Northumbria Research Link

Northumbria University
NEWCASTLE

UniversityLibrary

Screening tools for Autism Spectrum Disorder, used with people with an Intellectual Disability: A systematic review

Authors

Dale Metcalfe[1], Karen McKenzie[1], Kristofor McCarty[1], George Murray[2]

[1]Department of Psychology, Northumbria University, Newcastle-upon-Tyne

[2]NHS Lothian, Edinburgh, UK

Corresponding author

Dale Metcalfe, Department of Psychology, Northumbria University, Northumberland Building, Newcastle-upon-Tyne, NE1 8ST, UK.

Email: dale.r.metcalfe@northumbria.ac.uk

Abstract

Background: A diagnosis of autism spectrum disorder (ASD) can be beneficial in ensuring the person receives appropriate support. People with intellectual disability often have undiagnosed co-occurring ASD, due to the specific diagnostic challenges that having intellectual disability can present. Screening tools can be useful to indicate those who are likely to require full diagnostic assessment of ASD.

Method: We conducted a systematic review of the literature. The databases ProQuest, PsycArticles, PubMed, and Web of Science were searched for articles published before July 2019. When duplicates were removed 3068 articles were retained. Articles were removed in stages and were retained if there was a possibility that the content was relevant. In total, 14 articles were reviewed fully.

Results: The articles covered eight ASD screening instruments and were reviewed in respect of the quality of the available reliability and validity data when used with people with intellectual disability.

Conclusion: A few tools have psychometric properties that indicate they have potential to screen for ASD in people with intellectual disability, but overall research with this group is limited, particularly in terms of reliability. The implications for screening and diagnosis of ASD in people with intellectual disability are discussed.

Electronic Supplementary Materials (ESM)

The ESM are hosted on the Open Science Framework and are available at this link:

https://osf.io/tg58m/?view_only=9125261d327548e5abdfab86a2aea70b

Keywords

Autism spectrum disorders; Intellectual disability; Screening; Diagnosis

Highlights

Screening tools can be used to assist with the identification of ASD

People with intellectual disability often have undiagnosed co-occurring ASD

We systematically review ASD screening tools that have been used with people with intellectual disability

The review concludes that there are a lack of screening tools that have been used with this group

Introduction

Autism spectrum disorder (ASD) is a lifelong condition that first exhibits in early childhood (American Psychiatric Association [APA], 2013) and it affects an estimated 1 in 100 people (Allison, Auyeung, & Baron-Cohen, 2012; Brugha et al., 2012). It is diagnosed on the basis of social and communication difficulties, and restrictive and stereotyped behaviours, interests and activities (APA, 2013). People with ASD can have a range of difficulties including understanding emotions (Harms, Martin, & Wallace, 2010), empathising with others (Baron-Cohen & Wheelwright, 2004; Jones, Happé, Gilbert, Burnett, & Viding, 2010) and making eye contact (Dalton et al., 2005). People with ASD may experience social isolation (Humphrey & Lewis, 2008), bullying due to lack of social understanding (Church, Alisanski, & Amanullah, 2000; Humphrey & Lewis, 2008) and low rates of employment (The National Autistic Society, 2016). An early diagnosis is essential for people with ASD and their carers in order for them to receive appropriate support to address many of these issues (Goin & Myers, 2004), however, for many people with intellectual disability, ASD is often overlooked and not diagnosed (Matson & Shoemaker, 2009).

Intellectual disability is a neurodevelopmental disorder that has a childhood onset, with the person experiencing significant difficulties with adaptive and cognitive functioning (APA, 2013). People with intellectual disability form a heterogeneous group with levels of severity from mild to profound. Classification of severity was previously determined by IQ (mild = 50-55 to 70, moderate = 35-40 to 50-55, severe = 20-25 to 35-40, and profound = less than 20 to 25: American Psychiatric Association [APA], 2000). Today severity is classified in terms of adaptive functioning, in order to better inform the support needs of those with an intellectual disability, with severity and requirement for support increasing as daily living skills decrease (American Phychiatic Association, 2013; World Health Organisation [WHO], 2018).

The heterogeneity of the group makes assessment and diagnosis of ASD particularly challenging. In addition, there is a large degree of overlap between the symptoms of ASD and intellectual disability, making them difficult to attribute to one condition or the other. Intellectual disability often overshadows other conditions, meaning that ASD can be missed by a clinician (O'Brien & Pearson, 2004). Another complicating factor is that due to poorer literacy skills (for summary see Poncelas & Murphy, 2007) individuals may lack the skills to self-report symptoms or complete measures that require good literacy skills.

This is concerning as intellectual disability is possibly the most common co-occurring condition with ASD, with estimated prevalence rates increasing over time (Matson & Shoemaker, 2009). Examples of high comorbidity include a study by La Malfa, Lassi, Bertelli, Salvini and Placidi (2004) which concludes that 40% of those with intellectual disability had ASD and 70% of those with ASD have intellectual disability. Additionally when summarising previous research, Buescher, Cidav, Knapp and Mandell (2014) estimated that between 40% and 60% of people with ASD also have intellectual disability.

Screening is a method that helps differentiate between people who are likely and unlikely to have a particular condition, with those who screen positive being recommended for full diagnostic assessment (Glascoe, 2005). The importance of screening tools for ASD has been argued by Allison et al. (2012) as health professionals can refer those who likely need a full assessment to the relevant professionals. Screening can also be useful for research purposes, when there is only a requirement that a particular group is likely to have a particular condition and when there are not available resources to conduct full diagnostic assessment with all participants (McKenzie & Murray, 2015). Screening can be differentiated from assessment in that the former is designed to indicate the likely presence of a particular condition, whereas the latter is a more comprehensive process designed to clarify the nature of the condition and inform the diagnosis and subsequent intervention (Glascoe, 2005; Public Health England, 2019).

It is important that screening tools have good psychometric properties, to ensure that they identify those who do and do not have the condition of interest, as accurately as possible. Sensitivity is the ability of a screening tool to correctly identify those who do not have the condition, while specificity is the ability to correctly identify those who do not (Glascoe, 2005). Incorrectly classifying someone as having the condition could result in the person undergoing unnecessary further assessment, with the associated costs in time, resources, worry and potential stigma. Incorrectly classifying someone as not having the condition may mean they miss out on assessment and support that they would otherwise have benefitted from.

It can be difficult for clinicians to identify the most appropriate screening tool to use, as a particular screening tool may not be endorsed (e.g. Centers for Disease Control and Prevention, 2020) or an endorsed tool may not be suitable for a particular purpose. In respect of the latter, the

Autism-Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001) is the only ASD screening tool recommended by the National Institute for Health and Clinical Excellence (NICE; 2012), but as it is a self-report tool, where respondents read each item and indicate to what extent they agree with it, it is unsuitable for many people with intellectual disability who have poor or no literacy skills. Though screening tools do not give a full diagnosis they can be advantageous for numerous reasons. First, they can be typically used by a wider range of people than assessment measures. For example, the AQ does not require the person using it to have a particular professional background or training, whereas the Autism Diagnostic Interview Revised (ADI-R) can only be used by an appropriately qualified person (Lord, Rutter, & Le Couteur, 1994). Second, screening measures are typically quicker to use than diagnostic assessment measures making them less time intensive. As a result of both of these factors, there is a recognition by some professional bodies (British Psychological Society, 2003) that there can be pragmatic reasons for using screening measures, for example in order to facilitate timely diagnosis.

Recently, a systematic review was conducted that reviewed the current evidence in terms of psychometric properties of existing questionnaires and diagnostic measures for ASD in adults (see Wigham et al., 2018). However, this only briefly covered questionnaires accessible to people with intellectual disability and only included articles published since 2014. As a result, there is a need for a more comprehensive review of screening measures that have been developed as, or adapted to be, screening tools suitable for people with intellectual disability. The present review aims to address this need. In addition, as many researchers have used measures that were originally designed for children with both adults and children with intellectual disability, the review will include adult and child ASD screening questionnaires. The populations investigated will also comprise both children and adults with intellectual disability, with and without ASD. The measures will be reviewed in terms of their psychometric properties particularly the reliability, validity and standardisation when used with people with intellectual disability.

