeman and her colleagues⁶ reported ASQ use in only three countries and provided no evidence of normative data. We, in contrast, documented broad and varied use across a range of languages, countries, and contexts and we identified a range of psychometric evidence in support of the measure’s use in LMICs. Finally, there is a need for further validation studies across all of the cited regions. Despite the extensive evidence summarized in this article, construction of a validity argument is an on-going and iterative process that requires continuous updating as a measure is transported to a new country or to a new context within a country. Although measurement development may not follow a linear path from feasibility testing to large-scale validation, it is vital these steps occur and that evidence across studies is compiled and weighed, both for measures developed within and outside of LMICs.

References

1. Aly Z, Taj F, Ibrahim S. Missed opportunities in surveillance and screening systems to detect developmental delay: a developing country perspective. *Brain & Development* 2010; **32**: 90-97.
2. Scherzer AL, Chhagan M, Kauchali S, Susser E. Global perspective on early diagnosis and intervention for children with developmental delays and disabilities. *Developmental Medicine & Child Neurology* 2012; **54**: 1079-1084.
3. Bornstein MH, Hendricks C. Screening for developmental disabilities in developing countries. *Social Science & Medicine* 2013; **97**: 307-315.
4. Fischer VJ, Morris J, Martines J. Developmental screening tools: Feasibility of use at primary healthcare level in low- and middle-income settings. *J Health Popul Nutr* 2014; **32**: 314-326.
5. Robertson J, Hatton C, Emerson E, Yasamy MT. The identification of children with, or at significant risk of, intellectual disabilities in low- and middle-income countries: A review. *Journal of Applied Research in Intellectual Disabilities* 2012; **25**: 99-118.
6. 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 Neuropsychology* 2017; **23**: 761-802.
7. Cappa C, Petrowski N, Njelesani J. Navigating the landscape of child disability measurement: A review of available data collection instruments. *ALTER, European Journal of Disability Research* 2015; **9**: 317-330.
8. 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**: 482-488.
9. Fernald LCH, Prado E, Kariger P, Raikes A. *A toolkit for measuring early childhood development in low- and middle-income countries*. Washington DC: The World Bank; 2017.
10. Ertrem I. *Developmental difficulties in early childhood: Prevention, early identification, assessment and intervention in low- and middle-income countries: A review*. Geneva: World Health Organization; 2012.
11. Ertrem, I. The international Guide for Monitoring Child Development: Enabling individualized interventions. *Early Childhood Matters: Advances in early childhood development*. The Hague: Bernard van Leer Foundation; 2017.
12. Singh A, Yeh CJ, Blanchard SB. Ages and Stages Questionnaire: A global screening scale. *Boletín Médico del Hospital Infantil de México* 2017; doi 10.1016/j.bmhmx.2016.07.008 [E-pub ahead of print]
13. Velikonja T, Edbrooke-Childs J, Calderon A, Slead M, Brown A, Deighton J. The psychometric properties of the Ages & Stages Questionnaires for ages 2-2.5: A systematic review. *Child: Care, Health and Development* 2017; **43**: 1-17.
14. American Academy of Pediatrics Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee, Medical Home Initiatives for Children with Special Needs Project Advisory Committee. Identifying infants and young children with developmental disorders in the medical home: an algorithm for developmental surveillance and screening. *Pediatrics* 2006; **118**: 405-420.

