Screening for Intellectual Disabilities and/or Autism Amongst Older Children and Young Adults: a Systematic Review of Tools for Use in Africa

Eziafakaku Uchechukwu Nwokolo[1] · Peter E. Langdon[2] · Glynis H. Murphy[1]

Abstract

There are many well-developed screening tools for both intellectual disabilities and autism, but they may not be culturally appropriate for use within Africa. Our specific aims were to complete a systematic review to (1) describe and critically appraise short screening tools for the detection of intellectual disabilities and autism for older children and young adults, (2) consider the psychometric properties of these tools, and (3) judge the cultural appropriateness of these tools for use within Africa. Six screening tools for intellectual disabilities and twelve for autism were identified and appraised using the Consensus-based Standards for the Selection of Health Measurement Instruments (COSMIN) guidelines. We identified two screening tools which appeared appropriate for validation for use within African nations.

Keywords Adolescent · Screening · Diagnosis · Autism · Intellectual disabilities · Africa

Several studies (e.g. Eldevik et al., 2009; Luckasson & Schalock, 2013; Schalock & Luckasson, 2013; Steiner et al., 2012; Swinkels et al., 2006) have highlighted the benefits of early detection of developmental disabilities such as intellectual disabilities and autism. The benefits have included improved behavioural outcomes and family support, as well as earlier intervention. Other benefits included improved planning for educational needs and support, improved social skills, and greater cognitive and language development. These findings have emerged predominantly from Western and high-income countries with there having been very limited research from low- to medium-income countries (LMICs), as indexed by the published gross national income by the United Nations (Gladstone et al., 2010; Tomlinson et al., 2014; United Nations, 2014; United Nations Department of Economic & Social Affairs, 2021; World Bank, 2020). While the presentation of autism is the same regardless of economic status, the political climate and associated social burdens within LMICs, such as in the African countries, discourage the early detection of developmental disabilities as it is not seen as urgent, which increased the health disparities faced by this population (Emerson, 2012; Gladstone et al., 2014). The situation is similar for those with intellectual disabilities, with late identification leading to further delay of intervention.

Screening for developmental disabilities can be done in any setting, such as the community (Kopp & Gillberg, 2011), schools (Suhail & Zafar, 2008; Webb et al., 2003), primary care settings (Robins, 2008; Barton et al., 2012; Gura et al., 2011; Limbos et al., 2011), urban settings (Guevara et al., 2013), the criminal justice system (Murphy et al., 2017), and many others. In the African context, individuals with developmental disabilities are noticed either in schools or when parents seek medical attention for a severe illness or when researchers embark on studies targeted specifically at populations with disabilities (Gladstone et al., 2010; Knox et al., 2018; Saloojee et al., 2007; Scherzer et al., 2012).

Preliminary screening for intellectual disabilities or autism can occur through the use of a variety of methods, such as observation, informal and formal interviews, history taking, and the use of short screening tools. Irrespective of which method is used, the important factors to consider are the accuracy of results, validity, reliability, training requirements, ease of administration, and the simplicity and ease

✉ Eziafakaku Uchechukwu Nwokolo
eun5@kent.ac.uk

[1] Tizard Centre, University of Kent, Canterbury, UK

[2] Centre for Educational Development, Appraisal and Research (CEDAR), University of Warwick and Coventry and Warwickshire Partnership NHS Trust, Coventry, UK

of interpreting results (Cochrane & Holland, 1971; Westerlund & Sundelin, 2000). The accuracy of screening tools is vital, and Glascoe (2005) recommends that the sensitivity, or the true positive rate, should be between 70 and 80%, while specificity, or the true negative rate, should be at least 80%. Screening tools require validation when used outside the environment and population for which they were developed, and this process involves comparing the results of the screening tool to that of an accepted gold standard instrument (Maxim et al., 2014). Most screening tools in existence have been validated in the West, but evidence for their validation in Africa is scant (Soto et al., 2015; Van der Linde et al., 2015).

Another consideration is the adaptation of measures for use outside of the original design environment. A robust screening tool should be culturally sensitive and useable with multiple populations (Van der Linde et al., 2015). Given that almost all the measures were developed within Western countries, issues regarding cultural sensitivity and feasibility of using these screening tools in their original format with the African populace need investigating. Screening tools developed in high-income environments do not necessarily consider the application and understanding of the terminology in other environments. Screening results and reliability can be affected where the language of the screening tool differs in application or understanding (Soto et al., 2015). In Africa, some studies that measured developmental milestones and disabilities utilised screening tools developed in the West (Oshodi et al., 2016; Koura et al., 2013; Jinabhai et al., 2004). For example, Oshodi et al., (2016) used the Modified Checklist for Autism in Toddlers (M-CHAT) in a Nigerian urban setting where language and terminology were not barriers, thereby eliminating the need for translation. Jinabhai et al. (2004) adapted and substituted examples in both the Auditory Verbal Learning Test (AVLT) and Young's Group Mathematics Test (GMT) with more familiar items for Zulu participants. The AVLT instructions were given in Zulu with the items 'turkey' and 'ranger' replaced with the more culturally familiar Zulu words 'chicken' and 'herdboy', respectively. Jinabhai et al. (2004) made considerably more adaptations to the GMT and administered the test in Zulu. The adaptations centred on change of words and examples to more familiar items such as 'tarts' to 'cakes', 'marbles' to 'balls', 'engine' to 'truck' and the names 'Dick and Jim' were changed to 'Sipho and Thembi'. Koura et al. (2013) adopted a rigorous translation model to translate and adapt the Mullen Scales of Early Learning (MSEL) used in their study from English to French, while the parents' instruction was translated into Fon, the local language.

There are some other studies which highlight the importance of language and terminology in translated versions. Wild et al. (2012) translated the CBCL into six languages: Korean, Hebrew, Spanish, Kannada, and Malayalam. In the

Malayalam version, several cultural adaptations such as changing the 'milk delivery' to a more familiar job and giving different examples of sports and hobbies were needed, while in the Hebrew version, two sexually related items were removed from the measure. Koura et al. (2013) also used the 'Ten Questions' (TQ) to screen for disabilities and collect cognitive development information for their participants; however, the items were not translated into other languages.

The 'Ten Questions' is a disability screening tool which has been used widely in developing countries. The TQ was primarily designed as a stop-gap screening tool for numerous kinds of impairment in children aged 2–9 years old, including intellectual disabilities, and has been used to estimate prevalence within low-income and low-resource countries. The TQ is about cognitive skills, motor skills, hearing, epilepsy, and vision problems. Stein et al. (1986) used the measure as a screening tool in the first stage of their prevalence study across several countries, to identify children with moderate to profound intellectual disabilities. Intellectual disabilities were classified as an intelligence quotient less than or equal to 55 (IQ ≤ 55). Study samples were from eight countries (India, Philippines, Bangladesh, Sri Lanka, Malaysia, Pakistan, Brazil, and Zambia). No specific figures were reported for sensitivity and specificity. However, the team reported that most participants with intellectual disabilities were probably identified, while children with other conditions and IQs greater than 55 were also identified. These results from Stein et al. (1986) do not seem adequate to judge the psychometric properties of the TQ. Also, any consideration of cultural issues was not documented. Two other studies, Mung'ala-Odera et al. (2004) in Kenya and Kakooza-Mwesige et al. (2014) in Uganda, used the TQ with children in their early years. Kakooza-Mwesige and colleagues screened 1169 Ugandan children between the ages of 2 and 9 years using an adapted version of the TQ, which included 13 additional questions about autism. Questions about autism covered the three criteria: qualitative impairment in social interaction, qualitative impairments in communication, and restricted repetitive and stereotypical behaviours. The adapted version of the TQ was called the 23Q. The authors reported high negative predictive value (0.90) and specificity (0.90) with very low positive predictive value (0.22) and sensitivity (0.52) for participants with autism. As such Kakooza-Mwesige et al. (2014) concluded that the neither the TQ nor the 23Q met the criteria as useful screening tools for autism. While the TQ has been useful in identifying children with specific disabilities, its appropriateness for detecting more complex and hidden disabilities such as mild intellectual disabilities or autism are unclear (Durkin, 2001; Olusanya & Okolo, 2006). Also, suggestions have been made that the continued use of the TQ in Africa or LMICs may undermine efforts towards effective screening and early intervention (Olusanya & Okolo, 2006).

Additionally, screening tools are sometimes used to monitor the progress of interventions, where the same measure is re-administered to the same individual, to examine progress, indicative of the 'responsiveness' of a screening tool. McConachie et al. (2015) reviewed the measurement properties of some screening tools used to measure progress and outcomes in young children with autism spectrum disorder aged up to 6 years. Their reviewed focused on measuring the progress and improved quality of life post-intervention for participants in the West.

Soto et al. (2015), in their systematic review of 21 included studies, investigated efforts towards the cultural adaptation of screening tools for use outside of the environments in which they were primarily developed. With a specific emphasis on autism spectrum disorder only, the review examined the adherence to recommended adaptation procedures and the psychometric properties of the adapted instruments. Studies about people with intellectual disabilities were excluded. The adaptation studies included in the Soto et al. review had been carried out in nineteen countries and involved ten languages. Only two of those countries are in continental Africa: Egypt and Tunisia. The M-CHAT was used in the studies in both countries. Egypt and Tunisia are Arabic-speaking countries, and the M-CHAT was translated into Arabic. In LMICs, where resources are limited, the cost and burden of a rigorous translation and adaptation process are barriers to acquiring reliable screening tools.

Recently, attempts have been made towards developing screening tools in areas such as nutrition, neurodevelopmental disabilities, and mental health which are culturally sensitive for use within the African continent (Gladstone et al., 2010; Hasegawa et al., 2017; Vawda et al., 2017). While these efforts are commendable, study populations are often limited to early childhood, with children aged 2- to 9-year-olds. The focus upon young children (2- to 5-year-olds) would allow for the implementation of interventions earlier but would miss older children (10 years and above). Relative to studies on young children, there is very little data on studies with older children and adolescents; however, studies involving adolescents are emerging (Allison et al., 2012; Morales-Hidalgo et al., 2017Nijman et al., 2018). The paucity of adolescent studies is not peculiar to Africa. What appears to be unique to LMICs and Africa is the relatively low level of awareness, insufficient economic resources, insufficient numbers of professionals, and a culture of not seeking immediate help (Franz et al., 2017).

In African countries like Nigeria, Kenya, Ghana, and Uganda, awareness is growing, yet it is still common for families not to seek immediate help for individuals with autism or intellectual disability till later in life (Franz et al., 2017). It therefore remains the case that many of these children are not screened or diagnosed early in life. Such individuals are then brought to the attention of professionals around the onset of adolescence as this is the period when teenagers begin to spend an increasing amount of time away from the family home. Adolescents and young people are those aged 11 to 26 years of age, an age range which is consistent with the critical period of brain maturation associated with development during adolescence (Sawyer et al., 2012, 2018). To identify these older children and young adults who have been missed or not diagnosed in a time-efficient and effective way, an appropriate screening tool should be available. However, there is a marked absence of well-developed screening tools for use with adolescents amongst professionals and services in African countries (Hirota et al., 2018).

Overall, screening for either intellectual disabilities or autism in individuals in African countries requires the use of a validated and reliable measure which is accessible to front line professionals such as teachers, nurses, carers, family doctors, and those who are in primary healthcare services. While some screening tools have been developed and validated in the West, and investigated for use in Africa, the researchers have not always compared their study results against acceptable gold standard instruments, a crucial stage in measuring the validity of tools when used in new environments. For instance, Oshodi et al. (2016) and Koura et al. (2013) obtained reasonable results from their studies. However, they did not compare their results to that of an acceptable gold standard instrument, and this presents limitations. Besides selecting and validating a standardised screening instrument for use with adolescents, the tool ought to be culturally relevant for use within the African context. Through careful adaptation and translational work, screening tools developed in the West may be adopted for use in LMIC such as Nigeria, Ghana, and other African countries. By doing so, some of the costs and time to develop entirely new tools can be reduced.

To identify such tools, a systematic review was completed with the following aims: (1) to describe and critically appraise short screening tools for the detection of intellectual disabilities and autism in children and young people aged 11 to 26 years, (2) to consider the psychometric properties of these tools, and (3) to consider the appropriateness of using these tools across a range of cultures.

Method

Search Strategy

A literature search of the following electronic databases was carried out to identify relevant studies: Academic Search Complete, MEDLINE, CINAHL Plus, PsycINFO, and PsycArticles. The key search terms were 'intellectual', 'learning', and 'autism'. These key terms were then combined with disability and with screening and diagnosis. Truncated terms

were used as appropriate to ensure inclusion of variations of the words. Older words used to describe people with intellectual disabilities, such as 'mentally retarded' or 'mental retardation', were also included. Titles and abstracts were the focus of the initial search. The combined search terms are found in Table 1 in the supplementary material. Backward (ancestry) searching was used to identify other papers that may be relevant from references of eligible studies. The search was done using EBSCOhost and concluded on the 22nd of June 2018. To ensure that no new studies published, or tools developed were missed, the search was updated with the same terms on the 5th of November 2020.

To provide transparency of the review process and avoid duplication of the study, the review was registered with Research Registry (https://www.researchregistry.com/; registration code, reviewregistry798). Research Registry is an international database for registering all types of research studies such as case reports, observational and interventional studies, and systematic reviews and meta-analyses.

