The Autism Symptom Interview, School-Age: A Brief Telephone Interview to Identify Autism Spectrum Disorders in 5-to-12-Year-Old Children

Somer L. Bishop[1], **Marisela Huerta**[2], **Katherine Gotham**[3], **Karoline Alexandra Havdahl**[4], **Andrew Pickles**[5], **Amie Duncan**[6], **Vanessa Hus Bal**[1], **Lisa Croen**[7], and **Catherine Lord**[2]

[1]University of California, San Francisco, Department of Psychiatry

[2]Weill Cornell Medical College, Center for Autism and the Developing Brain

[3]Vanderbilt University School of Medicine, Department of Psychiatry

[4]Lovinsberg Hospital and the Norwegian Institute of Public Health

[5]King's College London, Department of Biostatistics

[6]Cincinnati Children's Hospital Medical Center, Division of Developmental and Behavioral Pediatrics

[7]Kaiser Permanente Northern California, Division of Research

Abstract

Lay Abstract—Recent advances in multiple areas of autism research, including genetics and epidemiology, have increased the need for large numbers of participants with autism spectrum disorders (ASD). The Autism Symptom Interview (ASI) is a brief phone interview that was designed to facilitate rapid ascertainment of children with ASD for research studies. The ASI is based on questions from the Autism Diagnostic Interview-Revised (ADI-R), a comprehensive, semi-structured parent interview, but the ASI is designed to be administered in approximately 20 minutes by interviewers with minimal training. This study reports on the initial validation of the ASI, School-Age, for children ages 5 to 12 years. Children with previous diagnoses or suspicion of ASD or another neurodevelopmental disorder participated in a comprehensive diagnostic assessment as part of the study and were classified as ASD or non-ASD following the assessment. The ASI scores of children with and without ASD were then compared. For verbal children (defined as using phrases or better on a daily basis), the ASI showed reasonable accuracy in identifying children with ASD (sensitivity=.87), but specificity was low (.62). However, when ASI scores were considered together with scores from the Autism Diagnostic Observation Schedule (ADOS), sensitivity was maintained at .82, and specificity improved to .92. These findings suggest that the ASI school age may serve as a useful tool to more quickly classify children with ASD for research purposes.

Correspondence to: Somer L. Bishop, Ph.D., University of California, San Francisco, 401 Parnassus Avenue, Box- CAS 0984, San Francisco, CA 94143-0984. Tel: (415) 502-3349. Fax: (415) 502-6361. Somer.Bishop@ucsf.edu.

Scientific Abstract—This study reports on the initial validation of the Autism Symptom Interview (ASI), School-Age, a brief (15–20 minute) phone interview derived from questions from the Autism Diagnostic Interview-Revised (ADI-R). The ASI, School-Age was administered by interviewers with minimal training to parents of children ages 5 to 12 who had all been previously identified with (or referred for assessment of) ASD or another neurodevelopmental disorder. Children then underwent a comprehensive assessment to determine a best-estimate clinical diagnosis of ASD (n=159) or non-ASD (e.g., language disorder, intellectual disability, ADHD; n=130). Clinicians who conducted the assessments were blind to ASI results. ROC analyses compared ASI scores to clinical diagnosis. Due to the small number of participants with non-ASD diagnoses who were classified as nonverbal (i.e., not yet using phrases on a daily basis), it was not possible to assess sensitivity and specificity of the nonverbal algorithm in this sample. The verbal algorithm yielded a sensitivity of .87 (95% CI=.81–.92) and a specificity of .62 (95% CI=.53–.70). When used in conjunction with the Autism Diagnostic Observation Schedule (ADOS), sensitivity and specificity were .82 (95% CI=.74–.88) and .92 (95% CI=.86–.96), respectively. Internal consistency and test-retest reliability were both excellent. Particularly for verbal school age children, the ASI may serve as a useful tool to more quickly ascertain or classify children with ASD for research or clinical triaging purposes. Additional data collection is underway to determine the utility of the ASI in children who are younger and/or nonverbal.

Keywords

assessment; rapid ascertainment; screening; neurodevelopmental disorders

Introduction

During the past 20 years, the advent of new assessment technology has significantly improved our ability to reliably identify children with autism spectrum disorders (ASD) (Gotham, Bishop, & Lord, 2011; Lord & Risi, 1998). Particularly when used together, standardized parent interviews and child observation measures exhibit high levels of sensitivity and specificity in differentiating children with ASD from children with a range of other developmental disabilities (e.g. Bishop, Gahagan, & Lord, 2007; Risi et al., 2006; Zander, Sturm, & Bolte, 2015). While the use of comprehensive assessment batteries is preferred in most circumstances, conducting lengthy diagnostic evaluations is not always feasible for ascertaining research participants, nor is it necessarily appropriate depending on the specific research or clinical question. Many standardized diagnostic instruments require substantial examiner training and are time consuming to administer, which limits the number of participants that can be accurately classified into ASD and non-ASD diagnostic groups within a typical study period. Therefore, especially for epidemiological and etiological investigations requiring very large numbers of participants, there is a need for instruments that can be used to more efficiently identify children with ASD.

The current study reports on the development of a brief parent interview that was designed primarily as a case confirmation tool for ASD. The Autism Symptom Interview (ASI) was based on questions from the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Couteur, 1994), a widely used and well-established standardized parent interview that is generally considered to be the first choice for parent-report ASD diagnostic instruments

(Lord & Corsello, 2005; Yonan et al., 2003). Unlike the ADI-R, the ASI was designed to be administered on the telephone, in less than 15–20 minutes, by interviewers with very minimal training. Also, whereas the ADI-R is a diagnostic measure that elicits detailed information about symptoms and provides domain scores in the areas of Communication, Reciprocal Social Interaction, and Restricted and Repetitive Behaviors and Interests, the ASI was initially conceived as a research tool that could be employed to quickly identify individuals with a high probability of ASD "caseness." The ASI does not yield domain scores, severity indicators, or other metrics necessary to provide detailed phenotypic characterization of participants. Thus, the ASI was not designed as a replacement for a full diagnostic evaluation, but rather as a more efficient means of ascertaining participants who would be likely to meet standard diagnostic criteria for ASD and who could undergo more detailed phenotyping at a later time.