The aim of the present paper, is therefore, to provide a detailed overview of the psychometric properties of ASD screening tools that are currently available for use with adults and/or children with an intellectual disability. This is in order to help inform clinicians about the most appropriate screening measure available for their purpose, population and individual being screened, and to inform future

directions for the development of screening tools for people with an intellectual disability who may have ASD.

Search strategy

The criteria for the literature search terms are shown in table 1. English language papers which referred to the following in either title, abstract and/or keywords were included: ASD or related term (column one); a screening instrument (column two) and a keyword related to an instrument or scale (column 3). The terms 'intellectual disability,' 'learning disability,' and 'mental retardation' were not included as some articles include this group but may not note them in either the title, abstract, or keywords and so may be missed.

Literature searching was conducted in four databases: ProQuest, PsycArticles, PubMed and Web of Science with publication dates up to July 2019.With duplicates removed a total of 3068 articles were retrieved. The titles of identified articles were initially screened for relevance, then abstracts were read to determine if they were relevant to the review, paying specific attention to the participants and statistical approaches. Any article that the first author had uncertainty about was reviewed by the second author and a consensus between the two was reached about whether to retain the article or not. Full texts of the remaining 44 papers were read, alongside their reference sections in order to identify articles not identified in the initial search. Articles were excluded if they were a review or paper which did not detail specific reliability and validity of screening tools (e.g. Reilly, 2009), stated that DSM-III or earlier criteria were used in respect of diagnosis as clinicians are required to use the most up to date diagnostic methods (e.g. Teal & Wiebe, 1986), did not compare people with intellectual disability and ASD with people who only have intellectual disability (e.g. Li et al., 2018) or outlined a tool which was not a screening instrument for ASD per se, but instead screened for additional challenges people with ASD/intellectual disability may face (e.g. Matson, Fodstad, & Mahan, 2009), for further details of inclusion criteria see table 3. Where it was unclear if an article satisfied these criteria, it was read in full by the first and second authors, who then reached agreement about whether to retain it or not. There was no strict cut-off point regarding date, however the requirement to use at least DSM-IV diagnostic methods meant that articles published pre-1994 would not be included.

The remaining articles were read in full by the first and second authors. Each article had to evince a good quality diagnosis for both ASD and intellectual disability. For ASD this meant diagnosis was consistent with the recommendations of the National Institute for Health and Care Excellence (NICE, 2016; NICE, 2011), while for intellectual disability diagnosis was in line with recommendations by The British Psychological Society (2015). Articles were also included if the participants had been recruited from a setting that was specific to people with intellectual disability e.g. an intellectual disability hospital or had a genetic condition that often results in intellectual disability, such as Down syndrome. The retained articles were scored in terms of quality of diagnosis of participants (see table 2). These scores were agreed upon by the first and second authors.

The 10 retained articles were included in the final review. A search of reference lists identified three further papers that met the search criteria and a final paper was identified from the review by Wigham et al. (2018). The final review contained 14 articles, all of which were scored for quality of diagnosis as outlined above. A full breakdown of numbers of articles identified throughout each stage of screening is given in figure 1. A summary of samples, including how participants were recruited and how ASD and intellectual disability were diagnosed can be found in table 2.

[INSERT TABLE 1 HERE]

[INSERT TABLE 2 HERE]

[INSERT TABLE 3 HERE]

[INSERT FIGURE 1 HERE]

Method of classifying results

As the review had a particular focus on the psychometric properties of the screening tools, an adapted version of the Critical Skills Appraisal Programme (CASP, 2018) checklist was used to guide the quality appraisal, in addition to recommendations from previous researchers about the rating of psychometric vales (see table 4). The review reports on the reliability, validity and standardisation of a range of screening tools and table 4 provides an overview of how reliability and validity were categorised, and the source of the classification. In addition, results that indicate the presence of variance or invariance, or significant or non-significant differences are stated as such. Further details

about psychometric properties can be found in the electronic supplementary materials (ESM:

https://osf.io/tg58m/?view_only=9125261d327548e5abdfab86a2aea70b)

[INSERT TABLE 4 HERE]

Results

The articles reviewed related to 8 screening tools, the majority of which were not designed specifically to screen for ASD in people with intellectual disability but have been adapted for this purpose. The term 'intellectual disability' will be used throughout this article to replace any terms used to indicate intellectual disability in the original articles and ASD to replace any terms used to indicate ASD in the original article (e.g. Autism).

Background information about each screening tool is provided, followed by information about their psychometric properties *in relation to people with intellectual disability.* The former was sourced from general literature about the measures, while the latter was obtained from the papers identified by the systematic search. While it is likely that further relevant information about some of the measures is available, only information about directly using these measures with a sample of people with intellectual disability is included here.

Screening tools only evaluated with children

Autism Behaviour Checklist (ABC: Krug, Arick, & Almond, 1980)

The ABC is an observational instrument designed to screen for ASD in a large population (Bravo Oro, Navarro-Calvillo, & Esmer, 2014). The scale is a checklist of non-adaptive behaviours that reflect an individual's response to challenges in everyday life. The tool consists of 57 items (each scored 1-4) and has five subscales. The cut-point of 58 was proposed by Oswald and Volkmar (1991), scores greater than 58 indicated a high chance of ASD and those below less chance of ASD.

Reliability and validity data

De Bildt et al. (2003) compared the ABC and the Scale of Pervasive Developmental Disorder in Mentally Retarded Persons (PDD-MRS: see below), against existing clinical classification of ASD. The ABC and PDD-MRS showed agreement in 44.8% of cases, and the ABC identified 42.1% of the cases that the PDD-MRS identified, showing very poor agreement between the two. Odds ratios were significant between the ABC and both the ADI-R and clinical classification, but not when compared to

the ADOS-G. Receiver operating characteristic (ROC) analysis of the ABC, compared against clinical classification found an average area under curve (AUC).

Conclusion

The ABC shows agreement with ASD classification when compared with the ADI-R and clinical classification, however, it does not show significant agreement with the ADOS-G and has low agreement when compared to the PDD-MRS. Higher scores are found in those with greater levels of intellectual impairment, which may lead to false positives in those with more severe intellectual disability. No reliability information was available in relation to its use with people with intellectual disability. Overall, caution should be exercised when using this tool.

Modified Checklist for Autism in Toddlers (M-CHAT: For development see Robins, Fein, Barton, & Green, 2001)

Designed to screen for ASD in toddlers, the M-CHAT is a questionnaire completed by parents/carers, with the final version comprising 23 yes/no items (Robins, Fein, Barton, & Green, 2001). A positive screen (indicating a high likelihood of ASD) is determined by either failing on any three items, or at least 2 of the 6 critical items.

Reliability and validity

DiGuiseppi et al. (2010) assessed the M-CHAT with children with Down syndrome. It demonstrated good sensitivity, but inadequate specificity. When combined with the Social Communication Questionnaire (SCQ), false positive results were most common in children with a hearing or a persistent visual problem.

Conclusion

While there is a great deal of research on the M-CHAT in children without intellectual disability, only one study was found relating to children with intellectual disability, which did not report on reliability. Also, odds ratios indicated that factors other than ASD can affect the score of the M-CHAT. A follow-up interview is available which is designed to increase the tool's accuracy, although this was not developed with people with intellectual disability (Robins, Fein, & Barton, 2009). Overall, there is limited evidence that the M-CHAT is an effective screening tool for ASD in people with intellectual disability.

Screening tools evaluated only with adults

    Autism Checklist (ACL)

Based on ICD-10 (WHO, 1990) criteria, the ACL is an observational tool which aims to identify ASD in suspected cases. Each of the three ICD-10 domains are scored 0-4, based on the presence of each criteria. To screen positive, a person needs to have two points in domain one, one point in domains two and three, and six points across all three domains in total (Mutsaerts, Heinrich, Sterkenburg, & Sappok, 2016). No specialised training is needed to complete the checklist.

    Reliability and validity data

Sappok et al. (2014) found a good correlation between Diagnostic Behavioral Assessment for ASD – Revised (DiBAS-R) total score and ACL total score. Mutsaerts et al. (2016) later found the ACL showed significantly higher scores in people with ASD, compared to people without. ROC analysis found an average AUC. The ACL and DiBAS-R showed agreement in 75% of cases, with a poor Cohen's Kappa (κ) value. No information on reliability was found.