15. Marks KP, Glascoe FP, Macias MM. Enhancing the algorithm for developmental-behavioral surveillance and screening in children 0 to 5 years. *Clin Pediatr* 2011; **50**: 853-868.
16. Hix-Small H, Marks K, Squires J, Nickel B. Impact of implementing developmental screening at 12 and 24 months in a pediatric practice. *Pediatrics* 2007; **120**: 381-389.
17. Chung CY, Liu WY, Chang CJ, Chen CL, Tang SF, Wong, AM. The relationship between parental concerns and final diagnosis in children with developmental delay. *Journal of Child Neurology* 2011; **26**: 413-419.
18. Glascoe FP. Parents' evaluation of developmental status: How well do parents' concerns identify children with behavioral and emotional problems? *Clinical Pediatrics* 2003; **42**: 133-138.
19. Marshall J, Coulter ML, Gorski PA, Ewing A. Parent recognition and responses to developmental concerns in young children. *Infants & Young Children* 2016; **29**: 102-115.
20. Squires J, Twombly E, Bricker D, Potter L. ASQ-3 User's guide. Baltimore, MD: Paul H. Brookes Publishing Co, 2009.
21. Sireci S, Faulkner-Bond, M. Validity evidence based on test content. *Psicothema* 2014; **26**: 100 – 107.
22. Sireci S, Padilla J. Validating assessments: Introduction to the special section. *Psicothema* 2014; **26**: 97 – 99.
23. Downing SM. Validity: On the meaningful interpretation of assessment data. *Medical Education* 2003; **37**: 830 – 837.
24. Padilla J, Benítez I. Validity evidence based on response processes. *Psicothema* 2014; **26**: 136 – 144.
25. Beaton DE, Bombardier C, Guillemin F, Ferraz, MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *SPINE* 2000; **25**: 3186 – 3191.
26. Hambleton RK. Issues, designs, and technical guidelines for adapting tests into multiple languages and cultures. In: Hambleton RK, Merenda PF, Spielberger CD, editors. *Adapting Educational and Psychological Tests for cross-cultural assessment*. London: LEA; 2005. 3-38.
27. Rios J, Wells C. Validity evidence based on internal structure. *Psicothema* 2014; **26**: 108 – 116.
28. Lane S. Validity evidence based on testing consequences. *Psicothema* 2014; **26**: 127 – 135.
29. World Bank. World Bank Country and Lending Groups. Retrieved from <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups>.
30. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, Griffey R, Hensley M. Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health* 2011; **38**: 65-76.
31. Cizek GJ, Rosenberg SL, Koons HH. Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement* 2008; **68**: 397-412.
32. Boyle MH. Guidelines for evaluating prevalence studies. *Evidence-Based Mental Health* 1998; **1**: 37-39.
33. Hsiao C, Richter L, Makusha T, Matafwali B, Heerden A, Mabaso M. Use of the Ages and Stages Questionnaire adapted for South Africa and Zambia. *Child: Care, Health and Development* 2017; **43**: 59-66.

34. Muennig P, Boulmier-Darden P, Khouzam N, Zhu W, Hancock P. Predictors of health among refugee adults from Myanmar and the development of their children. *J Immigrant Minority Health* 2015; **17**: 1385-1390.
35. Tobin T, Boulmier P, Zhu W, Hancock P, Muennig P. Improving outcomes for refugee children: A case study on the impact of Montessori education along the Thai-Burma border. *The International Education Journal: Comparative Perspectives* 2015; **14**: 138-149.
36. van Heerden A, Hsiao C, Matafwali B, Louw J, Richter L. 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* 2017; **17**: 1-9.
37. Wallander JL, Bann CM, Biasini FJ et al. Development of children at risk for adverse outcomes participating in early intervention in developing countries: A randomized controlled trial. *J Child Psychol Psychiatry* 2014; **55**: 1251-1259.
38. Fernald LCH, Kariger P, Hidrobo M, Gertler PJ. Socioeconomic gradients in child development in very young children: Evidence from India, Indonesia, Peru, and Senegal. *PNAS* 2012; **109**: 17273-17280.
39. Bello AI, Quartey JNA, Appiah LA. Screening for developmental delay among children attending a rural community welfare clinic in Ghana. *BMC Pediatrics* 2013; **13**: 1-7.
40. Kvestad I, Taneja S, Kumar T, Bhandari N, Strand TA, Hysing M. The assessment of developmental status using the Ages and Stages Questionnaire-3 in nutritional research in north Indian young children. *Nutrition Journal* 2013; **12**: 1-11.
41. Charafeddine L, Sinno D, Ammous F, Yassin W, Al-Shaar L, Mikati MA. Ages and Stages Questionnaires: Adaptation to an Arabic speaking population and cultural sensitivity. *European Journal of Paediatric Neurology* 2013; **17**: 471-478.
42. Kyerematen V, Hamb A, Oberhelman RA, Cabrera L, Bernabe-Ortiz A, & Berry SJ. Exploratory application of the Ages and Stages (ASQ) child development screening test in a low-income Peruvian shantytown population. *BMJ Open* 2014; **4**: 1-4.
43. Abdoola S. Translating tools for better parent-based assessment: An exploratory study. *The South African Journal of Communication Disorders* 2015; **62**: 1-12.
44. Bornman J, Sevcik RA, Ronski M, Pae HK. Successfully translating language and culture when adapting assessment measures. *Journal of Policy and Practice in Intellectual Disabilities* 2010; **7**: 111-118.
45. Visser M, Nel M, Bronkhorst C, et al. Childhood disability population-based surveillance: Assessment of the Ages and Stages Questionnaire Third Edition and Washington Group on Disability Statistics/UNICEF module on child functioning in a rural setting in South Africa. *African Journal of Disability* 2016; **5**: 1-9.
46. Saihong P. International Conference on Learner Diversity 2010: Use of Screening Instrument in Northeast Thai Early Childcare Settings. *International Conference on Learner Diversity 2010, Procedia - Social and Behavioral Sciences* 2010; **7**: 97-105.
47. World Health Organization. Process of translation and adaptation of instruments. Retrieved from http://www.who.int/substance_abuse/research_tools/translation/en/.
48. Filgueiras A, Pires P, Maissonette S, Landeira-Fernandez J. Psychometric properties of the Brazilian-adapted version of the Ages and Stages Questionnaire in public child daycare centers. *Early Human Development* 2013; **89**: 561-576.