Approach to the Development of the Autism Screening Interview (ASI)

Development of the ASI was guided primarily by analyses of previously collected ADI-R data (see below). The ADI-R inquires both about "Current" (within the last 3 months) and past (either "Ever" or during the "Most Abnormal" period between the ages of 4 and 5 years) behaviors. The diagnostic algorithm for children over the age of 4 years is based on "Most Abnormal 4–5" and "Ever" behaviors, whereas the diagnostic algorithm for children under 4 years is based on both "Current" and "Ever" behaviors. The focus on past behaviors was intended to minimize the effects of chronological age and language ability on ADI-R scores by having all parents report on their children's social and communication behaviors during the ages of 4 and 5 years, an age by which even children with significant developmental and/or language delays should have acquired the types of social-communication skills inquired about in the ADI-R (e.g., the ability to share enjoyment, use gestures, offer comfort, smile in response to another person, etc.) (American Psychiatric Association, 2000; Lord et al., 1994). For restricted and repetitive behaviors and atypical language symptoms (e.g., stereotyped speech, neologisms, pronoun reversal), "Ever" scores are used in the diagnostic algorithm because these behaviors are not typically expected at any point in development, though they are commonly seen in children with non-ASD diagnoses, and in some typically developing young children (Richler, Bishop, Kleinke, & Lord, 2007).

Focusing on past behaviors offers advantages in terms of potentially minimizing age and language effects, but there are many important reasons why current descriptions of behavior, as opposed to retrospective reports, may be of value. These include problems with validity/reliability of retrospective reporting on the ADI-R (Hus, Taylor, & Lord, 2011; Jones et al., 2015), as well as recent findings that a small proportion of individuals who meet criteria for ASD as young children make such significant improvements that they no longer qualify for this diagnosis in later childhood or adolescence (Anderson, Liang, & Lord, 2014; Fein et al., 2013). The process of obtaining information about past behaviors during the ADI-R also relies on the interviewer's ability to help the parent remember what else was happening when the child was 4 years old, which requires more time and interviewing skills.

Because the ASI is intended to identify individuals whose *current* behavior is consistent with a diagnosis of ASD, we chose to focus on behaviors that had been observed during the previous three-month period. This required us to consider different algorithms for children of different ages and language levels, given substantial evidence that the concordance between specific ASD symptoms and an ASD diagnosis varies according to developmental characteristics of the individual (e.g. Gray, Tonge, & Sweeney, 2008; Lord, Storoschuk, Rutter, & Pickles, 1993; Nordin & Gillberg, 1998; Ventola et al., 2006).

The current study was completed in two phases. As described in more detail below, initial development of the ASI (Measure Development Phase) was guided by analyses of existing ADI-R item data, past research on ASD instrument development, and focus groups with parents and expert clinicians. During the Initial Validation Phase, the ASI was administered to parents of children who then underwent comprehensive testing to confirm their diagnostic status of ASD or another disorder. Two versions of the ASI were developed: a preschool version for children 2 years, 0 months to 4 years, 11 months, and a school-age version for children 5 years, 0 months to 12 years, 11 months. This paper reports on the initial validation of the ASI School-age version.

Method

Measure Development Phase: Item Creation

Previously collected ADI-R scores from assessments of 3,126 children with ASD diagnoses and 471 with non-ASD diagnoses were obtained from existing datasets available through the University of Michigan Autism and Communication Disorders Center (UMACC), Kaiser Permanente Northern California, and the Simons Simplex Collection (SSC; Fischbach & Lord, 2010). Children were divided into six age by language groups (see Table 1) on the basis of previous literature about developmental differences in ASD symptom manifestation (Gotham, Risi, Pickles, & Lord, 2007; Lord & Pickles, 1996; Luyster et al., 2009). Language divisions were primarily based on ADOS modules, with the hope that we might be able to construct an instrument with more fine grained language divisions than the ADI-R (i.e., separations going beyond phrase speech or better vs. less than phrase speech). Age divisions were largely driven by existing scoring conventions for the ADI-R (i.e., certain items are only administered to parents of children under 10, regardless of language level).

ADI-R item scores are assigned on a 0–3 scale, with higher numbers indicating more definite presence or greater severity of symptoms. Within each age by language group, we examined the sensitivity and specificity of individual ADI-R items when either a cut-off of 1 was employed (i.e., comparing those who received a score of 0 to those who received a score of 1, 2, or 3), or when a cut-off of 2 was employed (i.e., comparing those who received a score of 0 or 1 to those who received a score of 2 or 3); a complete table of ADI-R item distributions in the development sample is available upon request from the corresponding author. These analyses provided information about how well each existing item discriminated between diagnostic groups within the various age by language groups. Information from these item analyses was also used to determine whether items with new content were needed for certain age by language groups, or whether existing items needed to be modified in particular ways to increase sensitivity and/or specificity.

We chose not to employ stringent criteria for determining whether or not an item would be adapted for inclusion in the ASI, because we wanted to ensure that concepts that are known to be diagnostically important would be represented to at least some extent in the initial draft of the ASI (recognizing the possibility that these items might later be omitted). In addition, because many of the age by language groups in the development sample contained small (and in some cases very small) numbers of non-ASD participants, it was not advisable to make decisions about item selection based only on numerical cut-offs. We also knew that most of the ADI-R items would need to undergo at least some modification to make them appropriate for the new format, and of course no empirical data existed for these modified items. Thus, decisions about which items to include were both empirically and conceptually based. The overall approach that guided ASI item selection was to use ADI-R item analyses within each age by language group to 1) identify items that provided good differentiation between ASD and non-ASD, 2) identify items that were clearly *not* likely to be useful for differentiating ASD from non-ASD, and 3) identify items that appeared to have potential utility and/or that were highly conceptually relevant, and then use information from the item distributions to make revisions to the items.

For the majority of ADI-R items, scores of 0, 1, 2, or 3 are determined based not only on the type or quality of the behavior, but also on the frequency and/or severity of the behavior. Because the ASI was designed to be quickly administered by examiners with minimal training, we attempted to write questions in which the quality/type of behavior was clearly defined by the question itself. We then chose a Likert scale of response options to elicit information about whether and to what extent that behavior was present. For example, the ASI *Pointing to Express Interest* item reads: *Some children point to request things. Others also point to show something of interest, such as pointing to an airplane in the sky. How often does* _____ *use his/her finger to point out something of interest?* The hope was that specificity of the original ADI-R item (in this case *Pointing*) would be maintained or enhanced by clearly operationalizing the behavior of interest in each question, and sensitivity would be maintained or enhanced by the availability of multiple frequency ratings.