Eligibility Criteria and Study Selection

Titles and abstracts were initially screened for inclusion based on the following criteria: (1) the article was written in English, (2) validated screening tools were used, or the study involved developing a screening tool, (3) little or no extra training was required to administer the tool, (4) the tool did not take longer than 1 h to administer, (5) some or the majority of the participants were aged 11 years and younger than 27 years, and (6) participants in the validation sample for intellectual disabilities or autism were diagnosed by a duly qualified healthcare professional. Some articles which had multiple studies and participants across a broad age range (Baron-Cohen et al., 2006; McKenzie et al., 2012c; Nijman et al., 2018; Deb et al., 2009; Kraijer & De Bildt, 2005) were included because of their relevance in at least part of their research. Studies were excluded if any of the following criteria were met: (1) the study was related to other health issues such as diabetes, cancer, visual, and any other medical condition in persons with intellectual disabilities or autism; (2) the study was about linguistic and speech-related conditions; (3) the study was about developmental learning disorders/difficulties (e.g. impairments in reading or writing); (4) the tools were not for screening but diagnostic tools; (5) publications were letters, correspondences, editorials, or recommendations to the editors; (6) studies had missing information on age; (7) full text was not available; and (8) additional skills or training were required to administer the tool. Due to the paucity of research with adolescents, and in order not to miss any potential screening tools, there was no restriction on publication date. Studies done in both clinical and non-clinical settings were considered. Also, the inclusion of English only articles was based on the authors'
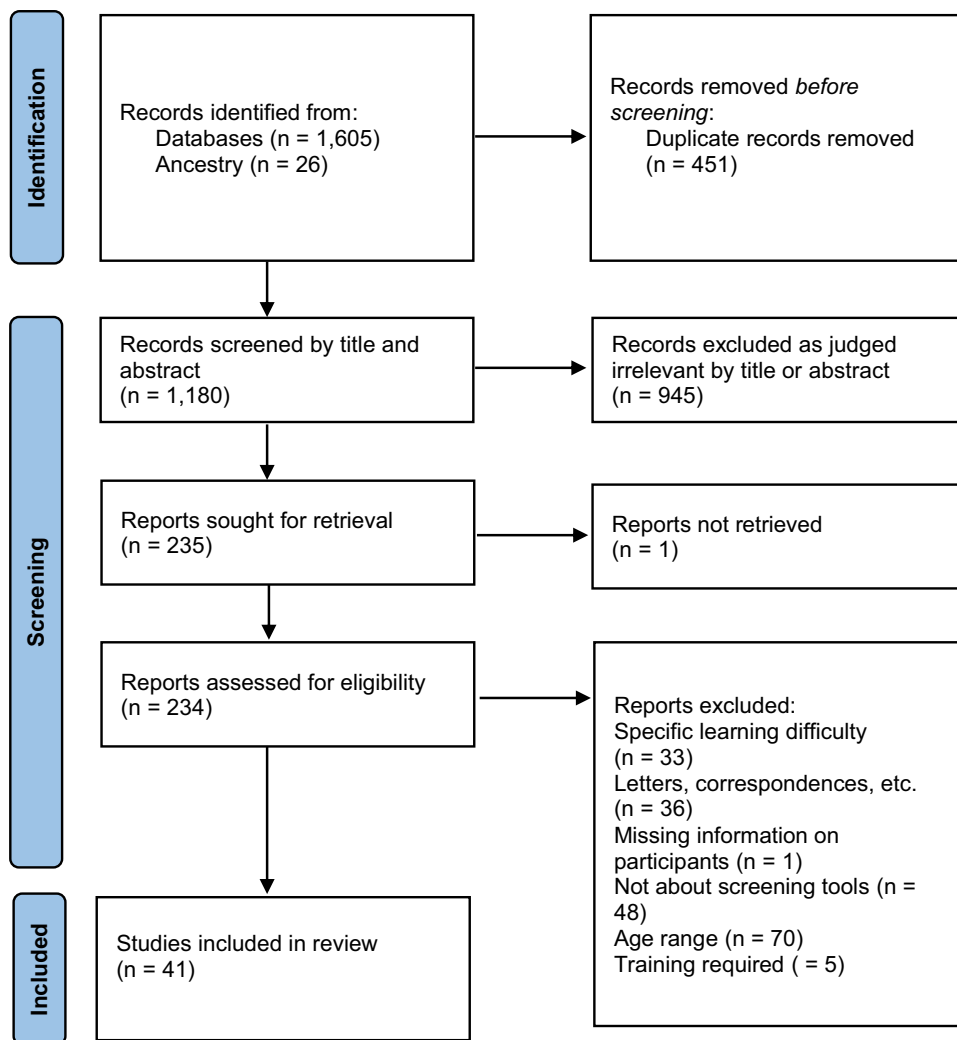
language proficiency. This initial search produced over 1000 potential articles.

After removing duplicates and completing a title and abstract screen against the eligibility criteria, a total of 235 articles were retrieved for full-text screening. This led to the exclusion of a further 194 papers. Studies were excluded due to the ages of participants ($n = 70$) or the fact that the article was not about screening tools ($n = 48$). One of the papers excluded at this stage had no participants in the study (Al Mamun et al., 2016), another had no details of the author and the full text could not be accessed, and thirty-three were about specific learning disorders/difficulties. See Fig. 1 for the PRISMA flowchart of the study selection process (Page et al., 2021). The remaining 41 studies met eligibility criteria. The eligibility criteria were applied independently by two members of the research team (EN & GM) with excellent agreement, $k = 1$.

Quality Appraisal

A quality appraisal of the included studies was conducted using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) risk of bias checklist (Mokkink et al., 2018a; Prinsen et al., 2018; Terwee et al., 2018) and the manual as guides. The COSMIN manual was developed for the systematic review of patient-reported outcome measures (PROMs). Although this review did not consider PROMs, the COSMIN checklist was adopted due to its robustness. The relevant aspects of the COSMIN for this review include its usefulness for assessing the methodological quality of studies, development and design of measurement tools, psychometric properties, and cultural validity. The appraisal was done for all papers by EN and was independently checked by a second member of the team (PL) for 40% of the papers. Following the review of the ratings and correction of errors, the agreement was $k = 1$. Based on the COSMIN guidelines, the quality of included studies was rated (Tables 2 and 3 in the Appendix). For each study, the quality was assessed based on a four-point rating system where each standard within the COSMIN box can be rated as 'very good', 'adequate', 'doubtful' or 'inadequate'. This overall rating of the study quality contributed to grading the quality of the evidence for each tool. The quality of evidence and methods were scored on a four-point rating scale, that is, sufficient, insufficient, indeterminate, or inconsistent. The overall score for quality of evidence is according to the 'lowest score counts' method, and the categories used were high, moderate, low, and very low. Overall ratings for the study methodologies, quality of tool development, and quality of evidence for the measurement properties using the COSMIN checklist are in Tables 4 to 7 (in the Appendix).

One key component of the COSMIN is its usefulness in evaluating cross-cultural validity of tools. Cross-cultural

Fig. 1 PRISMA flow diagram
of the study selection process

Identification

Records identified from:
Databases (n = 1,605)
Ancestry (n = 26)

Records removed *before screening*:
Duplicate records removed
(n = 451)

Screening

Records screened by title and abstract
(n = 1,180)

Records excluded as judged irrelevant by title or abstract
(n = 945)

Reports sought for retrieval
(n = 235)

Reports not retrieved
(n = 1)

Reports assessed for eligibility
(n = 234)

Reports excluded:
Specific learning difficulty
(n = 33)
Letters, correspondences, etc.
(n = 36)
Missing information on participants (n = 1)
Not about screening tools (n = 48)
Age range (n = 70)
Training required (= 5)

Included

Studies included in review
(n = 41)

validity refers to 'the degree to which the performance of the items on a translated or culturally adapted PROM (Patient Reported Outcome Measures) are an adequate reflection of the performance of the items of the original version of the PROM' (Prinsen et al., 2018, p. 1154). Cross-cultural validity is assessed when a tool is used with at least two different groups. Such populations could differ in language, diagnosis, gender, age groups, or ethnicity (Mokkink et al., 2018b).

Additionally, the COSMIN manual suggests areas for adaptation by the review team. Some of the adaptations made for this review were related to the hypothesis testing for responsiveness (criterion validity) and construct validity. In the case of Box 9 which is hypothesis testing for construct validity, we used it to assess the convergent validity and discriminative validity where applicable. Regarding responsiveness, Box 10a for criterion approach, we assessed the diagnostic accuracy of diagnostic tools used in the studies rather than change scores. Outcome measures of specificity and sensitivity were also assessed. For Boxes 10b and c, construct approach are studies which utilised similar

measurement instruments or where the study design was between groups (children, adults, or those with and without ID or ASD sub-groups). Box 10d was not utilised for any studies as we did not look at interventions. Ratings of insufficient, inadequate, or doubtful were given in instances where there was insufficient information reported in the study for a higher rating as required by the COSMIN checklist. For clarification and completeness, manuals, where available, and authors of the tools were consulted for further evidence. This is discussed further in the Result section.

Data Extraction and Synthesis

Relevant information about the aims of each included study, along with the tool used, the design, participants, time to administer the tool, and outcomes were extracted and are reported in Tables 8 and 9 (in the Appendix). The tables were arranged alphabetically by the first author and chronologically when the first author co-authored more than one study. All included studies were quantitative.

Results

Search Results

Forty-one papers met the eligibility criteria (Fig. 1), and 22 of these were about screening tools for autism, while 19 focused upon screening tools for intellectual disabilities. The quality ratings for the included studies are found in Tables 2 and 3 (in the Appendix). Additionally, sensitivity (true positive rate), specificity (true negative rate), positive predictive value or precision (the probability of screening positive and being correct), and negative predictive values (the probability of screening negative and being correct) for the tools were extracted and are reported in Tables 8 and 9 in the Appendix.

Description and Characteristics of Studies

Autism Spectrum Disorder

There were a total of 12,240 participants across the 22 autism studies with age ranging from 15 months to 80 years. Studies with children younger than 11 years or older than 27 years old were only included if most of their participants were within the specified range of the inclusion criteria. Of the 12,240 participants, a little over 9000 involved proxy respondents such parents, teachers, or caregivers of people with autism. Most of the studies were conducted in the UK ($n = 7$) with others spread across various countries, including the US ($n = 5$), Spain ($n = 2$), and one each from Germany, Sweden, Netherlands, Qatar, Australia, Turkey, Singapore, and Argentina (Table 8 in the Appendix). A variety of screening tools were used across the studies, including the Autism Screening Quotient (AQ-10) adolescent and adult versions ($n = 3$), Autism Screening Quotient (AQ-50) ($n = 1$), Autism Spectrum Screening Questionnaire (ASSQ) ($n = 1$), Autism Spectrum Screening Questionnaire-Revised Extended Version (ASSQ-REV) ($n = 1$), Social Communication Questionnaire (SCQ) ($n = 7$), Developmental Behaviour Checklist-Autism Screening Algorithm (DBC-ASA) ($n = 1$), Childhood Autism Rating Scale (CARS) ($n = 1$), Mobile Autism Risk Assessment (MARA) ($n = 1$), Pervasive Developmental Disorder in Mentally Retarded Persons Scale (PDD-MRS) ($n = 2$), EDUTEA ($n = 1$), Child Behaviour Checklist (CBCL) ($n = 1$), Adapted Autism Behaviour Checklist (AABC) ($n = 1$), and Autism Diagnostic Inventory-Telephone Screening in Spanish (ADI-TSS) ($n = 1$).

Fifteen studies were between-group designs, one within-group design, and six single group designs. One study was longitudinal and included data collected over 15 years. Across the included 22 studies, two broad aims were discerned: (a) designing a short screening tool and (b) validating the discriminative ability of tools. Those that focused on designing short tools were further categorised in two ways: (a) adapting existing tools into shorter versions or (b) the development of entirely new tools. Eighteen out of 22 (approximately 82%) of the papers reviewed based their studies on existing tools developed over 10 years ago. The remaining four (18%) considered tools that were developed in the last 2 to 3 years. The existing tools were mainly used with children, while the other studies reviewed focused upon adapting the tools for adolescents and adults.

Autism Screening Tools

The Autism Spectrum Quotient (AQ) The Autism Spectrum Quotient (Baron-Cohen et al., 2001) is a short, easy to use and score, self-administered screener for adults with Asperger syndrome or high-functioning autism. It is comprised of 50 questions divided into five subsets of 10 questions each covering five domains — social skills, attention to detail, attention switching, communication, and imagination. Over time, the AQ was adapted and modified to include adolescents (Baron-Cohen et al., 2006) while maintaining the original 50-item format. The AQ-50 child, AQ-50 adolescent, and AQ-50 adult measures were adapted to create shorter versions by selecting the ten most discriminating items from each and validating the short tool (Allison et al., 2012). The AQ in different variations was used in three different studies (Allison et al., 2012; Baron-Cohen et al., 2006; Booth et al., 2013). The short version of the adolescent tool, the AQ-10 (Allison et al., 2012), had a sensitivity of 0.93, a specificity of 0.95, and a positive predictive value (PPV) of 0.86, while the AQ-50 (Baron-Cohen et al., 2006) tested with adolescents had a sensitivity of 0.89 and specificity of 1. Baron-Cohen et al. (2006) reported no PPV but commented that future research should explore this.

For adult participants (includes participants older than 18 years and/or 16 years of age in some instances) that employed the short AQ-10, Allison et al. (2012) found a sensitivity of 0.88 and specificity of 0.91, while Booth et al. (2013) found a sensitivity of 0.80 with a specificity of 0.87. All three studies (Allison et al., 2012; Baron-Cohen et al., 2006; Booth et al., 2013) included participants with a previous diagnosis of autism.

While all the three studies that employed the AQ defined the constructs to be measured, the quality of evidence was rated as low. Specifically, content validity was rated as low since participant and expert involvement in the studies was unclear. Structural validity, internal consistency, reliability, construct validity, cross-cultural validity, and criterion validity were all examined by Allison et al. (2012). Baron-Cohen

et al. (2006) examined internal consistency and reliability with moderate evidence for cross-cultural validity. Booth et al. (2013) provided evidence for structural validity, while reliability and cross-cultural validity were undetermined. As such, the evidence for reliability was rated as low, and the overall rating for cross-cultural validity was found to be low. To ensure that the psychometric properties of the AQ-10 were accurately captured, the authors were contacted for the manual, who responded that the tests and 'manuals' were those on the authors' website. In summary, although the psychometric results met the criteria for good tools (Glascoe, 2005) following the COSMIN guidelines, where the lowest score counts, the overall quality of evidence for the tool was rated as low.

Autism Spectrum Screening Questionnaire (ASSQ) and the Autism Spectrum Screening Questionnaire-Revised (ASSQ-REV) Preliminary development of the ASSQ took place in Sweden for use within a prevalence study for high-functioning autism and Asperger syndrome in mainstream schools (Ehlers & Gillberg, 1993). The ASSQ is a 27-item checklist that can be completed by laypersons such as teachers or parents and was developed further in later studies (Ehlers et al., 1999). An extended version of the ASSQ-REV was developed for the early identification of girls with autism (Kopp & Gillberg, 2011). The original Swedish version of the ASSQ has been translated into multiple languages — Mandarin Chinese (Guo et al., 2011), English (Ehlers & Gillberg, 1993), Norwegian (Posserud et al., 2006), Finnish (Mattila et al., 2009), and Lithuanian (Lesinskiene, 2000).

Cederberg et al. (2018) examined the diagnostic accuracy of the ASSQs in adolescents previously diagnosed with high-functioning autism. While participant gender and the psychometric properties of the measure were not reported, the authors reported that the ASSQ appeared sensitive to correctly identifying autism. Kopp and Gillberg (2011) examined the validity and accuracy of individual items for detecting autism in girls and boys aged 6–16 years. Different items showed considerable discriminative ability (AUC > 0.70) (see Kopp & Gillberg, 2011) for those with autism versus typically developing children across genders. Both studies used participants who had a previous diagnosis of autism. Like Cederberg et al. (2018), Kopp and Gillberg (2011) reported no sensitivity, specificity, PPV, or negative predictive value (NPV).