49. Santana CMT, Filgueiras A, Landeira-Fernandez J. Ages and Stages Questionnaire – Brazil – 2011: Adjustments on an early childhood development screening measure. *Global Pediatric Health* 2015; **2**: 1-12.
50. Bian X, Yao G, Squires J, et al. Translation and use of parent-completed developmental screening test in Shanghai. *Journal of Early Childhood Research* 2012; **10**: 162-175.
51. Zirakashvili M, Gabunia M, Tatishvili N, et al. Cultural adaptation and psychometric validation of the Ages and Stages Questionnaires for use in Georgia. *Journal of Child and Family Studies* 2017; doi.org/10.1007/s10826-017-0917-z
52. Juneja M, Mohanty M, Jain R, Ramji S. (2012). Ages and Stages Questionnaire as a screening tool for developmental delay in Indian children. *Indian Pediatrics* 2012; **49**: 457-461.
53. Vameghi R, Sajedi F, Mojembari AK, Habiollahi A, Lornezhad HR, Delavar B. Cross-cultural adaptation, validation and standardization of Ages and Stages Questionnaire (ASQ) in Iranian children. *Iranian Journal of Public Health* 2013; **42**: 522-528.
54. Chong KC, Zhou VL, Tarazona D, et al. ASQ-3 scores are sensitive to small differences in age in a Peruvian infant population. *Child: Care, health and development* 2017; **43**: 556-565.
55. Kapci EG, Kucuker S, Uslu RI. How applicable are Ages and Stages Questionnaires for use with Turkish children? *Topics in Early Childhood Special Education* 2010; **30**: 176-188.

Table 1. Categorization criteria for articles.

Article category	Categorization Criteria
Feasibility study	<ul style="list-style-type: none"> • The article should meet at least one of the following criteria: <ul style="list-style-type: none"> ○ Includes information on the translation and adaptation process for the ASQ <i>and</i> included preliminary reliability data; ○ Includes information on the translation and adaptation process for the ASQ <i>and</i> explicitly discussed or presented data on the feasibility or utility of implementing the ASQ in a new context or culture; ○ Situates collection of the ASQ in the context of developmental delay and early identification, implemented the ASQ with a sample of participants, and reported preliminary prevalence rates or prevalence rates on a small sub-sample of the population.
Psychometric study	<ul style="list-style-type: none"> • The article presents at least one form of validity evidence (e.g., content, response process, internal structure, relationship to other variables, or consequences). • The article should also discuss the adaptation and translation process and reliability evidence or should reference research describing this process (e.g., another article, technical report, or manual).
Prevalence study	<ul style="list-style-type: none"> • The article adheres to guidelines for prevalence studies including analysis of a clearly defined target population identified via probability sampling and reports ASQ domain and/or age by domain prevalence rates using valid, country-specific cutoffs.
Research study	<ul style="list-style-type: none"> • The article uses the ASQ within a broader research study to examine child development.

Table 2. Identified studies by economy and study type.

