Published in final edited form as:

J Autism Dev Disord. 2015 December ; 45(12): 3919–3931. doi:10.1007/s10803-015-2586-3.

Replication and Comparison of the Newly Proposed ADOS-2, Module 4 Algorithm in ASD without ID: A Multi-site Study

Cara E. Pugliese¹, Lauren Kenworthy¹, Vanessa Hus Bal², Gregory L Wallace³, Benjamin E Yerys^{4,5}, Brenna B. Maddox^{4,6}, Susan W. White⁵, Haroon Popal⁷, Anna Chelsea Armour¹, Judith Miller^{4,5}, John D. Herrington^{4,5}, Robert T. Schultz^{4,5}, Alex Martin⁷, and Laura Gutermuth Anthony¹

¹Children's National Medical Center, Rockville, Maryland ²UCSF School of Medicine, San Francisco, CA ³George Washington University, Washington, DC ⁴Children's Hospital of Philadelphia, Philadelphia, PA ⁵Perelman School of Medicine - University of Pennsylvania, Philadelphia, PA ⁶Virginia Tech, Blacksburg, VA ⁷National Institute of Mental Health

Abstract

Recent updates have been proposed to the Autism Diagnostic Observation Schedule-2 Module 4 diagnostic algorithm. This new algorithm, however, has not yet been validated in an independent sample without intellectual disability (ID). This multi-site study compared the original and revised algorithms in individuals with ASD without ID. The revised algorithm demonstrated increased sensitivity, but lower specificity in the overall sample. Estimates were highest for females, individuals with a verbal IQ below 85 or above 115, and ages 16 and older. Best practice diagnostic procedures should include the Module 4 in conjunction with other assessment tools. Balancing needs for sensitivity and specificity depending on the purpose of assessment (e.g., clinical vs. research) and demographic characteristics mentioned above will enhance its utility.

Correspondence concerning this article should be addressed to Cara Pugliese, Center for Autism Spectrum Disorders, Children's National Health System, 15245 Shady Grove Road #350, Rockville, MD 20850. CPugliese@ChildrensNational.org, Phone: 1.301.765.5424.

Cara E. Pugliese, Center for Autism Spectrum Disorders, Children's Research Institute, Washington, DC; Lauren Kenworthy, Center for Autism Spectrum Disorders, Children's Research Institute, Washington, DC; Vanessa Hus Bal, Department of Psychiatry, University of California San Francisco, San Francisco, CA; Gregory L. Wallace, Department of Speech and Hearing Sciences, George Washington University, Washington, DC; Benjamin Yerys, Children's Hospital of Philadelphia, Philadelphia, PA, Perelman School of Medicine – University of Pennsylvania, Philadelphia, PA; Brenna B. Maddox, Department of Child and Adolescent Psychiatry and Behavioral Sciences, Children's Hospital of Philadelphia, Philadelphia, PA, and Department of Psychology, Virginia Tech, Blacksburg, VA; Susan W. White, Department of Psychology, Virginia Tech, Blacksburg, VA; Haroon Popal, Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD; Anna Chelsea Armour, Center for Autism Spectrum Disorders, Children's Research Institute, Washington, DC; Judith Miller, Children's Hospital of Philadelphia, Philadelphia, PA, Perelman School of Medicine – University of Pennsylvania, Philadelphia, PA; John Herrington, Children's Hospital of Philadelphia, Philadelphia, PA, Perelman School of Medicine – University of Pennsylvania, Philadelphia, PA; Robert Schultz, Children's Hospital of Philadelphia, Philadelphia, PA, Perelman School of Medicine – University of Pennsylvania, Philadelphia, PA; Alex Martin, Laboratory of Brain and Cognition, National Institute of Mental Health, Bethesda, MD; Laura Gutermuth Anthony, Center for Autism Spectrum Disorders, Children's Research Institute, Washington, DC.

Disclosure of potential conflicts: There are no financial or non-financial conflicts of interest to declare.

Compliance with Ethical Standards: This project involved human subjects and was conducted in compliance with standards established by all institution's IRB including procedures for informed consent. This research did not involve the use of animals.

Keywords

autism spectrum disorder; adults; adolescents; diagnosis; autism diagnostic observation schedule; sensitivity; specificity

Of the increasing number of individuals identified with an Autism Spectrum Disorder (ASD), the fastest growing subgroup is those without co-occurring intellectual disability (ID; Baio, 2014). To receive a diagnosis of ASD, symptoms must be present from early childhood, though the disorder may not be detected until later in life. Later diagnosis may be due to the presence of milder ASD symptoms, the individual's ability to compensate for limitations through high intellectual ability, or the clinical difficulty disentangling ASD symptoms from other psychiatric disorders (Bastiaansen et al., 2011). Core difficulties of ASD generally persist into adulthood (Seltzer et al., 2003; Shattuck et al., 2007; Taylor & Seltzer, 2010), and the majority of individuals with ASD will require varying levels of life-long supports (Hus & Lord, 2014). Even for those without ID, ASD often has negative impacts on learning and independence in adulthood (Farley et al., 2009; Howlin, 2013; Pugliese et al., 2014). Thus, it is important to understand ASD symptoms in adults with ASD without ID in order for accurate diagnosis and subsequent supports to promote positive outcomes (Interagency Autism Coordinating Committee, 2011). The purpose of this paper is to understand the performance of the Autism Diagnostic Observation Schedule-2 (ADOS-2) Module 4 as a diagnostic tool and measure of symptom severity in an independent sample of adults with ASD without ID.

Despite a growing demand, there is currently no established tradition for diagnosing ASD in adulthood, particularly because developmental information from a parent-informant may not always be available or accurately reported upon (Volkmar, Langford Booth, McPartland, & Wiesner, 2014). Because diminished self-awareness and social understanding may impair an individual's self-report of symptoms, observation of the individual during social interaction is important (Bishop & Seltzer, 2012). The ADOS (Lord et al., 2000) is one of the most widely used diagnostic observation instruments in the field of autism research. The most recent edition, the ADOS-2 (Lord et al., 2012) includes five modules suited for individuals with different developmental and language levels. It is a reliable and valid instrument to assess the presence of ASD in children (de Bildt et al., 2004; Gray, Tonge, & Sweeney, 2008; Lord et al., 2000; Papanikolaou et al., 2009), though Module 4, developed for adolescents and adults with fluent language, has received less psychometric evaluation.

In the original version of the ADOS Module 4 (ADOS-G; Lord et al., 2000), the normative group included a small Non-ASD group with a wide range of diagnoses and cognitive abilities, making specificity estimates less generalizable to the larger population. In the initial validation study (Lord et al., 2000), the Module 4 algorithm demonstrated an overall sensitivity of 90.0% and specificity of 93.0% across cognitive ability levels for participants with autism and PDD-NOS compared to non-spectrum individuals. Replication in a small independent study by Bastiaansen and colleagues (2011) comprised exclusively of adults without ID, however, indicated the ADOS Module 4 was only able to correctly classify 74.2% of the cases in their sample as having ASD or not based on clinical DSM-IV-TR

(American Psychiatric Association, 2000) diagnosis. Because specificity was high (82%) and sensitivity was quite low (61%) using the original cut-offs, Bastiaansen and colleagues (2011) cautioned that the ADOS Module 4 failed to classify ASD in a significant proportion of their sample. Specifically, the authors highlighted the need for an updated algorithm consistent with the newly updated DSM-5 (American Psychiatric Association, 2013) criteria for ASD, particularly to reflect both domains of Social Affect and Restricted Repetitive Behavior.