Although the ASSQ was originally in Swedish and has been translated into different languages, cross-cultural validity was rated as low using COSMIN due to insufficient evidence of its effectiveness in different cultures. Criterion validity, construct validity, internal consistency, and reliability were rated as insufficient based on the combined evidence from both studies (Cederberg et al., 2018; Kopp

& Gillberg, 2011). Neither of the two studies examined the content nor structural validity of the ASSQ. To ensure that all relevant evidence and information on the tool's development were examined, efforts were made to access the manual but were unsuccessful. No other studies utilising the ASSQ outside the West were found. The overall quality of the tool was rated as very low.

Social Communication Questionnaire (SCQ) The SCQ formerly known as the Autism Screening Questionnaire (ASQ) (Berument et al., 1999) was initially designed as a companion screening tool for the Autism Diagnostic Interview (ADI) (Snow, 2013). The SCQ is a brief 40-item parent or caregiver report screening measure modelled after the ADI-R and has been used widely in research (Berument et al., 1999; Rutter et al., 2003). The measure has two versions, the lifetime version and the current version, both focusing on symptoms of autism most likely to be observed by the individual's principal caregiver. The caregiver must be familiar with the individual's developmental history and current behaviour. The SCQ is a screening tool and cannot be used for the diagnosis of autism. The measure is used for anyone 4 years old and above. The design allows for the comparison of symptoms across different groups of individuals such as children with language delays and those with medical conditions co-existing with autism. The SCQ is currently available in seventeen languages (Danish, Dutch, English, Finnish, French, German, Greek, Hebrew, Hungarian, Italian, Japanese, Korean, Norwegian, Romanian, Russian, Spanish, and Swedish) and is used widely in research.

Seven studies (Aldosari et al., 2019; Brooks & Benson, 2013; Berument et al., 1999; Charman et al., 2007; Corsello et al., 2007; Mouti et al., 2019; Ung et al., 2016) utilised the SCQ. Five studies (Aldosari et al., 2019; Charman et al., 2007; Corsello et al., 2007; Mouti et al., 2019; Ung et al., 2016) included samples of adolescents, and one included adults with intellectual disabilities (Brooks & Benson, 2013), while one (Berument et al., 1999) was a development study and included children, teenagers, and adults (age range, 4–40 years). Berument and colleagues (1999) recommended an optimal cut-off of 15 for differentiating those with and without autism. Using this cut-off, they reported a sensitivity of 0.85, specificity of 0.75, PPV of 0.93, and NPV of 0.55. In the other studies, the cut-off was varied to generate optimal values, depending on the age of the participants. For instance, Brooks and Benson (2013) using a cut-off of 15 reported that the sensitivity was 0.71, specificity 0.77, PPV 0.58, and NPV 0.86. However, when the cut-off was lowered to 12, the sensitivity was 0.86, specificity 0.60, PPV 0.49, and NPV 0.91. Similarly, Corsello et al. (2007) reported finding sensitivity of 0.71, specificity = .71, PPV = .88, and NPV = .45 at a cut-off of 15 while at a cut-off of 12 sensitivity was 0.82, specificity = .56, PPV = .84,

and NPV = .51. However, as is typical with screening tools, lower cut-off scores will improve sensitivity but at the expense of specificity.

Recently, Mouti et al. (2019) examined the optimal cut-off for differentiating between ASD, attention deficit and hyperactivity disorders (ADHD), and typically developing individuals. Their result showed that at a cut-off of score of 9, the SCQ showed excellent discriminative ability between ASD and non-ASD with a sensitivity of 0.1 and specificity of 0.84. Additionally, Mouti et al. (2019) showed that at the cut-off of 13, ASD was clearly discriminated in individuals who were diagnosed as ASD only (sensitivity = .96, specificity = .87) or a combination of both ASD and ADHD (sensitivity = .87, specificity = .85). In the Arabic validation study, Aldosari et al. (2019) reported sensitivity and specificity of 0.80 and 0.97, respectively, at the recommended cut-off score of 15. However, for a cut-off range between 11 and 15, the sensitivity varied between 0.90 and 0.80, while specificity varied between 0.85 and 0.97. Aldosari et al. (2019) also reported internal consistency of $\alpha = .92$.

Apart from Ung et al. (2016), who validated the SCQ against the Childhood Autism Rating Scale (CARS-2) only, all the other studies validated the SCQ against either the Autism Diagnostic Observation Schedule-2 (ADOS-2) or Autism Diagnostic Interview-Revised (ADI-R) or a combination of the CARS, ADOS-2, and ADI-R. Overall, the psychometric properties of the SCQ met the guidelines (Glascoe, 2005) for good tools, and the SCQ correlated well with the ADI-R (Berument et al., 1999).

Out of the seven studies reviewed, four (Berument et al., 1999; Corsello et al., 2007; Mouti et al., 2019; Aldosari et al., 2019) examined the structural validity with sufficient outcomes reported. Criterion validity and reliability were rated as excellent across all seven studies. All seven studies had clear constructs with five (Charman et al., 2007; Corsello et al., 2007; Ung et al., 2016; Mouti et al., 2019; Aldosari et al., 2019) providing sufficient evidence for the construct validity. There was an excellent outcome on the criterion validity across all seven studies. Five studies (Berument et al., 1999; Corsello et al., 2007; Mouti et al., 2019; Aldosari et al., 2019; Charman et al., 2007) rated positive had sufficient evidence for cross-cultural validity, while the remaining two (Brooks & Benson, 2013; Ung et al., 2016) were rated negative with insufficient evidence. Soto et al. (2015) in their review of culturally adapted tools reported that the Chinese validation study (Gau et al., 2011) of the SCQ had good test–retest reliability (r_{ICC} = .77–0.78) and internal consistency (α = .73–0.91). The authors (Gau et al., 2011) reported excellent concurrent validity ($r \leq 0.65$). Given that the SCQ is available in 17 languages; has been used across countries including Africa (Bozalek, 2013) and across ethnicities, genders, and ages; and widely employed in research, it meets several of the qualities for good

cross-cultural validity as defined by COSMIN. The SCQ was rated overall as medium based on the evidence from the seven studies reviewed (Brooks & Benson, 2013; Berument et al., 1999; Charman et al., 2007; Corsello et al., 2007; Ung et al., 2016; Mouti et al., 2019; Aldosari et al., 2019) and previous work done by McConachie et al. (2015). Given the above results, the SCQ seems an appropriate tool to be considered for use within African nations, especially as very little training is required to score it.

Childhood Autism Rating Scale (CARS) The CARS was developed by Schopler et al. (1980) as a diagnostic tool for children with autism. However, this measure, while meant to be diagnostic, was included because Mesibov et al. (1989) used it as a screening tool with adolescent participants, suggesting the CARS' potential utility as a screening instrument for autism. Nevertheless, Mesibov et al. (1989) did comment that the CARS was meant to be used as a diagnostic tool. The CARS is a 15-item rating scale that assesses behaviours associated with autism. The measure is meant to ease the identification of children with autism for parents, educators, clinicians, and other healthcare providers. The scale is available in English, Brazilian Portuguese, Lebanese, Japanese, Swedish, and French. The second edition now includes a scale for identifying high-functioning autism and a parent information form. Some training is required to administer the tool.

Although the CARS was initially validated for use with children, Mesibov et al. (1989), in their longitudinal study, examined its suitability for use with adolescents and adults with autism. Fifty-nine participants with a previous autism diagnosis were re-assessed, and the results showed that 81% ($n = 48$) retained their diagnosis. In comparison, 19% ($n = 11$) of them received a revised diagnosis of no autism based on a cut-off score of 30. However, moving the score to a cut-off of 27 (to account for the mean difference in scores between the younger and older sample), 92% ($n = 54$) were accurately diagnosed. As a result of the improved diagnostic outcomes, Mesibov and colleagues recommended 27 as the cut-off for persons over the age of 13 years.

Based on COSMIN guidelines, content validity, internal consistency, and construct validity were rated as not determined, since it was unclear from the study whether these were tested. There was insufficient evidence for structural validity and criterion validity. Cross-cultural validity was rated as positive with moderate evidence due to the availability of the measure in different translations. The evidence for reliability was moderate; however, this was based on the evidence from the only study found (Mesibov et al., 1989). The authors were contacted for more information on the development of the tool or for access to the relevant portion of the manual, unsuccessfully. A search was done to find other studies that reported the development of the measure

or studies in which the CARS was used. One such study was identified (Schopler et al., 1980) which reported an internal consistency coefficient of $\alpha = .94$ and interrater reliability of 0.71. Two other studies (Breidbord & Croudace, 2013; DiLalla & Rogers, 1994) were also identified: DiLalla and Rogers (1994) presented the results of an exploratory factor analysis of the CARS, while Breidbord and Croudace (2013) examined the interrater reliability and internal consistency from various studies. Based on the results of these studies (Breidbord & Croudace, 2013; DiLalla & Rogers, 1994; Schopler et al., 1980) and evidence from McConachie et al. (2015), internal consistency, structural validity, and reliability were rated as moderate. The overall rating for the measure was medium based on COSMIN guidelines.

Additionally, as per the publisher's guidance, some training and specific educational qualification are required before using the CARS. Thus, it seems inappropriate for further consideration for screening adolescents in Africa.

Developmental Behaviour Checklist-Autism Screening Algorithm (DBC-ASA) The DBC-ASA (Brereton et al., 2002) is a 29-item autism screening measure derived from the Developmental Behaviour Checklist (DBC). The DBC was revised and updated to the DBC2 in 2018. The parent version of the DBC is available in the following languages: Chinese, Arabic, Croatian, Dutch, French, Finnish, German, Greek, Hindi, Norwegian, Portuguese (Brazilian), Italian, Japanese, Spanish, Swedish, Turkish, and Vietnamese.

Deb et al. (2009) screened a total of 109 children aged 3–17 years with intellectual disabilities for autism using the instrument. Forty-four of the children were between 3 and 9 years old, 50 of them between 10 and 15 years old, and 15 participants were older than 15 years. A cut-off score of 19 for the 3–9 years olds yielded a sensitivity of 1 and specificity of 0.71, while a cut-off of 26 for the 10–15-year-olds yielded a sensitivity of 0.70 and specificity of 0.75. When a total population cut-off score of 20 was applied, sensitivity was 0.90 and specificity 0.60. The figures generated by Deb et al. (2009) differ from those obtained in Brereton et al. (2002) where a cut-off score of 14 yielded sensitivity of 0.86 and specificity of 0.55 and a cut-off score of 17 yielded a sensitivity of 0.79 and specificity of 0.63. Perhaps, this could be attributed to the characteristics of the participants as noted by Deb et al. (2009); they screened for autism in children with intellectual disabilities only, while Brereton et al. (2002) examined the validity of the tool amongst individuals with and without intellectual disabilities. Neither study reported a PPV or NPV. There was no validation against an accepted gold standard tool; rather, the participants received a clinical diagnosis of autism based on the ICD-10-DCR (International Classification of Diseases 10th Revision, Research Diagnostic Criteria) in the Deb et al. (2009) study.

Appraising the quality of the reviewed study (Deb et al., 2009), the content validity, structural validity, cross-cultural validity, internal consistency, construct validity, and reliability were all rated as undetermined. Criterion validity was rated as sufficient based on the evidence. As peer-reviewed studies do not always provide sufficient information, the authors of the DBC were contacted to confirm which of the validities were examined. Based on the evidence provided by the authors and excerpts from the manual, reliability, internal consistency, convergent validity, criterion validity, discriminative validity, and concurrent validity were all rated as positive. Since the DBC-ASA is not an independent measure but an algorithm within the DBC, the relevant psychometric (discriminative validity) property of the DBC-ASA was assessed. Brereton et al. (2002) and Deb et al. (2009) both reported that the DBC-ASA had very good discriminative ability. However, there remains inadequate information on cross-cultural validity, placing a limitation on its use in an African context. The overall rating for the tool based on the COSMIN checklist was medium.

Pervasive Developmental Disorder in Mentally Retarded Persons (PDD-MRS) The PDD-MRS is a 12-item questionnaire designed for clinician screening for autism amongst those with intellectual disabilities. It has dichotomous items spread across the following domains: communication, social behaviour, and stereotyped behaviour. It was designed to be used with children and adults ages 2–55 years old. The original Dutch version, the Autisme- en Verwante kontakt-stoornissenschaal voor Zwakzinnigen (AVZ), was developed specifically for use with people with intellectual disabilities (Kraijer, 1990) with a revision in 1994 (Kraijer, 1994). The instrument is based upon the DSM-III-R criteria for pervasive developmental disorders and has been widely used in the Netherlands and Belgium.

Kraijer and de Bildt (2005) described and discussed the construction of the scale and its validation. The psychometric properties were tested on a sample of 1230 participants with varying levels of intellectual disabilities. The resulting sensitivity at a cut-off score range of 10–19 was 0.92, while specificity was 0.92, but neither the PPV nor NPV was reported. Internal consistency for participants with functional speech was reported as $\alpha = .86$ and for those without speech $\alpha = .81$. Cortés et al. (2018) developed and validated the Escala de Valoración del Trastorno del Espectro Autista en Discapacidad Intelectual (EVTEA-DI), the Spanish version of the PDD-MRS. Reported results were $r = .78$ for convergent validity between the EVTEA-DI and the CARS; the internal consistency measured by the Kuder-Richardson-20 (KR-20) was 0.71. At a cut-off score of 30, sensitivity was 0.71, specificity of 0.90, PPV of 0.73, and NPV of 0.90. To assess the discriminative validity of the EVTEA-DI, Cortés

et al. ([2018](#)) utilised the Youden Index (YI). At a cut-off score of 8, sensitivity = .84 and specificity = .83.

For the PDD-MRS, content validity was rated as moderate based on the evidence from reviewed studies. Structural validity, internal consistency, criterion validity, and construct validity were all rated as positive as there was sufficient methodological evidence found to support the rating. There was moderate evidence for cross-cultural validity since individuals with varying disabilities from different populations were participants. Studies were completed with Dutch- and Spanish-speaking participants. Reliability was rated as insufficient based on the COSMIN rating of lowest score counts. Authors were contacted for further evidence without success. The overall COSMIN rating for this tool was medium.

Child Behaviour Checklist (CBCL) The CBCL (Achenbach & Rescorla, [2001](#)) is now a component of the Achenbach System of Empirically Based Assessment (ASEBA). The CBCL is a caregiver report questionnaire on which children and teenagers (2–18 years) are rated for various behavioural and emotional difficulties. Associated with disorders from the DSM-5, it measures difficulties on a scale made up of eight categories — rule-breaking behaviour, anxious/depressed, social problems, somatic complaints, thought problems, attention problems, withdrawn/depressed, and aggressive behaviour. The form consists of 118 items that take between 30 min to an hour to complete. The CBCL has been translated into 60 different languages. Previous versions of the checklist were not designed to screen for autism in young children older than 4 years and 6 years in the current revision (Mazefsky et al., [2011](#)).