    Conclusion

Information on the scale, as used with people with intellectual disability is limited and further research is needed to determine if it would be a useful screening tool for ASD with this group.

    Diagnostic Behavioral Assessment for ASD – Revised (DiBAS-R: For development see Sappok et al., 2014)

The DiBAS-R assesses social communication and interaction in people with intellectual disability and can be used to detect ASD. The assessment is completed by someone who knows the person well and comprises 19 items, each rated 0 to 3, which indicates how often each is true. Higher scores indicate an item is true more often. There is a maximum possible score of 57 and an overall score of 29 or more is used as an indicator of likely ASD, provided that the cut-points of subscales are met. The items are split across two domains in line with the DSM-V criteria of 'Social communication and interaction' and 'Stereotyped and restrictive behaviours and repetitive interests.' The maximum scores of each subscale are 36 and 21 respectively, and the cut-points are 21 and 5 respectively (Mutsaerts et al., 2016).

    Reliability and validity

Sappok et al. (2014) proposes two factors: Social Communication and Interaction (SCI) and Stereotypy Rigidity and Sensory Abnormalities (SRS). Both factors and overall score had good internal consistency. The DiBAS was found to discriminate between those with and without ASD on both subscales and overall scores. A ROC analysis of the total scale showed an average (nearly good, .89) AUC. The best overall cut-point was found to be 29, requiring a score of 21 on the SCI subscale and 5 on the SRS subscale; this showed good sensitivity and adequate specificity but very poor Kappa. The DiBAS showed good correlations with the SCQ, PDD-MRS, and ACL. Inter-rater reliability was good ($r = .88$, $p < .001$, $N = 36$).

In a later study, Mutsaerts et al. (2016) found that DiBAS total score was significantly higher in people with ASD compared to people without ASD. A good correlation between DiBAS-R scores and ACL scores was shown, but Kappa between the two measures was very poor. Those with a milder intellectual disability had a higher chance of a false positive result (Fishers' exact test: $\phi = .31$, $p = .017$).

Heinrich, Böhm and Sappok (2017) found, when the DiBAS-R was assessed in the whole group, it had average AUC, adequate sensitivity but inadequate specificity. When only participants who had mild to moderate intellectual disability were included, AUC was again shown to be average and sensitivity to be adequate, but specificity could be considered good. When only those with severe to profound intellectual disability were included, AUC was shown to be poor and specificity to be very poor, yet sensitivity was considered adequate. The overall percentage accuracy of correctly identifying someone was 70.3% in the whole group; it was notably higher in those with a mild to moderate intellectual disability (83.3%) compared with those with a severe to profound intellectual disability (51.0%).

Conclusion

The DIBAS-R was designed for the detection of ASD through observable social behaviour. Overall, the suitability of this measure shows mixed results, from studies indicating good validity to some finding it to be only adequate or even inadequate on some indices. The available evidence generally suggests that it is a reliable measure, but this evidence is limited. In all, while some findings show that this tool may be an appropriate screening tool for ASD in people with intellectual disability, more research is required before it can be recommended for wider use.

Music-based Autism Diagnostics (MUSAD: Bergmann et al., 2015)

The MUSAD was developed as a diagnostic tool built upon a music framework. It was specifically developed for adults with a lower level of functioning, including those with severe language impairments and is completed by an observer. The test differs slightly if the person is non-verbal. The MUSAD uses music to elicit behaviours that are indicative of ASD symptom severity. It encompasses ten musical interactional situations and the final measure is a 37-item checklist scored 0 to 3, consisting of 3 factors (Social interaction; stereotypies and sensory issues; motor coordination).

Reliability and validity

The MUSAD had a good correlation with the PDD-MRS and modules 1 and 2 of the ADOS-G, an average correlation with the SCQ and a poor correlation with the ABC. Inter-rater reliability was shown to be good between two raters ($r = .71$, 95% CI [.59, .82]) and also good between three ($r = .67$, 95% CI [.62, .72]). Additionally, it showed good test-retest reliability across four tests ($r = .69$) (Bergmann et al., 2015).

Conclusion

The MUSAD is the only reviewed measure that is not an informant measure or questionnaire based. It has good validity, generally showing strong relationships with other ASD screening and diagnostic tools and fair reliability. Overall, it shows potential to be an effective screening tool but more studies are needed with people with intellectual disability.

Screening tools evaluated with both children and adults

Autism Spectrum Disorder-Diagnostic scale (ASD-DA: Matson & Minshawi, 2006)

The ASD-DA is a questionnaire completed by a third party rater which attempts to differentiate people with intellectual disability into those with and without ASD on the basis of observable behaviour. The scale includes 31 items which are scored as either "not different, no impairment" (0) or "different, some impairment" (1). Level of impairment is compared with others of the same age as the target individual (Matson, Wilkins, Boisjoli, & Smith, 2008). Scores greater than or equal to 19 indicate the likely presence of ASD. The scale can be split into three factors, social

impairment, communication impairment, and restricted behaviour (Matson, Wilkins, & González, 2007).

Matson et al. (2008) found the ASD-DA had a good correlation with both the DSM-IV/ICD-10 checklist and the Matson Evaluation of Social Skills for individuals with Severe Retardation (Matson, 1995), and an average correlation with the Socialisation domain of the VABS (Sparrow, Balla, & Cicchetti, 1984).

Conclusion

The ASD-DA was developed for use with people with intellectual disability and shows potential to be used as an ASD screening tool for this group. The limited research indicates that it has good validity when compared with other measures, but reliability information was not available. Further research is needed.

Scale of Pervasive Developmental Disorder in Mentally Retarded Persons (PDD-MRS: Original development see Kraijer, 1990)

The PDD-MRS is a tool completed by a clinician which is specifically designed to detect Pervasive Developmental Disorder Not Otherwise Specified (PDD-NOS) and 'Autism Disorder' in children with intellectual disability. The scale has 12 dichotomous items, indicating presence or absence of ASD. Each item is weighted, with scores of 10 or more, out of a possible maximum of 19, indicating a high likelihood, 7 to 9 indicating a doubtful category and 6 or less indicating a low likelihood (de Bildt et al., 2003).

Reliability and validity

De Bildt et al. (2003) found that the PDD-MRS and ABC agreed for 44.8% of participants, yielding an average correlation between the two, but a very poor kappa coefficient. Odds ratios between the PDD-MRS and the ADOS-G, ADI-R, and clinical classification were significant. AUC was average for detection of PDD and ASD compared with outcomes based on the ADOS-G and clinical classification, and poor for detection of ASD compared to ADI-R results. Specificity was especially high when compared to clinical classification (.917)

Kraijer and de Bildt (2005) made comparisons between people with and without PDD and a 'doubtful' group, where clinical diagnosis was not clear. Significant differences were found between people with and without PDD in all subgroups[1], apart from those with hearing deficits. When comparing those with PDD and the 'doubtful' group, significant differences were found between all subgroups, aside from those with hearing deficits and with Down syndrome. The sensitivity of the PDD-MRS was good, specificity was adequate, and the misclassification rate was low at 10.6. When compared to outcomes based on the ADOS-G, the PDD-MRS had good sensitivity but inadequate specificity.

In a later study, Sappok et al. (2014), found a good correlation between the PDD-MRS and the DiBAS-R.

Pandolfi, Magyar and Dill (2018) investigated the performance of the PDD-MRS with a sample of children with Down syndrome and ASD. Using a cut score of 1.5 an average AUC, good sensitivity, and inadequate specificity was found.

Cortes et al. (2018) adapted the PDD-MRS for a Spanish speaking sample. The internal consistency was shown to be .71 using a Kuder-Richardson-20 test. ROC analysis was conducted with data from the whole sample and those with different levels of intellectual disability. For the whole sample, mild and moderate intellectual disability the AUC was good, for those with severe and profound intellectual disability it was average. The sensitivity of the PDD-MRS was found to be good for the mild, moderate and profound intellectual disability groups, and adequate for the whole sample and severe intellectual disability group. Specificity was adequate for all groups, except for those with a profound intellectual disability where it was inadequate.

Conclusion

The PDD-MRS has undergone more research than most of the measures included in this review, most likely due to it being developed earlier than the others. Overall, it shows good validity and high sensitivity when compared with the ADOS-G, although specificity values were low. It showed good agreement with a range of other screening tools, with the exception of the ABC. Additionally,

---

[1] Subgroups: Profound, severe, moderate, mild and borderline intellectual disability; male, female; speaking, non-speaking; age 2-9, 10-19, 20-39, 40-49, 50-80; blind/severe visual impairment, deaf/severe hearing loss; Down syndrome, and fragile X.