Hus and Lord (2014) proposed an updated algorithm for the ADOS-2 Module 4 to align with the DSM-5 two-domain symptom partition and to increase its consistency with the algorithms from other ADOS modules. Items were selected based on those included in the algorithms for other ADOS modules (for uniformity) and their ability to distinguish individuals with DSM-IV clinical diagnoses of Autism from those with Non-ASD diagnoses. In this algorithm, the Social Affect domain includes the following items: *Unusual Eye Contact, Amount of Social Communication, Facial Expressions, Quality of Rapport, Communication of Own Affect, Quality of Social Overtures, Conversation, Emphatic Gestures, Quality of Social Response, and Insight*. The Restricted Repetitive Behaviors Domain consisted of: *Speech Abnormalities Associated with Autism, Stereotyped Language, Unusual Sensory Interests, Highly Specific Topics, and Hand Mannerisms*. Cut-off scores of 8 and above on this new algorithm are indicative of an instrument classification of ASD, per Hus and Lord (2014).

In a sample of 437 cases (393 participants) aged 9-62 years with and without ID, the newly proposed algorithm had an overall sensitivity of 90.5% (ranging from 87.1-95% across cognitive ability levels) and specificity of 82.2% (ranging from 77.3-92.9% across cognitive ability levels; Hus & Lord, 2014). Interestingly, while the sensitivity of this algorithm increased from the original algorithm, the specificity decreased for those with average cognitive abilities. However, because the Non-ASD sample with average IQs was quite small (N=44), it is difficult to interpret whether this decrease (which reflects 2 participants incorrectly classified by the new algorithm) is meaningful. Given the growing proportion of individuals with ASD without ID, there is a clear need to further investigate the diagnostic validity of the newly proposed algorithm in independent samples and determine how it functions across IQ, gender, and age.

The need for tools that can assess ASD symptom severity on a continuous scale across development periods is well established (Gotham, Pickles, & Lord, 2009). Along with the newly proposed algorithm, Hus and Lord (2014) developed the Calibrated Severity Metric (CSS) for the Module 4 to address this need. The CSS is a standardized version of the ADOS-2 score that is intended to capture severity of core ASD symptoms and can be used to track progression of ASD symptoms across time as well as across different modules of the ADOS-2. As such, Hus and Lord (2014) have indicated the CSS should be correlated with other measures of ASD symptom severity, but not with age or IQ. The authors caution that the CSS should not be used as an indicator of functional impairment and should not take the place of thorough assessment of language, adaptive functioning, and other psychiatric features to determine functional impairment. To our knowledge, independent studies have not yet examined patterns of association between the CSS and these variables.

The purpose of the present study is to 1) examine item level discrimination and internal consistency of the new (2014) Module 4 algorithm in an independent sample, 2) replicate the Hus and Lord (2014) findings using a matched subsample and extend existing knowledge by exploring the 2014 algorithm across subgroups of participants that differ demographically (e.g., gender, age, IQ), and 3) investigate the utility of the CSS of Module 4 as an unbiased measure of symptom severity through associations with demographic variables as well as other measures of ASD symptom severity and adaptive functioning in an independent sample. In order to accurately compare psychometrics of the newly proposed algorithm in individuals without ID, as well as to explore item level differences not previously reported, participants without ID from Hus and Lord's (2014; $n=307$) study were used as a comparison sample. We hypothesized that:

1. The 2014 algorithm items would discriminate the ASD and Non-ASD samples effectively and algorithm domains (SA and RRB) would demonstrate adequate internal consistency.
2. Sensitivity and specificity estimates for the 2000 and 2014 algorithm in the current sample would replicate sensitivity and specificity findings in Hus & Lord's (2014) subsample without ID. The 2000 and 2014 algorithms were explored across subgroups of participants who differed on IQ, gender, and age.
- 3a. Replicating prior findings, the CSS would not significantly correlate with age and cognitive ability.
- 3b. Extending knowledge on performance of the CSS, the CSS would positively correlate with measures of ASD symptom severity, but not adaptive behavior.

Methods

Participants

Current Sample—The current sample included data from 321 research participants and clinical referrals for assessment of possible ASD to: the Center for Autism Spectrum Disorders at Children's National Health System ($n=80$), the National Institute of Mental Health, Laboratory of Brain and Cognition ($n=92$), the Center for Autism Research at the Children's Hospital of Philadelphia ($n=88$), and the Child Study Center and Psychosocial Interventions Lab at Virginia Tech ($n=61$). These four sites each maintain a rigorous program of research reliable ADOS administration and scoring. Participants were included only if they possessed either a verbal or full scale IQ at or above 70. Those with significant hearing, vision or motor problems that interfered with standardized testing were excluded. The sample was predominantly male ($n=246$, 77%) and ages at the time of assessment ranged from 11 to 61 years ($M = 18.91$, $SD = 7.64$). The majority of the sample was between 11-15 ($n=110$) and 16-20 ($n=150$) years of age, with smaller numbers of participants aged 21-25 ($n=35$), 26-30 ($n=7$), and over 30 ($n=19$). Of those who reported their race ($n=251$), 84.5% were Caucasian, 6.4% were African American, 4.0% identified as other, 3.2% were Asian, and 2% were American Indian/Alaskan Native. Approximately 7.2% of the sample was Hispanic/Latino/a. Of those who provided information on maternal education ($n=155$), the majority was well-educated, with 68.4% completing college or graduate school.

Of the 321 participants, 253 had clinical diagnoses of a DSM-IV-TR Asperger's Disorder, Autistic Disorder, or Pervasive Developmental Disorder-Not Otherwise Specified (PDD-NOS), or DSM-5 Autism Spectrum Disorder from ASD specialty clinics. Similar to the stipulation in DSM-5 that allows individuals with Asperger's Disorder or PDD-NOS to be grandfathered into DSM-5, individuals with DSM-IV-TR ASD diagnoses were combined with individuals diagnosed under DSM-5. The remaining 68 participants did not receive a diagnosis of ASD. The Non-ASD sample was similarly comprised of both clinical referrals (n=26) and individuals recruited to research studies (n=42). Of the 56 Non-ASD participants for whom diagnostic data were available, 43 received a primary diagnosis of a DSM-IV-TR or DSM-5 disorder: 17 participants had a primary diagnosis of ADHD, 15 had an anxiety disorder, 3 had Obsessive Compulsive Disorder, 3 had a mood disorder (1 participant had mood disorder with psychotic features), 2 had schizophrenia/schizoaffective disorder, 1 had a substance abuse disorder, 1 had Oppositional Defiant Disorder, and 1 had bipolar disorder. The majority of Non-ASD participants had two Axis-I diagnoses (range 0-5). The remainder of the Non-ASD sample for whom diagnostic data were available (n=12) did not meet criteria for any DSM-IV-TR/DSM-5 diagnosis at the time of assessment. Three participants had a previous diagnosis of ASD, but did not meet criteria for any psychiatric diagnosis at the time of assessment.