However, Ooi et al. ([2011](#)) aimed to derive and test an autism scale that could significantly differentiate children and adolescents with and without autism using the CBCL. The study participants were between 4 and 18 years old. The researchers considered whether eight scale factors could significantly differentiate individuals with and without autism, and they reported a sensitivity range of 48–78% and a specificity range of 59–87%. Following this, Ooi et al. ([2011](#)) derived and tested an autism scale comprised of items taken from the CBCL that significantly differentiated autistic children from other groups. Results showed that nine specific items were predictive of autism with sensitivity ranging from 0.68 to 0.78 and specificity range of 0.73–0.92. The PPV and NPV were not reported. The CBCL scores falling below the 93rd percentile are considered normal, and scores between the 93rd to 97th percentile are borderline clinical, while scores above the 97th percentile are in the clinical range. Results of Ooi et al. ([2011](#)) are consistent with findings from previous studies (Mazefsky et al., [2011](#)). Both Ooi et al. ([2011](#)) and Mazefsky et al. ([2011](#)) reported that the CBCL scales with more effective discriminative abilities

between the typical and autistic school-aged children were the 'thought problems, social problems, and withdrawn/depressed' categories.

Regarding the quality appraisal from the reviewed study Ooi et al. ([2011](#)), the content validity for the CBCL was rated as indeterminate, while structural validity was rated as positive, given the quality of the evidence reviewed. Criterion validity, construct validity, and internal consistency were all rated as undetermined as there was not sufficient evidence. There was moderate evidence for reliability, with sufficient evidence to rate cross-cultural validity as positive. The scale which was originally developed in English was used with participants in three different languages (English, Malay, and Tamil) and five different groups (Ooi et al., [2011](#)). The authors were contacted for more evidence or access to relevant portions of the manual. Based on the author's response, content validity, reliability, criterion validity, construct validity, internal consistency, and discriminative validity were all rated as sufficient. The overall rating for the CBCL was medium, based on the level of evidence using the COSMIN checklist. Although work has gone into translating the tool into different languages and deriving a potential autism-specific screening subscale from the CBCL, some training is required. The level of training depends on how the data are to be used. For LMICs such as Nigeria, Ghana, Kenya, and other African countries, these requirements are potential barriers.

Mobile Autism Risk Assessment (MARA) Duda et al. ([2016](#)) described the MARA, a new 7-item parent or caregiver questionnaire designed to screen for individuals at risk of autism. The MARA was developed based on the analysis of a pool of ADI-R score sheets of individuals with and without autism. An alternating decision tree algorithm was used to generate the questions and responses. The tool is administered and scored electronically, and the reported sensitivity was 0.90 and the specificity was 0.80. Given that the data used for testing the measure were taken from the ADI-R, it should follow that the discriminatory ability and construct validity should be good. The MARA was validated against the Autism Diagnostic Observation Schedule (ADOS), and the PPV was 0.67, and NPV was 0.95. Duda et al. ([2016](#)) reported no specific cut-off scores; however, they referenced Wall et al., ([2012](#)) where they used a categorical variable with two options — autistic or not autistic. Although the MARA looks promising, more large-scale reliability and validity studies with participants of differing developmental abilities are needed.

Based on the reviewed study (Duda et al., [2016](#)), there was adequate evidence to rate structural validity as positive. Internal consistency, reliability, criterion validity, and construct validity were rated negative due to insufficient evidence. Content validity was rated as insufficient as the

involvement of experts and users was unclear. Evidence for cross-cultural validity was insufficient and was rated as very low. The authors were contacted for more information and possible access to the manual if available. Based on feedback from one of the authors, content validity was revised to a positive rating. However, other studies provided were not on the MARA but on detecting ASD through machine learning. Participants in those studies were children younger than 5 years of age, thus not meeting the inclusion criteria for this review. Based on the COSMIN standard, the overall rating for the measure was low. Also, using this tool in Africa could be challenging, given that not everyone has Internet access or personal computers.

EDUTEA: a DSM-5 Teacher Screening Questionnaire for Autism and Social Communication Disorders (EDUTEA)

The EDUTEA was developed in Spain as a brief autism screening tool for use by teachers and school professionals who had limited time (Morales-Hidalgo et al., 2017). The EDUTEA is an 11-item questionnaire based upon DSM-5 diagnostic criteria and was designed to enable teachers to gain information about the social interactions, behaviours, and communication skills of children. The tool was validated against the ADOS-2 and ADI-R and compared to the CBCL, Childhood Autism Spectrum Test (CAST), and Schedule for Affective Disorders and Schizophrenia (K-SADS-PL). Scoring of items is on a 4-point Likert scale, resulting in a minimum score of 0 to a maximum score of 33.

In evaluating the discriminatory ability and psychometric properties of the tool, Morales-Hidalgo et al. (2017) recommended a cut-off score of 10. At the recommended cut-off, the EDUTEA successfully discriminated between those with autism and related disorders and those with ADHD with an associated sensitivity of 0.83 and specificity of 0.73. For differentiating individuals at risk of autism or social pragmatic communication disorder (SCD), the authors reported good discriminatory abilities at the cut-off score of 10, with sensitivity = .87 and specificity = .91NPV of 0.99 and a PPV of 0.87. The two-factor internal consistency for the measure was $\alpha = .95$ for social communication impairments and $\alpha = .93$ for restricted behaviour patterns. Overall internal consistency was $\alpha = .97$. No other studies using the instrument were found from the literature search.

Content validity was rated as positive as teachers were involved in the development of the instrument. The structural validity, internal consistency, criterion validity, reliability, and construct validity were all positive with moderate evidence. However, cross-cultural validity was judged as having insufficient evidence. The overall rating based on COSMIN standards was medium.

Autism Diagnostic Inventory-Telephone Screening in Spanish (ADI-TSS)

Vranic et al., (2002) developed the ADI-TSS as a semi-structured interview administered over the telephone. ADI-TSS was modelled upon the Autism Diagnostic Interview-Revised (ADI-R) with forty-seven questions in three areas. The final version used in the study was administered to 59 participants and had a sensitivity of 1 and a specificity of 0.66 with no PPV or NPV reported. Although this tool was developed over 15 years ago, no other studies validating its use and properties were found. Interrater reliability for the subscales were as follows: social reciprocity $\alpha = .94$, verbal communication $\alpha = .93$, non-verbal communication $\alpha = .94$, and repetitive behaviour $\alpha = .94$.

Content validity for the subscales was rated positive, while the overall content validity was rated low due to insufficient evidence for end-user input in the development of the tool. Structural validity and internal consistency were rated insufficient. Cross-cultural validity was rated insufficient as the translation methodology was unclear. Although interrater reliability for the subscales was shown, there was insufficient evidence for the reliability of the total tool; thus, this was rated insufficient. Based on the COSMIN checklist, the tool was rated as low overall. The feasibility of using the ADI-TSS in Africa, where there are high costs associated with mobile telephone use would be a challenge.

Diagnostic Behavioural Assessment for Autism Spectrum Disorder-Revised (DiBAS-R)

The DiBAS was developed by Sappok and colleagues (2014b) to help with screening autism amongst adults with intellectual disabilities. It was designed to be administered by caregivers or individuals knowledgeable about the person, but who also lacked specific knowledge about autism. The 20-item questionnaire was derived from the ICD-10 and DSM-5 criteria for autism. To improve its diagnostic validity further, a single item was deleted following a pilot study and item-revision of the DiBAS (Sappok et al., 2014a). The resulting 19-item screening tool can be completed in 5 min by a caregiver, family member, staff, or any person who is familiar with the individual.

Heinrich et al. (2018) assessed the diagnostic validity of the DiBAS-R in 381 adolescents and adults with intellectual disabilities, some of who had autism. Study participants ages ranged between 16 and 75 years. Based on the recommended cut-off score of 29, the reported results were sensitivity = .82, specificity = .67, PPV = .44, and NPV = .92. The participant's diagnosis was confirmed using the ADOS and ADI-R.

Based on the reviewed study (Heinrich et al. (2018), content validity was rated as undetermined. Expert clinicians participated in the development, but the item reduction process was unclear. Assessment of comprehensibility and comprehensiveness was also unclear. Evidence for cross-cultural validity, structural validity, and internal consistency were also insufficient. Reliability was rated as insufficient, while criterion validity and construct validity had

sufficient evidence to rate them positive. The authors were contacted for access to the manual or further evidence on the tool's development. Since the manual is in German, the authors provided Sappok et al. (2014a) in which the relevant information was reported. Following this, the content validity, structural validity, internal consistency, reliability, convergent, and discriminative validity were all rated positive. However, evidence for cross-cultural validity remained insufficient. DiBAS-R was rated as medium based on the additional evidence using the COSMIN checklist. DiBAS-R is currently available only in German, thereby limiting the feasibility of using it in Africa.

Adapted Autism Behaviour Checklist (AABC) The AABC which is based on the Autism Behaviour Checklist (Krug et al., 1980) is a 57-item measure developed in Turkey by Özdemir and Diken (2019). Modifications were made to the original form to include the ICD-10 and DSM-5 criteria for autism. The measure was designed to be completed by a parent, primary caregiver, or a teacher familiar with the individual and then scored and interpreted by a trained professional.

Özdemir and Diken (2019) assessed the diagnostic validity of the AABC in 1133 children and adolescents with autism and intellectual disabilities. Study participants ages ranged between 3 and 15 years. Reported results were $r = .73$ between the AABC and the Gilliam Autism Rating Scale-2 Turkish Version (GARS-2 TV), internal consistency measured by the Kuder-Richardson-21 (KR-21) was 0.89, test–retest reliability was $r = .82$, and correlation between the two-factors (social limitations and problematic/repetitive behaviours) was $r = .46$. At a cut-off score of 13, the measure was discriminated between the ASD and ID groups reliably with a sensitivity of 0.87 and specificity of 0.82.

Based on the COSMIN checklist, content validity, structural validity, internal consistency, reliability, criterion validity, and construct validity were all rated as positive. Cross-cultural validity was rated as insufficient based on the evidence. The tool has only been used in Turkey. Since this measure is only available in Turkish, the feasibility of using it in Africa is limited as substantial resources would be required for translation. The overall rating for the measure was medium.

Intellectual Disability

The nineteen studies identified focused upon people with intellectual disabilities and included a total of 3129 participants with age ranging from 3 to 74 years. Like autism, studies with participants younger than 11 years or older than 26 years old were included when some or the majority of their participants were within the specified age range of the inclusion criteria. The number of studies by country was as

follows: the UK ($n = 7$), the USA ($n = 4$), Norway ($n = 5$), and one each from Australia, the Netherlands, and Belgium (Table 9 in the Appendix). Three of the studies (Ford et al., 2008; McKenzie et al., 2012b; Trivedi, 1977) involved adolescents only, while fifteen studies involved a combination of children, adolescents, and adults. The screening tools used in the studies were the Slosson Intelligence Test (SIT), Learning Disability Screening Questionnaire (LDSQ), Child and Adolescent Intellectual Disability Screening Questionnaire (CAIDS-Q), Screener for Intelligence and Learning Disabilities (SCIL), Hayes Ability Screening Index (HASI), and the Quick Test (QT). Validation of these screening tools was against full-length tests considered as the gold standard, such as the different editions or versions of the Weschler scales and the Adaptive Behaviour Assessment System-Second Edition (ABAS II). In some instances, screening tools were compared to other full-length scales, such as the Stanford-Binet Intelligence Scale or to similar short measures. For example, the HASI was compared with the KBIT and the SIT with the Peabody Picture Vocabulary Test (PPVT).

The quality ratings for the included studies are found in Tables 6 and 7 (in the Appendix). Five of the included studies made use of a single group of participants, while eight used a between-group design and six a within-group design. Each screening tool is considered in turn below.

Intellectual Disabilities Screening Tools

Slosson Intelligence Test (SIT) and Slosson Intelligence Test-Revised (SIT-R) The original SIT was developed by Richard Slosson in 1963 and used as part of an assessment to determine whether an individual has an intellectual disability, measured as IQ. At the time of this review, no studies utilising the third and fourth versions of the SIT were found. Rotatori and Epstein (1978) assessed the ability of special education teachers without previous psychological testing experience, to reliably administer the SIT. Reported test–retest reliability results ($r = .94$) appeared excellent, indicating that the test was reliable over time when administered by special education teachers. To examine the concurrent validity of the revised SIT, Kunen et al. (1996) compared the SIT R to the Stanford-Binet Intelligence Scale, Fourth Edition. The correlation was high ($r = .92$), but the consistency of the IQ classification between the two instruments for those who had intellectual disabilities was poor. In comparison to the Stanford-Binet Intelligence Scale, the SIT-R had insufficient evidence of construct validity due to discrepancies in match rates between the SIT and the Stanford-Binet. For instance, for the entire study sample with IQs ranging from 36 to 110 (Kunen et al., 1996), there was a 50% match rate between the Stanford-Binet and the SIT for all the classifications, mild, moderate, average, and low average IQs. Nevertheless, for the mild to moderate categories out of the 38 participants

categorised as mild on the Stanford-Binet, SIT categorised them as 1, low; 2, slow; 9, mild; and 26, moderate. Trivedi (1977) meanwhile examined the comparability of the SIT against the Wechsler Intelligence Scale for Children (WISC) in adolescents. He found significant correlations between the WISC and SIT when compared on mental age ($r = .87$) and IQ ($r = .86$). Trivedi (1977) concluded that the SIT reliably approximated the WISC as a screening tool.

Blackwell and Madere (2005) commented that the SIT-R demonstrated and fulfilled its stated purpose of 'being a valid, reliable, individual screening test of general verbal cognitive ability' (p. 184) but have also suggested problems with the reliability and validity of the SIT-R. Reviews by other authors have also raised concerns about the reliability and validity of the SIT-R (Campbell & Ashmore, 1995). Potential challenges regarding the use of the SIT-R with those from multicultural backgrounds or where English is a second language were reported by Blackwell and Madere (2005). Other limitations of the SIT-R are its inability to measure functioning levels of other intellectual areas such as perceptual motor functioning. There is also the difficulty of comparing SIT scores with those of other IQ tests for persons older than 16 years of age due to the unclear and insufficient methodological information given by the developers (Campbell & Ashmore, 1995). Although the SIT has the above limitations, one advantage is that persons with limited psychometric training and knowledge can administer it.

Based on the COSMIN checklist, the SIT (or SIT-R) was rated as low overall. There was sufficient evidence for reliability from the studies reviewed for it to be rated as moderate. Content validity and structural validity were rated undetermined. Both criterion validity and construct validity were rated as inconsistent. Internal consistency and cross-cultural validity were rated as negative, based on the poor amount of evidence.

