



HHS Public Access

Author manuscript

J Am Acad Child Adolesc Psychiatry. Author manuscript; available in PMC 2016 December 01.

Published in final edited form as:

J Am Acad Child Adolesc Psychiatry. 2016 December ; 55(12): 1054–1063.e3. doi:10.1016/j.jaac.2016.09.490.

Multidimensional influences on autism symptom measures: Implications for use in etiological research

Karoline Alexandra Havdahl, Ph.D.,

Norwegian Institute of Public Health and Lovisenberg Hospital, Oslo, Norway

Vanessa Hus Bal, Ph.D.,

University of California San Francisco, USA

Marisela Huerta, Ph.D.,

Weill Cornell Medical College, New York, USA

Andrew Pickles, Ph.D.,

King's College London, UK

Anne-Siri Øyen, Ph.D.,

Norwegian Institute of Public Health and Lovisenberg Hospital, Oslo, Norway

Camilla Stoltenberg, Ph.D.,

Norwegian Institute of Public Health, Oslo, Norway and University of Bergen, Norway

Catherine Lord, Ph.D., and

Weill Cornell Medical College, New York, USA

Somer L. Bishop, Ph.D.

University of California San Francisco, USA

Abstract

Objective—Growing awareness that symptoms of autism spectrum disorder (ASD) transcend multiple diagnostic categories, and major advances in the identification of genetic syndromes associated with ASD, has led to widespread use of ASD symptom measures in etiological studies of neurodevelopmental disorders. Insufficient consideration of potentially confounding factors such as cognitive ability or behavior problems can have important negative consequences in interpretation of findings, including erroneous estimation of associations between ASD and etiological factors.

Corresponding Author: Somer L. Bishop, 401 Parnassus Avenue, Langley Porter, San Francisco, CA 94143, Telephone: +1 (415) 502-3349, Somer.Bishop@ucsf.edu.

financial disclosures

Drs. Lord, Bishop, and Pickles receive royalties for publication of diagnostic instruments they have co-authored; the Autism Diagnostic Observation Schedule, Generic (Pickles and Lord), the Autism Diagnostic Observation Schedule, 2nd Edition (Bishop and Lord), and the Autism Diagnostic Interview-Revised (Lord). Ms. Havdahl and Drs. Hus Bal, Huerta, Øyen, and Stoltenberg report no potential conflicts of interest.

A subset of results was presented at the Gatlinburg Conference on Research and Theory in Intellectual and Developmental Disabilities in New Orleans, LA, on April 2, 2015.

Method—Participants were 388 children aged 2–13 years with diagnoses of ASD or another neurodevelopmental disorder without ASD. Receiver operating characteristics methods were used to assess the influence of IQ and emotional/behavioral problems on the discriminative ability of three widely used ASD symptom measures; the Social Responsiveness Scale (SRS), the Autism Diagnostic Interview-Revised (ADI-R), and the Autism Diagnostic Observation Schedule (ADOS).

Results—IQ influenced the discriminative thresholds of the SRS and ADI-R, and emotional/behavioral problems affected the discriminative thresholds of the SRS, ADI-R, and ADOS. This resulted in low specificity of ASD cut-offs on the SRS and ADI-R for children with intellectual disability without ASD (27–42%), and low specificity across all three instruments for children without ASD with elevated emotional/behavioral problems (36–59%). Adjustment for these characteristics resulted in improved discriminative ability for all three ASD measures.

Conclusion—The findings indicate that scores on ASD symptom measures reflect far more than ASD symptoms. Valid interpretation of these measures requires steps to account for the influences of IQ and emotional/behavioral problems.

Keywords

measurement; discriminative ability; autism diagnostic observation schedule; autism diagnostic interview-revised; social responsiveness scale

Introduction

Numerous standardized instruments designed to measure symptoms of autism spectrum disorders (ASD) have been developed during the past two decades. These instruments have had a major influence on the field, changing the way that ASD assessments are conducted for clinical and research purposes across many parts of the world. The use of ASD symptom measures has extended to a wide range of populations with complex behavioral presentations, as awareness has grown about the frequent co-occurrence of ASD, or ASD symptoms, with other disorders. In clinical settings, standardized instruments are commonly used to assess ASD symptoms among children with medical, developmental, and psychiatric concerns not specific to ASD.^{1, 2} Current practice parameters recommend routine screening for symptoms of ASD as part of general child psychiatric and developmental assessments, and comprehensive diagnostic evaluation for ASD if symptoms are detected.³ There has also been a surge in research studies using standardized instruments to determine the prevalence and correlates of ASD in individuals with a range of psychiatric diagnoses (e.g., ADHD) and neurogenetic syndromes (e.g., Fragile X syndrome, Cornelia de Lange syndrome).^{4–8}

A foundational assumption is that scores on ASD symptom measures actually reflect symptoms of ASD; higher or lower scores are thought to correspond to higher or lower levels of ASD symptomatology. However, concerns have been raised that factors not specific to ASD, such as cognitive impairments and emotional/behavioral problems can lead to artificially inflated scores on ASD symptom measures.^{4, 9–13} For example, impulsivity and hyperactivity may be manifested as inappropriate social approaches or repetitive behaviors,

anxiety and mood problems as social avoidance or inflexibility, and intellectual disability as age-inappropriate peer interactions or interests.

Due to the challenges inherent in differentiating ASD symptoms from other developmental, emotional and behavioral problems, best practices dictate that ASD diagnosis be based on expert clinician judgment, informed by all available data from a comprehensive assessment including multiple sources of information and standardized instruments.^{3, 14, 15}

Nevertheless, the ways in which individual characteristics such as IQ and emotional/behavioral problems influence agreement between ASD symptom measures and clinical best-estimate diagnosis are not well understood. Additionally, in research studies, ASD symptom scores are commonly analyzed without properly considering how individual characteristics might affect score distributions.^{4-6, 16-22}

Insufficient attention to factors influencing ASD symptom measures can have important negative consequences, both in clinical service contexts (e.g., misdiagnoses, loss of time for appropriate interventions) and research (e.g., inflated prevalence rates, erroneous estimation of the strength of associations between ASD symptoms and etiological factors). As the composition of clinical referral and research populations continues to evolve, a thorough understanding of how individual characteristics affect the discriminative ability of ASD symptom measures becomes critical.^{23, 24}

The purpose of the current study was to examine how cognitive ability (IQ) and emotional/behavioral problems (EBP) influence the discriminative ability of three widely used ASD symptom measures; the Social Responsiveness Scale (SRS)^{25, 26}, the Autism Diagnostic Interview-Revised (ADI-R)²⁷, and the Autism Diagnostic Observation Schedule (ADOS)^{28, 29}. While absolute differences in discriminative accuracy between the instruments would be expected due to method variance (e.g., length, format, informant), the main goal was to investigate the influence of IQ and EBP on discriminative ability within and across multiple ASD measures. Discriminative ability can be affected in various alternative or complementary ways, with implications for interpretation of ASD symptom scores. For example, if cognitive ability is comparably associated with elevated ASD symptom scores in children both with and without ASD, this will only affect the validity of specific thresholds and not the overall discriminative capacity of the ASD measure. Thus, reasonable steps could be taken to adjust for IQ-related differences in discriminative thresholds. However, it is also possible that cognitive ability is differentially associated with ASD symptom scores among children without ASD than in children with ASD, thereby compromising the inherent capacity of the measure to distinguish between these groups. We aimed to assess both forms of covariate influence on the discriminative abilities of the ASD measures in a sample of children with ASD and a control group of children with other neurodevelopmental disorders without ASD.