Cortes et al. (2018) showed that a translated version appears to be a useful tool in Spanish speaking samples. In all, little information regarding reliability was found, therefore more research is required. Due to the age of this tool, it should be investigated more closely to ascertain whether the items included are consistent with a more up to date understanding of ASD.

Social Communication Questionnaire (SCQ: Original development see Berument et al., 1999)

Based upon the ADI-R, the SCQ is an informant completed measure which can be used to screen for ASD. There are various versions of the SCQ, but Sappok et al. (2015) reports that the "Lifetime Version" is a 40-item rating scale with two factors (Social communication; Stereotyped behaviour and unusual interests). The items are scored according to whether 'abnormal' behaviour is present or not. Higher scores indicate an increased likelihood of the presence of ASD, with different cut-off scores being used with different versions of the measure and recommended for different groups (Berument et al., 1999).

Reliability and validity

DiGuiseppi et al. (2010) found sensitivity to be good (100%) but specificity to be inadequate. Magyar, Pandolfi and Dill (2012) assessed the performance of the SCQ with participants with Down syndrome, using a two-factor version of the SCQ: Social-Communication (SC) and Stereotyped Behaviour and Unusual Interests (SBUI). Factor analysis yielded reliability coefficients of 0.96 for SC and 0.83 for SBUI. A verbal and non-verbal version of the SCQ was used in this study, depending upon ability of participants. Where possible, items common to both were analysed together. T-tests showed that SCQ scores were significantly higher for those with ASD. ROC analyses were run on both the verbal and non-verbal version of the SCQ and both had an average AUC. The non-verbal version showed good sensitivity and adequate specificity, the verbal version showed adequate sensitivity and specificity.

Sappok et al. (2014) found an average correlation between scores on the SCQ and scores on the DiBAS-R. In a later study, Sappok, Diefenbacher, Gaul and Bölte (2015) tested a number of cut-points of the SCQ-Current score. Using a cut-point of 15, AUC was shown to be average, with good sensitivity but below adequate specificity. Increasing cut-points to 16 and 18, classifications of sensitivity and specificity did not change and AUC was not reported. Kappa was very poor for all three cut-points. The results were broadly the same when the SCQ-Lifetime score was used with cut-points

of 15 and 20. Good correlations were found between SCQ Current scores and the PDD-MRS and ADOS, while an average correlation was found with the ADI-R. Using the SCQ-Lifetime score, a good correlation was found with the ADI-R, but correlations with the PDD-MRS and ADOS were not significant.

Further research by Sappok, Brooks, Heinrich, McCarthy and Underwood (2017) looked at the SCQ across cultures. This found that scores were lower on the SCQ in females compared to males and were also significantly affected by the country the person was recruited from. ROC analysis identified the optimum cut-point as 13, which yielded an average AUC, good sensitivity and inadequate specificity. Three further papers examined the performance of the SCQ, however, participants were stratified by IQ only or described in terms of 'delay,' rather than diagnosis of intellectual disability.

Derks et al. (2017) assessed the SCQ with both a training and validation sample. In most cases, results were the same for both samples. Those with ASD and intellectual disability scored significantly higher than those with intellectual disability only; a cut-point of 15 yielded an average ROC, sensitivity was good while specificity was inadequate. When 5 SCQ items were removed (which were deemed inappropriate for participants) and a cut point of 9 was used, an average AUC, good sensitivity and inadequate specificity were found. Kappa showed very poor/poor agreement between final diagnostic classification and SCQ scores, in both the complete and reduced sets of items.

Conclusion

Of all the measures included in this review, the SCQ is the most widely researched. Overall the scale shows good concurrent validity with other ASD measures (although some low kappa values are reported) and appears to be able to identify ASD well in people with intellectual disability. The specificity is not always adequate, but the sensitivity is consistently high. The limited information that is available about the reliability of the SCQ, suggests it has good internal consistency, but no information about the test-retest or interrater reliability was identified. Overall, while this is the most researched of all measures in this review and generally shows good validity, reliability data are still lacking.

Discussion

There are a number of instruments which have been used to aid the detection and diagnosis of ASD in people with intellectual disability. The ABC and M-CHAT both only had psychometric information available pertaining to children. Both instruments had limited available information about validity and no reliability information, in relation to their use with people with an intellectual disability. The ABC showed good agreement with clinician opinion, but was poorer in other areas of validity, while the M-CHAT scores were influenced by factors unrelated to ASD, which may lead to false positive results. The ACL, DiBAS-R and MUSAD only had limited information available, and this was only in relation to use with adults with an intellectual disability. The ASD-DA, PDD-MRS, and SCQ all had research relating to both children and adults. While the ASD-DA appeared to have good validity, the available research was extremely limited and no information on reliability was provided. Both the PDD-MRS and SCQ were better researched. The validity of the former was generally quite good, but little information on reliability was available. The SCQ showed a mixed picture, with some studies indicating good validity, although specificity was found to be less than adequate by some researchers. The reliability information provided was limited to internal-consistency which was good, but other types of reliability information should be investigated in future research.

Some overarching points to consider are that the majority of measures have limited research specific to people with intellectual disability (with the SCQ appearing to have the most research relevant to this group). Also, while many of the measures have assessed reliability with other populations (e.g. Robins, Fein, Barton, & Green, 2001) limited research was found relating to the reliability of the measures as used with people with intellectual disability.

Another issue relates to the diagnosis of both ASD and intellectual disability. A number of articles were identified as being potentially relevant for the review, but on closer examination, the diagnostic processes were somewhat unclear or not sufficiently robust. For example, Matson, Wilkins and González (2007) investigated the Autism Spectrum Disorders Diagnostic Scale for Intellectually Disabled adults, however in the article they outline that they used checklists based on DSM-IV and ICD-10, to classify individuals, rather than a full diagnostic assessment. Similarly, Arun and Chavan (2018) outline the development of the Chandigarh Autism Screening Instrument, however, it is unclear whether the sample includes people who have intellectual disability and ASD or whether these are two discreet groups.

The retained articles use ASD diagnostic criteria comparable to those recommended by NICE (2012), for instance using the ADOS-G and ADI. That said, these recommended tools were not designed specifically for diagnosing ASD in people with intellectual disability and can be influenced by factors relevant to this group including IQ, complex or more subtle presentation (see Wigham et al., 2018). These authors recommend combining the ADOS-G and ADI-R for better detection of ASD and many of the reviewed articles do so (e.g. De Bildt et al., 2003). More research into the psychometric properties of ASD diagnostic assessments, as used with people with intellectual disability, is needed to ensure they are robust gold standard measures against which the outcomes of screening tools can be compared.

While out of the scope of the present review, a number of other conditions may affect and complicate accurate diagnosis of ASD in people with intellectual disability (Heinrich et al., 2017; Underwood, McCarthy, Chaplin, & Bertelli, 2015; Wigham et al., 2019) and should also be considered when using screening tools. For example, people with intellectual disability experience a higher rate of comorbid conditions compared to those without ASD and many experience more than one, including epilepsy, schizophrenia, anxiety, and alcohol misuse (Cooper et al., 2015).

The findings of this review should be considered alongside the methodology used to identify the articles. The authors chose not to include the term 'intellectual disability' or a synonym thereof in the search terms. This meant that the initial searches were not specific for this group, however, the strategy helped ensure that as many potential papers as possible were included, which, given the wide range of terminology that is used to refer to people with intellectual disability, may have been missed if specific search terms were used at the first stage of searching.

A limitation of this study relates to the variability of information that was available across the included studies in respect of the psychometric properties being examined. No one existing system was found that summarised all of the potential statistical results and a categorisation system had to be developed that was adapted from the Critical Skills Appraisal Programme (CASP, 2018) and recommendations from previous researchers. While this allowed all papers to be judged by the same criteria, it is acknowledged that researchers differ in the ways they categorise results and that classifications are, to some extent, subjective. Related to this, in some cases the statistics, which should have been classified as 'good' or 'poor' based on their values alone, would be rated the

opposite way when considering the implications of the result for the performance of the measure as a screen for ASD. For example, a positive correlation of .9 between a screening measure score and IQ, would be rated as a 'good' correlation on the classification system alone, but would be rated 'poor' in terms of screening measure, as it suggests that the screening score is associated with factors other than ASD i.e. intelligence. Where these conflicts have arisen, they have been noted in the text and in the ESM, however, this illustrates that the statistical properties of a measure should not be considered independently of the purpose of the measure and the context within which the measurement takes place.