Quick Test (QT) The QT is an intelligence test measuring verbal information processing and receptive vocabulary (Ammons & Ammons, 1962). It comprises three parallel forms with 50 items, each of which can be administered to children and adults. Verbal intelligence is measured by the ability to match words of increasing complexity to pictures. Sawyer and Whitten (1972) investigated the concurrent validity of the individual and combined scores of QT against the WISC sub-tests. Moderate correlations ($r = .33–0.52$) were reported for the picture arrangement, coding, performance scale score, and the full-scale score. For the verbal scale, the correlation between both the QT and the WISC was between $r = .31$ and 0.34 for both the individual and combined forms of the QT. One challenge with the QT is that it predominantly measures verbal skills. This limitation may have impacted the Sawyer and Whitten (1972) study, as most of the participants had limited verbal ability. Moreover,

the pictures used are rather old-fashioned and may not transfer well to the African context.

Based on the COSMIN checklist, the overall evidence for the QT was very low. Structural validity, internal consistency, and reliability were rated low based on insufficient amount of evidence both from the study and manual. There was sufficient evidence to rate the construct validity, content validity, and criterion validity as positive, while cross-cultural validity was undetermined. The overall rating for the QT was very low.

Hayes Ability Screening Index (HASI) The HASI is a brief screening tool for intellectual abilities comprised of four subtests covering background information, puzzle, clock drawing, and backward spelling (Hayes, 2000). The HASI has been used predominantly in criminal justice settings to identify vulnerable persons with intellectual disabilities. HASI is designed for use with people aged 13 to adulthood. For those aged 13–18 years, the cut-off score is 90, while for those older than 18 years, it is 85. Some training is required before its use.

Hayes (2002) reported on the construct validity of the HASI and the correlation with the Kaufman Brief Intelligence Test (KBIT), Wechsler Abbreviated Scale of Intelligence (WASI), and WISC-III. The total population sample correlation between the HASI and KBIT was reported as high ($r = .62$). The reported sensitivity for the study was 0.82, and the specificity was 0.72. Hayes (2002) suggested that the youth cut-off be maintained at 90. A different study (Ford et al., 2008) which had all adolescent (10–19-year-olds) participants found a correlation of $r = .55$ between the HASI and the FSIQ of the WISC-IV or the Wechsler Adult Intelligence Scale (WAIS-III) and $r = .38$ with the Vineland Adaptive Behaviour Scale (VABS). At the recommended cut-off score of 90 for those below 18 years of age and 85 for those over 18 years old, the authors reported a poor agreement ($k = .25$) between the HASI and the FSIQ from the Weschler scales when categorising as ID. Sensitivity at these cut-off scores was 0.66 and specificity of 0.51. Lowering the cut-off score to 80.2 yielded better agreement ($k = .54$), a sensitivity of 0.80 and specificity of 0.65. Søndenaa et al. (2007) translated the HASI to Norwegian and validated the construct and criteria of the screening tool against the Norwegian version of the WAIS-III. The study participants were between 17 and 60 years old. The authors found a high correlation between both instruments ($r = .81$) with an internal consistency of $\alpha = .76$. Søndenaa et al. (2007) also reported that scores on all HASI subtests, WAIS-III FSIQ, and the verbal and performance subscales were significantly correlated with r above 0.61. At the recommended cut-off score of 85 for indicating ID, the sensitivity was 1 and specificity 0.57. However, Søndenaa et al. (2007) adjusted the cut-off score to 81 for their sample to reduce the over-inclusion of

false positives. The alternative cut-off of 81 yielded a sensitivity of 0.95 and specificity of 0.72.

In the Søndenaa et al. (2008) prevalence study, the HASI was validated against the WASI as a screening tool. The HASI was found to be somewhat overly inclusive with a specificity of 72.4% and sensitivity of 93.3%. Correlations between the WASI full-scale and HASI were significant with $r = .72$, verbal tests $r = .63$, and performance tests $r = .74$. In Søndenaa et al. (2011), the criterion validity of the HASI was examined against the WASI with a psychiatric population. The study reported the over-categorisation by the HASI with a sensitivity of 1 and specificity of 0.35 at the recommended cut-off score as previously mentioned. However, the authors argued that the HASI is designed to be overly inclusive, since it is better to identify everyone who may need full assessments rather than miss some people. Also, Søndenaa et al. (2011) reported moderate correlations between the subtests of the WASI and HASI ($r = .67$). However, when the 'background information' subtest was eliminated, correlation increased to $r = .71$ and internal consistency of $\alpha = .67$.

To et al. (2015) examined the discriminative and convergent validities of the Dutch version of the HASI against the WASI-III in persons with substance abuse problems. Convergent validity between the HASI and WAIS-III FSIQ scores was significantly correlated ($r = .69$). There was also a correlation between the HASI subtests and the WAIS-III as follows, background information $r = .58$, spelling $r = .50$, puzzle $r = .46$, clock drawing $r = .45$, verbal subscale $r = .70$, and the performance subscale was $r = .63$. Discriminant validity was reported as significant from the receiver operating characteristics (ROC), with an area under the curve (AUC) of 0.95 yielding a sensitivity of 0.91 and specificity of 0.80 at the cut-off score of 85. In Braatveit et al. (2018), the convergent and discriminative validities of the Norwegian version of the HASI were examined in a population of persons with a substance abuse history. At the cut-off of 85, sensitivity was reported as 1 and specificity of 0.65. Braatveit et al. (2018) also reported that lowering the cut-off score to 80.7 yielded increased specificity of 0.81 without affecting the sensitivity. Similar to Søndenaa et al. (2011), Braatveit et al. (2018) also mentioned that the over-categorisation by the HASI was intended to be a means of detecting other persons with/without intellectual disabilities but who may benefit from further evaluation. Regarding convergent validity, Braatveit et al. (2018) correlated the HASI against the full-scale WAIS-IV with a significant correlation ($r = .70$).

Based on the reviewed studies, and the COSMIN checklist, the overall rating for the HASI was low. Reliability was rated as negative due to insufficient evidence. Structural validity had inadequate evidence and was rated as undetermined. Content validity was rated as low due to insufficient evidence. Criterion validity and construct validity were rated positive with excellent evidence. There was moderate evidence for a positive rating on the cross-cultural validity based on the use of the Norwegian and Dutch versions, as well as the original Australian version. To ensure that all relevant properties of the tool were properly rated, the manual was consulted. Based on the manual, additional ratings employing the COSMIN were made. Content validity remained low as there was no evidence on expert clinicians or end users involvement in the development. Evidence for internal consistency was not in the manual, thus a rating of insufficient was given. Reliability was rated as insufficient, while criterion validity and construct validity had sufficient evidence to retain their positive rating. The overall rating of the HASI was revised to medium following the combined evidence from the studies and the manual. Although the HASI has been adapted for use in two further languages and environments outside of the original development area, most of the studies used the tool in the criminal justice system. Studies that employed the tool with adolescents outside of the CJS would have been more useful for forming a decision on adapting it for use in Africa and countries like Nigeria.

Learning Disability Screening Questionnaire (LDSQ) McKenzie and Paxton (2006) developed this 7-item screener for the identification of adults with intellectual disabilities to assist in deciding eligibility for community services. The LDSQ has also been used in criminal justice and forensic settings. Areas assessed include literacy, living situation, and employment. The LDSQ has been reported to have both criterion and convergent validity when compared to the WAIS-III (McKenzie & Paxton, 2006). McKenzie et al. (2012a) examined the convergent and discriminative validities of the LDSQ in forensic settings. Convergent validity between the FSIQ and the LDSQ was reported as highly significant with a correlation coefficient of $r = .71$. The authors also reported good discriminative ability of the LDSQ with a sensitivity of 82.3% and specificity of 87.5% based on the receiver operating characteristics analysis (AUC $= .898$). PPV and NPV were reported as 92.9% and 73.7%, respectively. McKenzie et al. (2014) validated the LDSQ's criteria against a standardised tool, the WAIS-IV FSIQ and reported a good correlation between them with a sensitivity of 0.92 and specificity of 0.92 (AUC $= .945$). Convergent validity was reportedly significant for the WAIS-IV FSIQ and LDSQ total performance with a coefficient of $r = .71$. Significant correlations were also reported for the subtests — verbal comprehension ($r = .54$), perceptual reasoning ($r = .69$), working memory ($r = .58$), and processing speed ($r = .58$). Although these studies by McKenzie et al., (2012a; 2014) reported excellent psychometric properties for the LDSQ, the independent study by Stirk et al. (2018) reported a sensitivity of 0.67 and specificity of 0.71 at the threshold given by McKenzie et al. (2014), showing that the LDSQ may require

more investigation to align the properties with recommended standards (Glascoe, 2005).

Based on the evidence from the studies reviewed, criterion validity and construct validity were rated as moderate. Content validity was rated as insufficient since there was not enough evidence of user participation in the development of the tool. Structural validity, internal consistency, cross-cultural validity, reliability were all rated low due to insufficient evidence. The manual was obtained to confirm which of the tool's properties were examined during development. From the manual, there was moderate evidence for content validity, discriminative validity, and convergent validity. Interrater reliability was assessed, while there was no evidence for internal consistency. Combining the evidence from the studies and the manual, the overall quality of the LDSQ was rated as medium using the COSMIN checklist. Like the HASI, this measure has been used primarily with adults in the CJS and forensic services. However, unlike the HASI, evidence to support the cross-cultural application was not apparent, and so the feasibility of its use with African adolescents is limited.

The Child and Adolescent Intellectual Disability Screening Questionnaire (CAIDS-Q) The CAIDS-Q, modelled after the LDSQ, was developed by McKenzie and Paxton in 2012 as a short 7-item screening questionnaire for detecting intellectual disabilities in children and adolescents in mental health and forensic services. It is designed for use with individuals aged 8–18 years. According to McKenzie and Paxton, the CAIDS-Q can discriminate between those with and without intellectual disabilities with 97% accuracy. Four studies, McKenzie et al. (2014; 2012b, 2012c; 2019) evaluated and validated the psychometric properties of the tool and reported values within recommended standards. McKenzie et al. (2014) assessed the discriminatory ability of the CAIDS-Q against a short form of the WISC-IV with a sample of children aged 10 to 11 years with and without intellectual disabilities (who had been fully assessed for this on either the WISC-IV or the WAIS-III). Overall, the WISC-IV short form itself led to the correct classification of 91% of the participants. When broken down, the classification of those with intellectual disabilities was 92% correct, while those without intellectual disabilities was 91% correct, using the WISC-IV short form. AUC for the WISC-IV was 0.98 which gave a PPV of 0.87 and NPV of 0.95. The CAIDS-Q led to the correct classification of 89% of children with intellectual disabilities and 88% of those without; the PPV was 0.92, and the NPV was 0.85 based on an AUC = .94. Overall, the CAIDS-Q correctly classified 88% of the participants. McKenzie et al. (2012c) evaluated the face, construct, criterion, convergent, and discriminative validity of the CAIDS-Q with comparisons made to either the WISC-IV FSIQ or the WAIS-III FSIQ depending on the participant's age. Results obtained from the study showed high internal consistency ($\alpha = .88$), significant correlations between the CAIDS-Q and the WISC-IV FSIQ ($r = .78$), and significant correlations between the CAIDS-Q and the WAIS-III FSIQ ($r = .79$). At a cut-off of 62 for the children (8–11 years), the measure had a sensitivity of 0.97 and specificity of 0.86, and at a cut-off of 64 for the adolescents (12–18 years), the sensitivity was 0.96 and specificity 0.85. McKenzie et al. (2012c) reported that there was no significant difference between age and the CAIDS-Q score for the total population ($r = .02$). McKenzie et al. (2012b) evaluated the convergent and discriminative validity of the CAIDS-Q against the WISC-IV in a forensic setting. Reported outcomes were significant correlations between the CAIDS-Q and the FSIQ ($r = .76$), with correlations between the CAIDS-Q score and the subtests as follows: verbal comprehension ($r = .54$), perceptual reasoning ($r = .65$), working memory ($r = .52$), and processing speed ($r = .74$). Other results include a PPV of 1, NPV of 1, and good internal consistency ($\alpha = .72$). McKenzie et al. (2019) examined the convergent validity, test–retest reliability, interrater reliability, sensitivity, specificity, PPV, and NPV in a paediatric neurodevelopmental setting based on previously determined cut-off scores. Convergent validity of the CAIDS-Q was examined against the WISC-IV and Adaptive Behaviour Assessment System, Second/Third Edition (ABAS II/III). Reported correlations between the CAIDS-Q and the FSIQ ranged between $r = .62$ and 0.79, with correlations between the CAIDS-Q and the ABAS GAC ranging between $r = .48$ and 0.60. Other results include a PPV of 1, NPV of 0.78, sensitivity = 1, and specificity = .88 for the total sample. A 2-week time frame yielded a test–retest correlation of $r_{32} = .90$, while interrater reliability k was between 0.26 and 1 for the four items (time, read, write, and laces) tested.

From the studies reviewed, content validity was undetermined as user participation in the development was unclear. Assessment of comprehensibility was also unclear. Structural validity was rated as negative due to insufficient evidence. Evidence for reliability was moderate from the studies. Cross-cultural validity was rated as moderate since the measure was used with two different age groups: children and adolescents. Criterion validity was also rated as moderate. However, incorporating information from the manual, content validity, structural validity, internal consistency, reliability, criterion validity, and construct validity were all rated as positive with moderate evidence. Based on the COSMIN checklist, the overall rating of the tool was medium.

Screener for Intelligence and Learning Disabilities (SCIL) The SCIL is a tool for identifying persons with a level of general intellectual functioning that falls within and below the 'borderline' range (Nijman et al., 2018). The SCIL comprises elements of social adaptive skills, language comprehension,

education, arithmetic, reading, and writing abilities. Geijsen et al. (2018) examined the predictive validity of the SCIL for identifying intellectual disabilities amongst adolescents and adults in a criminal justice setting (police detention). Reported results from the study showed that the SCIL total score correlated moderately with the WAIS-III short form ($r = .56$) with a sensitivity of 0.72 and specificity of 0.70. The PPV and NPV were not reported. Additionally, reliability was reported as $\alpha = .64$ and $\alpha = .84$ in a previous study (Kaal et al., 2015). Nijman et al. (2018) conducted two further studies, split into adults and adolescents, and investigated the predictive validity of the SCIL. Participants in both groups included persons with and without intellectual disabilities. At the suggested cut-off of 19, sensitivity was 0.82 and specificity was 0.89 for adults, and for adolescents, the suggested cut-offs varied according to age. For those aged between 16 and 17 years old, a cut-off score of 18 resulted in a sensitivity that was 0.80 and specificity that was 0.84; for those aged 14 to 15-years old, a cut-off score of 16 resulted in a sensitivity of 0.85 and specificity of 0.82. The AUC for adolescents as a total group was 0.91 and 0.93 for adults. The SCIL had test–retest reliability of $r = .92$. Nijman et al. (2018) analysed the internal consistency using the split-half method which yielded high correlations; $\alpha = .84$ in the first half, and $\alpha = .82$ in the second half.