Methods and Materials

Participants

Participants included in the current analyses were assessed as part of a study aiming to validate an ASD screening interview.³⁰ The sample consisted of 407 children aged 2–13

years who were recruited from two clinics based on previous diagnosis of ASD or another neurodevelopmental disorder without ASD (i.e., ADHD, language disorders, intellectual disability and/or early developmental delays, or mood/anxiety disorders). The only exclusion criteria were known genetic syndromes, or severe sensory (i.e., blindness, deafness) or motor impairments (i.e., not walking) that would render the administration of certain instruments invalid. For the present analyses, 19 children were excluded due to not fulfilling criteria for any DSM-IV-TR diagnosis (n=10), or not having measures of IQ (n=1) or parent-rated emotional/behavioral problems (n=8). Therefore, the final sample consisted of 388 children with best-estimate clinical diagnoses of ASD (n=225) or other neurodevelopmental disorders without ASD (n=163; ADHD n=62, language disorder n=48, intellectual disability n=26, mood/anxiety disorder n=27). Sample characteristics are presented in Table 1. Co-occurring diagnoses were assigned to 54.7% of children in the ASD group and 29.5% of children in the non-ASD group (Table S1, available online).

Participants were recruited through clinic referral/intake, professional referral, and flyers or website communication, either from the Divisions of Developmental and Behavioral Pediatrics and Behavioral Medicine and Clinical Psychology at a large Children's hospital in Ohio, or at a university-based clinic specializing in ASD in Michigan. The majority of study participants with ASD were seen at the ASD specialty clinic (80.0%; n=180 of 225), whereas most of the non-ASD participants were seen at the Children's hospital (88.3%; n=144 of 163). The Institutional Review Boards at the respective sites approved study procedures, and caregivers provided informed consent prior to participation.

Measures

The SRS²⁵ is a 65-item parent-completed questionnaire designed to measure ASD symptoms in 4–18 year-old children (applicable to 319 of 388 participants in the current study). The raw total (range 0–195) was used in analyses of continuous SRS scores. The SRS-Second Edition (SRS-2)²⁶ is equivalent to the original SRS for this age group, and provides T scores and cutoffs for clinically significant ASD symptoms according to population-based norms by age and sex (Mild range=T 60, Moderate range=T 66, Severe range=T 75). The SRS-2 Moderate and Severe cutoffs were used in the present analyses, as scores in the Mild range are reported to have less discriminative utility.²⁶

The ADI-R²⁷ is a standardized caregiver interview that collects information about both past and current ASD symptoms. In order to examine continuous ADI-R scores based on current symptoms, as for the SRS and the ADOS, ADI-Current total scores (range 0–64) were used. ADI-Current comprised the sum of scores from ADI-R items that have been mapped to DSM-5 criteria in previous studies.^{7, 31} The ADI-R formal classification of autism (i.e., Autistic Disorder), as well as criteria recommended for classifying broader ASD,³² are based on past behavior items from the diagnostic algorithm. Therefore, results of analyses are also reported for the total sum of diagnostic algorithm items (ADI-diagnostic, range 0–68).

The ADOS is a clinician-administered observational assessment designed to elicit behaviors relevant to the diagnosis of ASD at particular ages and developmental levels.^{11, 28, 29} The ADOS-2 algorithm raw total (range 0–28) and calibrated severity scores (CSS, standardized

according to age and language level) were calculated. Raw totals are used to generate classifications of autism, ASD, and non-ASD. The raw totals were used when examining continuous ADOS scores given the restricted range (1–10) of the CSS. In addition to items measuring ASD-related behaviors, the ADOS includes ratings of over-activity, aggressive behavior, and anxiety that occurred during the ADOS session (“other abnormal behaviors”, not included in the ASD algorithm).²⁸ The sum of these three items was used to measure clinician-observed emotional/behavioral problems (EBP-ADOS, range 0–6).

Parent-rated emotional/behavioral problems were measured by the Internalizing and Externalizing Problems scales from the Child Behavior Checklist (CBCL).^{33, 34} T scores of 64 or higher are in the clinical range. Table 1 shows the proportions of children with clinical-range emotional/behavioral problems. Nonverbal IQ (NVIQ) was used as a continuous measure of cognitive ability, measured with standard scores from the Differential Ability Scales-II³⁵ (DAS-II; n=330), or ratio scores (mental age/chronological age*100)³⁶ from the Mullen Scales of Early Learning³⁷ (MSEL; n=58). Among the children without ASD, 26 (16.0%) were diagnosed with intellectual disability (ID). NVIQ below 70 was used to define intellectual disability in children with ASD (n=72, 32.0%).

Procedure

The study assessment protocol included questionnaires (CBCL, SRS), parent interview (ADI-R, Vineland-II³⁸), and direct child assessment (cognitive and language testing, ADOS, self-report questionnaires for older, verbal children) (see ³⁰). The ADOS and ADI-R were administered by experienced clinicians who had demonstrated research reliability on the instruments, and reliability was continuously monitored within and across the two sites. There was always at least one Ph.D. level clinical psychologist assigned to each case. Clinicians were blinded to the diagnostic status or other previous information about the participant. To maintain blindness during the assessment protocol, the parent interview and child assessment were conducted by different clinicians whenever possible (72% of cases). In accordance with current best practices for diagnosing ASD,^{3, 14, 39} consensus clinical judgment informed by all information obtained during the evaluation was used to assign a best-estimate diagnosis (DSM-IV-TR criteria). Total/algorithm scores from the SRS, ADI-R, and ADOS were not calculated until after the diagnosis had been assigned, but the clinicians reviewed all other information from the assessment (including qualitative and item-level information elicited from these instruments).