Information from the ADI-R item distributions was used to identify critical cut-points that could be used to maximize sensitivity and specificity for items that showed promise or were of particular clinical importance but that required more substantial content revisions. For example, if an ADI-R item provided optimal differentiation at a cut-off of 2, then the ASI item was modified to highlight the features embedded in an ADI-R code of 2 or 3. It was also sometimes necessary to break apart complex ADI-R items into multiple questions to ensure that the behavior of interest could be reasonably captured by the question itself. For example, the ADI-R item Conversation is rated with respect to several aspects of conversational ability, including initiating conversation, responding to others' conversational bids, and conversing about a range of topics, etc. For the ASI, each of these skills was reflected in its own item. In modifying ADI-R items for inclusion in the ASI, we incorporated input from experienced clinicians who were also trainers on the ADI-R and Autism Diagnostic Observation Schedule (Lord et al., 2000) about how items could be re-worded so that they reflected the most salient features of the original ADI-R item. These

discussions, together with analyses of ADOS items and literature reviews, also informed the creation of new ASI items.

Measure Development Phase: Construction of the ASI

Similar to the ADI-R, there are additional questions in the ASI about use of language for parents of children with flexible phrase speech or fluent speech (i.e., children who meet the ASI definition of "verbal"). In the School-age form, some questions are only administered to parents of children ages 5 to 10, whereas other questions are administered to parents of children ages 10 to 12, based on previously described analyses of ADI-R items within age and language level groups. Once a draft of each form was completed, parent focus groups and clinician meetings were held to solicit feedback about the ASI items. Two parent focus groups were held. Each group included between 6 and 8 parents who were selected to represent parents of children of different ages, both sexes, and with different levels of verbal ability. Parents were asked to read through the draft ASI items and provide feedback about items that were confusing. They were also asked for suggestions about symptoms their children exhibited that were not reflected in the ASI. Clinicians groups included 2 or 3 expert clinicians and 1 or 2 of the ASI authors. These clinicians were experienced in ASD diagnostic assessment and were trainers on the ADI-R and ADOS, meaning that they were familiar with the original items and constructs from which the ASI items were derived. Clinicians were asked to review the draft instrument and provide feedback about items that were difficult to understand in terms of what information they were attempting to ascertain. They also provided suggestions about symptoms that they thought were not adequately represented in the ASI. Feedback from parents and clinicians was used to make additional modifications to the draft instrument.

Initial Validation Phase: Participants

Participant characteristics of the validation sample are presented in Table 2. Children with ASD were recruited primarily from clinic referrals and ongoing research projects at the University of Michigan Autism and Communication Disorders Center (UMACC), whereas children with non-ASD diagnoses were recruited mostly from the Divisions of Developmental and Behavioral Pediatrics and Behavioral Medicine and Clinical Psychology at Cincinnati Children's Hospital. Based on a growing body of literature that children with certain non-ASD diagnoses are often misclassified by ASD screening and diagnostic measures, including the ADI-R (Chandler et al., 2007; Gray et al., 2008; Lord et al., 1993; Molloy, Murray, Akers, Mitchell, & Manning-Courtney, 2011; Towbin, Pradella, Gorrindo, Pine, & Leibenluft, 2005), non-ASD controls were specifically recruited from diagnostic groups characterized by high levels of ASD symptom overlap (i.e., intellectual disability (ID), Attention-Deficit/Hyperactivity Disorder (ADHD), language disorder, anxiety/mood disorder). Children with known genetic syndromes or abnormalities were excluded from the initial validation study, because these children would normally be excluded from genetics investigations, and the ASI was originally conceived as a measure to quickly ascertain participants for genetics studies.

Initial Validation Phase: Procedure

Prior to the in-person assessment, parents completed both the ASI over the telephone and a questionnaire packet (the ASI and the questionnaire packet were administered in counter-balanced order across participants). The ASI includes questions about previous diagnoses, but these questions are asked at the end of the interview, so interviewers had no information about diagnostic history while administering items about specific behaviors. The ASI was administered by research assistants with limited knowledge of ASD and no prior training on or exposure to the ADI-R. ASI item distributions were first examined mid-way through data collection, at which point the interview was longer than desired, taking closer to 30 minutes in many cases. Items that appeared to be performing poorly based on preliminary analyses (see below) were dropped from the instrument for the remainder of data collection. This had the effect of significantly reducing the administration time. For all the interviews performed during the study period, the mean administration time was 23.30 minutes (SD=5.72 minutes; range: 14–54 minutes) for parents of verbal children and 18.95 minutes (SD=4.78; range: 12–32 minutes) for parents of nonverbal children. In its current form, the interview takes approximately 20 minutes or less in most cases. In addition to tracking time of administration, research assistants administering the ASI were also asked to carefully record any instances where parents required clarification or did not provide a direct response to the question. Based on these examples, a list of "Standard Prompts" was generated (see Supplementary File A) and used by the research assistants to increase ease and standardization of ASI administration throughout data collection.

The questionnaire packet included measures designed to assess ASD symptoms and measures relevant for establishing non-ASD diagnoses, including the Conners' Parent Rating Scale-Revised (CPRS-R; Conners, Sitarenios, Parker, & Epstein, 1998), the Spence Children's Anxiety Scale (SCAS; Spence, 1998), and the Child Behavior Checklist (CBCL; Achenbach & Rescorla, 2001). The parent and child in-person assessments were each completed in approximately 3 to 4 hours. Parents were administered the ADI-R and the Vineland-II (Sparrow, Cicchetti, & Balla, 2005). Children completed a cognitive test, either the Mullen Scales of Early Learning (MSEL; Mullen, 1995) or the Differential Ability Scales-2nd edition (DAS-II; Elliott, 2007), the ADOS, and additional language testing as necessary to determine language impairment (e.g., subtests of the Clinical Evaluation of Language Fundamentals-4th edition; CELF-4, (Semel, Wiig, & Secord, 2003); Peabody Picture Vocabulary Test-4th edition; PPVT-4, (Dunn & Dunn, 2007). Children over age 8 who were capable of self-reporting completed the Multidimensional Anxiety Scale for Children (MASC; March, Parker, Sullivan, Stallings, & Conners, 1997).