Based on the COSMIN checklist, the overall rating for SCIL was moderate. Content validity was rated positive with moderate evidence based on the involvement of experts and end-user. Structural validity, criterion validity, and internal consistency were all rated as highly positive with enough evidence. Reliability had moderate evidence with consistent findings in both studies. Cross-cultural validity was positive as participants were recruited from different cities, police stations, and refugee sites. The SCIL showed promising results, but more studies to validate the tool are required.

Discussion

Identification and selection of a user friendly, accessible, time-efficient, and useful screening tool for use in Africa require careful thought and consideration. The focus of this review was on identifying potentially useful tools for screening African adolescents and younger adults with autism and/or intellectual disabilities. This age range was the focus as many who have intellectual disabilities and/or autism are noticed as they become more independent and when they begin to interact more often with others outside of their immediate family, for example, in secondary school settings and the wider community. In interacting with these environments, disabilities and challenges become more obvious. Such adolescents may not have received a diagnosis earlier

in their life because of a lack of awareness, insufficient or inadequate resources, limited numbers of professionals, and the families sometimes not seeking immediate help for those individuals with autism or intellectual disabilities till later in life. To begin to address this gap, appropriate and suitable screening tools need to be designed or identified for use in Africa. Towards this, the aims of this review were to (1) describe and critically appraise a range of short screening tools for the detection of intellectual disabilities and autism, (2) consider the psychometric properties of these tools, and (3) consider the appropriateness of using these tools across a range of cultures. A discussion of the review findings is presented below.

Description and Appraisal of Short Screening Tools

A total of 12 tools screening tools for autism were identified through this review. The tools are the ADI-TSS, EDUTEA, PDD-MRS, DiBAS-R, AQ-10, ASSQ-REV, SCQ, CARS, CBCL, DBC-ASA, AABC, and MARA. Apart from the AQ-10 adolescent version, all the other tools were designed to be used across a wide age range. The CARS and CBCL were not originally developed as screening tools, but the studies (Mesibov et al., 1989; Ooi et al., 2011) reviewed utilised them as such with the intent of developing subscales for autism screening. Moreover, the CBCL has over 100 items and takes between 30 min to an hour to complete, which goes against the timing for brief tools. Both the CARS and CBCL require some training and specific qualifications before use. Given the socioeconomic climate of African countries, the resources required to gain specific administrative qualifications for these tools may not be readily and widely available. As such, there will be challenges associated with routine use within Africa.

Both the PDD-MRS and DiBAS-R can be used across a wide age range from 2- to 80-year-olds and administered in 5 to 20 min. The wide age range allows for their use with adolescents, while the short administration time qualifies them as short and time-efficient tools. The PDD-MRS and DiBAS-R were designed for use with persons who are known to have intellectual disabilities. Limiting the measures to those with known intellectual disabilities presumes those individuals have been diagnosed; this is not entirely the situation in Africa. Considering this design limitation, the feasibility of their use in Africa will be challenging.

The MARA, ADI-TSS, and SCQ were modelled after the ADI-R. While the SCQ and MARA take between 5 and 10 min to administer, the ADI-TSS takes between 20 and 40 min. The lengthy administration time of the ADI-TSS may be because of the telephone administration. As an over-the-telephone screening tool, the usefulness of ADI-TSS in Africa, where not everyone may have access to a telephone, poses immediate limitations. Similar constraints

are associated with the MARA, which is a computer-based parent or carer administered screening tool. The number of persons with immediate access to either a smart device, personal computer, or constant electricity is likely to be low in the African continent or individual African countries. This lack of immediate access to smart devices poses a limitation to the usefulness of the MARA in Africa. Meanwhile, for the SCQ, seven (32%) out of the 22 studies reviewed employed this tool, and this observation is consistent with findings from previous studies that concluded the SCQ was used more widely in research (Bozalek, 2013). The SCQ comprises two forms: lifetime and current. The SCQ current form is used to assess an individual's behaviour during the past 3 months, while the lifetime form assesses the developmental history. One advantage of the SCQ is the availability of the lifetime form, which enables information gathering for adolescents who have never been screened. This feature, amongst others, makes the SCQ a viable option for use with African adolescents.

The DBC-ASA, which is a subset of the Developmental Behaviour Checklist, is limited to those under age 18 years, which is about the midpoint of the adolescent age range (11–26 years). The upper age limit of the screening tool poses a current challenge for routine use of the tool. On this basis, adopting the tool for use in Africa does not seem practicable without further standardisation work inclusive of a wider age range.

Both the ASSQ and AQ were developed for persons with HFA. While one of the reasons tools are developed is to bridge a gap or meet a need, in the African setting where screening is still in its infancy, using such disability-specific tools will not yield optimal results. EDUTEA, an 11-item questionnaire, was developed for use by teachers and school professionals. The study by Morales-Hidalgo et al. (2017) did not provide any information on administration time. Similarly, there was no information on administration time provided for the AABC, a 57-item questionnaire developed to be completed by parents, teachers, primary caregivers, or persons familiar with the individual (Özdemir & Diken, 2019). Estimating the administration time based on the 11 or 57 questions introduces subjectivity when compared to the SCQ, which has 40 questions and takes 10 min. The EDUTEA and AABC are emerging tools, and having more comprehensive information would have aided in forming an opinion about their usefulness in Africa.

For intellectual disabilities, a total of 6 tools were identified: the HASI, LDSQ, CAIDS-Q, SIT, SCIL, and QT. Two of these tools (SIT and QT) focus solely on IQ scores to determine the presence of intellectual disabilities. Moreover, the QT is rather outdated and also measures mostly verbal skills, based on old-fashioned pictures which may not be culturally relevant to African settings. For individuals not verbally able, in Africa, the QT will not be very useful. The

original SIT was considered outdated and not on a par with the Wechsler scales (Kunen et al., 1996) and was revised to address some of the concerns. However, the new SIT-R still focuses on verbal cognitive ability. In addition, as previously mentioned, other reviews of the SIT-R have mentioned problems associated with the reliability and validity of the tool. The LDSQ, meanwhile, is adult-specific, and studies that used the LDSQ had participants aged 18 years and above; 18 years is considered the legal adult age in most developed economies. The challenge posed lies in the lower age limit of 18, implying that the LDSQ cannot be used with persons younger than 18 years old.

To close the gap, the CAIDS-Q was developed, by the authors of the LDSQ. CAIDS-Q is used for 8–18-year-olds. For screening adolescents, as defined by age 11–26 years, a more encompassing single measure is required. Two screening tools met this criterion, the HASI and the SCIL. HASI can be used with persons as young as 10 years, as there are two different cut-off scores: one for those below 18 years and another for those above 18 years. Given that HASI requires some training to use it and is also used largely in the CJS and forensic services, two areas that are underrepresented in the African context, these may impact on its usefulness in the African environment. The SCIL was developed and examined with adults (18–63 years) and adolescents (12–17 years). The SCIL also incorporates test items that assess social adaptive skills in line with the current diagnostic criteria for intellectual disabilities, as per the DSM-5, but is currently only available in the Dutch language.

A combined total of 18 screening tools were reviewed for autism and intellectual disabilities. The quality of the tool's design, studies employing them, and overall evidence provided were analysed using the COSMIN Risk of Bias checklist (Tables 2 to 7 in the Appendix). Examples of the areas analysed were the concept elicitation, clearly describing the construct of interest, target population, and context of use. Based on the results of the review using the COSMIN checklist and additional information from manuals, the overall ratings for twelve tools (SCQ, CARS, PDD-MRS, EDUTEA, AABC, DiBAS-R, DBC-ASA, CBCL, LDSQ, CAIDS-Q, HASI, and SCIL) were moderate. For four tools (AQ, MARA, SIT/SIT-R, and ADI-TSS) the rating was low, and the remaining 2 tools (ASSQ and QT) were rated as very low.

Psychometric Properties

For autism, clinical samples with a previous diagnosis participated in most of the studies, leading to a focus upon discriminative validity, differentiating those with and without autism. Using clinical data also meant that a comparison of the outcomes from the screening tools was not necessarily compared to those of an acceptable gold standard

instrument. Regarding sensitivity, specificity, PPV, and NPV, most of the studies reported values for specificity and sensitivity only. Psychometric properties from the studies reviewed were quite varied (Table 8 in the Appendix). The variations could be due to the heterogeneity of the participants across age, gender, severity, or the adjustment in cut-off scores. Other factors that can impact outcomes are study methodology and proxy informants (Ehlers & Gillberg, 1993). Deriving a cut-off score that is associated with precision is part of the development of instruments; however, in some studies, these adjustments resulted in marked variations. This variability was exemplified in studies that utilised the LDSQ. The studies by McKenzie et al., (2012a, 2014) reported sensitivities of 0.82 and 0.92, respectively, while Stirk et al. (2018) reported a sensitivity of 0.67.

Applying the guidelines from the COSMIN checklist, the quality of studies on measurement properties and the evidence for those properties were analysed (details in Tables 2 and 3 in the Appendix). The properties included content validity (this includes relevance of the items in the tool, comprehensiveness, and comprehensibility), structural validity, internal consistency, reliability, criterion validity, and construct validity. In rating the content validity, expert and end-user input are considered. COSMIN ratings are based on the 'lowest score' counts, as previously mentioned, and this formed the basis for the overall rating of the studies and outcomes. Eight tools (PDD-MRS, DiBAS-R, CBCL, DBC-ASA, MARA, AABC, SCQ, and EDUTEA) had moderate evidence for content validity with the remaining four rated low. There was moderate evidence for structural validity for eight tools (SCQ, CARS, CBCL, DBC-ASA, MARA, EDUTEA, DiBAS-R, and PDD-MRS). AABC had high evidence for structural validity, while the remaining 3 had low or very low evidence. Only 82% (18) of the studies examined criterion validity, and these were studies that used the EDUTEA, PDD-MRS, ADI-TSS, DiBAS-R, AQ-10, ASSQ, SCQ, AABC, and DBC-ASA. Out of these, the PDD-MRS, EDUTEA, and SCQ were rated high, while the DiBAS-R, AABC, and ADI-TSS were rated moderate. Evidence from the remaining three tools was inadequate, and they received ratings of low. There was enough evidence to give a rating of moderate to the ADI-TSS, EDUTEA, SCQ, and DiBAS-R for construct validity, while the PDD-MRS received a rating of high. Reliability was high in the EDUTEA and moderate for SCQ, CBCL, and CARS. Internal consistency was found to be high in the EDUTEA and moderate for the SCQ, PDD-MRS, AABC, CARS, DiBAS-R, and AQ-10. Some ratings for the DBC-ASA and the CBCL were based on the manuals not on the studies. When all components of the psychometric properties are considered, the SCQ, CBCL, DBC-ASA, and PDD-MRS met most of the COSMIN criteria.

Turning to consider intellectual disabilities, out of the nineteen studies reviewed, fifteen of them which used the

HASI, CAIDS-Q, LDSQ, and SCIL incorporated the current DSM-5 criteria for intellectual disabilities by using both IQ and adaptive behaviour in the tools. The other four based on the SIT and QT focused on making comparisons with full-scale IQ as a basis for identifying participants with intellectual disabilities. Seventeen of the studies reviewed (89%) validated their results against the age-appropriate Wechsler scales, the most widely used assessment of general intellectual functioning and often regarded as the gold standard. One study (Kunen et al., 1996) compared the SIT to the Stanford-Binet, while Rotatori and Epstein (1978) focused on test–retest reliability.

All studies involving people with intellectual disabilities had evidence of explicit constructs for the development of the tools, and like the autism studies, these studies examined mainly the discriminative and predictive validities of the measures. Criterion validity was examined in all the studies with the HASI, and SCIL rated as high, while those with SIT, CAIDS-Q, LDSQ, and QT were rated moderate. Evidence for construct validity was high for the HASI and CAIDS-Q and moderate for SIT, SCIL, LDSQ, and QT. Internal consistency was high in the SCIL; moderate for CAIDS-Q; low for the HASI, LDSQ, and QT; and very low for the SIT. The quality of evidence for content validity was moderate for the SCIL, CAIDS-Q, LDSQ, and QT, while very low for SIT and low for the HASI.

Regarding reliability, the HASI, QT, and LDSQ were rated low, while the SCIL, SIT, and CAIDS-Q were moderate. Structural validity was rated high for the SCIL, moderate for CAIDS-Q, and low to very low for the remaining four. Sensitivity, specificity, PPV, and NPV values from the studies were also varied but generally within acceptable ranges (Glascoe, 2005). Sensitivity was between 0.67 and 1, while specificity was between 0.35 and 0.92. Based on this review, none of the intellectual disabilities screening tools identified through this review seemed to have been used in Africa. The SCIL and CAIDS-Q were found to have better overall psychometric properties and scored better on the COSMIN checklist (Tables 3 and 7 in the Appendix). Not all studies incorporated adaptive behaviour scores alongside IQ, and overall, in the future, there needs to be more of a shift from IQ testing as a measure of intellectual disabilities, to incorporating adaptive skills during screening and eventual diagnosis by using a tool that captures both.

Cultural Adaptation

A key element for any of the tools selected for use within African nations is cross-cultural validity. Cross-cultural validity based on the COSMIN checklist includes the sample size, agreement between the original and translated versions, use with different populations, diagnoses, and ethnicities. For example, a Spanish version was compared

to an English version or Dutch participants compared to German participants or adults to adolescents.

All the screening tools identified through the review were used for both male and female participants. Comparisons were made between those with and without autism or intellectual disabilities. Concerning the use of different versions of each tool, the AQ in English was used only in the UK, the English ASSQ in the USA, and the Swedish version in Sweden. The DiBAS-R which is in German was used in Germany, the MARA was used in the USA, the ADI-TSS Spanish version was developed and used in Argentina, the PDD-MRS in the Netherlands where it originated as well as the Spanish version was used in Spain, and the EDUTEA in Spain where it was developed was used. Four different versions of the CBCL (English version completed by 60% of the participants, the Chinese version 30%, Malay 8%, Tamil 2%) were used by Ooi et al. (2011) and CARS in the USA and Spain, while the DBC-ASA was used only in the USA. The SCQ was used in the UK, Qatar, Australia, and the USA. The AABC was used only in Turkey. Out of the 12 screening tools for autism, the SCQ was used across a wider age range and across more disabilities and comorbidities (Ung et al., 2016). The validity of the SCQ has also been examined in a small sample of children aged between 2.5 and 14 years in a South African study (Bozalek (2013). When all assessment criteria for cross-cultural validity were examined, the overall rating for the autism tools was as follows: very low for the ADI-TSS, AQ, and MARA; low for the ASSQ, DBC-ASA, AABC, and DiBAS-R; and medium for EDU-TEA, PDD-MRS, CBCL, CARS, and SCQ.