Data analyses

All analyses were undertaken in Stata 13 (Statacorp, 2013), in parallel for the SRS (raw), the ADI-R (current and diagnostic), and the ADOS (raw). Emotional/behavioral problems were examined both as rated by parents (CBCL-Internalizing and CBCL-Externalizing) and clinicians (EBP-ADOS) because previous research suggests that these problems are often context-specific, with modest associations across rating contexts.⁴⁰ In the total sample, EBP-ADOS (sum of the non-algorithm items of over-activity, anxiety and aggressive behavior) correlated significantly with CBCL-Externalizing (0.15, $p=0.003$), but not with CBCL-Internalizing ($r=0.06$, $p=0.27$). Child’s age and sex have been found to be associated with both ASD symptoms and emotional/behavioral problems,^{41, 42} and were therefore included

in all analyses. Assessment site and racial composition (which was associated with site) differed between the ASD and non-ASD diagnostic groups, but were not included in the models given no significant associations with ASD instrument scores when accounting for diagnosis, and no interaction effects of site/race and diagnosis (all $p > 0.05$).

ROC analysis is a well-established method of assessing the discriminative ability of a continuously scored instrument compared against a dichotomously defined gold standard (e.g., clinical diagnosis).⁴³ The ROC curve is a plot of the true positive rate, the proportion of individuals with ASD correctly classified as ASD, against the false positive rate, the proportion of individuals without ASD misclassified by the instrument as ASD, across the full range of possible thresholds (i.e., cut-offs). In assessment of the discriminative ability of ASD instruments, confounding occurs when a covariate such as IQ is associated both with instrument scores and diagnostic group. For example, if lower IQ is associated with higher scores on an ASD instrument, then discriminative ability will be overestimated when mean IQ is lower in the ASD group than in the control group.

Pepe and colleagues have developed methods to assess the influence of covariates on the discriminative ability of instruments.^{43, 44} The first step is to examine whether the covariate is associated with scores on the instrument in the control group, thereby affecting the threshold (i.e., cut-off) of optimal separation between ASD cases and non-ASD controls. Thus, we used multiple linear regression analysis with NVIQ, emotional/behavioral problems, age and sex entered as predictors of scores on each ASD instrument (SRS, ADI-R, ADOS) (Table 2). Power analysis showed at least 80% power for the detection of small associations ($\eta^2_p = 0.01$; Cohen's $f = 0.10$) (6 predictors, $\alpha < .05$, smallest subsample $n=142$). The second step is to assess whether the covariate is associated with the overall separation (i.e., the ROC curve) between ASD cases and non-ASD controls, above and beyond a need to use covariate-adjusted thresholds. ROC regression methods were used to test this, using the Stata procedure *rocreg* with the linear covariate adjustment option (Table 3).

In all regression analyses, the continuous predictors (i.e., NVIQ, emotional/behavioral problems, age) were centered at the mean, and bootstrapping was used for robust estimation of standard errors and confidence intervals (1,000 resamples). Significance level was set at $\alpha < .05$ (two-tailed), using raw p values given that all analyses were planned. We also report the p values adjusted for the false discovery rate⁴⁵ (denoted as q values). As measures of effect size, we report partial eta squared (η^2_p , small: 0.01–0.05, medium: 0.06–0.13, large: 0.14) from multiple linear regression analyses, and sensitivity, specificity and area under the curve (AUC) scores from ROC analyses. Pearson's correlations showing the bivariate associations among the child characteristics and the ASD instruments are presented in Table S2, available online.

Results

Influence on discriminative thresholds

The multiple linear regressions showed that lower NVIQ among children without ASD was associated with elevated scores on the SRS ($\eta^2_p = 0.16$, $q < 0.01$) and the ADI-R (ADI-diagnostic $\eta^2_p = 0.09$, $q < 0.01$; ADI-current $\eta^2_p = 0.04$, $q = 0.05$), but not on the ADOS (Table

2). Thus, the specificity of the SRS and ADI-R was dependent on NVIQ. When comparing children with and without intellectual disability, specificity was 27% versus 65% for the SRS (Moderate cutoff), 42% versus 71% for the ADI-R (ASD cutoff), and 77% versus 76% for the ADOS (ASD cutoff) (see confidence intervals in Table 4).

Emotional/behavioral problems were associated with elevated scores on all three ASD measures, although the particular associations varied according to the rating context. Parent-rated EBP were associated with elevated scores on the parental report-based SRS (CBCL-Internalizing $\eta^2_p=18$, $q<0.01$; CBCL-Externalizing $\eta^2_p=18$, $q<0.01$) and ADI-R (CBCL-Internalizing: ADI-current $\eta^2_p=07$, $q<0.01$, ADI-diagnostic $\eta^2_p=04$, $q=0.05$; CBCL-Externalizing: ADI-current $\eta^2_p=0.03$, $q=0.05$, ADI-diagnostic $\eta^2_p=02$, $q=0.17$). When comparing children with and without clinically significant parent-rated EBP, specificity of the SRS (Moderate cutoff) was 36–39% versus 68–75%, and specificity of the ADI-R (ASD cutoff) was 56–59% versus 72%. Clinician-observed EBP were not significantly associated with scores on these parental report-based ASD measures over and above parent-rated EBP (Table 2).

Parent-rated EBP were not associated with scores on the clinician observation-based ADOS, and fairly stable specificity was found across subgroups differing in parent-rated EBP (75–78%). In contrast, clinician-observed EBP were significantly associated with elevated ADOS scores ($\eta^2_p=0.05$, $q<0.05$) (Table 2). When comparing children with and without marked emotional/behavioral problems as observed by clinicians, specificity was 56% versus 83% for the ADOS (ASD cutoff), 51% versus 72% for the ADI-R (ASD cutoff), and 41% versus 63% for the SRS (Moderate cutoff).

In addition to NVIQ and EBP, the child's age was associated only with the discriminative threshold of the ADI-R (ADI-current $\eta^2_p=0.11$, $q<0.01$, ADI-diagnostic $\eta^2_p=0.06$, $q<0.01$), such that younger age was associated with elevated ADI-current scores among children without ASD. The child's sex was not significantly associated with the discriminative threshold of any of the ASD measures.

Influence on overall discriminative capacity (the ROC curve)

The multiple regression analyses above examined whether the covariates (i.e., NVIQ and EBP) were associated with shifted distributions of ASD instrument scores, suggesting the need for covariate-adjusted thresholds. ROC regression tested whether the covariates were associated with the inherent capacity of the instruments to separate between children with and without ASD when using covariate-adjusted thresholds. Table 3 presents the regression coefficients of the covariates on the ROC curve for each of the ASD instruments (SRS, ADI-R and ADOS). Neither NVIQ nor EBP were found to significantly affect the ROC curve of any of the three ASD measures. As shown in Table 4, subgroups stratified by level of NVIQ and EBP had AUC scores with overlapping confidence intervals, and the reductions in specificity associated with intellectual disability (SRS, ADI-R) and elevated EBP (SRS, ADI-R, ADOS) were generally accompanied by increases in sensitivity. Figure 1 illustrates that the score distributions on the ASD measures were shifted according to level of EBP in both the ASD and non-ASD diagnostic groups (EBP are shown by parent-rating for the SRS

and ADI-R and by clinician-observation for the ADOS because these were significantly associated; other combinations are shown in Figure S1, available online).