Although children were only eligible for the study if they had a previous diagnosis of ASD or one of the targeted non-ASD diagnoses, or if a parent or professional had significant concerns about ASD, the designation of ASD vs. non-ASD used for the current analyses was based on the comprehensive diagnostic assessment conducted as part of the research project. For example, if a child was recruited into the study because of a previous diagnosis of ADHD, but he/she ultimately received a diagnosis of ASD following our assessment, then he/she was classified as ASD for the current study.

Clinicians conducting the in-person assessments were blind to algorithm scores and classifications from the standardized diagnostic instruments until after they had assigned a best-estimate clinical diagnosis. In addition, whenever possible (72%), separate clinicians were assigned to conduct the parent and child in-person assessments, and these clinicians were kept blind to all previous diagnostic information about the participant until after the evaluations were completed. Introductory questions about diagnosis that are normally included in the ADI-R were moved to the end of the interview so that the clinician could assign ADI-R item ratings without knowledge of the child's previous diagnoses. However, in 19% of parent assessments, the examiner was given some information by the parent or another professional about the child's diagnostic status prior to beginning the ADI-R. In 14% of child assessments, the examiner was not blind to the child's previous diagnoses because he/she had also conducted the ADI-R administration for that participant. ASI scores were not reviewed by the clinicians at any point during the diagnostic process. Although individual ASI items were examined on a group level during preliminary data analyses mid-way through the project (see below), individual participants were not viewed in relation to their ASI scores (before, during, or after the study diagnostic assessment) until after data collection was complete.

Following the completion of all measures, clinicians met to discuss their impressions and assign a consensus clinical best estimate diagnosis and corresponding diagnostic certainty rating. Impressions from the ADI-R and ADOS were considered together with information from other measures, but algorithm total scores were not calculated until after the best-estimate clinical diagnosis had been assigned. This was done to ensure that best-estimate diagnosis was not tied specifically to ADI-R and/or ADOS classifications (though clinicians obviously would have been able to draw on impressions obtained during the administrations and/or may have remembered individual item scores they had assigned as part of their administrations).

Analyses

Each ASI item is rated on a four point scale corresponding to the response options: Not at all; Occasionally, Often; Very Frequently. The parent is asked to select one response that best describes his/her child's behavior during the past 3 months. ASI item scores ranged from 0 to 3, with higher scores indicating a greater level of abnormality on the item. Thus, for items assessing behaviors that are expected to occur (e.g., making eye contact, responding to name, answering questions), a score of 3 corresponds to "Not at all," whereas for items assessing behaviors that are not expected to occur (e.g., sensory abnormalities, repetitive mannerisms), a score of 3 corresponds to "Very frequently." Distributions and odds ratios derived from ordinal regressions (best-estimate diagnosis of ASD vs. non-ASD predicting ASI item scores) were examined for each individual ASI item. Analyses were conducted separately for children who were classified as "verbal" (i.e., reported by their parents to use simple phrases or complex sentences on a daily basis) and those classified on the ASI as "nonverbal" (reported to use single words or no words).

At the conclusion of data collection, items that had been retained following preliminary analyses (46 for verbal children under 10, 42 for verbal children over 10, 31 for nonverbal

children under 10, and 27 for nonverbal children over 10) were rank ordered by odds ratios, and item distributions were examined. In general, only items with odds ratios over 2 were considered for inclusion in the algorithm. ASI items were judged as acceptable for algorithm inclusion if fewer than 25% of children with ASD had received a 0 and fewer than 25% of children with non-ASD diagnoses had received a 3. As has been reported in previous ASD measure development efforts (see Gotham et al., 2007), repetitive behavior items had lower sensitivity. To increase sensitivity of these items, alternative scores were created so that the higher of the two scores between Hand Mannerisms and Complex Mannerisms and between Unusual Preoccupations and Circumscribed Interests was summed in the algorithm total score.

Items selected for inclusion in the algorithms were totaled, and Area under the curve (AUC) from receiver operating characteristic (ROC) analyses was used to measure overall agreement between the ASI algorithm total scores and best-estimate clinical diagnosis.

Results

As indicated above, we had initially hoped to construct separate algorithms pertaining to children with no speech vs. single word speech vs. phrase speech vs. fluent speech, mirroring the age by language groups depicted in Table 1. However, comparing parent-reported language level to clinician-selected ADOS module indicated that parents were not able to accurately report at this level of detail. On the other hand, when responses from the ASI language level item were collapsed (no speech or single word speech vs. phrase speech or fluent speech), agreement was high. As shown in Table 2, 91% of children in the ASI nonverbal sample received a Module 1 (designed for children who are not yet using flexible phrase speech), and 94% of the ASI verbal sample received either a Module 2 (for children with flexible phrase speech) or Module 3 (for children who speak in complex sentences). In addition, whereas 264 (99%) of the 268 children in the ASI verbal sample were administered the DAS-II to assess their cognitive abilities, 81% (17 out of 21) children in the ASI nonverbal sample were administered the MSEL. This reflects the fact that children in the nonverbal sample also had significantly lower cognitive abilities.

Verbal Algorithm

On the basis of the criteria detailed above, 29 items were selected for the School-age verbal algorithm (see Table 3). To calculate an ASI total score, item scores from the selected algorithm items were summed. Scores ranged from 2 to 63 for children with non-ASD diagnoses (M=35.29, SD=13.32) and from 17 to 81 for children with ASD diagnoses (M=50.10, SD=12.15); F(1, 266)=100.65, p<.001. Score distributions by diagnosis are presented in Figure 1.

Internal consistency of the algorithm items, as measured using Cronbach's alpha, was excellent (.92) (Cicchetti, 1994). Correlations were also calculated between all verbal algorithm items and age, VIQ, and NVIQ and did not exceed .29.

ROC analysis showed that overall agreement between the ASI verbal algorithm total score and best-estimate clinical diagnosis of ASD vs. non-ASD was good; AUC=.80, 95%

confidence interval=.75–.86 (Tape, 1999). A cut-off of 38 yielded a sensitivity of .87 and a specificity of .62. Positive predictive value (PPV) for a cutoff of 38 was .72 while negative predictive value (NPV) was .81. Table 4 shows the sensitivity and specificity of the ASI verbal algorithm as compared to and in combination with the ADI-R and ADOS.

Nonverbal Algorithm

Because of the exceedingly small size of the nonverbal non-ASD group, we did not calculate the AUC, as we did not feel comfortable recommending an algorithm cut-off based on such a small sample. However, in our sample, 13 items were identified out of a pool of 27 (for children over 10) or 31 (for children under 10) as being the best differentiating items. These items are shown in Table 3 to allow comparison with other samples. Summed scores on these 13 items ranged from 4 to 18 for the non-ASD group (M=12.90, SD=6.07) and from 14 to 34 for the ASD group (M=26.29, SD=5.68); F(1, 19)=17.64, p<.001.