Concurrent IQ data were available for 314 of 321 participants. All participants had either a full scale IQ (FSIQ) score or verbal IQ (VIQ) score at or above 70 as measured by the Wechsler Abbreviated Scale of Intelligence-four subtest version (Wechsler, 1999; n=157), Wechsler Intelligence Scale for Children-III (Wechsler, 1991; n=10), Wechsler Intelligence Scale for Children-IV (Wechsler, 2003; n=47), Wechsler Adult Intelligence Scale-III (Wechsler, 1997; n=7), Wechsler Adult Intelligence Scale-IV (Wechsler, 2008; n=30), or the Differential Ability Scales-II (Elliot, 2007; n=63). Table 1 provides further description of the sample.

Hus and Lord's 2014 Comparison Subsample—Because the Hus and Lord (2014) sample included individuals with a wide range of cognitive ability, including ID, a subsample of participants within the age range and IQ criteria stated above were selected from Hus and Lord's data to promote a more accurate comparison to the current sample. This subsample was comprised of 307 individuals diagnosed with a DSMIV-TR Pervasive Developmental Disorder (e.g., Autistic Disorder, Asperger's Syndrome or PDD-NOS) and 66 individuals who did not have an ASD. More details regarding the comparison sample can be found in Hus and Lord (2014). To our knowledge, demographic variables were highly comparable between the present sample and that of Hus and Lord. Although mean IQ scores are equivalent, the Hus and Lord subsample included a wider range of scores.

Measures

Adaptive Behavior Assessment System, Second Edition (ABAS-II; Harrison & Oakland, 2003)—The ABAS is a questionnaire that measures adaptive functioning in the home and community. The overall General Adaptive Composite is divided into the Conceptual (communication, functional academics, self-direction), Social, and Practical (self-care, home living, community use, health and safety) domains. The ABAS

demonstrates acceptable internal consistency and test-retest reliability (Harrison & Oakland, 2003). Complete ABAS data was available for 143 participants.

Autism Diagnostic Interview and Autism Diagnostic Interview-Revised (ADI/ADI-R; Le Couteur et al., 1989; Lord, Rutter, & Le Couteur, 1994)—The ADI/ADI-R is a structured interview given to caregivers to assess the occurrence of ASD symptoms in early childhood in the domains of communication, social interaction, and restricted interests and repetitive behaviors. The ADI/ADI-R provides a historical overview of a child's symptoms in the early development whereas the ADOS provides direct observation of an individual's current symptoms. ADI (n=26) and ADI-R (n=204) data were available for 230 participants.

Autism Diagnostic Observation Schedule, Generic, First, and Second Editions (ADOSG/ADOS/ADOS-2; Lord et al., 2000; Lord, Rutter, DiLavore, & Risi, 2002; Lord et al., 2012)—The ADOS is a semi-structured standardized observational assessment in which a clinician rates a participant's communication, reciprocal social behavior, and repetitive behaviors and stereotyped interest patterns. The ADOS consists of five modules and selection of the appropriate module is based on an individual's chronological age and expressive language level. Module 4 emphasizes interview questions and conversations and is appropriate for use with verbally fluent older adolescents and adults who possess at least a minor level of independence in terms of relationships and goals. Interrater reliability for the ADOS algorithm ranged from 83-100% for exact agreement, with κ ranging from .41-.84. In the current sample, all ADOSes were administered and scored by a clinical psychologist or trainee under direct supervision of a clinical psychologist (e.g., graduate student or postdoctoral fellow) who met standard requirements for research reliability.

Autism Spectrum Quotient (AQ; Baron-Cohen, Wheelwright, Skinner, Martin, & Clubley, 2001)—A self-report measure of characteristics associated with ASD, the AQ has 50 items, with higher scores indicating a greater degree of characteristics associated with ASD. Previous studies have reported good test-retest reliability and acceptable internal consistency in adult samples for the total AQ score (Baron-Cohen et al., 2001). AQ data were available for 102 participants.

Social Communication Questionnaire (SCQ; Rutter, Bailey, Berument, Le Couteur, & Lord, 2003)—The SCQ is a 40-item parent-report questionnaire based on the ADI used to assess the likelihood of an ASD diagnosis, with higher scores indicative of greater severity. The ASD cut-off for screening purposes is 15. There are two versions of the SCQ: current and lifetime. The current version is suitable for children younger than 4 years, while the lifetime version is used with older individuals. This study utilized the lifetime version; data were available for 83 participants.

Social Responsiveness Scale, First and Second Editions (SRS, SRS-2; (Constantino & Gruber, 2005, 2012)—The SRS is a 65-item parent-rating scale of social impairment associated with ASD symptoms, with greater scores indicating more impairment. The SRS has excellent internal consistency and good test-retest reliability for

the total score (Constantino & Gruber, 2005, 2012). Parent-reported SRS data were available for 234 participants.

Vineland Adaptive Behavior Scales, First and Second Editions (VABS, VABS-II; Sparrow, Balla, & Cicchetti, 1984; Sparrow, Cicchetti, & Balla, 2005)—The VABS is a standardized, structured parent/caregiver interview that assesses adaptive skills from birth through 90 years, 11 months. The Adaptive Behavior Composite is composed of the Communication, Daily Living, and Socialization domain standard scores. The VABS has demonstrated strong reliability and validity (Sparrow et al., 2005) and is widely used to examine adaptive behavior in ASD. Complete VABS data were available for 138 participants.

Procedure

The ADOS was given as either part of a research study or clinical evaluation in an ASD specialty clinic. The ADI/ADI-R (Le Couteur et al., 1989; Lord et al., 1994) was administered in 72% of evaluations to obtain developmental history. The ADI/ADI-R or SCQ was available for 76% of participants. In all cases, clinicians had direct observation of the individual through at least the ADOS and cognitive testing. Best estimate clinical diagnoses based on DSM-IV-TR or DSM-5 criteria were made by a supervising or assessing clinical psychologist after review of all assessment information. While scores on the ADOS factored into the clinical decision, diagnoses were not based solely on ADOS cut-offs. This study was conducted in compliance with standards established by all institutions' IRB including procedures for informed consent. Refer to Hus and Lord (2014) for information regarding procedures used to ascertain the comparison sample. The Hus and Lord's subsample was also comprised of clinic and research referrals.

Data Analyses

Chi-square analyses were conducted to determine which items significantly differed between the ASD and Non-ASD groups. Distribution of item code percentages was inspected to determine whether notable differences existed between Hus and Lord's (2014) subsample and the current sample. Internal consistency of the SA and RRB scales were assessed using non-parametric inter-item correlations. Non-parametric correlations were used because ADOS items are categorical and scaled differently. Sensitivity and specificity were calculated using the 2000 and 2014 ADOS algorithms for the overall sample. Sensitivity and specificity estimates were also calculated for (1) individuals with the non-ID verbal IQ ranges investigated by Hus and Lord (2014), (2) the sample divided by gender and, (3) the sample divided by age: under 16 (n=110), 16-20 (n=150), and over 20 (n=61). Finally, correlations were calculated between the three CSS scores proposed by Hus and Lord (2014) - total, social affect (SA-CSS) and repetitive behavior (RRB-CSS) - and measures of cognitive ability, age, ASD symptomatology, and adaptive behavior.