The child's sex was not significantly associated with the ROC curve of any of the ASD measures (Table 3). However, child age was found to be significantly associated with the ROC curve of the ADI-R (ADI-diagnostic coefficient=0.21, 95% CI=0.11,0.31, $q<0.01$, ADI-current $q=0.06$), such that ADI-R scores showed increased accuracy for ASD in older compared to younger children.

Sensitivity analyses

Results for the ADI-R and ADOS were similar when restricting the sample to children who also had completed the SRS (i.e., age 4 years or older). Results were also very similar when using the ADOS calibrated severity scores instead of the raw scores, or when using full scale or verbal IQ instead of NVIQ. Controlling for clinic site did not affect the results. Due to high correlations between NVIQ and continuous measures of language functioning ($r=0.64$ with Vineland Expressive Communication), we could not adequately address language as an additional factor. However, adjusting for whether children had attained fluent speech (i.e., whether or not they were capable of completing Module 3 of the ADOS) did not affect the findings, and this measure of language level was not significantly associated with the ASD measures beyond IQ. Given that CBCL-Internalizing covers withdrawn behaviors that may overlap with core ASD symptoms, we re-analyzed the data using CBCL-Anxiety, with very similar findings. Measurement of EBP by parental interview (i.e., ADI-R current aggressive behavior) instead of the questionnaire-based CBCL also resulted in similar results (significant medium sized associations with discriminative thresholds only for the SRS and ADI-R, non-significant associations with ROC curves). Medication use was not significantly associated with ASD symptom scores, and did not affect any of the models tested. The results of these analyses are available upon request.

Discussion

This study provides several insights into the performance of ASD symptom measures when used with children with differing levels of cognitive ability and emotional/behavioral problems (EBP). To our knowledge, this is the first study to examine the influence of individual characteristics on several ASD measures based on multiple sources of information, and to use a control group with non-ASD neurodevelopmental disorders who had undergone careful assessment to rule out ASD. The results indicate that the discriminative abilities of three commonly used ASD measures (the SRS, ADI-R, and ADOS) are substantially affected by individual characteristics beyond ASD symptoms. Among children without ASD, lower NVIQ and more parent-rated EBP were significantly associated with elevated scores on the SRS and ADI-R, while clinician-observed EBP were significantly associated with elevated scores on the ADOS. The established instrument cutoffs for ASD showed low specificity among children with intellectual disability (SRS, ADI-R), elevated parent-rated EBP (SRS, ADI-R), and elevated clinician-observed EBP (SRS, ADI-R, ADOS).

We found that NVIQ was less associated with the discriminative threshold of the ADOS than of the SRS or ADI-R. This could be partially explained by the fact that ADOS administration and coding procedures are specifically designed to take developmental level into account (i.e., through the use of a Module system).¹¹ In line with previous studies,^{12, 46, 47} parent-rated EBP affected the threshold validity of parental report-based ASD measures (SRS, ADI-R), but not the clinician observation-based ADOS. The present study extends these findings by demonstrating that clinician-observed EBP significantly influenced discriminative thresholds on the ADOS. This highlights the importance of considering the context and informant in assessment of relationships between ASD symptoms and other psychiatric symptoms or behavior problems. Further study using an independent measure of clinician-observed EBP is needed in order to clarify their influence on the validity of the ADOS and other ASD symptom measures.

For all three ASD measures, adjustment for the influence of NVIQ and EBP on the discriminative threshold resulted in stable discriminative accuracy among children with differing levels of cognitive ability and behavior problems. For example, the ASD measures provided similar discriminative ability for ASD among children with low versus elevated EBP (Figure 1). However, much weaker differentiation was found when comparing children with ASD and low EBP to children without ASD and elevated EBP. Notably, SRS scores did not distinguish between these subgroups. These findings illustrate the potential danger of making unadjusted comparisons of ASD symptom scores between children who vary with respect to EBP or IQ. The results also lend support to earlier recommendations that reasonable steps could and should be taken to facilitate more valid interpretation of scores on ASD measures,^{10, 12, 48} such as using scores and thresholds standardized according to IQ or EBP, or applying statistical adjustment.

The use of scores from ASD measures without adequately considering the influence of other child-level variables remains widespread. Studies of genetic syndromes often report the proportions of children meeting ASD instrument cutoffs without taking IQ or EBP into account.^{4, 6, 16–18} Unadjusted comparisons of ASD symptom scores are also commonly made between children with ASD and children with other psychiatric disorders or genetic syndromes who differ significantly on these variables.^{4, 5, 19} Furthermore, many studies assessing the relationships between ASD symptoms (or autistic traits) and other behavioral, genetic, and neurobiological variables do not carefully take into account phenotypic characteristics, such as age, IQ, or EBP, as potential confounding factors.^{20–22} These practices have potentially very serious consequences, as they increase the likelihood of drawing erroneous conclusions about associations between ASD symptoms and etiological mechanisms.¹²

Another implication of these findings is that when large-scale research studies of ASD (e.g., Simons Simplex Collection, SSC) base inclusion/exclusion criteria on classifications from ASD measures, research samples could become biased toward children who are more likely to meet standardized instrument cut-offs (e.g., those with lower IQ and/or more emotional/behavioral problems). This could potentially exaggerate differences between research samples and clinical or general population samples of children with ASD. In the current sample, a post-hoc linear regression analysis showed that meeting SSC inclusion criteria was

significantly associated with having more parent-rated internalizing problems ($p < 0.01$) and clinician-observed emotional/behavioral problems ($p < 0.01$) (parent-rated externalizing problems $p = 0.07$; IQ $p = 0.11$).

The results of this study should be interpreted in light of relevant methodological factors. The influence of IQ and EBP on the discriminative ability of ASD symptom measures was examined in the context of clinical best-estimate diagnosis. Thus, although total/algorithm scores from the SRS, ADI-R and ADOS were not calculated, information elicited from these instruments (and all other information from the assessment) was considered in the diagnostic decision making process. On the other hand, keeping clinicians blind to other diagnostic information while they were rating the ADI-R and ADOS (including assigning separate clinicians to administer these measures whenever possible), ensuring that clinicians were blind to ASD instrument total scores and classifications while making best-estimate diagnoses, and continuously checking and maintaining reliability on the instruments within and across sites throughout the study period, are significant methodological strengths.