Test-retest reliability

Test-retest reliability of the algorithms was assessed in a separate sample of 10 children with ASD and non-ASD diagnoses recruited after the conclusion of the initial validation study (due to time constraints during the initial study period). The ASI was administered twice over the course of 2 weeks, with mean time interval of 11.8 days (range 8–15). In all cases, the two interviews were conducted by different interviewers. Intra-class correlation coefficient (ICC) estimates were interpreted according to the guidelines suggested by Cicchetti (1994) (excellent: 0.75, good: 0.60–0.74, fair: 0.40–0.59, poor: <0.40). ICC was 0.96, with 95% confidence interval of 0.85–0.99, p<0.001. When the three nonverbal participants were excluded, ICC for the verbal algorithm was 0.97, with 95% confidence interval of 0.82–0.99, p=<0.001.

Discussion

Results of this initial validation study indicate that the ASI School-Age may serve as a useful tool to more quickly classify children with ASD for research purposes. The verbal algorithm showed acceptable predictive validity in comparison to the most commonly used ASD questionnaires and checklists (see Table 2 of Charman & Gotham, 2013), and did so against a non-ASD comparison sample purposefully recruited to have symptom overlap with ASD. An advantage of the ASI compared to other brief ASD symptom measures is that it was developed for use over the phone. Telephone interviews offer the opportunity to clarify certain questions and routing rules (e.g., the interviewer only asks questions relevant to the individual's age and language level), as well as to record additional information that the respondent wishes to provide (Tsuchiya et al., 2013; Ward-King, Cohen, Penning, & Holden, 2010). For research studies, some families may also be easier to reach by telephone than by mail.

When used in combination with the ADOS ASD cut-offs, the ASI yielded similar levels of sensitivity and specificity in this sample than the much lengthier ADI-R. However, unlike the ADI-R or other semi-structured parent interviews, such as the Diagnostic Interview for Social and Communication Disorders (DISCO; Wing, Leekam, Libby, Gould, & Larcombe,

2002) or Developmental, Dimensional and Diagnostic Interview (3Di; Skuse et al., 2004), the ASI yields only a yes vs. no classification. The ASI does not provide the level of detail that would be necessary to describe the symptom profile or severity of a child with ASD. Furthermore, even if the only goal is to classify a participant as ASD or non-ASD, the ASI School-age algorithms did not reach a high enough level of specificity to justify its use in isolation. Some type of standardized direct observation, such as the ADOS, is necessary to improve the accuracy of classification, though that is true of all brief parent-report instruments available at this time.

Research in measurement of ASD symptoms clearly indicates that no one measure can "do it all." Different measures are required for different purposes, and individual study and clinical goals must be carefully considered when selecting assessment tools (Lord & Jones, 2012). Researchers are under increasing pressure to quickly ascertain large groups of children with ASD, but accuracy and speed in diagnostic assessment overlap to only a limited extent. As shown in Figure 2, results of this study suggest that it may be possible to relatively quickly screen out a sizable proportion of children with non-ASD diagnoses using the ASI School-age or a similar parent report instrument (e.g., for the purposes of case confirmation). However, the largely overlapping distributions of ASI scores in children with ASD (17 – 81) vs. non-ASD diagnoses (2 – 63) further illustrate the relatively poor discriminative ability of the ASI when used in isolation. Thus, some sort of in-person assessment by trained examiners is likely to be necessary to achieve accuracy levels that approximate those obtained through comprehensive clinical diagnostic assessments. This may be even more likely to be true for children who are referred specifically for ASD-related concerns, unlike the children in the non-ASD group in the current study, who were mostly comprised of children with previously established non-ASD diagnoses. Given that a direct assessment of some kind may be necessary in any case, it is possible that researchers might have as much success using parent-reported previous diagnosis of ASD as a means of initially screening children into a study, rather than going through the process of administering the ASI or a similar measure of ASD symptoms.

Limitations

The current study represents only a first attempt at validating the ASI School-age. Replication in independent samples, including clinic referral or general population samples, is necessary in order to understand its utility for case confirmation or other purposes. Unfortunately, our sample was not of sufficient size that we could identify algorithm items in one half and test on the other half, which would have been an optimal strategy. This lack of cross-validation is a major weakness of the current study. For the verbal items, it will be necessary to verify that items selected for the verbal algorithm based on our data demonstrate diagnostic validity in other samples. For the nonverbal items, our nonverbal sample was so small that it was not possible to calculate the AUC or corresponding sensitivity and specificity for the items that were found to differentiate best in this sample. Thus, more research is needed to understand how well these nonverbal items perform in other samples. Difficulties recruiting nonverbal participants with non-ASD diagnoses who are over the age of 5 is a testament to the relative rarity of such extreme language difficulties in children who do not have ASD and/or a known genetic syndrome (see also Lord &

Pickles, 1996). Also, the relatively small number of ASD participants with low verbal abilities in this group may reflect changes in the epidemiology of ASD, with more and more children with ASD acquiring functional speech by school age (Tager-Flusberg, Paul, & Lord, 2005). Nevertheless, particularly for research purposes, there is a need for instruments that can accurately differentiate ASD from non-ASD among children with severe language impairments. Such children usually have co-occurring ID and may also have genetic syndromes; therefore, our exclusion of children with known genetic syndromes represents another limitation of these data, as ASD symptom measures are increasingly being utilized in children with identified genetic abnormalities.

Conclusions

The ASD field has seen tremendous growth in the development, refinement, and application of standardized diagnostic tools for clinical practice and research. At this point, a priority is to determine how best to employ different tools, or combinations of tools, for different purposes. The ASI School-Age is a newly developed parent interview that may serve as a useful option for clinicians and researchers who wish to employ phone screening to identify children at high risk for ASD and then conduct in-person assessments to verify ASD caseness.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

Achenbach, TM., Rescorla, LA. Child Behavior Checklist. Burlington, VT: University of Vermont Research Center for Children, Youth & Families; 2001.