Economy	Countries	Study Type*			
		Feasibility N(%)	Psychometric N(%)	Prevalence N(%)	Research N(%)
Low-income (n = 3)	<ul style="list-style-type: none"> • Nepal (n = 1) • Senegal (n = 1) • Uganda (n = 1) 	0 (0.0)	0 (0.0)	0 (0.0)	3 (100.0)
Lower middle-income (n = 20)	<ul style="list-style-type: none"> • Bangladesh (n = 1) • Georgia (n = 1) • Ghana (n = 1) • India (n = 10) • Indonesia (n = 1) • Myanmar (n = 2) • Pakistan (n = 1) • Zambia (n = 3) 	2 (10.0)	4 (20.0)	0 (0.0)	14 (70.0)
Upper middle-income (n = 39)	<ul style="list-style-type: none"> • Albania (n = 1) • Brazil (n = 2) • China (n = 2) • Colombia (n = 1) • Ecuador (n = 4) • Iran (n = 13) • Lebanon (n = 1) • Mexico (n = 1) • Peru (n = 3) • South Africa (n = 6) • Thailand (n = 3) • Turkey (n = 2) 	6 (15.4)	8 (20.5)	9 (23.1)	16 (41.0)

*Although there are only 53 articles in the review, four articles referenced two countries; one article referenced three countries; and one referenced four countries. Thus, the total number of articles by country reported in this table equals 62 rather than 53.

Table 3. Characteristics of feasibility studies (n = 8).

Study	Country	Sample	Sample size	Language	Age intervals	Study purpose
Bello et al. ³⁹	Ghana	“mothers of children whose ages were less than 5 years and who were attending the out-patient department of a rural welfare clinic in Ghana (p. 2)	330	Not specified	2 – 60 months	“to screen children under five years of age for developmental delay in a rural Welfare Clinic with special consideration to their socio-demographic risk factors” (p. 2)
Kvestad et al. ⁴⁰	India	Children from the “low and middle socioeconomic settings of Tigri and Dakshinpuri in New Delhi” (p. 3)	422	Hindi	12 – 36 months	“to assess the feasibility of the ASQ-3 ‘home procedure’ for measuring developmental status in young children in a field trial in New Delhi, India” (p. 3)
Charafeddine et al. ⁴¹	Lebanon	Parents and children “from health care centers (private and public sectors) and day care centers located in all Lebanese provinces” (p. 472)	733	Arabic	4 – 60 months	“the aim of this study was not to validate but to test the adequacy of administering ASQ in Arabic and to assess its reliability and its cultural acceptance in a sample of Arabic speaking population” (p. 472)
Kyerematen et al. ⁴²	Peru	Children from Pampas de San Juan de Miraflores in Lima, Peru	129	Spanish	4 – 60 months	“to (1) implement the Ages and Stages Questionnaires...in children aged 3 months to 5 years in a low-income Peruvian population and (2) to correlate outcomes of the ASQ with risk factors” (p. 1)

Study	Country	Sample	Sample size	Language	Age intervals	Study purpose
Abdoola et al. ⁴³	South Africa	“South African preschool children of Indian origin” (p. 3)	15	Hindi	60 months	“ to determine the appropriateness of translations of the ASQ instrument (60 month age group) from the English to Hindi language and culture” (p. 2)
Bornman et al. ⁴⁴	South Africa	“children from three preschools in the same geographical area of Gauteng, South Africa” (p. 113)	47	Afrikaans	36 – 60 months	“to examine the utility and functionality of translations of instruments from English to Afrikaans’ language and culture” (p. 113)
Visser et al. ⁴⁵	South Africa	Children from “six towns in the Kopanong Municipal area situated within the Xhariep District” (p. 4)	50	Afrikaans and Sesotho	24 – 48 months	“to report on a pilot study that investigated the sensitivity and specificity of translated versions of the ASQ-III and the WG/UNICEF module as parent-reported measures, to identify early childhood disabilities in children 24-48 months...to determine their appropriateness for population-based studies in similar contexts” (p. 4)
Saihong ⁴⁶	Thailand	“children between the ages of 24 to 36 months...and their parents and early childcare staff/teachers” (p. 99)	267	Thai	24 – 36 months	“ to determine and investigate the psychometric properties and utility of a screening system using an adapted version of...the Ages and Stages Questionnaire: Thai (ASQ: Thai), in early childcare settings in Northeast Thailand” (p. 99)