Results

Hypothesis 1

The 2014 algorithm items would discriminate the ASD and Non-ASD samples effectively and algorithm domains (SA and RRB) demonstrate adequate internal consistency.

Group Comparison at Item Level—Results of the chi-square analyses demonstrated that the following low-incidence items did not differ significantly between the ASD and Non-ASD groups in the current sample: Overall Level of Non-echoed Language, Immediate Echolalia, Unusual Sensory Interest in Play Material of Person, Self-Injurious behavior, Overactivity/Agitation, Tantrums/Aggression/Negative or Disruptive Behavior, and Anxiety. Consistent with prediction, all other items were significantly different, with the ASD group receiving higher scores compared to the Non-ASD group.

As hypothesized, all non-parametric inter-item correlations for the SA scale were significant ranging from $\rho=.18$ to $.61$. Significant non-parametric inter-item correlations on the RRB scale ranged from $\rho=.11$ to $\rho=.31$ (see Table 2). *Excessive Interest in Unusual or Highly Specific Topics/Objects or Repetitive Behaviors* was not significantly correlated with *Unusual Sensory Interest in Play Material/Person* or *Hand and Finger and Other Complex Mannerisms*.

Hypothesis 2

Sensitivity and specificity estimates for the 2000 and 2014 algorithm in the current sample would replicate sensitivity and specificity findings in Hus & Lord's (2014) sub-sample without ID.

Sensitivity and Specificity Analyses—The distribution of item code percentages was inspected to determine whether notable differences existed between Hus and Lord's (2014) subsample and the current sample, which may have affected specificity estimates. Comparison of the distribution of scores on item level data in the Non-ASD sample between the current sample and the subsample from Hus and Lord (2014) indicated our sample received a higher percentage of “2” or “3” scores (e.g., indicating more impairment) on two algorithm items (e.g., *Unusual Eye Contact*, *Speech Abnormalities Associated with Autism*) and “1” scores (e.g., behavior that is mildly abnormal or slightly unusual) on almost all other items (data available from authors upon request). Though not tested statistically, there appeared to be some notable differences between the two samples, indicating that a higher percentage of our Non-ASD sample was rated as having more impairment.

The 2000 ADOS algorithm was able to correctly classify 80.4% (76.1%-85.2% across sites) of the cases in our sample as having an ASD or not, based on clinical diagnosis assigned, while the 2014 algorithm correctly classified 81.9% (76.1%-86.9%) of the current sample. In the current sample, sensitivity (84.6%) was higher but specificity (72.1%) was lower on the 2014 algorithm compared to the 2000 algorithm (81.0% and 77.9%, respectively). These estimates were lower than those found by Hus and Lord (2014) in their larger sample. Contrary to hypothesis, they are also lower than what was found in the Hus and Lord (2014) subsample used for the present study, which resulted in sensitivity and specificity estimates

of 82.7% and 83.3% for the 2000 algorithm and 90.2% and 81.8% on the 2014 algorithm. Although the rate of false negatives decreased in our sample with utilization of the 2014 algorithm, a sizeable minority (15.4%) of the sample clinically diagnosed with ASD was not classified as having ASD on the ADOS Module 4. Sensitivity ranged from 76.6%-85.9% and specificity ranged from 69.0%-91.7% across sites for the 2000 algorithm. For the 2014 algorithm, sensitivity estimates improved across sites (80.3%-89.1%) while specificity generally decreased (62.1%-90.9%).

Hypothesis 2

Exploration of the 2000 and 2014 algorithm across subgroups of participants who differed on IQ, gender, and age.

IQ—The pattern of higher sensitivity but lower specificity was consistent across IQ levels, but with higher estimates for those with a VIQ below 85 or above 115 compared to participants with a VIQ in the Average range, between 85 and 115.

Gender—When the sample was divided according to gender, sensitivity and specificity estimates on both algorithms were lower in the current sample when compared to Hus and Lord's subsample. However, estimates were higher for females in both samples on the 2014 algorithm. Sensitivity and specificity estimates for females in the current sample generally reached 80% on both algorithms, though specificity was improved with the 2014 algorithm. In contrast, specificity estimates tended to be lower for males in both samples on the 2014 algorithm. When different cut points were examined between males and females, the 2014 algorithm performed better at a lower cut point in females. In contrast, there was no cut-off where both sensitivity and specificity reached an 80% benchmark in males.

Age—When participants were divided by age, both algorithms produced higher sensitivity and specificity estimates for participants aged 16 and above in our sample compared to those below 16 in our sample. While sensitivity and specificity estimated in Hus and Lord's subsample were above 80% on both the 2000 and 2014 algorithm, specificity estimates in our small sample of 23 individuals younger than 16 without ASD fell at or below 65%. Sensitivity and specificity estimates for the 16-20 and over 20 age groups were above 80% on the 2000 algorithm, and greater for those in the over 20 group. Sensitivity for these two groups were also higher on the 2014 algorithm, but specificity dropped below 80%. In general, the 2014 algorithm performed better with increasing age in our sample.

Hypothesis 3a

Replicating prior findings, the CSS would not significantly correlate with age and cognitive ability.

Within the ASD group, the CSS demonstrated small, but statistically significant, negative correlations with FSIQ ($r(251)=-.13, p<.05$) and Verbal IQ ($r(251)=-.18, p<.01$), but not age ($r(253)=-.09, p=ns$) providing partial support for our hypothesis.

Hypothesis 3b

Extending knowledge on performance of the CSS, the CSS would positively correlate with measures of ASD symptom severity, but not adaptive behavior.

This hypothesis was not supported by the findings. The CSS did not correlate significantly with the ADI/ADI-R social, communication, or repetitive behaviors domains ($r(192)=-.02$, $r(192)=.04$, $r(192)=.00$, all $ps=ns$), and there was a moderate *negative* correlation with the SCQ ($r(56)=-.31$, $p<.05$). A small positive correlation with the SRS ($r(186)=.18$, $p<.05$) was consistent with hypothesis. With regard to self-report measures of ASD symptoms, the CSS was not correlated with total AQ score ($r(91)=-.02$, $p=ns$). Finally, as predicted, the CSS was not significantly associated with the VABS Communication, Daily Living, Socialization, or Adaptive Behavior Composite scores ($r(97)=.03$, $r(98)=-.19$, $r(96)=-.05$, $r(97)=-.10$, all $ps=ns$), though contrary to prediction, it correlated negatively with the ABAS Conceptual, Practical, Social, and Global Adaptive composite scores ($r(123)=-.25$, $p<.01$; $r(123)=-.28$, $p<.01$; $r(121)=-.21$, $p<.05$; $r(125)=-.26$, $p<.01$, respectively). Correlations were similar in the subset of individuals who received the ABAS and VABS concurrently ($n=40$), suggesting that difference in association with measure was not due to differences in the participants administered those measures.