American Psychiatric Association. Diagnostic and statistical manual of mental disorders. Fifth. Washington, DC: Author; 2000. Revised

Anderson DK, Liang JW, Lord C. Predicting young adult outcome among more and less cognitively able individuals with autism spectrum disorders. Journal of Child Psychology and Psychiatry. 2014; 55(5):485–494. DOI: 10.1111/jcpp.12178 [PubMed: 24313878]

Bishop SL, Gahagan S, Lord C. Re-examining the core features of autism: A comparison of autism spectrum disorder and fetal alcohol spectrum disorder. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2007; 48(11):1111–1121. DOI: 10.1111/j.1469-7610.2007.01782.x

Chandler S, Charman T, Baird G, Simonoff E, Loucas T, Meldrum D, Pickles A. Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. Journal of Amer Academy of Child & Adolescent Psychiatry. 2007; 46(10):1324.doi: 10.1097/chi.0b013e31812f7d8d

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. DOI: 10.1111/j.1475-3588.2012.00664.x [PubMed: 23539140]

Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment. 1994; 6(4):284–290. DOI: 10.1037/1040-3590.6.4.284

Conners C, Sitarenios G, Parker J, Epstein J. The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability, and criterion validity. Journal of Abnormal Child Psychology. 1998; 26(4):257–268. [PubMed: 9700518]

Dunn L, Dunn D. PPVT-4: Peabody Picture Vocabulary Test. 2007

Elliott, CD. Differential Ability Scales. Second. New York: Harcourt Brace Jovanovich; 2007.

Fein D, Barton M, Eigsti IM, Kelley E, Naigles L, Schultz RT, Tyson K. Optimal outcome in individuals with a history of autism. Journal of Child Psychology and Psychiatry. 2013; 54(2):195–205. DOI: 10.1111/jcpp.12037 [PubMed: 23320807]

Fischbach G, Lord C. The Simons simplex collection: A resource for identification of autism genetic risk factors. Neuron. 2010; 68(2):192–195. DOI: 10.1016/j.neuron.2010.10.006 [PubMed: 20955926]

Gotham, K., Bishop, SL., Lord, C. Diagnosis of autism spectrum disorders. In: AmaralGeshwind, Dawson, editors. Autism Spectrum Disorders. New York: Oxford University Press; 2011. p. 30-43.

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. DOI: 10.1007/s10803-006-0280-1 [PubMed: 17180459]

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. DOI: 10.1007/s10803-007-0432-y [PubMed: 17690967]

Hus V, Bishop S, Gotham K, Huerta M, Lord C. Factors influencing scores on the social responsiveness scale. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2013; 54(2):216–224. DOI: 10.1111/j.1469-7610.2012.02589.x

Hus V, Taylor A, Lord C. Telescoping of caregiver report on the Autism Diagnostic Interview–Revised. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2011; 52(7):753–760. DOI: 10.1111/j.1469-7610.2011.02398.x

Jones RM, Risi S, Wexler D, Anderson D, Corsello C, Pickles A, Lord C. How interview questions are placed in time influences caregiver description of social communication symptoms on the ADI-R. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2015; 56(5):577–585. DOI: 10.1111/jcpp.12325

Lord, C., Corsello, C. Diagnostic instruments in autistic spectrum disorders. In: Volkmar, FR.Klin, A., Paul, R., editors. Handbook of autism and pervasive developmental disorders. 3. 2005.

Lord C, Jones RM. Annual research review: re-thinking the classification of autism spectrum disorders. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2012; 53(5):490–509. DOI: 10.1111/j.1469-7610.2012.02547.x

Lord C, Pickles A. Language level and nonverbal social-communicative behaviors in autistic and language-delayed children. Journal of the American Academy of Child and Adolescent Psychiatry. 1996; 35(11):1542–1550. DOI: 10.1097/00004583-199611000-00024 [PubMed: 8936922]

Lord C, Risi S. Frameworks and methods in diagnosing autism spectrum disorders. Mental retardation and developmental disabilities research reviews. 1998; 4(2):90–96. DOI: 10.1002/(SICI)1098-2779(1998)4:2<90::AID-MRDD5>3.0.CO;2-0

Lord C, Risi S, Lambrecht L, Cook EH, Leventhal BL, DiLavore PC, Rutter M. The autism diagnostic observation schedule—generic: A standard measure of social and communication deficits

associated with the spectrum of autism. Journal of Autism and Developmental Disorders. 2000; 30(3):205–223. doi:0162-3257/00/0600-0205. [PubMed: 11055457]

Lord C, Rutter M, 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(5):659–685. DOI: 10.1007/BF02172145 [PubMed: 7814313]

Lord C, Storoschuk S, Rutter M, Pickles A. Using the ADI-R to diagnose autism in preschool children. Infant Mental Health Journal. 1993; 14:234–234. DOI: 10.1002/1097-0355(199323)14:3<234::AID-IMHJ2280140308>3.0.CO;2-F

Luyster R, Gotham K, Guthrie W, Coffing M, Petrak R, Pierce K, Lord C. The Autism Diagnostic Observation Schedule-toddler module: a new module of a standardized diagnostic measure for autism spectrum disorders. Journal of Autism and Developmental Disorders. 2009; 39(9):1305–1320. DOI: 10.1007/s10803-009-0746-z [PubMed: 19415479]

March J, Parker J, Sullivan K, Stallings P, Conners C. The Multidimensional Anxiety Scale for Children (MASC): factor structure, reliability, and validity. Journal of the American Academy of Child and Adolescent Psychiatry. 1997; 36(4):554–565. [PubMed: 9100431]

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. DOI: 10.1177/1362361310379241 [PubMed: 21339248]

Mullen, EM. Mullen scales of early learning. Circle Pines, MN: American Guidance Service; 1995.

Nordin V, Gillberg C. The long-term course of autistic disorders: Update on follow-up studies. Acta Psychiatrica Scandinavica. 1998; 97(2):99–108. DOI: 10.1111/j.1600-0447.1998.tb09970.x [PubMed: 9517902]

Richler J, Bishop SL, Kleinke JR, Lord C. Restricted and repetitive behaviors in young children with autism spectrum disorders. Journal of Autism and Developmental Disorders. 2007; 37(1):73–85. DOI: 10.1007/s10803-006-0332-6 [PubMed: 17195920]

Risi S, Lord C, Gotham K, Corsello C, Chrysler C, Szatmari P, Pickles A. Combining information from multiple sources in the diagnosis of autism spectrum disorders. Journal of the American Academy of Child and Adolescent Psychiatry. 2006; 45(9):1094–1103. DOI: 10.1097/01.chi.0000227880.42780.0e [PubMed: 16926617]

Semel, E., Wiig, EH., Secord, WA. Clinical Evaluation of Language Fundamentals - Fourth Edition (CELF - 4). Pearson Assessments; 2003.