The sample consisted of 2–13 year old children who had been recruited based on identified ASD or another neurodevelopmental disorder. The parents of participants received compensation for their time and a report from a comprehensive diagnostic evaluation that was free of charge, and this could have contributed to self-selection bias. Future studies should examine the generalizability of these findings to children of different ages or those recruited in other ways (e.g., epidemiologically ascertained or exclusively clinic referred samples). Given the small number of non-ASD participants with intellectual disabilities and mood and anxiety disorders, replication is particularly warranted in these groups. Importantly, differences in the score ranges of the ASD and emotional/behavioral problem measures could have contributed to variable power to detect associations with child characteristics. It is also likely that there are reciprocal relationships between ASD symptoms and impairments of intellectual, emotional and behavioral functioning; longitudinal studies will be particularly useful in elucidating these relationships across development.

Taken together, results of this study indicate that ASD symptom measures capture far more than symptoms of ASD. Elevated scores on these measures can reflect impairments in behavioral dimensions other than the social-communication deficits and repetitive behaviors they were designed to capture. The findings underscore the need to carefully consider factors not specific to ASD, including IQ and behavior problems, in analyses of ASD symptom measures in clinical and research settings. Overlooking these factors in genetic and neurobiological research studies places the field at risk for drawing incorrect conclusions about etiological and phenotypic relationships relevant to ASD and other neurodevelopmental disorders. As ASD symptom measures are adopted for wider use and applied to the study of populations not necessarily represented in the original validation samples, it is essential that clinicians and researchers take appropriate precautions to ensure valid interpretation of scores.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to all of the participating families, and to the clinical and research staff contributing to the study. The project was supported by grants from the South-Eastern Norway Regional Health Authority (2012101 to Øyen, fellow: Havdahl), the National Institute of Child Health and Human Development (R01HD065277 to Bishop), and the National Institute of Mental Health (RC1MH089721 and R01MH081873-01A1 to Lord).

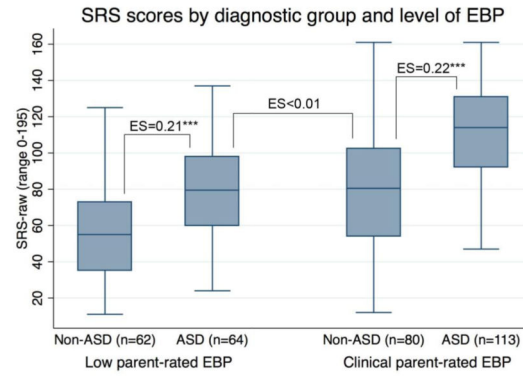
References

1. Molloy CA, Murray DS, Akers R, Mitchell T, Manning-Courtney P. Use of the Autism Diagnostic Observation Schedule (ADOS) in a clinical setting. *Autism*. 2011; 15(2):143–162. [PubMed: 21339248]
2. Zander E, Sturm H, Bolte S. The added value of the combined use of the Autism Diagnostic Interview-Revised and the Autism Diagnostic Observation Schedule: Diagnostic validity in a clinical Swedish sample of toddlers and young preschoolers. *Autism*. 2015; 19(2):187–199. [PubMed: 24413849]
3. Volkmar F, Siegel M, Woodbury-Smith M, et al. Practice parameter for the assessment and treatment of children and adolescents with autism spectrum disorder. *J Am Acad Child Adolesc Psychiatry*. 2014; 53(2):237–257. [PubMed: 24472258]
4. Abbeduto L, McDuffie A, Thurman AJ. The fragile x syndrome-autism comorbidity: What do we really know? *Front Genet*. 2014; 5:355. [PubMed: 25360144]
5. Richards C, Jones C, Groves L, Moss J, Oliver C. Prevalence of autism spectrum disorder phenomenology in genetic disorders: A systematic review and meta-analysis. *Lancet Psychiatry*. 2015; 2(10):909–916. [PubMed: 26341300]
6. Polyak A, Kubina RM, Girirajan S. Comorbidity of intellectual disability confounds ascertainment of autism: Implications for genetic diagnosis. *Am J Med Genet B Neuropsychiatr Genet*. 2015; 168(7):600–608. [PubMed: 26198689]
7. Grzadzinski R, Dick C, Lord C, Bishop S. Parent-reported and clinician-observed autism spectrum disorder (ASD) symptoms in children with attention deficit/hyperactivity disorder (ADHD): Implications for practice under DSM-5. *Mol Autism*. 2016; 7:7. [PubMed: 26788284]
8. McDuffie A, Thurman AJ, Hagerman RJ, Abbeduto L. Symptoms of autism in males with fragile X syndrome: A comparison to nonsyndromic ASD using current ADI-R scores. *J Autism Dev Disord*. 2015; 45(7):1925–1937. [PubMed: 24414079]
9. Charman T, Baird G, Simonoff E, et al. Efficacy of three screening instruments in the identification of autistic-spectrum disorders. *Br J Psychiatry*. 2007; 191(6):554–559. [PubMed: 18055961]
10. Frazier TW, Youngstrom EA, Embacher R, et al. Demographic and clinical correlates of autism symptom domains and autism spectrum diagnosis. *Autism*. 2014; 18(5):571–582. [PubMed: 24104512]
11. Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J Autism Dev Disord*. 2009; 39(5):693–705. [PubMed: 19082876]
12. Hus V, Bishop S, Gotham K, Huerta M, Lord C. Factors influencing scores on the Social Responsiveness Scale. *J Child Psychol Psychiatry*. 2013; 54(2):216–224. [PubMed: 22823182]
13. Warren Z, Vehorn A, Dohrmann E, Nicholson A, Sutcliffe JS, Veenstra-Vanderweele J. Accuracy of phenotyping children with autism based on parent report: What specifically do we gain phenotyping “rapidly”? *Aut Res*. 2012; 5(1):31–38.
14. Falkmer T, Anderson K, Falkmer M, Horlin C. Diagnostic procedures in autism spectrum disorders: A systematic literature review. *Eur Child Adolesc Psychiatry*. 2013; 22(6):329–340. [PubMed: 23322184]