Table 4. Characteristics of psychometric studies (n = 10).⁵

Study	Country	Sample	Sample size	Language	Age intervals	Study purpose
Filgueiras et al. ⁴⁸	Brazil	“children enrolled in all of the 468 public daycare centers in the city of Rio de Janeiro” (p. 562)	45,640	Brazilian Portuguese	6 – 60 months	“to translate all of the ASQ-3 questionnaires into Brazilian Portuguese and explore their psychometric characteristics...to develop a reliable assessment instrument that might be used to help evaluate Brazilian public child daycare programs and allow the screening of children for possible developmental delays” (p. 562)
Santana et al. ⁴⁹	Brazil	“children enrolled in 972 public daycare centers and public preschools in Rio de Janeiro” (p. 2)	67,522	Brazilian Portuguese	10 – 60 months	“to improve the ASQ-BR by proposing changes based on Filgueiras et al study in order to try to solve the problems pointed out in the first adaptation of the ASQ-3 to Brazilian Portuguese for a public daycare center sample” (p. 2)
Bian et al. ⁵⁰	China	Parents/caregivers and children from “18 districts, health care clinics, and institutes in greater Shanghai” (p. 164)	8,472	Chinese	2 – 60 months	“to evaluate a...Chinese translation of the ASQ-3...in Shanghai, China, and to study the validity and reliability in terms of identification of developmental delays in preschool children” (p. 163)
Zirakashvili et al. ⁵¹	Georgia	Parents and children recruited from	2,974	Georgian	2 – 60 months	“to culturally adapt, standardize, and validate the Georgian version of the ASQ-3

⁵ Twelve psychometric studies were reported in ten manuscripts. For sake of clarity and concision, we only report Hsiao et al. and van Heerden et al. once in this table.

Study	Country	Sample	Sample size	Language	Age intervals	Study purpose
		regional primary health care centers in five regions of Georgia				for children between one and 66 months of age, including the assessment of test parameters and factors associated with domain area scores” (p. 5)
Juneja et al. ⁵²	India	Children from the Child Development Clinic of Lok Nayak Hospital in New Delhi, India	200	Hindi	4, 10, 18 and 24 months	“to evaluate ASQ for detecting developmental delay in Indian children” (p. 457)
Vameghi et al. ⁵³	Iran	Children from 4 – 60 months old “in selected cities throughout the country” (p. 522)	10,516	Persian	4 – 60 months	“for the purpose of cultural adaptation, validation and standardization of the ASQ questionnaire for 4 – 60 months old Iranian children and children with similar socio-cultural backgrounds, such as all those living in the Middle East” (p. 523)
Chong et al. ⁵⁴	Peru	“infants under 2 years old that were evaluated in August 2013 for the baseline survey of <i>Cuna Más</i> ” (p. 558)	5,850	Quechua & Spanish	2 – 24 months	“aims to determine if an age-dependent fluctuation in scores within a given ASQ-3 screening interval exists within children evaluated for <i>Cuna Más</i> ” (p. 557)
Kapci et al. ⁵⁵	Turkey	Children recruited from preschools, special education schools, hospitals, and community health	978	Turkish	4 – 60 months	To assess the cultural appropriateness of the Turkish ASQ and assesses the reliability and validity of the measure

Study	Country	Sample	Sample size	Language	Age intervals	Study purpose
		centers in Ankara and Denizli				
Hsiao et al. ³³	South Africa & Zambia	Caregivers and children “from well-baby clinics, daycare centres and home visits in Lusaka, a peri-urban environment” (p. 60)	422 & 431	Zulu & Nyanja	2 – 60 months	“to test the psychometric properties and appropriateness of the ASQ-3 in South Africa and Zambia...to determine how the instrument may need to be adapted to render age-appropriate assessment in the region” (p. 60)
Van Heerdan et al. ³⁶	South Africa & Zambia	Caregivers and children “from well-baby clinics, daycare centres and home visits in peri-urban settlements close to Lusaka” (p. 2)	422 & 431	Zulu & Nyanja	2 – 60 months	“for the purpose of cultural adaptation, validation and standardization of the ASQ questionnaire for 2 – 60-month-old children living in Southern Africa” (p. 2)