Skuse DH, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, Place M. The developmental, dimensional and diagnostic interview (3di): a novel computerized assessment for autism spectrum disorders. Journal of the American Academy of Child and Adolescent Psychiatry. 2004; 43(5):548–558. DOI: 10.1097/00004583-200405000-00008 [PubMed: 15100561]

Sparrow, SS., Cicchetti, DV., Balla, DA. Vineland adaptive behavior scales, (Vineland-II). Circle Pines, MN: American Guidance Services; 2005.

Spence SH. A measure of anxiety symptoms among children. Behaviour Research and Therapy. 1998; 36(5):545–566. doi:http://dx.doi.org/10.1016/S0005-7967(98)00034-5. [PubMed: 9648330]

Tager-Flusberg, H., Paul, P., Lord, C. Handbook of autism and pervasive developmental disorders. 3. Hoboken, NJ: John Wiley; 2005. Language and communication in autism; p. 335-364.

Tape, TG. The area under an ROC curve. Interpreting Diagnostic Tests. 1999. Retrieved from http://gim.unmc.edu/dxtests/Default.htm

Towbin KE, Pradella A, Gorrindo T, Pine DS, Leibenluft E. Autism spectrum traits in children with mood and anxiety disorders. Journal of Child and Adolescent Psychopharmacology. 2005; 15(3):452–464. DOI: 10.1089/cap.2005.15.452 [PubMed: 16092910]

Tsuchiya KJ, Matsumoto K, Yagi A, Inada N, Kuroda M, Inokuchi E, Takei N. Reliability and validity of autism diagnostic interview-revised, Japanese version. Journal of Autism and Developmental Disorders. 2013; 43(3):643–662. DOI: 10.1007/s10803-012-1606-9 [PubMed: 22806002]

Ventola P, Kleinman J, Pandey J, Barton M, Allen S, Green J, Fein D. Agreement among four diagnostic instruments for autism spectrum disorders in toddlers. Journal of Autism and Developmental Disorders. 2006; 36(7):839–847. DOI: 10.1007/s10803-006-0128-8 [PubMed: 16897398]

Ward-King J, Cohen IL, Penning H, Holden JJ. Brief report: telephone administration of the autism diagnostic interview–revised: reliability and suitability for use in research. Journal of Autism and Developmental Disorders. 2010; 40(10):1285–1290. DOI: 10.1007/s10803-010-0987-x [PubMed: 20195733]

Wing L, Leekam SR, Libby SJ, Gould J, Larcombe M. The diagnostic interview for social and communication disorders: Background, inter-rater reliability and clinical use. Journal Of Child Psychology And Psychiatry, And Allied Disciplines. 2002; 43(3):307.doi: 10.1111/1469-7610.00023

Yonan AL, Alarcon M, Cheng R, Magnusson PK, Spence SJ, Palmer AA, Gilliam TC. A genomewide screen of 345 families for autism-susceptibility loci. American Journal of Human Genetics. 2003; 73(4):886–897. [PubMed: 13680528]