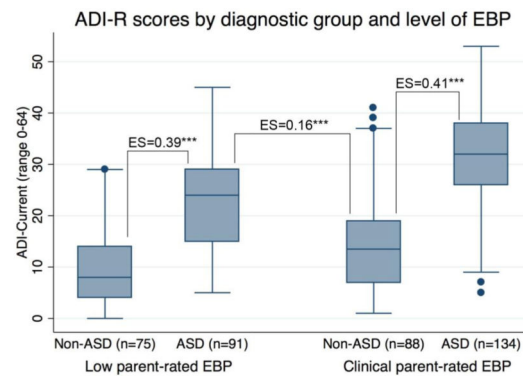
15. Zwaigenbaum L, Bryson S, Lord C, et al. Clinical assessment and management of toddlers with suspected autism spectrum disorder: Insights from studies of high-risk infants. *Pediatrics*. 2009; 123(5):1383–1391. [PubMed: 19403506]
16. Constantino JN, Zhang Y, Holzhauser K, et al. Distribution and within-family specificity of quantitative autistic traits in patients with neurofibromatosis type I. *J Pediatr*. 2015; 167(3):621–626. [PubMed: 26051969]
17. Sheth K, Moss J, Hyland S, Stinton C, Cole T, Oliver C. The behavioral characteristics of Sotos syndrome. *Am J Med Genet A*. 2015; 167(12):2945–2956.
18. Garg S, Lehtonen A, Huson SM, et al. Autism and other psychiatric comorbidity in neurofibromatosis type 1: Evidence from a population-based study. *Dev Med Child Neurol*. 2013; 55(2):139–145. [PubMed: 23163236]
19. Moss J, Howlin P, Magiati I, Oliver C. Characteristics of autism spectrum disorder in Cornelia de Lange syndrome. *J Child Psychol Psychiatry*. 2012; 53(8):883–891. [PubMed: 22490014]
20. Coon H, Villalobos ME, Robison RJ, et al. Genome-wide linkage using the Social Responsiveness Scale in Utah autism pedigrees. *Mol Autism*. 2010; 1(1):8. [PubMed: 20678250]
21. Duvall JA, Lu A, Cantor RM, Todd RD, Constantino JN, Geschwind DH. A quantitative trait locus analysis of social responsiveness in multiplex autism families. *Am J Psychiatry*. 2007; 164(4):656–662. [PubMed: 17403980]
22. Harrison AJ, Gamsiz ED, Berkowitz IC, Nagpal S, Jerskey BA. Genetic variation in the oxytocin receptor gene is associated with a social phenotype in autism spectrum disorders. *Am J Med Genet B Neuropsychiatr Genet*. 2015; 168(8):720–729. [PubMed: 26365303]
23. Thurm A, Tierney E, Farmer C, et al. Development, behavior, and biomarker characterization of Smith-Lemli-Opitz syndrome: An update. *J Neurodev Disord*. 2016; 8:12. [PubMed: 27053961]
24. Bishop SL, Havdahl KA, Huerta M, Lord C. Subdimensions of social-communication impairment in autism spectrum disorder. *J Child Psychol Psychiatry*. 2016 [Epub ahead of print].
25. Constantino, JN.; Gruber, CP. *Social Responsiveness Scale (SRS)*. Los Angeles, CA: Western Psychological Services; 2005.
26. Constantino, JN.; Gruber, CP. *Social Responsiveness Scale. Second*. Los Angeles, CA: Western Psychological Services; 2012.
27. Rutter, M.; Le Couteur, A.; Lord, C. *Autism Diagnostic Interview-Revised (ADI-R)*. Los Angeles, California: Western Psychological Services; 2003.
28. Lord C, Risi S, Lambrecht L, et al. The autism diagnostic observation schedule-generic: a standard measure of social and communication deficits associated with the spectrum of autism. *J Autism Dev Disord*. 2000; 30(3):205–223. [PubMed: 11055457]
29. Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S.; Gotham, K.; Bishop, SL. *Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Modules 1–4*. Los Angeles, California: Western Psychological Services; 2012.
30. Bishop SL, Huerta M, Gotham K, et al. The autism symptom interview, school-age: A brief telephone interview to identify autism spectrum disorders in 5-to-12-year-old children. *Autism Res*. 2016
31. Lord C, Bishop S, Anderson D. Developmental trajectories as autism phenotypes. *Am J Med Genet C Semin Med Genet*. 2015; 169(2):198–208. [PubMed: 25959391]
32. Risi S, Lord C, Gotham K, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. 2006; 45(9):1094–1103. [PubMed: 16926617]
33. Achenbach, TM.; Rescorla, LA. *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth & Families; 2000.
34. Achenbach, TM.; Rescorla, LA. *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth & Families; 2001.
35. Elliott, CD. *Differential Ability Scales - Second Edition*. New York: Harcourt Brace Jovanovich; 2007.
36. Bishop SL, Guthrie W, Coffing M, Lord C. Convergent validity of the Mullen Scales of Early Learning and the Differential Ability Scales in children with autism spectrum disorders. *Am J Intellect Dev Disabil*. 2011; 116(5):331–343. [PubMed: 21905802]

37. Mullen, EM. Mullen scales of early learning. Circle Pines, MN: American Guidance Service; 1995.
38. Sparrow, SS.; Cicchetti, VD.; Balla, AD. Vineland adaptive behavior scales. 2nd. Circle Pines, MN: American Guidance Service; 2005.
39. Kim SH, Macari S, Koller J, Chawarska K. Examining the phenotypic heterogeneity of early Autism Spectrum Disorder: subtypes and short-term outcomes. *J Child Psychol Psychiatry*. 2015 [Epub ahead of print].
40. van der Ende J, Verhulst FC, Tiemeier H. Agreement of informants on emotional and behavioral problems from childhood to adulthood. *Psychol Assess*. 2012; 24(2):293–300. [PubMed: 21928910]
41. Charman T, Gotham K. Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders - Lessons from research and practice. *Child Adolesc Ment Health*. 2013; 18(1): 52–63. [PubMed: 23539140]
42. Frazier TW, Georgiades S, Bishop SL, Hardan AY. Behavioral and cognitive characteristics of females and males with autism in the Simons Simplex Collection. *J Am Acad Child Adolesc Psychiatry*. 2014; 53(3):329–340. [PubMed: 24565360]
43. Janes H, Longton G, Pepe M. Accommodating covariates in ROC analysis. *The Stata Journal*. 2009; 9(1):17–39. [PubMed: 20046933]
44. Janes H, Pepe MS. Adjusting for covariates in studies of diagnostic, screening, or prognostic markers: an old concept in a new setting. *Am J Epidemiol*. 2008; 168(1):89–97. [PubMed: 18477651]
45. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol*. 1995; 57(1):289–300.
46. Bal VH, Lord C. Replication of standardized ados domain scores in the Simons Simplex Collection. *Aut Res*. 2015; 8(5):583–592.
47. Sprenger L, Buhler E, Poustka L, et al. Impact of ADHD symptoms on autism spectrum disorder symptom severity. *Res Dev Disabil*. 2013; 34(10):3545–3552. [PubMed: 23973801]
48. Havdahl KA, von Tetzchner S, Huerta M, Lord C, Bishop SL. Utility of the Child Behavior Checklist as a screener for autism spectrum disorder. *Aut Res*. 2016; 9(1):33–42.