Out of the 19 studies reviewed for the intellectual disabilities screening tools, five studies used between groups designs, including samples of people without intellectual disabilities, six were within-subject designs, while the remaining eight were cross-sectional designs. Utilising the tool with different groups is a criterion for cross-cultural validity in the COSMIN, so studies that have not demonstrated this adequately were rated low in that area. HASI was used in two within-subject studies and five cross-sectional studies. SIT was used in three within-subject studies, while the QT was used in one within-subject study. The LDSQ was employed in two between-subject studies, the CAIDS-Q in 3 between-subject, and one cross-sectional studies, while the SCIL was used in two cross-sectional studies. The HASI was used in 4 different countries and languages: Norway, the UK, Australia, and Belgium. LDSQ was used in the UK and Scotland, while the CAIDS-Q and QT were used in the UK and the USA, respectively. The SCIL was used in Norway and the Netherlands. Putting together all the criteria for evaluating cross-cultural validity, the overall rating for the tools was moderate for CAIDS-Q and HASI, high for SCIL, low for the LDSQ, and very low for both the SIT and QT.

Finally, given that one of the aims of this review was to consider the appropriateness of using these tools across a range of cultures, it is important to note that there are diverse cultures in Africa. These include a variety of spoken languages, beliefs, and behaviours; therefore, whichever tools are identified through this review will require additional contextual adaptation and may perhaps benefit from further ethnological research.

Limitations

There are limitations to this review. By limiting the search to studies in English only, it is possible that some studies with adolescents, and potentially other tools, may have been missed. This in turn may limit the generalisability of the findings of this review, as there are some African countries whose official languages are not English. Manuals for some of the screening tools (seven in total) identified were not readily accessible. This meant that some information on validation reported in the studies could not be compared. Additionally, some of the administration and training requirements could not be examined in detail.

Conclusion

There are two main challenges. The first relates to cultural adaptations and use of the tools outside of the development environment. Whichever tool is identified for use in Africa, it must be sensitive to local differences and expression. The language of the tool must be simple enough to understand, allowing for ease with translation or substitution where required. Validation of selected tools will require time, expertise, and financial resources as determining the psychometric properties in a nouvelle environment requires capacity. As such, the less complex the tool is the easier it may be to assemble the required resources. These challenges are not to suggest the screening tools developed in the West are irrelevant to Africa or LMICs but that careful research and translational work may need to be done to ensure that a tool can be used successfully with people from other countries and cultures. A second challenge is that the tools selected for use with the adolescent and young adult population need to apply to a wide age range while remaining flexible and sensitive. Finally, the limited number of studies involving adolescents identified through this review has presented challenges (as in Hirota et al., 2018). Without a large body of knowledge about adolescents and continental Africa, particularly, the choice of tools is limited.

Developing and validating a continent-specific or country-specific tool for screening autism or intellectual disabilities will take considerable time, effort, and resources. Such resources as time, training, and personnel may not be readily

available. Given the socio-economic and political climate of most African countries, the process could place a considerable financial burden on the economies. In summary, of the 18 tools (6 for intellectual disabilities and 12 for autism) identified through the review, except for the SCQ, none had been utilised in Africa. The SCQ was designed to be used with a wide age range, 4 years and above, and has two versions (current and lifetime) which makes it a good fit for use with adolescents. The SCIL, meanwhile, was validated for adolescents and adults and includes test questions for intellectual abilities as well as social adaptive skills. The broad age range and inclusion of DSM-5 items place it above the other tools reviewed. Additionally, any tool that requires training and more than 20 min of administration time will add to the burden. Thus, to begin the process of validating the screening tools for autism and intellectual disabilities in African adolescents, two tools seem particularly appropriate from the review. These are the SCQ for autism and the SCIL for intellectual disabilities.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40489-022-00342-6.

Declarations

Conflict of Interest The authors declare no competing interests.

References

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms & profiles*. University of Vermont. Research Center for Children, Youth, & Families.

Al Mamun, K. A., Bardhan, S., Ullah, M. A., Anagnostou, E., Brian, J., Akhter, S., & Rabbani, M. G. (2016). Smart autism - A mobile, interactive and integrated framework for screening and confirmation of autism. Conference Proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. *Annual Conference, 2016*, 5989–5992.

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). Toward brief "red flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(2), 202–212.

Ammons, R. B., & Ammons, C. H. (1962). The quick test (QT): Provisional manual. *Psychological Reports, 11*(1), 111–161.

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The Autism Spectrum Quotient (AQ)-Adolescent version. *Journal of Autism & Developmental Disorders, 36*(3), 343–350.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The Autism Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism & Developmental Disorders, 31*(1), 5–17.

Barton, M. L., Dumont-Mathieu, T., & Fein, D. (2012). Screening young children for autism spectrum disorders in primary practice. *Journal of Autism and Developmental Disorders, 42*(6), 1165–1174.

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism Screening Questionnaire: Diagnostic validity. *British Journal of Psychiatry, 175*(5), 444–451.

Blackwell, T. L., & Madere, L. N. (2005). Slosson Intelligence Test-Revised. *Rehabilitation Counseling Bulletin, 48*(3), 183.

Booth, T., Murray, A. L., McKenzie, K., Kuenssberg, R., O'Donnell, M., & Burnett, H. (2013). Brief report: An evaluation of the AQ-10 as a brief screening instrument for ASD in adults. *Journal of Autism and Developmental Disorders, 43*(12), 2997–3000.

Bozalek, F. (2013). Autism screening in children: using the Social Communication Questionnaire in a Western Cape population. (Thesis). University of Cape Town, Faculty of Humanities, Department of Psychology.

Braatveit, K. J., Torsheim, T., & Hove, O. (2018). Screening for intellectual disabilities: A validation of the Hayes Ability Screening Index for in-patients with substance use disorder. *Nordic Journal of Psychiatry, 72*(5), 387–392.

Breidbord, J., & Croudace, T. J. (2013). Reliability generalization for Childhood Autism Rating Scale. *Journal of Autism and Developmental Disorders, 43*(12), 2855–2865.

Brereton, A. V., Tonge, B. J., Mackinnon, A. J., & Einfeld, S. L. (2002). Screening young people for autism with the Developmental Behavior Checklist. *Journal of the American Academy of Child & Adolescent Psychiatry, 41*(11), 1369–1375.

Brooks, W. T., & Benson, B. A. (2013). The validity of the Social Communication Questionnaire in adults with intellectual disability. *Research in Autism Spectrum Disorders, 7*(2), 247–255.

Campbell, C. A., & Ashmore, R. J. (1995). Test review: The Slosson Intelligence Test-Revised (SIT-R). *Measurement and Evaluation in Counseling and Development, 28*(2), 116–118.

Cederberg, C. D., Gann, L. C., Foley-Nicpon, M., & Sussman, Z. (2018). ASD screening measures for high-ability youth with ASD: Examining the ASSQ and SRS. *Gifted Child Quarterly, 62*(2), 220–229.

Charman, T., Baird, G., Simonoff, E., Loucas, T., Chandler, S., Meldrum, D., & Pickles, A. (2007). Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *British Journal of Psychiatry, 191*(6).

Cochrane, A., & Holland, W. (1971). Validation of screening procedures. *British Medical Bulletin, 27*(1), 3–8.

Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Jr., Leventhal, B. L., & Lord, C. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry, 48*(9), 932–940.

Cortés, M. J., Orejuela, C., Castellví, G., Folch, A., Rovira, L., Salvador-Carulla, L., Irazábal, M., Muñoz, S., Haro, J. M., Vilella, E. & Martínez-Leal, R. (2018). Psychometric properties of Spanish adaptation of the PDD-MRS scale in adults with intellectual developmental disorders: The EVTEA-DI scale. *Journal of Autism & Developmental Disorders, 48*(5), 1566–1578.

Deb, S., Dhaliwal, A. J., & Roy, M. (2009). The usefulness of the DBC-ASA as a screening instrument for autism in children with intellectual disabilities: A pilot study. *Journal of Applied Research in Intellectual Disabilities, 22*(5), 498–501.

DiLalla, D. L., & Rogers, S. J. (1994). Domains of the Childhood Autism Rating Scale: Relevance for diagnosis and treatment. *Journal of Autism and Developmental Disorders, 24*(2), 115–128.

Duda, M., Daniels, J., & Wall, D. (2016). Clinical evaluation of a novel and Mobile Autism Risk Assessment. *Journal of Autism & Developmental Disorders, 46*(6), 1953–1961.

Durkin, M. S. (2001). Measurement of childhood disabilities in population studies. In *United Nations Seminar on measurement of disability* (pp. 4–6).

Ehlers, S., & Gillberg, C. (1993). The epidemiology of Asperger syndrome: A total population study. *Child Psychology & Psychiatry & Allied Disciplines, 34*(8), 1327–1350.

Ehlers, S., Gillberg, C., & Wing, L. (1999). A screening questionnaire for Asperger syndrome and other high-functioning autism spectrum disorders in school age children. *Journal of Autism and Developmental Disorders, 29*(2), 129–141.

Eldevik, S., Hastings, R. P., Hughes, J. C., Jahr, E., Eikeseth, S., & Cross, S. (2009). Meta-analysis of early intensive behavioral intervention for children with autism. *Journal of Clinical Child & Adolescent Psychology, 38*(3), 439–450.

Emerson, E. (2012). Deprivation, ethnicity and the prevalence of intellectual and developmental disabilities. *Journal of Epidemiology and Community Health, 66*(3), 218–224.

Ford, G., Andrews, R., Booth, A., Dibdin, J., Hardingham, S., & Kelly, T. P. (2008). Screening for learning disability in an adolescent forensic population. *The Journal of Forensic Psychiatry & Psychology, 19*(3), 371–381.

Franz, L., Chambers, N., von Isenburg, M., & de Vries, P. J. (2017). Autism spectrum disorder in sub-Saharan Africa: A comprehensive scoping review. *Autism Research, 10*(5), 723–749.

Gau, S. S., Lee, C., Lai, M., Chiu, Y., Huang, Y., Kao, J., & Wu, Y. (2011). Psychometric properties of the Chinese version of the Social Communication Questionnaire. *Research in Autism Spectrum Disorders, 5*(2), 809–818.

Geijsen, K., Kop, N., & de Ruiter, C. (2018). Screening for intellectual disability in Dutch police suspects. *Journal of Investigative Psychology and Offender Profiling, 15*(2), 200–214.

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.

Gladstone, M., Mallewa, M., Jalloh, A. A., Voskuijl, W., Postels, D., Groce, N., Kerac, M., & Molyneux, E. (2014). Assessment of neurodisability and malnutrition in children in Africa. *Seminars in Pediatric Neurology, 21*(1), 50–57.

Glascoe, F. P. (2005). Screening for developmental and behavioral problems. *Mental Retardation and Developmental Disabilities Research Reviews, 11*(3), 173–179.

Guevara, J. P., Gerdes, M., Localio, R., Huang, Y. V., Pinto-Martin, J., Minkovitz, C. S., & Pati, S. (2013). Effectiveness of developmental screening in an urban setting. *Pediatrics, 131*(1), 30.

Guo, Y., Tang, Y., Rice, C., Lee, L., Wang, Y., & Cubells, J. F. (2011). Validation of the Autism Spectrum Screening Questionnaire, Mandarin Chinese version (CH-ASSQ) in Beijing. *China. Autism, 15*(6), 713–727.

Gura, G. F., Champagne, M. T., & Blood-Siegfried, J. (2011). Autism spectrum disorder screening in primary care. *Journal of Developmental & Behavioral Pediatrics, 32*(1), 48–51.

Hasegawa, J., Ito, Y. M., & Yamauchi, T. (2017). Development of a screening tool to predict malnutrition among children under two years old in Zambia. *Global Health Action, 10*(1), 1339981.

Hayes, S. C. (2000). *Hayes ability screening index: HASI: Manual.* University of Sydney.

Hayes, S. C. (2002). Early intervention or early incarceration? Using a screening test for intellectual disability in the criminal justice system. *Journal of Applied Research in Intellectual Disabilities, 15*(2), 120–128.

Heinrich, M., Böhm, J., & Sappok, T. (2018). Diagnosing autism in adults with intellectual disability: Validation of the DiBAS-R in an independent sample. *Journal of Autism & Developmental Disorders, 48*(2), 341–350.

Hirota, T., So, R., Kim, Y. S., Leventhal, B., & Epstein, R. A. (2018). A systematic review of screening tools in non-young children and adults for autism spectrum disorder. *Research in Developmental Disabilities, 80*, 1–12.

Jinabhai, C. C., Taylor, M., Rangongo, M. F., Mkhize, N. J., Anderson, S., Pillay, B. J., & Sullivan, K. R. (2004). Investigating the mental abilities of rural Zulu primary school children in South Africa. *Ethnicity & Health, 9*(1), 17–36.

Kaal, H. L., Nijman, H. L. I., & Moonen, X. M. H. (2015). Identifying offenders with an intellectual disability in detention in The Netherlands. *Journal of Intellectual Disabilities and Offending Behaviour, 6*(2), 94–101.

Kakooza-Mwesige, A., Ssebyala, K., Karamagi, C., Kiguli, S., Smith, K., Anderson, M. C., & Smith, D. (2014). Adaptation of the "Ten Questions" to screen for autism and other neurodevelopmental disorders in Uganda. *Autism, 18*(4), 447–457.

Knox, J., Arpadi, S. M., Kauchali, S., Craib, M., Kvalsvig, J. D., Taylor, M., Bah, F., Mellins, C., & Davidson, L. L. (2018). Screening for developmental disabilities in HIV positive and HIV negative children in South Africa: Results from the Asenze Study. *PLoS ONE, 13*(7), e0199860.

Kopp, S., & Gillberg, C. (2011). The Autism Spectrum Screening Questionnaire (ASSQ)-Revised Extended Version (ASSQ-REV): An instrument for better capturing the autism phenotype in girls? A preliminary study involving 191 clinical cases and community controls. *Research in Developmental Disabilities, 32*(6), 2875–2888.

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

Kraijer, D. W. (1990). AVZ, Autisme-en Verwante kontaktstoornisschaal voor Zwakzinnigen. Handleiding (Autism and related contact disorders scale for the mentally retarded. Manual). Swets and Zeitlinger, Lisse, The Netherlands.

Kraijer, D. W. (1994). AVZ-R, Autisme-en Verwante stoornissenschaal voor Zwakzinnigen-Revisie. Handleiding. Sterk herziene en uitgebreide uitgave (Autism and related disorders scale for the mentally retarded-Revision. Manual. Swets and Zeitlinger Lisse, The Netherlands.

Kraijer, D., & de Bildt, A. (2005). The PDD-MRS: An instrument for identification of autism spectrum disorders in persons with mental retardation. *Journal of Autism and Developmental Disorders, 35*(4), 499–513.

Krug, D. A., Arick, J., & Almond, P. (1980). Behavior checklist for identifying severely handicapped individuals with high levels of autistic behavior. *Journal of Child Psychology and Psychiatry, 21*(3), 221–229.

Kunen, S., Overstreet, S., & Salles, C. (1996). Concurrent validity study of the Slosson Intelligence Test-Revised in mental retardation testing. *Mental Retardation, 34*(6), 380–386.