Finally, the review focused on the psychometric properties of the screening measures as they related to people with intellectual disability. It is acknowledged that a number of the measures may, for example, show good reliability with other groups, but have no data in respect of people with intellectual disability. This may be underestimating the reliability of measures in this respect. As Raykov (2002) notes, however, good reliability underpins good validity and there is a need to explore any potential group differences when a measure is used with different populations.

In conclusion, there are several screening tools that have been used with people with intellectual disability. We hope that the information helps guide clinical decision making if professionals are considering screening for ASD in people with an intellectual disability. The review shows that, while some tools have some evidence indicating that they may be effective at screening for ASD, no one tool can be recommended based upon the evidence presented here. Any results, therefore, must be treated with some caution and considered in the light of the psychometric properties of the tool, the individual with whom it is being used, and the purpose of using it. The evidence shows that there is a need for further research into all of the tools reviewed here, especially concerning the reliability of screening measures with people with intellectual disability. Additionally, in order to have confidence in the performance of any screening tool, when used with people with an intellectual disability, researchers need to use robust methods for, and provide clear information about, the diagnosis of ASD and intellectual disability in the participant groups.

Funding

Conflict of interest

None

References

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 for autism in toddlers in 1,000 cases and 3,000 controls [corrected]. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*(2), 202–212. https://doi.org/10.1016/j.jaac.2011.11.003

American Phychiatic Association. (2013). DSM-5 Intellectual disability fact sheet. *American Psychiatric Association*, 2.

American Psychiatric Association [APA]. (1994). *Diagnostic and statistical manual of mental disorders (DSM-4)* (4th ed.). Washington DC: American Psychiatric Association.

American Psychiatric Association [APA]. (2000). *Diagnostic and statistical manual of mental disorders (DSM-4, Text Revision)*. Washington DC: American Psychiatric Association.

American Psychiatric Association [APA]. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Washington DC: American Psychiatric Association.

Aron, A., Coups, E. J., & Aron, E. N. (2013). *Statistics for Psychology*. (Q, Ed.) (Sixth). New Jersey: Pearson.

Arun, P., & Chavan, B. S. (2018). Development of a screening instrument for autism spectrum disorder: Chandigarh Autism Screening Instrument. *Indian Journal of Medical Research*, *147*, 369–375. https://doi.org/10.4103/ijmr.IJMR_1968_16

Baron-Cohen, S., & Wheelwright, S. (2004). The Empathy Quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

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 and Developmental Disorders*, *31*(1), 5–17.

https://doi.org/10.1023/A:1005653411471

Bergmann, T., Sappok, T., Diefenbacher, A., Dames, S., Heinrich, M., Ziegler, M., & Dziobek, I. (2015). Music-based Autism Diagnostics (MUSAD) – A newly developed diagnostic measure for adults with intellectual developmental disabilities suspected of autism. *Research in Developmental Disabilities*, *43–44*, 123–135. https://doi.org/10.1016/J.RIDD.2015.05.011

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Clinical Psychiatry*, *175*(5), 444–451. https://doi.org/10.1192/bjp.175.5.444

Brace, N., Kemp, R., & Snelgar, R. (2016). *SPSS For Psychologists (And Everybody Else)* (Sixth). New York: Routledge.

Bravo Oro, A., Navarro-Calvillo, M. E., & Esmer, C. (2014). Autistic Behavior Checklist (ABC) and its applications. In *Comprehensive Guide to Autism* (pp. 2787–2798). New York, NY: Springer New York. https://doi.org/10.1007/978-1-4614-4788-7_164

Brugha, T., Cooper, S. A., McManus, S., Purdon, S., Smith, J., Scott, F. J., … Tyrer, F. (2012). Estimating the prevalence of Autism Spectrum Conditions in adults: Extending the 2007 adult psychiatric morbidity survey. London: NHS, The Health and Social Care Information Centre.

Buescher, A. V. S., Cidav, Z., Knapp, M., & Mandell, D. S. (2014). Costs of Autism Spectrum Disorders in the United Kingdom and the United States. *JAMA Pediatrics*, *168*(8), 721–728. https://doi.org/10.1001/jamapediatrics.2014.210

Centers for Disease Control and Prevention. (2020). Screening and diagnosis of Autism Spectrum Disorder for healthcare providers. Retrieved from https://www.cdc.gov/ncbddd/autism/hcp-screening.html

Church, C., Alisanski, S., & Amanullah, S. (2000). The social, behavioral, and academic experiences of children with asperger syndrome. *Focus on Autism and Other Developmental Disabilities*, *15*(1), 12–20. https://doi.org/10.1177/108835760001500102

Cohen, B. H. (2008). *Explaining Psychological Statistics* (Third). Hoboken, New Jersey, New Jersey: John Wiley & Sons, Inc.

Cooper, S.-A., McLean, G., Guthrie, B., McConnachie, A., Mercer, S., Sullivan, F., & Morrison, J. (2015). Multiple physical and mental health comorbidity in adults with intellectual disabilities: Population-based cross-sectional analysis. *BMC Family Practice*, *16*(1), 110. https://doi.org/10.1186/s12875-015-0329-3

Cortes, M. J., Orejuela, C., Castellvi, G., Folch, A., Rovira, L., Salvador-Carulla, L., … Martinez-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 and Developmental Disorders*, *48*(5), 1566–1578. https://doi.org/10.1007/s10803-017-3416-6

Critical Skills Appraisal Programme (CASP). (2018). CASP Diagnostic Test Study Checklist. Retrieved from https://casp-uk.net/wp-content/uploads/2018/03/CASP-Diagnostic-Checklist-2018_fillable_form.pdf

Dalton, K. M., Nacewicz, B. M., Johnstone, T., Schaefer, H. S., Gernsbacher, M. A., Goldsmith, H. H., … Davidson, R. J. (2005). Gaze fixation and the neural circuitry of face processing in autism. *Nature Neuroscience*, *8*(4), 519–526. https://doi.org/10.1038/nn1421

de Bildt, A., Sytema, S., Ketelaars, C., Kraijer, D., Volkmar, F., & Minderaa, R. (2003). Measuring pervasive developmental disorders in children and adolescents with mental retardation: A comparison of two screening instruments used in a study of the total mentally retarded population from a designated area. *Journal of Autism and Developmental Disorders*, *33*(6), 595–605. https://doi.org/10.1023/B:JADD.0000005997.92287.a3

Derks, O., Heinrich, M., Brooks, W., Sterkenburg, P., McCarthy, J., Underwood, L., & Sappok, T. (2017). The Social Communication Questionnaire for Adults with Intellectual Disability: SCQ-AID. *Autism Research*, *10*(9), 1481–1490. https://doi.org/10.1002/aur.1795

DiGuiseppi, C., Hepburn, S., Davis, J. M., Fidler, D. J., Hartway, S., Lee, N. R., … Robinson, C. (2010). Screening for autism spectrum disorders in children with down syndrome: Population prevalence and screening test characteristics. *Journal of Developmental and Behavioral Pediatrics*, *31*(3), 181. https://doi.org/10.1097/DBP.0b013e3181d5aa6d

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

https://doi.org/10.1002/mrdd.20068

Goin, R. P., & Myers, B. J. (2004). Characteristics of infantile autism: Moving toward earlier detection. *Focus on Autism and Other Developmental Disabilities*, *19*(1), 5–12.