A:



B:



C:

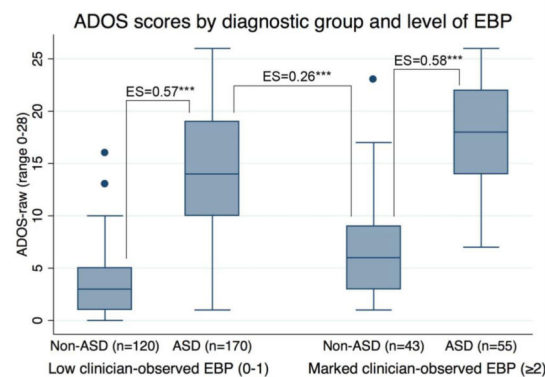


Figure 1.

(A, B, C): Box-plot diagrams showing the distributions of autism symptom scores for children with and without autism spectrum disorder (ASD) by level of emotional/behavioral problems (EBP), and effect sizes of the mean differences between diagnostic groups (non-ASD versus ASD).

Note: ASD=autism spectrum disorder, SRS=Social Responsiveness Scale, ADI=Autism Diagnostic Interview-Revised, ADOS=Autism Diagnostic Observation Schedule, Parent-rated EBP=Child Behavior Checklist Externalizing or Internalizing scale in clinical range

(T 64). Clinician-observed EBP=ADOS Other Abnormal Behaviors section. ES=effect size (partial eta squared), * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Sample characteristics [range]

	ASD (n=225)		Other diagnoses (n=163)	
Male, n (%)	175 (77.78)		111 (68.1)	
Female, n (%)	50 (22.22)		52 (31.9)	
Age in years, m (SD)	7.04 (3.06)	[2–13]	7.57 (2.87)	[2–13]
Maternal education ^a : Some college, n (%)	179 (87.32)		129 (86)	
White race ^a , n (%)	168 (74.67)		96 (59.26)	
Language level (ADOS module)				
Nonverbal or Single words, n (%)	69 (30.67)		14 (8.59)	
Phrase-speech, n (%)	37 (16.44)		33 (20.25)	
Complex sentences, n (%)	119 (52.89)		116 (71.17)	
Verbal IQ, m (SD)	75.72 (31.39)	[10–139]	91.66 (22.02)	[14–157]
Nonverbal IQ, m (SD)	82.25 (27.45)	[12–141]	93.09 (20.21)	[19–150]
CBCL Internalizing T, m (SD)	62.5 (10.6)	[33–87]	59.06 (12.03)	[29–84]
Clinical range, n (%)	117 (52)		68 (41.72)	
CBCL Externalizing T, m (SD)	60.78 (12.32)	[33–97]	59.09 (12.38)	[28–85]
Clinical range, n (%)	88 (39.11)		54 (33.13)	
EBP-ADOS total, m (SD)	1.02 (1.02)	[0–4]	0.99 (0.94)	[0–4]
Marked (≥2), n (%)	55 (24.44)		43 (26.38)	
Medication use ^a , n (%)	50 (23.92)		50 (34.72)	
Vineland adaptive composite, n (%)	76.41 (14.48)	[38–117]	86.14 (13.34)	[43–118]
SRS raw (age 4 years) ^b , m (SD)	100.54 (29.9)	[24–161]	69.2 (31.88)	[11–161]
SRS T (age 4 years) ^b , m (SD)	77.52 (12.49)	[46–105]	65.1 (12.96)	[41–101]
ADI-current, m (SD)	27.77 (10.47)	[5–53]	12.35 (8.93)	[0–41]
ADI-diagnostic, m (SD)	36.9 (12.39)	[9–64]	17.61 (11.53)	[1–57]
ADOS raw, m (SD)	15.11 (5.41)	[1–26]	4.52 (3.76)	[0–23]
ADOS CSS, m (SD)	7.25 (2.02)	[1–10]	2.56 (1.97)	[1–10]

Note: ASD=autism spectrum disorder, CBCL=Child Behavior Checklist, SRS=Social Responsiveness Scale, ADI=Autism Diagnostic Interview-Revised, ADOS=Autism Diagnostic Observation Schedule, CSS=calibrated severity scores, standardized by age and language level. EBP-ADOS=clinician-observed emotional/behavioral problems,

^aMissing: maternal education n=33, race n=1, medication use n=35.

^bSRS: n=177 ASD, Other diagnoses n=142.

Table 2 Results of the multiple linear regression analyses using child characteristics as predictors of ASD instrument thresholds (SRS, ADI-R and ADOS)

Child characteristics	SRS Raw		ADI-R Current		ADOS Raw	
	B	(95% CI)	B	(95% CI)	B	(95% CI)
IQ (NVIQ)	-0.50***	(-0.68,-0.31)	-0.08*	(-0.14,-0.01)	-0.03	(-0.06,0.01)
Parent-rated EBP						
CBCL-Internalizing	1.08***	(0.68,1.48)	0.22***	(0.09,0.34)	-0.05	(-0.11,0.01)
CBCL-Externalizing	1.00***	(0.62,1.38)	0.13*	(0.02,0.25)	0.02	(-0.03,0.07)
Clinician-observed EBP ^a						
	-0.05	(-4.85,4.75)	0.53	(-0.99,2.05)	0.89*	(0.17,1.62)
Sex (female)	-6.17	(-14.21,1.88)	-1.53	(-3.92,0.86)	-0.90	(-2.02,0.23)
Age (years)	-0.43	(-1.96,1.11)	-1.02***	(-1.49,-0.55)	-0.19	(-0.41,0.03)
Constant	75.58***	(70.76,80.40)	14.27***	(12.62,15.93)	4.99***	(4.27,5.70)
R ²		0.53		0.29		0.16
n (all non-ASD)		142		163		163

Note: NVIQ=nonverbal IQ, CBCL=Child Behavior Checklist, EBP=emotional/behavioral problems, SRS=Social Responsiveness Scale, ADI-R=Autism Diagnostic Interview-Revised, ADOS=Autism Diagnostic Observation Schedule.

^a ADOS “Other Abnormal Behaviors” section total. The SRS was included for children aged 4y.