The SA-CSS demonstrated a similar pattern and magnitude of correlations as the overall CSS, except that the SA-CSS was not correlated with FSIQ. The RRB-CSS was not significantly correlated with any participant characteristics.

Discussion

The present study is the first attempt to evaluate the newly proposed algorithm for the ADOS-2 Module 4 in an independent sample comprised exclusively of individuals with ASD without ID ($n=321$). In this study we: 1) investigated whether the 2014 algorithm items discriminated between ASD and non-ASD samples and demonstrated adequate internal consistency, 2) sought to replicate Hus and Lord's findings using an independent sample of adolescents and adults with ASD without ID, and 3) examined the utility of the CSS as an unbiased measure of symptom severity. We found several significant differences from the validation study (Hus & Lord, 2014) indicating the need for further refinement of the Module 4 algorithm, as well as caution with respect to its diagnostic application for individuals without ID.

In support of our first hypothesis, item level analyses indicated ADOS items effectively discriminated the ASD from Non-ASD sample on all items except for low frequency items most commonly associated with ID (e.g., echolalia, unusual sensory interests). This is to be expected given the low frequency of these problems in our sample with average and above-average cognitive abilities. There was mixed support for internal consistency, with the SA scale demonstrating greater consistency than the RRB scale. This is consistent with the Hus and Lord (2014) study indicating higher internal consistency for the SA than RRB scale. All items on the SA scale demonstrated significant but variable (e.g., negligible to strong) associations with other items. Items on the RRB scale were negligibly to moderately correlated, though the *Unusual Sensory Interests* item had the lowest correlations overall

and was not correlated with several other scale items. Our sample demonstrated robust item level discrimination on relevant symptoms of ASD, and cohesion among items on the SA scale, but these results suggest that items on the RRB scale may not be a cohesive representation of this symptom domain in ASD without ID. This suggests that we need to think broadly about the manifestation of RRBs in adults with ASD without ID, and give consideration as to whether additional items are needed in the Module 4 to improve discriminability in individuals without ID (e.g., self-reporting of experience via interview rather than actual observation inside the room).

There was mixed support for our second hypothesis. In our overall sample, we found lower sensitivity and specificity estimates on both the 2000 and 2014 algorithms compared to a non-ID subsample from Hus and Lord (2014). Although there were modest improvements in sensitivity using the 2014 algorithm in our current sample, there were also slight decreases in specificity – a pattern that was replicated in Hus and Lord's subsample. Although interpretation of specificity is somewhat limited by our small Non-ASD sample, item level exploration suggests that our Non-ASD sample showed differences in their social-communication profiles. It is noteworthy that Hus and Lord (2014) selected items for the 2014 algorithm based on high frequency in individuals with DSM-IV diagnoses of autism (as opposed to PDD-NOS or Asperger's, who may be assumed to present with milder symptoms) and low frequency in NonASD diagnoses. Comparisons between our sample and Hus and Lord's (2014) subsample indicated that in our sample the Non-ASD participants were rated as more impaired (e.g., scores of “2” or “3” on two items in the SA domain as well as rated as mildly impaired (e.g., “1” scores) on almost all items in the algorithm. This may reflect study differences in the Non-ASD samples; our sample had a higher proportion of adults with ADHD, whereas Hus & Lord's sample had very few individuals with ADHD. Future studies with larger samples of specific Non-ASD diagnoses will be important to better understand the utility of the Module 4 for distinguishing different psychiatric groups.

Compared to the original algorithm, the 2014 algorithm provided higher estimates of sensitivity, but lower estimates of specificity, for key demographic subsets within our sample. Generally, use of the 2014 algorithm resulted in a higher sensitivity estimate at the expense of lower specificity across groupings based on VIQ, with both estimates being the lowest in individuals with Average verbal abilities (i.e., verbal IQ estimate between 85-115). In particular, specificity was strikingly low for this group in our sample (64.9%). It is important to note the small sample sizes of these subgroups; for example, the average IQ group was comprised of only 37 participants and the reduction in specificity of the 2014 algorithm is a result of only 2 additional participants being misclassified. This finding is in line with (albeit lower than) Hus and Lord (2014) who also found lowest specificity for this group (77.3%). Nonetheless, the lower specificity in our sample compared to their study warrants further investigation in larger Non-ASD samples. As noted above, different psychiatric groups may exhibit different symptom profiles that could provide insight into future modifications to improve specificity of the Module 4.

A notable finding of the present study was better sensitivity and specificity estimates for females compared to males. This result was found in both the current sample and Hus and Lord's subsample on the 2014 algorithm. Both sensitivity and specificity estimates were

greater than 80% for females in the present study, and the estimates were even higher in Hus and Lord's subsample. In contrast, sensitivity was roughly equivalent for males, but specificity was below the 80% benchmark in both the current sample and in Hus and Lord's subsamples. In the current sample, specificity estimates were considerably lower for males (65%) compared to females (86%). Future research should address whether different cut-off scores (or different questions) are needed for males and females. Such practices are not uncommon with other measures of social difficulties related to ASD (e.g., SRS; Constantino & Gruber, 2005). In the current multi-site sample, sensitivity and specificity estimates for females on the 2014 algorithm were firmly above the 80% benchmark at a lower cut-off. However, for males there was no cut-off for which sensitivity and specificity estimates were both above 80%. We elect not to publish the cut-off scores because a larger sample is needed to replicate these findings. If replicated, this may indicate that females can be reliably diagnosed at a lower threshold. This finding is bolstered by recent literature indicating that females with a clinical diagnosis of ASD tend to be underrepresented at the higher ends of IQ distribution and demonstrate increased functional social behavior and less repetitive behaviors compared to males (Halladay et al., 2015). Paradoxically, fewer females may present for a clinical evaluation for ASD, but when they do, these results suggest that the ADOS can discriminate ASD symptoms better for females than males.

Regarding age groupings, both the 2000 and 2014 algorithm in the current sample performed better for individuals aged 16 and above (and perform similarly for those aged 16-20 and above 20) compared to those under 16. The ADOS-2 manual indicates that Module 4 is optimal for individuals 16 and above, but provides some flexibility regarding appropriateness to participant's interests and level of independence. Some clinicians may wish to administer Module 4 for a younger adolescent who is clearly uncomfortable with Module 3 materials, such as the *Make-Believe Play* activity. However, these results caution against doing so due to markedly lower specificity in individuals under 16. Instead, the clinician may choose to follow ADOS-2 Module 3 recommendations, such as encouraging the individual to show the examiner how they would have played when they were younger (Lord et al., 2012).