Harms, M. B., Martin, A., & Wallace, G. L. (2010). Facial emotion recognition in autism spectrum disorders: A review of behavioral and neuroimaging studies. *Neuropsychology Review*, *20*(3), 290–322. https://doi.org/10.1007/s11065-010-9138-6

Heinrich, M., Boehm, J., Sappok, T., Böhm, J., Sappok, T., Boehm, J., & Sappok, T. (2017). Diagnosing autism in adults with intellectual disability: Validation of the DiBAS-R in an independent sample. *Journal of Autism and Developmental Disorders*, *48*(2), 341–350. https://doi.org/10.1007/s10803-017-3336-5

Humphrey, N., & Lewis, S. (2008). `Make me normal'. *Autism*, *12*(1), 23–46. https://doi.org/10.1177/1362361307085267

Jones, A. P., Happé, F. G. E., Gilbert, F., Burnett, S., & Viding, E. (2010). Feeling, caring, knowing: Different types of empathy deficit in boys with psychopathic tendencies and autism spectrum disorder. *Journal of Child Psychology and Psychiatry*, *51*(11), 1188–1197. https://doi.org/10.1111/j.1469-7610.2010.02280.x

Kline, P. (1998). *The New Psychometrics* (First). London: Routledge.

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

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. https://doi.org/http://dx.doi.org/10.1007/s10803-005-5040-0

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. https://doi.org/10.1111/j.1469-7610.1980.tb01797.x

La Malfa, G., Lassi, S., Bertelli, M., Salvini, R., & Placidi, G. F. (2004). Autism and intellectual

disability: A study of prevalence on a sample of the Italian population. *Journal of Intellectual Disability Research*, *48*(3), 262–267. https://doi.org/10.1111/j.1365-2788.2003.00567.x

Leung, K., Trevena, L., & Waters, D. (2012). Development of an appraisal tool to evaluate strength of an instrument or outcome measure. *Nurse Researcher*, *20*(2), 13–19. https://doi.org/10.7748/nr2012.11.20.2.13.c9436

Li, C., Zhou, H., Wang, T., Long, S., Du, X., Xu, X., … Wang, Y. (2018). Performance of the Autism Spectrum Rating Scale and Social Responsiveness Scale in identifying Autism Spectrum Disorder among cases of intellectual disability. *Neuroscience Bulletin*, *34*(6), 972–980. https://doi.org/10.1007/s12264-018-0237-3

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/BF02172145

Magyar, C. I., Pandolfi, V., & Dill, C. A. (2012). An initial evaluation of the social communication questionnaire for the assessment of autism spectrum disorders in children with down syndrome. *Journal of Developmental and Behavioral Pediatrics*, *33*(2), 134. https://doi.org/10.1097/DBP.0b013e318240d3d9

Matson, J. L. (1995). *The Matson evaluation of social skills for individuals with severe mental retardation.* Baton Rouge, LA: Scientific Publishers.

Matson, J. L., Fodstad, J. C., & Mahan, S. (2009). Cutoffs, norms, and patterns of comorbid difficulties in children with developmental disabilities on the Baby and Infant Screen for Children with aUtIsm Traits (BISCUIT-Part 2). *Research in Developmental Disabilities*, *30*(6), 1221–1228. https://doi.org/10.1016/j.ridd.2009.04.004

Matson, J. L., & Minshawi, N. F. (2006). *Early intervention for autism spectrum disorders: A critical analysis* (1st ed.). Oxford: Elsevier.

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities*, *30*(6), 1107–1114. https://doi.org/10.1016/J.RIDD.2009.06.003

Matson, J. L., Wilkins, J., Boisjoli, J. A., & Smith, K. R. (2008). The validity of the autism spectrum disorders-diagnosis for intellectually disabled adults (ASD-DA). *Research in Developmental Disabilities*, *29*(6), 537–546. https://doi.org/10.1016/J.RIDD.2007.09.006

Matson, J. L., Wilkins, J., & González, M. (2007). Reliability and factor structure of the Autism Spectrum Disorders— Diagnosis scale for intellectually disabled Adults (ASD—DA). *Journal of Developmental and Physical Disabilities*, *19*(6), 565–577. https://doi.org/10.1007/s10882-007-9070-8

McKenzie, K., & Murray, A. L. (2015). Evaluating the use of the Child and Adolescent Intellectual Disability Screening Questionnaire (CAIDS-Q) to estimate IQ in children with low intellectual ability. *Research in Developmental Disabilities*, *37*, 31–36. https://doi.org/10.1016/j.ridd.2014.11.001

Mutsaerts, C. G., Heinrich, M., Sterkenburg, P. S., & Sappok, T. (2016). Screening for ASD in adults with ID--moving toward a standard using the DiBAS-R and the ACL. *Journal of Intellectual Disability Research*, *60*(5), 512–522. https://doi.org/http://dx.doi.org/10.1111/jir.12290

National Institute for Health and Care Excellence (NICE). (2016). Autism spectrum disorder in adults: diagnosis and management. Retrieved from https://www.nice.org.uk/guidance/CG142/chapter/1-Guidance#interventions-for-autism

National Institute for Health and Clinical Excellence (NICE). (2011). NICE Clinical Guideline: Autism: recognition, referral and diagnosis of children and young people on the autism spectrum. *National Collaborating Centre for Women's and Children's Health*, *85*(1), 22–24.

National Institute for Health and Clinical Excellence [NICE]. (2012). *Autism: Recognition, referral, diagnosis and management of adults on the autism spectrum [CG23]*. London: NICE.

NICE. (2012). Autism spectrum disorder in adults: Diagnosis and management. *National Institute of Clinical Excellence*. Retrieved from https://www.nice.org.uk/guidance/cg142/chapter/1-guidance

O'Brien, G., & Pearson, J. (2004). Autism and learning disability, *8*(2), 125–140. https://doi.org/10.1177/1362361304042718

Oswald, D. P., & Volkmar, F. R. (1991). Signal detection analysis of items from the Autism Behavior

Checklist. *Journal of Autism and Developmental Disorders*, *21*(Dec 91), 543–549. https://doi.org/10.1007/BF02206876

Pandolfi, V., Magyar, C. I., & Dill, C. A. (2018). Screening for autism spectrum disorder in children with Down syndrome: An evaluation of the Pervasive Developmental Disorder in Mental Retardation Scale. *Journal of Intellectual & Developmental Disability*, *43*(1), 61–72. https://doi.org/10.3109/13668250.2016.1271111

Poncelas, A., & Murphy, G. (2007). Accessible information for people with intellectual disabilities: Do symbols really help? *Journal of Applied Research in Intellectual Disabilities*, *20*(5), 466–474. https://doi.org/10.1111/j.1468-3148.2006.00334.x

Public Health England. (2019). NHS population screening explained - GOV.UK. Retrieved from https://www.gov.uk/guidance/nhs-population-screening-explained

Raykov, T. (2002). Examining group differences in reliability of multiple-component instruments. *British Journal of Mathematical and Statistical Psychology*, *55*(1), 145–158. https://doi.org/10.1348/000711002159743

Reilly, C. (2009). Autism spectrum disorders in Down syndrome: A review. *Research in Autism Spectrum Disorders*, *3*(4), 829–839. https://doi.org/10.1016/j.rasd.2009.01.012

Robins, D. L., Fein, D., & Barton, M. (2009). *Modified Checklist for Autism in Toddlers, Revised, with Follow-Up (M-CHAT-R/F)*. Retrieved from http://www.mchatscreen.com.

Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified CHecklist for Autism in Toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*(2), 131–144. https://doi.org/10.1023/A:1010738829569

Sappok, T., Brooks, W., Heinrich, M., McCarthy, J., & Underwood, L. (2017). Cross-cultural validity of the social communication questionnaire for adults with intellectual developmental disorder. *Journal of Autism and Developmental Disorders*, *47*(2), 393–404. https://doi.org/10.1007/s10803-016-2967-2

Sappok, T., Diefenbacher, A., Gaul, I., & Bölte, S. (2015). Validity of the Social Communication

Questionnaire in adults with intellectual disabilities and suspected autism spectrum disorder. *American Journal on Intellectual and Developmental Disabilities*, *120*(3), 203–214.

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. https://doi.org/10.1016/J.RASD.2013.12.016

Sparrow, S. S., Balla, D., & Cicchetti, D. V. (1984). *Vineland Adaptive Behaviour Scale* (Survey edn). Circle Pines, MN: American Guidance Service.

Teal, M. B., & Wiebe, M. J. (1986). A validity analysis of selected instruments used to assess autism. *Journal of Autism and Developmental Disorders*, *16*(4), 485–494.