* p<0.05,

** p<0.01,

*** p<0.001 (raw p values).

Results of the ROC regression analyses using child characteristics as predictors of discriminative ability when adjusting for their influence on instrument thresholds

Table 3

Child characteristics	SRS Raw		ADI-R Current		ADOS Raw	
	B	(95% CI)	B	(95% CI)	B	(95% CI)
IQ (NVIQ)	0.01	(-0.00,0.02)	<0.01	(-0.01,0.01)	-0.01	(-0.02,0.01)
Parent-rated EBP						
CBCL-Internalizing	0.01	(-0.02,0.04)	0.03	(-0.00,0.06)	<0.01	(-0.05,0.04)
CBCL-Externalizing	-0.02	(-0.04,0.01)	-0.01	(-0.04,0.02)	<0.01	(-0.04,0.03)
Clinician-observed EBP ^a	-0.03	(-0.33,0.28)	-0.02	(-0.27,0.22)	0.12	(-0.15,0.39)
Sex (female)						
	0.62	(-0.04,1.28)	0.06	(-0.50,0.62)	-0.26	(-0.93,0.42)
Age (years)						
	0.10	(-0.02,0.22)	0.12	(0.04,0.21)	0.04	(-0.07,0.15)
Constant - Intercept						
	0.85	(0.51,1.19)	1.48	(1.11,1.85)	2.71	(1.92,3.51)
Constant - Slope						
	1.10	(0.84,1.35)	0.98	(0.71,1.24)	1.26	(0.68,1.84)
n Non-ASD		142		163		163
n ASD		177		225		225

Note: NVIQ=nonverbal IQ, CBCL=Child Behavior Checklist, EBP=emotional/behavioral problems, SRS=Social Responsiveness Scale, ADI-R=Autism Diagnostic Interview-Revised, ADOS=Autism Diagnostic Observation Schedule.

^a ADOS “Other Abnormal Behaviors” section total. The SRS was included for children aged 4y.

* p<0.05,

** p<0.01,

*** p<0.001 (raw p values).

Discriminative ability of the ASD instruments by representative levels of intellectual ability and emotional/behavioral problems (95% confidence intervals)

Table 4

Social Responsiveness Scale (SRS)						
Stratification	n		AUC	Moderate cut-off (T = 66)		
	ASD	Other		SRS-raw	Sensitivity,%	Specificity,%
Total sample	177	142	0.76 (0.71,0.81)	84 (78,89)	58 (49,66)	
Intellectual disability	Yes	26	0.73 (0.60,0.85)	93 (82,98)	27 (12,48)	
	No	122	0.76 (0.70,0.82)	80 (72,87)	65 (55,73)	
CBCL-Internalizing: Clinical	Yes	96	0.74 (0.66,0.82)	96 (90,99)	36 (24,49)	
	No	81	0.78 (0.71,0.85)	70 (59,80)	75 (64,84)	
CBCL-Externalizing: Clinical	Yes	73	0.74 (0.66,0.83)	96 (89,99)	39 (25,54)	
	No	104	0.78 (0.71,0.84)	76 (67,84)	68 (57,77)	
EBP-ADOS: Marked	Yes	41	0.72 (0.61,0.84)	93 (80,99)	41 (25,59)	
	No	136	0.77 (0.71,0.83)	82 (74,88)	63 (53,72)	
Autism Diagnostic Interview-Revised (ADI-R)						
Stratification	n		AUC	ADI-R ASD criteria		
	ASD	Other		ADI-current	Sensitivity,%	Specificity,%
Total sample	225	163	0.86 (0.82,0.89)	86 (81,90)	66 (58,74)	
Intellectual disability	Yes	72	0.85 (0.77,0.94)	89 (79,95)	42 (23,63)	
	No	153	0.85 (0.81,0.89)	84 (78,90)	71 (62,78)	
CBCL-Internalizing: Clinical	Yes	117	0.86 (0.80,0.92)	94 (88,98)	59 (46,71)	
	No	108	0.85 (0.80,0.90)	77 (68,84)	72 (61,80)	
CBCL-Externalizing: Clinical	Yes	88	0.85 (0.78,0.92)	92 (84,97)	56 (41,69)	
	No	137	0.86 (0.82,0.91)	82 (74,88)	72 (62,80)	
EBP-ADOS: Marked	Yes	55	0.80 (0.71,0.88)	91 (80,97)	51 (36,67)	
	No	170	0.88 (0.84,0.92)	84 (78,89)	72 (63,80)	
Autism Diagnostic Observation Schedule (ADOS)						
Stratification	n		AUC	ADOS ASD criteria		
	ASD	Other		ADOS-raw	Sensitivity,%	Specificity,%
Total sample	177	142	0.86 (0.82,0.89)	86 (81,90)	66 (58,74)	
Intellectual disability	Yes	26	0.85 (0.77,0.94)	89 (79,95)	42 (23,63)	
	No	153	0.85 (0.81,0.89)	84 (78,90)	71 (62,78)	
CBCL-Internalizing: Clinical	Yes	117	0.86 (0.80,0.92)	94 (88,98)	59 (46,71)	
	No	108	0.85 (0.80,0.90)	77 (68,84)	72 (61,80)	
CBCL-Externalizing: Clinical	Yes	88	0.85 (0.78,0.92)	92 (84,97)	56 (41,69)	
	No	137	0.86 (0.82,0.91)	82 (74,88)	72 (62,80)	
EBP-ADOS: Marked	Yes	55	0.80 (0.71,0.88)	91 (80,97)	51 (36,67)	
	No	170	0.88 (0.84,0.92)	84 (78,89)	72 (63,80)	

Social Responsiveness Scale (SRS)						
		225	163	0.93 (0.91,0.96)	95 (91,97)	76 (69,82)
Total sample						
Intellectual disability	Yes	72	26	0.96 (0.93,0.99)	96 (88,99)	77 (56,91)
	No	153	137	0.92 (0.89,0.95)	94 (89,97)	76 (68,83)
CBCL-Internalizing: Clinical	Yes	117	68	0.93 (0.89,0.96)	93 (87,97)	78 (66,87)
	No	108	95	0.95 (0.92,0.98)	96 (91,99)	75 (65,83)
CBCL-Externalizing: Clinical	Yes	88	54	0.93 (0.89,0.97)	91 (83,96)	76 (62,87)
	No	137	109	0.94 (0.91,0.97)	97 (93,99)	76 (67,84)
EBP-ADOS: Marked	Yes	55	43	0.94 (0.89,0.99)	100 (94,100)	56 (40,71)
	No	170	120	0.95 (0.92,0.97)	93 (88,96)	83 (75,90)

Note: AUC=area under the curve, SRS=Social Responsiveness Scale, ADI-R=Autism Diagnostic Interview-Revised, ADOS=Autism Diagnostic Observation Schedule, CBCL=Child Behavior Checklist. See Table S3 for sensitivity/specificity of alternative cut-offs.