When considering these findings, it is important to recognize that priorities for the identification of ASD may be different for clinical and research purposes; researchers may prefer narrower formulations to reduce the likelihood of false positives and reduce overlap with control groups, while clinicians may prefer a broader diagnosis to ensure access to services for individuals with fewer discrete symptoms. In their original analyses, Hus and Lord indicated that researchers who hope to achieve a higher level of specificity, at the expense of lower sensitivity, may wish to use a cut-off of 10 rather than 8. However, for purposes of diagnosis in a clinical setting, a cut-off of 8 may be most appropriate due to higher sensitivity and lower rates of false negatives. Results from our sample confirm this suggestion. When the threshold is increased to a cut-off of 10, specificity in the current sample increases to 80.9%, though sensitivity decreases to 73.9%. When only the participants aged 16 and above were considered, sensitivity and specificity estimates were 74.7% and 86.7%, respectively. However, it should be noted that using the higher cut-off may exclude more mildly affected individuals who, nonetheless, have ASD.

A further aim of this study was to explore the utility of the CSS as a measure of symptom severity. Providing mixed support for our third hypotheses, in the current sample the CSS was not associated with age, but demonstrated small, negative correlations with cognitive ability. Regarding predicted positive relationships between the CSS and other measures of autism symptomatology, the data were mixed. The CSS demonstrated small positive correlations with the SRS. The CSS did not significantly correlate with childhood ASD symptoms on the ADI/ADI-R, though it was moderately and negatively associated with retrospective symptoms on the SCQ. It is reasonable to expect that current and historical ASD symptoms would not be correlated on the ADI (e.g., Kenworthy, Black, Harrison, Della Rosa, & Wallace, 2009), though the finding of a significant *negative* association with historical symptoms on the SCQ is counterintuitive and warrants further investigation. Differences in patterns of association between the measures may be due to difference in modalities (e.g., clinician-rated symptoms versus parent report; Hus, Bishop, Gotham, Huerta, & Lord, 2013) or the small sample size for the SCQ analyses. The CSS was not significantly associated with self-report of ASD symptoms on the AQ, which may be due to differences in self-report versus clinician observation of symptoms in ASD (Bishop & Seltzer, 2012). There were no significant associations with adaptive behavior as measured by the VABS. Interestingly, however, the CSS was negatively correlated with adaptive behavior on the ABAS. Clinical experience suggests that the ABAS assesses a wider range of adaptive behaviors expected in adulthood than does the VABS. These results suggest that the CSS for Module 4 is variably associated with other ASD symptom measures and may be confounded with functional impairment in those without ID. Future studies are warranted to better understand the relationships between CSS and other measures of ASD severity and/or broader functional impairment to inform the most appropriate use and interpretation of this metric.

Limitations

As noted by Gotham and colleagues (2007), sensitivity and specificity of the ADOS algorithm may vary across clinical and research settings as a result of differences in examiner skill, sequence of administration and other factors. Although all of the sites in this study had attained and maintained research reliability on the ADOS, specificity and sensitivity estimates were lower than Hus and Lord's overall sample and subsample. While the sample size of Non-ASD participants was roughly equivalent to Hus and Lord, it similarly represented a diagnostically diverse group. Though ADHD and anxiety disorders were the most common primary diagnoses in the Non-ASD participants in the current sample, many had two or three psychiatric diagnoses making their presentation clinically complex. As such, more research is needed with well-defined clinical comparison groups (e.g., ADHD, anxiety, mood disorders) to examine items that best discriminate ASD from other specific disorders. Importantly, the NonASD group as a whole was roughly $\frac{1}{4}$ the size of the ASD group, and specificity estimates may reflect bias due to a proportionally smaller Non-ASD sample. Future research should include larger, diagnostically diverse groups of participants to examine sensitivity and specificity estimates in other disorders.

Additionally, the majority of our sample consisted of adolescents and young adults, with a relatively small proportion of individuals in middle or later adulthood. While our sensitivity

and specificity estimates were quite similar for individuals ages 16-20 and those above 20, more extensive testing of the algorithm is warranted for older adults. Indeed, the original validation sample also included very few older adults; the data available for the few older participants suggested there may be differences in symptom presentation that require further exploration (Hus & Lord, 2014). Nonsignificant correlations with age may be an artifact of restricted range rather than a truly null finding. In addition to being young, the current sample was predominantly Caucasian and ill-suited to addressing questions regarding algorithm performance across different races and ethnicities. It is important that future research examine performance of the ADOS in diverse, historically underserved populations to inform clinical diagnostic procedures.

Conclusions

Taken together, these data suggest several important considerations when making clinical and research diagnoses of ASD in older adolescents and adults. Most importantly, there is not yet an ADOS algorithm that perfectly reaches the same conclusions that clinicians do when they consider all the available clinical information. Although specificity in the overall sample was comparable to specificity when using ASD cut-offs for other ADOS modules (Gotham et al., 2007), the lower estimates in the small subsamples of adults with average IQ and males warrants further study. If these findings are replicated, particularly in specific psychiatric subgroups, further revisions to the algorithm or activities presented in the Module 4 may be warranted. Additionally, our findings suggest that the ADOS Module 4 may be more accurate in classification of ASD in females than males. Finally, our findings confirm that the Module 4 performs best with participants who are 16 years of age or older and for those with verbal IQ estimates below 85 or above 115. Despite these limitations, the ADOS Module 4 remains the only observational measure for diagnosing ASD in verbally-fluent older adolescents and adults. As such, best practice diagnostic procedures should continue to include the ADOS in conjunction with other assessment tools and modalities (e.g., caregiver report, developmental history, self-report) to inform differential clinical diagnoses, and balancing needs for sensitivity and specificity, depending on the purpose of the assessment and the demographic characteristics of the person being evaluated, will only enhance its utility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported in part by an award from the Isadore and Bertha Gudelsky Family Foundation and the Fredrick and Elizabeth Singer Foundation, and grants by the Organization for Autism Research, Autism Speaks (Dennis Weatherstone Predoctoral Fellowship), the Philadelphia Foundation, Pfizer, Shire Pharmaceuticals, Robert Wood Johnson Foundation (#6672), the Pennsylvania Department of Health (SAP #4100042728, SAP #4100047863), and the National Institutes of Health and the Intramural Research Program at NIMH (1-ZIA-MH002920-05, P30HD040677, T32 HD046388-01A2, 1K01MH079945-01, K23MH086111, R21MH092615, RC1MH088791, and P30HD026979). We gratefully acknowledge Catherine Lord, PhD for contributing data for comparative analyses; those data were funded by R01MH081873 to CL.