The British Psychological Society (BPS). (2000). *Learning disability: Definitions and contexts.* Leicester. Retrieved from https://www.bps.org.uk/sites/bps.org.uk/files/Member Networks/Faculties/Intellectual Disabilities/Learning Disability Definitions and Contexts %282000%29.pdf

The British Psychological Society (BPS). (2015). *Guidance on the assessment and diagnosis of intellectual disabilities in adulthood.*

The National Autistic Society. (2016). The autism employment gap: Too much information. The National Autistic Society. Retrieved from http://www.autism.org.uk/~/media/nas/get-involved/tmi/tmi employment report 24pp web.ashx?la=en-gb

Underwood, L., McCarthy, J., Chaplin, E., & Bertelli, M. O. (2015). Assessment and diagnosis of psychiatric disorder in adults with autism spectrum disorder. *Advances in Mental Health and Intellectual Disabilities*, *9*(5), 222–229. https://doi.org/10.1108/AMHID-05-2015-0025

Wigham, S., Rodgers, J., Berney, T., Le Couteur, A., Ingham, B., & Parr, J. R. (2019). Psychometric properties of questionnaires and diagnostic measures for autism spectrum disorders in adults: A systematic review. *AUTISM*, *23*(2), 287–305. https://doi.org/10.1177/1362361317748245

World Health Organisation [WHO]. (1990). *The ICD-10 Classification of Mental and Behavioural*

Disorders. Geneva: World Health Organisation.

World Health Organisation [WHO]. (2018). *No The ICD-11 Classification of Mental and Behavioural Disorders*. Geneva: World Health Organisation.

Zhu, W., Zeng, N., & Wang, N. (2010). Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS ® implementations. In *Proceeings Nesug*. Retrieved from https://www.lexjansen.com/nesug/nesug10/hl/hl07.pdf

Table 1

Search strategy, one from each column must be present in the result

OR	Autis*	AND	Screen*	AND	Instrument
OR	Asperger*	AND	'Red flag'	AND	Tool
OR	Pervasive developmental disorder	AND		AND	Detect*
OR				AND	Questionnaire
OR				AND	Quotient
OR				AND	Procedure
OR				AND	Scale
OR				AND	Indicato*
OR				AND	Identif*
OR				AND	Diagnos*

Table 2

Articles, measures used, details of sample and ratings of sample quality (organised in alphabetical order according to name of tool).

Article and measures	Participants	Area Under Curve (AUC), Sensitivity, and Specificity	ASD diagnosis				Intellectual disability diagnosis				
			Overall	Appropriate Clinician	ADOS-G	ADI-R	Previous diagnosis	Appropriate Clinician	IQ Test	Adaptive Functioning	Dev. History
De Bildt et al. (2003) ABC PDD-MRS	Children with intellectual disability Country: The Netherlands Recruited from: Facilities for children and adolescents with intellectual disability $N = 827$; $N_M = 521$; $N_F = 306$ Age < 12 $N = 437$ Age ≥ 12 $N = 390$ Profound $N = 80$ Severe $N = 102$ Moderate $N = 185$ Mild $N = 460$	*ABC* *Groups identified according to clinical judgement.* *'Autism Disorder'* AUC = .76 Sensitivity = .71 Specificity = .70 *'Pervasive Developmental Disorder'* AUC = .75 Sensitivity = .58 Specificity = .78	5	1	1	1	1	0	1	0	0
Mutsaerts et al. (2016) ACL DiBAS-R	Adults with intellectual disability Country: Germany	*ACL* AUC = .85 Sensitivity = .58 Specificity = 78	5	1	1	1	1	0	1	0	0

31

	Recruited from: Department of psychiatry that specialised in intellectual disability $N = 148$ *ASD group* $N_{ASD} = 84$ *M* age = 38.3; *SD* = 12.2 Severe/profound $N = 42$ Moderate $N = 27$ Mild $N = 15$ $N_{ID} = 64$ *M* age = 34.1; *SD* = 11.7 Severe/profound $N = 30$ Moderate $N = 36$ Mild $N = 34$										
Matson, Wilkins, Boisjoli, and Smith (2008) ASD-DA	Adults and adolescents with intellectual disability Country: USA Recruited from: Intellectual disability centres $N = 307$; $N_M = 168$; $N_F = 139$ *M* age = 52; range 16 – 88 $N_{ASD} = 156$ $N_{ID} = 151$ Profound $N = 235$ Severe $N = 40$ Moderate $N = 16$ Mild $N = 2$	Not available	4	1	0	0	1	0	1	1	0
Sappok et al. (2014)	Adults with intellectual disability	*DiBAS-R* AUC = .89 Sensitivity = .82	5	1	1	0	1	0	1	1	0

Main: DiBAS-R Extra: PDD-MRS ACL SCQ	Country: Germany Recruited from: Inpatient and outpatient services at a psychiatric clinic N = 219; N_M = 125; N_F = 94 M age = 35; SD = 12 N_{ASD} = 77 Severe/Profound N = 37 Moderate N = 26 Mild N = 14 N_{ID} = 142 Severe/Profound N = 31 Moderate N = 57 Mild N = 54	Specificity = .87									
Heinrich, Böhm, and Sappok (2017) DiBAS-R	Adults with intellectual disability Country: Germany Recruited from: Inpatient and outpatient services at a psychiatric clinic N = 381; N_F = 161 M age = 40.5; SD = 13.4 N_{ID} = 289; N_F = 131 M age = 40.8; SD = 13.9 Mild/Moderate N = 189 Severe/Profound N = 100 N_{ASD} = 92; N_F = 30 M age = 39.6; SD = 12.1; Mild/Moderate N = 39	*DiBAS-R* AUC = .81 Sensitivity = .82 Specificity = .67 *Scores given are for the whole group, scores for subgroups are available in ESM**.*	5	1	1	1	1	0	0	1	0

	Severe/Profound *N* = 53										
DiGuiseppi et al. (2010) M-CHAT SCQ	Children with Down syndrome Country: USA Recruited from: A registry of birth defects *N* = 123; N_M = 80; N_F = 43 Age range 2 – 11 Children were born between 1996 and 2003, numbers per year are given within the paper N_{ASD} = 52 N_{Con} = 71	*M-CHAT* Sensitivity = .82 Specificity = .47 *SCQ* Sensitivity = 1.0 Specificity =.57	5	1	1	1	1	0	0	1	0
Bergmann et al. (2015) MUSAD	Adults with intellectual disability Country: Germany Recruited from: Psychiatric department specialising in intellectual disability *N* = 76 *M* age = 38.3; *SD* = 11.7; Range 18 – 66 N_{ASD} = 50; N_M = 42 N_{Con} = 26; N_M = 5	Not available	6	1	1	1	1	0.5	0.5	0.5	0.5

Kraijer and de Bildt (2005) PDD-MRS	Large sample with intellectual disability Country: The Netherlands Recruited from: Previous research *N* = 1230 Age range 2 – 80 2 years *N* = 71 2 – 9 *N* = 379 10 – 19 *N* = 101 20 – 29 *N* = 168 30 – 39 *N* = 273 40 – 49 *N* = 238 50 – 80 *N* = 71	*Compared to diagnostic status* Sensitivity = .92 Specificity = .92 *Compared to ADOS-G* Sensitivity = .81 Specificity = .47								
	Two-year olds with all levels of intellectual disability *N* = 71 Persons in institutes/group homes *N* = 781 Profound intellectual disability *N* = 63 Persons who attend day centres *N* = 374 Persons who were attending a specialist clinic for observation and treatment *N* = 75 Subgroups listed: Profound, severe, moderate, mild and borderline intellectual disability; male, female; speaking, non-speaking; age 2-9, 10-19,		2	1	0	0	1	0	0	0	0

20-39, 40-49, 50-80; blind/severe visual impairment, deaf/severe hearing loss; Down syndrome, and fragile X.