Lesinskiene, S. (2000). Vilniaus miesto vaiku autizmas, Vilniaus Universitetas, Daktaro disetacijos santrauka, Biomedicinos mokslai, medicina 07B, *Psichiatrija* B650. Vilnius, Lithuania.

Limbos, M. M., & Joyce, D. P. (2011). Comparison of the ASQ and PEDS in screening for developmental delay in children presenting for primary care. *Journal of Developmental & Behavioral Pediatrics, 32*(7), 499–511.

Luckasson, R., & Schalock, R. L. (2013). What's at stake in the lives of people with intellectual disability? Part II: Recommendations for naming, defining, diagnosing, classifying, and planning supports. *Intellectual and Developmental Disabilities, 51*(2), 94–101.

Mattila, M.-L., Jussila, K., Kuusikko, S., Kielinen, M., Linna, S.-L., Ebeling, H., et al. (2009). When does the Autism Spectrum Screening Questionnaire (ASSQ) predict autism spectrum disorders in primary school-aged children? *European Child and Adolescent Psychiatry, 18*(8), 499–509.

Maxim, L. D., Niebo, R., & Utell, M. J. (2014). Screening tests: A review with examples. *Inhalation Toxicology, 26*(13), 811–828.

Mazefsky, C. A., Anderson, R., Conner, C. M., & Minshew, N. (2011). Child Behavior Checklist scores for school-aged children with autism: Preliminary evidence of patterns suggesting the need for referral. *Journal of Psychopathology and Behavioral Assessment, 33*(1), 31–37.

McConachie, H., Parr, J. R., Glod, M., Hanratty, J., Livingstone, N., Oono, I. P., et al. (2015). Systematic review of tools to measure outcomes for young children with autism spectrum disorder. *Health Technology Assessment, 19*(41), 1–506.

McKenzie, K., & Paxton, D. (2006). Promoting access to services: The development of a new screening tool. *Learning Disability Practice, 9*(6), 17–21.

McKenzie, K., & Paxton, D. (2012). *Child and Adolescent Intellectual Disability Screening Questionnaire*. GCM Records, Edinburgh

McKenzie, K., Michie, A., Murray, A., & Hales, C. (2012). Screening for offenders with an intellectual disability: The validity of the Learning Disability Screening Questionnaire. *Research in Developmental Disabilities, 33*(3), 791–795.

McKenzie, K., Murray, A. L., Murray, K. R., & Murray, G. C. (2014). Assessing the accuracy of the WISC-IV seven-subtest short form and the child and adolescent intellectual disability screening questionnaire in identifying intellectual disability in children. *Child Neuropsychology, 20*(3), 372–377.

McKenzie, K., Murray, G., Murray, A., Delahunty, L., Hutton, L., Murray, K., & O'Hare, A. (2019). Child and Adolescent Intellectual Disability Screening Questionnaire to identify children with intellectual disability. *Developmental Medicine & Child Neurology, 61*(4), 444–450.

McKenzie, K., Paxton, D., Michie, A., Murray, G., Murray, A., & Curtis, J. (2012). Screening with young offenders with an intellectual disability. *Journal of Forensic Psychiatry & Psychology, 23*(5–6), 676–688.

McKenzie, K., Paxton, D., Murray, G., Milanesi, P., & Murray, A. L. (2012). The evaluation of a screening tool for children with an intellectual disability: The Child and Adolescent Intellectual Disability Screening Questionnaire. *Research in Developmental Disabilities, 33*(4), 1068–1075.

Mesibov, G. B., Schopler, E., Schaffer, B., & Michal, N. (1989). Use of the Childhood Autism Rating Scale with autistic adolescents and adults. *Journal of the American Academy of Child and Adolescent Psychiatry, 28*(4), 538–541.

Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018a). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1171–1179.

Mokkink, L. B., Prinsen, C., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C., & Terwee, C. B. (2018b). COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs). *User manual.* https://www.cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018b-1.pdf. Accessed 28 Nov 2018

Morales-Hidalgo, P., Hernández-Martínez, C., Voltas, N., & Canals, J. (2017). EDUTEA: A DSM-5 teacher screening questionnaire for autism spectrum disorder and social pragmatic communication disorder. *International Journal of Clinical Health & Psychology, 17*(3), 269–281.

Mouti, A., Dryer, R., & Kohn, M. (2019). Differentiating autism spectrum disorder from ADHD using the Social Communication Questionnaire. *Journal of Attention Disorders, 23*(8), 828–837.

Mung'ala-Odera, V., Meehan, R., Njuguna, P., Mturi, N., Alcock, K., Carter, J. A., & Newton, C. R. (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.

Murphy, G. H., Gardner, J., & Freeman, M. J. (2017). Screening prisoners for intellectual disabilities in three English prisons. *Journal of Applied Research in Intellectual Disabilities, 30*(1), 198–204.

Nijman, H., Kaal, H., van Scheppingen, L., & Moonen, X. (2018). Development and testing of a Screener for Intelligence and Learning Disabilities (SCIL). *Journal of Applied Research in Intellectual Disabilities, 31*(1), e59–e67.

Özdemir, O., & Diken, I. H. (2019). Reliability and validity studies of the adapted autism behaviour checklist in Turkey. *Journal of Developmental and Physical Disabilities, 31*(3), 359–376.

Ooi, Y. P., Rescorla, L., Ang, R. P., Woo, B., & Fung, D. S. S. (2011). Identification of autism spectrum disorders using the Child Behavior Checklist in Singapore. *Journal of Autism and Developmental Disorders, 41*(9), 1147–1156.

Olusanya, B. O., & Okolo, A. A. (2006). Revisiting the ten questions questionnaire for developing countries. *International Journal of Epidemiology, 35*(4), 1103–1103.

Oshodi, Y. O., Olagunju, A. T., Oyelohunnu, M. A., Campbell, E. A., Umeh, C. S., Aina, O. F., & Adeyemi, J. D. (2016). Autism spectrum disorder in a community-based sample with neurodevelopmental problems in Lagos, Nigeria. *Journal of Public Health in Africa, 7*(2), 559–559.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & et. Al. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ, 372*, n71.

Posserud, M. B., Lundervold, A. J., & Gillberg, C. (2006). Autistic features in a total population of 7–9-year-old children assessed by the ASSQ (Autism Spectrum Screening Questionnaire). *Journal of Child Psychology and Psychiatry, 47*(2), 167–175.

Prinsen, C. A. C., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C. W., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*(5), 1147–1157.

Robins, D. L. (2008). Screening for autism spectrum disorders in primary care settings. *Autism, 12*(5), 537–556.

Rotatori, A. F., & Epstein, M. (1978). The Slosson Intelligence Test as a quick screening test of mental ability with profoundly and severely retarded children. *Psychological Reports, 42*(3), 1117–1118.

Rutter, M., Le Couteur, A., & Lord, C. (2003). Autism Diagnostic Interview-Revised. Los Angeles, CA: *Western Psychological Services, 29*(2003), 30.

Saloojee, G., Phohole, M., Saloojee, H., & IJsselmuiden, C. (2007). Unmet health, welfare and educational needs of disabled children in an impoverished South African peri-urban township. *Child Care, Health and Development, 33*(3), 230–235.

Sappok, T., Gaul, I., Bergmann, T., Dziobek, I., Bölte, S., Diefenbacher, A., & Heinrich, M. (2014). The Diagnostic Behavioral Assessment for Autism Spectrum Disorder—Revised: A

screening instrument for adults with intellectual disability suspected of autism spectrum disorders. *Research in Autism Spectrum Disorders, 8*(4), 362–375.

Sappok, T., Gaul, I., Dziobek, I., Bölte, S., Diefenbacher, A., & Bergmann, T. (2014). Der Diagnostische Beobachtungsbogen für Autismus Spektrumstörungen (DiBAS): Ein Screeninginstrument für Erwachsene mit Intelligenzminderung bei Autismusverdacht. *Psychiatrische Praxis, 41*, 1–7.

Sawyer, S. M., Afifi, R. A., Bearinger, L. H., Blakemore, S., Dick, B., Ezeh, A. C., & Patton, G. C. (2012). Adolescence: A foundation for future health. *The Lancet, 379*(9826), 1630–1640.

Sawyer, S. M., Azzopardi, P. S., Wickremarathne, D., & Patton, G. C. (2018). The age of adolescence. *The Lancet Child & Adolescent Health, 2*(3), 223–228.

Sawyer, R. N., & Whitten, J. R. (1972). Concurrent validity of the quick test. *Psychological Reports, 30*(1), 64–66.

Schalock, R. L., & Luckasson, R. A. (2013). What's at stake in the lives of people with intellectual disability? Part I: The power of naming, defining, diagnosing, classifying, and planning supports. *Intellectual and Developmental Disabilities, 51*(2), 86–93.

Scherzer, A. L., Chhagan, M., Kauchali, S., & Susser, E. (2012). Global perspective on early diagnosis and intervention for children with developmental delays and disabilities. *Developmental Medicine & Child Neurology, 54*(12), 1079–1084.

Schopler, E., Reichler, R. J., DeVellis, R. F., & Daly, K. (1980). Toward objective classification of childhood autism: Childhood Autism Rating Scale (CARS). *Journal of Autism and Developmental Disorders, 10*(1), 91–103.

Snow, A. (2013). Social communication questionnaire. *Encyclopedia of autism spectrum disorders* (pp 2893–2895). Springer.

Søndenaa, E., Bjørgen, T. G., & Nøttestad, J. A. (2007). Validation of the Norwegian version of Hayes Ability Screening Index for mental retardation. *Psychological reports, 101*(3), 1023–1030.

Søndenaa, E., Nygård, Ø., Nøttestad, J. A., & Linaker, O. M. (2011). Validation and adaptation of the Norwegian version of Hayes Ability Screening Index for intellectual difficulties in a psychiatric sample. *Nordic Journal of Psychiatry, 65*(1), 47–51.

Søndenaa, E., Rasmussen, K., Palmstierna, T., & Nøttestad, J. (2008). The prevalence and nature of intellectual disability in Norwegian prisons. *Journal of Intellectual Disability Research, 52*(12), 1129–1137.

Soto, S., Linas, K., Jacobstein, D., Biel, M., Migdal, T., & Anthony, B. J. (2015). A review of cultural adaptations of screening tools for autism spectrum disorders. *Autism, 19*(6), 646–661.

Stein, Z., Durkin, M., & Belmont, L. (1986). "Serious" mental retardation in developing countries: An epidemiologic approach. *Annals of the New York Academy of Sciences, 477*(1), 8–21.

Steiner, A. M., Goldsmith, T. R., Snow, A. V., & Chawarska, K. (2012). Practitioner's guide to assessment of autism spectrum disorders in infants and toddlers. *Journal of Autism and Developmental Disorders, 42*(6), 1183–1196.

Stirk, S., Field, B., & Black, J. (2018). An independent investigation of the utility of the Learning Disability Screening Questionnaire (LDSQ) within a community learning disability team. *Journal of Applied Research in Intellectual Disabilities, 31*(2), e223–e228.

Suhail, K., & Zafar, F. (2008). Prevalence of autism in special education schools of Lahore. *Pakistan Journal of Psychological Research, 23*(3–4), 45–64.

Swinkels, S. H., Dietz, C., van Daalen, E., Kerkhof, I. H., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum in children aged 14 to 15 months I: The development of the Early Screening of Autistic Traits Questionnaire (ESAT). *Journal of Autism and Developmental Disorders, 36*(6), 723–732.

Terwee, C. B., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research, 27*(5), 1159–1170.

To, W. T., Vanheule, S., Vanderplasschen, W., Audenaert, K., & Vandevelde, S. (2015). Screening for intellectual disability in persons with a substance abuse problem: Exploring the validity of the Hayes Ability Screening Index in a Dutch-speaking sample. *Research in Developmental Disabilities, 36*, 498–504.

Tomlinson, M., Yasamy, M. T., Emerson, E., Officer, A., Richler, D., & Saxena, S. (2014). Setting global research priorities for developmental disabilities, including intellectual disabilities and autism. *Journal of Intellectual Disability Research, 58*(12), 1121–1130.

Trivedi, A. (1977). A comparison of three intelligence tests for the assessment of mental retardation. *Journal of Mental Deficiency Research, 21*(4), 289–297.

Ung, D., Johnco, C., McBride, N. M., Howie, F., Scalli, L., & Storch, E. A. (2016). Optimizing the screening of autism spectrum disorders in outpatient clinics: An examination of the social communication questionnaire-lifetime. *Research in Autism Spectrum Disorders, 27*, 21–28.

United Nations. (2014). Country classification: Data sources, country classifications and aggregation methodology. *World Economic Situation and Prospects, 2014*, 143–150.

United Nations. Department of Economic and Social Affairs, & United Nations Conference on Trade and Development. (2021). *World Economic Situation and Prospects 2021*. United Nations Publications.

Van der Linde, J., Swanepoel, D. W., Glascoe, F. P., Louw, E. M., & Vinck, B. (2015). Developmental screening in South Africa: Comparing the national developmental checklist to a standardized tool. *African Health Sciences, 15*(1), 188–196.

Vawda, N., Milburn, N. G., Steyn, R., & Zhang, M. (2017). The development of a screening tool for the early identification of risk for suicidal behavior among students in a developing country. *Psychological Trauma: Theory, Research, Practice, and Policy, 9*(3), 267.

Vrancic, D., Nanclares, V., Soares, D., Kulesz, A., Mordzinski, C., Plebst, C., & Starkstein, S. (2002). Sensitivity and specificity of the Autism Diagnostic Inventory-Telephone Screening in Spanish. *Journal of Autism & Developmental Disorders, 32*(4), 313–320.

Wall, D. P., Dally, R., Luyster, R., Jung, J-Y., DeLuca, T. F. (2012) Use of Artificial Intelligence to Shorten the Behavioral Diagnosis of Autism. *PLoS ONE 7*(8):e43855.

Webb, E., Morey, J., Thompsen, W., Butler, C., Barber, M., & Fraser, W. (2003). Prevalence of autistic spectrum disorder in children attending mainstream schools in a Welsh education authority. *Developmental Medicine & Child Neurology, 45*(6), 377–384.

Westerlund, M., & Sundelin, C. (2000). Screening for developmental language disability in 3-year-old children experiences from a field study in a Swedish municipality. *Child Care, Health and Development Original Articles, 26*(2), 91–110.

Wild, D., Furtado, T., & Angalakuditi, M. (2012). The translation and cultural adaptation of the Child Behavior Checklist for use in Israel (Hebrew), Korea, the US (Spanish), India (Malayalam and Kannada), and Spain. *Psychology Research and Behavior Management, 5*, 51–56.

World Bank. (2020). *World Bank country and lending groups*. Retrieved from https://datahelpdesk.worldbank.org/knowledgebase/articles/906519. Accessed 31 Mar 2022

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.