References

- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th., text rev. Washington, DC: Author; 2000.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th. Washington, DC: Author; 2013.
- Baio, J. Morbidity and Mortality Weekly Report Surveillance Summaries. Vol. 63. Centers for Disease Control and Prevention; 2014. Prevalence of autism spectrum disorder among children aged 8 years Autism and Developmental Disabilities Monitoring Network, 14 Sites, United States, 2010.
- Baron-Cohen S, Wheelwright S, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders*. 2001; 31(1):5–17. [PubMed: 11439754]
- Bastiaansen, Ja; Meffert, H.; Hein, S.; Huizinga, P.; Ketelaars, C.; Pijnenborg, M.; et al. de Bildt, A. Diagnosing Autism Spectrum Disorders in Adults: the Use of Autism Diagnostic Observation Schedule (ADOS) Module 4. *Journal of Autism and Developmental Disorders*. 2011; 41(9):1256–1266.10.1007/s10803-010-1157-x [PubMed: 21153873]
- Bishop S, Seltzer M. Self-reported autism symptoms in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2012; 42(11):2354–2363. [PubMed: 22361924]
- Constantino, JN.; Gruber, CP. Social responsiveness scale (SRS). Los Angeles, CA: Western Psychological Association; 2005.
- Constantino, JN.; Gruber, CP. Social Responsiveness Scale. Second. Los Angeles, CA: Western Psychological Services; 2012.
- De Bildt A, Sytema S, Ketelaars C, Kraijer D, Mulder E, Volkmar F. Interrelationship between Autism Diagnostic Observation Schedule-Generic (ADOS-G), Autism Diagnostic Interview-Revised (ADI-R), and the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR) classification in children and adolescents with mental disorders. *Journal of Autism and Developmental Disorders*. 2004; 34(2): 129–137. [PubMed: 15162932]
- Elliot, CD. Differential Abilities Scales-II. San Antonio, TX: Harcourt Assessment; 2007.
- Farley, Ma; McMahon, WM.; Fombonne, E.; Jenson, WR.; Miller, J.; Gardner, M.; et al. Coon, H. Twenty-year outcome for individuals with autism and average or near-average cognitive abilities. *Autism Research: Official Journal of the International Society for Autism Research*. 2009; 2(2): 109–18.10.1002/aur.69 [PubMed: 19455645]
- Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *Journal of Autism and Developmental Disorders*. 2009; 39(5):693–705. [PubMed: 19082876]
- Gotham K, Risi S, Pickles A, Lord C. The Autism diagnostic observation schedule: Revised algorithms for improved diagnostic validity. *Journal of Autism and Developmental Disorders*. 2007; 37(4):613–627.
- Gray KM, Tonge BJ, Sweeney DJ. Using the Autism diagnostic interview-revised and the Autism diagnostic observation schedule with young children with developmental delay: Evaluating diagnostic validity. *Journal of Autism and Developmental Disorders*. 2008; 38(4):657–667. [PubMed: 17690967]
- Halladay AK, Bishop S, Constantino JN, Daniels AM, Koenig K, Palmer K, et al. Szatmari P. Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. *Molecular autism*. 2015; 6(1):36. [PubMed: 26075049]
- Harrison, PL.; Oakland, T. Adaptive behavior assessment system. second. San Antonio, TX: Harcourt Assessment; 2003.
- Howlin P. Social disadvantage and exclusion: adults with autism lag far behind in employment prospects. *Journal of the American Academy of Child and Adolescent Psychiatry*. 2013; 52(9): 897–9.10.1016/j.jaac.2013.06.010 [PubMed: 23972691]
- Hus V, Bishop S, Gotham K, Huerta M, Lord C. Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry*. 2013; 54(2):216–224. [PubMed: 22823182]

- Hus V, Lord C. The autism diagnostic observation schedule, module 4: revised algorithm and standardized severity scores. *Journal of Autism and Developmental Disorders*. 2014; 44(8):1996–2012.10.1007/s10803-014-2080-3 [PubMed: 24590409]
- Interagency Autism Coordinating Committee. The 2011 interagency autism coordinating committee strategic plan for autism spectrum disorder research. 2011
- Kenworthy L, Black DO, Harrison B, Della Rosa A, Wallace GL. Are executive control functions related to autism symptoms in high-functioning children? *Child Neuropsychology*. 2009; 15:425–440.10.1080/09297040802646983 [PubMed: 19173090]
- Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, McLennan J. Autism Diagnostic Interview: A Standardized Investigator-Based Instrument. 1989; 19(3):363–387.
- Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, Dilavore PC, et al. Rutter M. The Autism Diagnostic Observation Schedule–Generic: A Standard Measure of Social and Communication Deficits Associated with the Spectrum of Autism. 2000; 30(3):205–223.
- Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S. Autism Diagnostic Observation Schedule. Los Angeles: Western Psychological Services; 2002.
- Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S.; Gotham, K.; Bishop, SL. Autism Diagnostic Observation Schedule. 2nd (ADOS-2). Torrance, CA: Western Psychological Services; 2012.
- Lord C, Rutter M, Le Couteur A. 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*. 1994; 24:659–685. [PubMed: 7814313]
- Papanikolaou K, Paliokosta E, Houliaras G, Vgenopoulou S, Giouroukou E, Pehlivanidis A. Using the Autism diagnostic interview-revised and the Autism diagnostic observation schedule-generic for the diagnosis of Autism spectrum disorders in a Greek sample with a wide range of intellectual abilities. *Journal of Autism and Developmental Disorders*. 2009; 39(3):414–420. [PubMed: 18752062]
- Rutter, M.; Bailey, A.; Berument, SK.; Le Couteur, A.; Lord, C. Social communication questionnaire (SCQ) manual. Los Angeles, CA: Western Psychological Services; 2003.
- Seltzer MM, Krauss MW, Shattuck PT, Orsmond G, Swe A, Lord C. The symptoms of autism spectrum disorders in adolescence and adulthood. *Journal of Autism and Developmental Disorders*. 2003; 33(6):565–81. [PubMed: 14714927]
- Shattuck PT, Seltzer MM, Greenberg JS, Orsmond GI, Bolt D, Kring S, et al. Lord C. Change in autism symptoms and maladaptive behaviors in adolescents and adults with an autism spectrum disorder. *Journal of Autism and Developmental Disorders*. 2007; 37(9):1735–47.10.1007/s10803-006-0307-7 [PubMed: 17146700]
- Sparrow, S.; Balla, DA.; Cicchetti, D. Vineland Adaptive Behavior Scales (expanded form). Circle Pine, MN: American Guidance Service; 1984.
- Sparrow, S.; Cicchetti, D.; Balla, DA. Vineland Adaptive Behavior Scales Second Edition:Survey Forms Manual 2005. Minneapolis, MN: Pearson; 2005.
- Taylor JL, Seltzer MM. Changes in the autism behavioral phenotype during the transition to adulthood. *Journal of Autism and Developmental Disorders*. 2010; 40(12):1431–1446. [PubMed: 20361245]
- Volkmar FR, Langford Booth L, McPartland JC, Wiesner LA. Clinical Evaluation in Multidisciplinary Settings. *Handbook of Autism and Pervasive Developmental Disorders* (4th). 2014:661–672.
- Wechsler, D. Manual for the Wechsler Intelligence Scale for Children. Third. San Antonio, TX: The Psychological Corporation; 1991.
- Wechsler, D. Wechsler adult intelligence scale. 3rd. San Antonio, TX: The Psychological Corporation; 1997.
- Wechsler, D. The Wechsler Abbreviated Scale of Intelligence. San Antonio, TX: The Psychological Corporation; 1999.
- Wechsler, D. Wechsler Intelligence Scale for Children-Fourth Edition, administration and scoring manual. San Antonio, TX: Harcourt Assessment, Inc; 2003.
- Wechsler, D. Wechsler adult intelligence scale: Technical and interpretive manual. 4th. San Antonio, TX: Pearson; 2008.

Table 1

Demographic information.