Pandolfi, Magyar, and Dill (2018) PDD-MRS	Children with Down syndrome Country: USA Recruited from: A previous study of comorbidities of Down syndrome *N* = 386 Used in some analyses Sample, for tests of interest N_{Con} = 38 N_{ASD} N = 33 IQ Score Con *M* = 52.38, *SD* = 14.57 ASD *M* = 41.93, *SD* = 6.74 VABS Composite scores Con *M* = 69.65, *SD* = 9.87 ASD *M* = 60.12, *SD* = 10.91	*PDD-MRS* AUC = .81 Sensitivity = .91 Specificity = ..61	8	1	1	1	1	1	1	1	1
Cortes et al. (2018) PDD-MRS (Spanish adaption)	Adults with intellectual disability Country: Spain Recruited from: A wider project conducted by mental health and	*PDD-MRS* AUC = .91 Sensitivity = .70 Specificity = .91	2	1	0	0	1	0	0	0	0

	neurodevelopmental disorder professionals

N = 979; $\%_M$ = 55.7; $\%_F$ = 44.3%
M age = 42.4, SD = 13.9

Live in:
Staffed residences = 52.2%
At home = 47.8%

Mild = 25.5%
Moderate = 28.1%
Severe = 26.9%
Profound = 19.5%

Intellectual disability with genetic cause = 18.3%
Down syndrome = 9% | | | | | | | | | | |
| **Magyar, Pandolfi and Dill (2012)**

SCQ | Children with Down syndrome

Country: USA

Recruited from: A previous ASD prevalence study

Group 1, exploratory factor analysis:
N = 188; N_M = 95; N_F = 93
M age = 9.26; SD = 3.13

Group 2, confirmatory factor analysis:
N = 188;
N_M = 101; N_F = 87
M age = 9.26; SD = 3.13 | *SCQ*
Non-verbal
Cut score = 6.5
AUC = .82
Sensitivity = .82
Specificity = .68

Verbal
Cut score = 10.5
AUC = .78
Sensitivity = .73
Specificity = .76 | 5 | 1 | 1 | 1 | 1 | 0 | 0.5 | 0.5 | 0 |

	Group 3 other analyses: *N* = 71: *N*$_{Con}$ = 38; *N*$_M$ = 17; *N*$_F$ = 21 *M* age = 7.92; *SD* = 3.19 *M* IQ = 54.38 *N*$_{ASD}$ = 33; *N*$_M$ = 23; *N*$_F$ = 15 *M* age = 8.97; *SD* = 2.51 *M* IQ = 41.93; *SD* = 6.74										
Sappok, Diefenbacher, Gaul and Bölte (2015) SCQ	Adults with intellectual disability Country: Germany Recruited from: A university affiliated Department of Psychiatry *N* = 151 *M* age = 37.2, *SD* = 12.8 *N*$_{ASD}$ = 83; *N*$_M$ = 62 *M* age = 35; *SD* = 11 Severe/Profound *N* = 33 Moderate *N* = 39 Mild *N* = 11 *N*$_{ID}$ = 68; *N*$_M$ = 48 *M* age = 40; *SD* = 14 Severe/Profound *N* = 20 Moderate *N* = 34 Mild *N* = 14	*SCQ* *Cut score = 15* AUC = .85 Sensitivity = .98 Specificity = .47 *Scores for other cut-points are available in ESM**.*	6	1	1	1	1	0	0	1	0
Derks et al. (2017) SCQ	Adults with intellectual disability Country: Germany, UK, USA	*SCQ* *'Validation Sample'* *Cut score = 15*	4	1	1	1	1	0	0	0	0

	Recruited from: Specialised intellectual disability and mental health services	AUC = .81 Sensitivity = .84 Specificity = .62 Scores for other cut-points are available in ESM**.								
	N = 451 (all male) Severe/profound *N* = 130 Moderate *N* = 178 Mild *N* = 143 Germany sample *N* = 261 ASD: *M* age = 37.3; *SD* = 11.35 No ASD: *M* age 37.19; *SD* = 13.49 UK sample *N* = 121 ASD: *M* age = 36.35; *SD* = 12.41 No ASD: *M* age = 43.64; *SD* = 12.08 USA sample *N* = 69 ASD: *M* age = 27.62; *SD* = 5.78 No ASD: *M* age = 30.10; *SD* = 5.78 Participants were randomly split into a training sample (*N* = 226), to develop a new scoring algorithm. And, a validation sample (*N* = 225), to validate the algorithm.									
Sappok, Brooks, Heinrich, McCarthy	Adults with intellectual disability Country: Germany, UK, USA	*SCQ* *Cut point = 13* AUC = .80 Sensitivity = .87	5.5	1	1	1	1	0	0.5	0.5	0.5

and Underwood (2017) SCQ	Recruited from: Specialised intellectual disability and mental health services Germany sample $N = 261$; $N_M = 181$ M age = 37.3; $SD = 12.3$ $N_{ASD} = 181$ Mild $N = 52$ Moderate $N = 118$ Severe/profound $N = 91$ UK sample $N = 121$; $N_M = 87$ M age = 40.6; $SD = 12.7$ $N_{ASD} = 51$ Mild $N = 60$ Moderate $N = 30$ Severe/profound $N = 31$ USA sample $N = 69$; $N_M = 50$ M age = 29.4; $SD = 6.4$ $N_{ASD} = 21$ Mild $N = 31$ Moderate $N = 30$ Severe/profound $N = 8$	Specificity = .58

Note* N = Total, N_M = Male, N_F = Female, N_{ASD} = Total ASD group, N_{ID} = Total intellectual disability group, N_{Con} = Total non-ASD control group. Age in years unless specified, M = mean, SD = standard deviation. Profound, severe, moderate, mild refer to level of intellectual disability reported in the article. Studies published prior to the publication of DSM V (APA, 2013) are likely to have categorised severity in terms of IQ; studies after that date are likely to have categorised according to adaptive functioning. For ratings of sample quality, a score of 1 indicates that the criterion is satisfied, a score of 0 indicates that it was not satisfied or not made clear within the article, and a score of 0.5 indicates that the criterion was partially satisfied. For appropriate clinician this means that the clinicians reported to have carried out the diagnoses of either ASD or intellectual disability are appropriate and qualified. The ADOS-G and ADI-R columns indicate that the tools were used, as recommended by NICE (2016; 2011). For intellectual disability it is possible that the sample were previously diagnosed by the institution they were recruited from or by other means. There are also three criteria which are desirable to satisfy to appropriately diagnose intellectual disability which are that an IQ test, a test of adaptive functioning were carried out and a developmental history was provided, according to recommendations by the BPS (2000).
*Note*** ESM is available at: https://osf.io/tg58m/?view_only=9125261d327548e5abdfab86a2aea70b

Table 3

Classification of psychometric properties, the rating and corresponding values

Inclusion	Exclusion
English language.	A review paper.
Evidence of a good quality diagnosis	DSM-III or earlier diagnostic criteria.
For ASD: In line with NICE (2016, 2011) guidelines.	Did not compare ASD with intellectual disability to an intellectual disability without ASD group.
For intellectual disability, either: In line with BPS (2000) guidelines, were recruited from a specific intellectual disability setting, or had a genetic condition associated with intellectual disability.	Outlined, investigated, or developed a tool designed to screen for additional challenges associated with ASD and intellectual disability.

Table 4

Classification of psychometric properties, the rating and corresponding values

Rating	Cronbach's alpha (α)	Pearson correlation (r)	Kappa	Area Under Curve (AUC)
Good	≥ .80	+/- .50-1	≥ .91	> .9
Average	.70 - .79	+/- .30-.49	.81-.90	.7 - .9
Poor	.60 - .69	+/- .10-.29	.61-80	.5 - .69
Very poor	≤ .59	+/- <.10	≤ .60	≤ .49
	(Aron, Coups, & Aron, 2013; Brace, Kemp, & Snelgar, 2016; Cohen, 2008; Kline, 1998)	(Leung, Trevena, & Waters, 2012)	(Zhu, Zeng, & Wang, 2010)*1	(Leung et al., 2012)

Rating	Sensitivity	Specificity
Good	≥ .80	
Adequate	.70 - .79	≥ .80
Inadequate	≤ .69	< .79
	(Glascoe, 2005)	(Glascoe, 2005)

Note *1 Originally there were five categories: 'Excellent', 'good', 'worthless', 'not good' and 'very poor'. Here, 'not good' and 'worthless' have been collapsed into one category. Names of categories have been made consistent with other measures.

Where a result is significant (p < .05) using a statistical approach which is not listed above, a rating of 'Good' will be given, otherwise a rating of 'Very poor' will be given (Leung et al., 2012). Other results may indicate the presence of variance or invariance and will be stated as such.

FINAL figure1.pdf

Figure 1: A PRISMA diagram showing the identification and selection procedure of included articles.