	ASD				Non-ASD			
	N	Mean	SD	Range	N	Mean	SD	Range
Hus & Lord Subsample								
Age	303	20.32	7.13	11-55	65	22.43	10.7	13-62
VIQ	303	105.80	18.74	70-150	65	103.51	16.71	70-144
FSIQ	254	103.22	18.50	40-154	63	104.98	17.31	59-137
Current Sample								
Age	253	18.81	7.31	11-61	68	19.31	8.84	11-54
Verbal IQ	251	109.10	18.01	60-153	63	105.14	18.28	71-157
FSIQ	251	107.16	17.72	64-154	63	101.78	17.80	71-155
ADI/R Social	192	18.41	6.24	0-29	38	13.45	7.43	2-25
ADI/R Communication	192	14.94	4.89	0-26	38	10.66	5.77	0-20
ADI/R RRB	192	5.19	2.38	0-12	38	3.32	2.71	0-10
SRS Total-Parent	186	78.23	14.25	10-114	48	69.38	18.21	37-115
SCQ Total	56	20.64	6.61	6-34	27	17.33	8.59	1-32
AQ Total	91	26.12	7.38	11-47	10	25.90	11.86	10-43
ADOS Soc+Comm	253	11.23	4.42	1-21	68	5.34	4.47	0-19
ADOS RRB	253	1.28	1.27	0-6	68	0.65	1.03	0-6
ADOS SA *	253	9.96	4.58	0-20	68	4.69	4.54	0-19
ADOS RRB*	253	2.76	1.84	0-10	68	1.24	1.53	0-8
VABS Communication	97	75.14	12.55	43-106	41	77.46	17.02	22-117
VABS Daily Living	98	76.35	13.17	51-115	42	79.88	16.06	48-119
VABS Socialization	96	68.53	12.06	37-106	41	75.70	15.23	47-113
VABS ABC	97	70.88	9.61	49-94	41	75.39	13.49	37-106
ABAS Conceptual	123	82.02	11.57	55-115	23	85.30	13.10	65-118
ABAS Practical	121	78.10	15.58	40-120	22	80.50	18.8	48-120
ABAS Social	123	74.24	10.76	55-110	23	77.04	15.40	60-118
ABAS GAC	125	74.48	13.20	43-118	23	79.57	15.83	56-117

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Note: Age reported in years, IQ reported in standard scores (100 ± 15), and SRS reported in t-scores. All other scores are raw scores. FSIQ=Full Scale IQ; ADI/R=Autism Diagnostic Interview/ADI-Revised, RRB=Restricted and Repetitive Behaviors; SRS-P=Social Responsiveness Scale, Parent Report, total t-score; SCQ=Social Communication Questionnaire, total t-score; AQ=Autism Spectrum Quotient, raw score; ADOS Soc+Comm=combined social and communication domain 2000 algorithm scores, RRB=restricted and repetitive behavior domain 2000 algorithm score, SA *=newly proposed 2014 social affect domain score, RRB*=newly proposed 2014 restricted and repetitive behavior domain score; VABS=Vineland Adaptive Behavior Scales; VABS ABC=VABS Adaptive Behavior Composite; ABAS=Adaptive Behavior Assessment System; ABAS GAC=ABAS Global Adaptive Composite.

Table 2
Inter-item correlations for the Module 4 Social Affect and Restricted Repetitive Behaviors Domains.

Social Affect										
	1	2	3	4	5	6	7	8	9	
1. Unusual Eye Contact	1									
2. Amount of Social Communication	.20**	1								
3. Facial Expressions	.29**	.48**	1							
4. Quality of Rapport	.32**	.52**	.47**	1						
5. Communication of Own Affect	.19**	.40**	.29**	.31**	1					
6. Quality of Social Overtures	.26**	.59**	.40**	.58**	.39**	1				
7. Conversation	.24**	.68**	.46**	.46**	.42**	.60**	1			
8. Emphatic Gestures	.26**	.47**	.41**	.39**	.35**	.40**	.40**	1		
9. Quality of Social Response	.27**	.45**	.44**	.51**	.34**	.54**	.49**	.36**	1	
10. Insight	.18**	.42**	.25**	.34**	.55**	.45**	.47**	.23**	.39**	
Restricted Repetitive Behaviors										
	1	2	3	4						
1. Speech Abnormalities	1									
2. Stereotyped Language	.31**	1								
3. Unusual Sensory Interests	.16**	.15**	1							
4. Highly Specific Topics	.22**	.24**	.08	1						
5. Hand Mannerisms	.25**	.13**	.11	.10						

Note:

* $p < .05$,

** $p < .01$.

Table 3

Sensitivity and specificity estimates across verbal IQ levels.

	VIQ 70-85		VIQ 85-115		VIQ >115	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
H&L Subsample	ASD=47	NS=7	ASD=155	NS=44	ASD=101	NS=14
2000 Algorithm	85.1	85.7	83.8	81.8	79.2	85.7
2014 Algorithm	91.5	85.7	91.6	77.3	87.1	92.9
Current Sample	ASD=22		ASD=134		ASD=95	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
2000 Algorithm	95.5	80.0	79.1	70.3	81.1	87.5
2014 Algorithm	100	70.0	82.8	64.9	84.2	81.3

Note: H&L=Hus and Lord (2014) subsample; ASD=sample size in ASD group used for sensitivity calculations; NS=sample size in non-spectrum group used for specificity calculations.

Table 4

Sensitivity and specificity estimates for males and females in the present sample.

	Males		Females	
	<i>Sensitivity</i>	<i>Specificity</i>	<i>Sensitivity</i> ⁴	<i>Specificity</i>
H&L Subsample	ASD=253	NS=47	ASD=54	NS=19
2000 Algorithm	83.4	78.7	79.6	94.7
2014 Algorithm	89.3	78.7	94.4	89.5
	<i>Sensitivity</i>	<i>Specificity</i>	<i>Sensitivity</i>	<i>Specificity</i>
Current Sample	ASD=200	NS=46	ASD=53	NS=22
2000 Algorithm	81.5	73.9	79.3	86.4
2014 Algorithm	84.5	65.2	84.9	86.4

Note: H&L=Hus and Lord (2014) subsample; ASD=sample size in ASD group used for sensitivity calculations; NS=sample size in non-spectrum group used for specificity calculations.

Table 5
Sensitivity and specificity estimates when sample is divided at age 16 in the present sample.

		Under 16		16-20		Over 20	
		Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
H&L Subsample		ASD=62	NS=7	ASD=147	NS=36	ASD=98	NS=23
2000 Algorithm		83.9	85.7	82.3	77.8	82.7	91.3
2014 Algorithm		87.1	85.7	91.8	77.8	89.8	87.0
Current Sample		ASD=87	NS=23	ASD=116	NS=34	ASD=50	NS=11
2000 Algorithm		78.2	65.2	81.0	82.4	86.0	90.9
2014 Algorithm		81.6	60.9	84.5	79.4	90.0	72.7

Note: H&L=Hus and Lord (2014) subsample; ASD=sample size in ASD group used for sensitivity calculations; NS=sample size in non-spectrum group used for specificity calculations.