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. DOI: 10.1177/1362361313516199 [PubMed: 24413849]
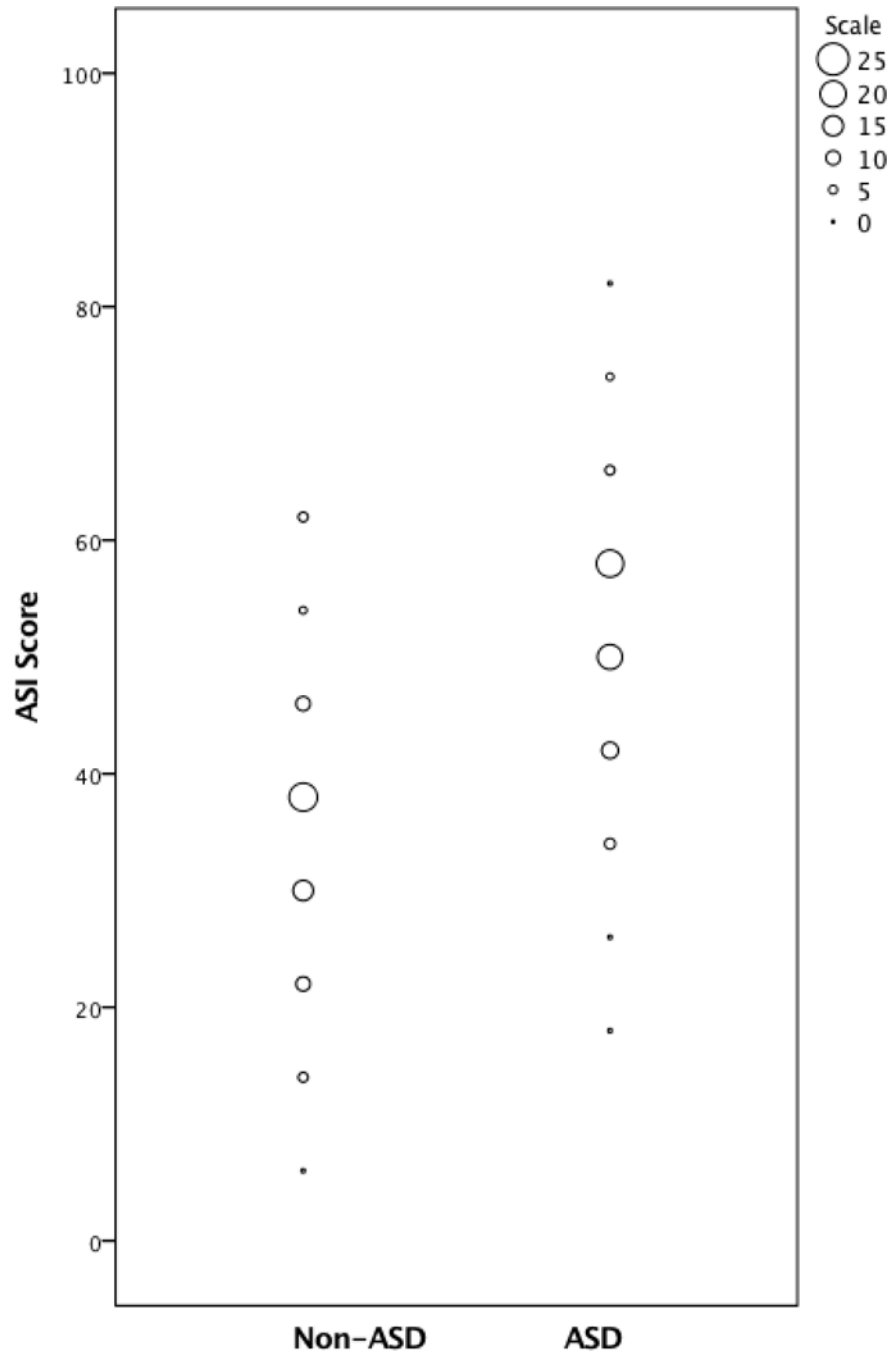
Figure 1.
ASI scores by best-estimate diagnosis
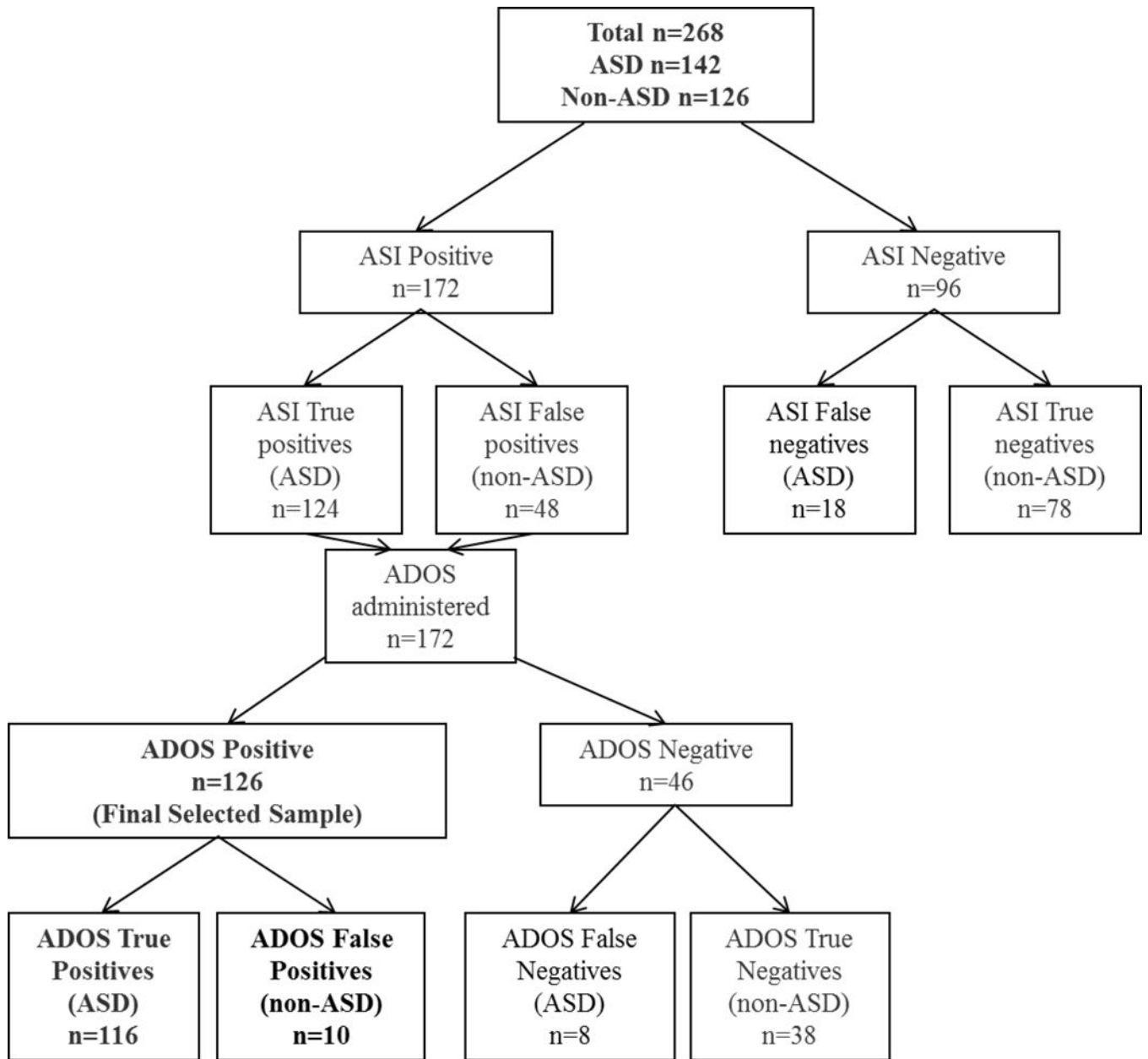
Figure 2.
Hypothetical use of the ASI in a case confirmation scenario

Table 1

Development Sample: Demographics

Age Range	2–4 years	5–17 years	2–4 years	5–17 years	2–10 years	10–17 years
Language Level	Single words or less	Single words or less	Phrases	Phrases	Fluent	Fluent
n ASD (%)	651 (88)	283 (92)	256 (73)	530 (92)	748 (87)	558 (90)
n Non-ASD (%)	91 (12)	24 (8)	134 (27)	47 (8)	111 (13)	64 (10)
n Male (%)	588 (77)	249 (81)	404 (80)	471 (81)	730 (84)	522 (84)
Mean Age (SD)	3.3 (0.8)	8.0 (2.9)	3.9 (0.7)	7.9 (2.8)	7.5 (1.6)	12.9 (2.2)
Mean VIQ (SD)	38.5 (20.3)	26.9 (16.1)	78.5 (20.6)	57.5 (21.1)	96.0 (19.1)	94.7 (25.1)
Mean NVIQ (SD)	61.8 (21.5)	44.4 (21.1)	86.6 (21.6)	71.5 (22.6)	98.1 (18.4)	94.5 (21.6)

Note. ASD=Autism Spectrum Disorder. VIQ=Verbal Intelligence Quotient. NVIQ=Nonverbal Intelligence Quotient.

Table 2

Validation Sample: Demographics

	ASD Verbal (n=142)		Non-ASD Verbal (n=126)		ASD Nonverbal (n=17)		Non-ASD Nonverbal (n=4)	
	Mean	n	Mean	n	Mean	n	Mean	n
Male		106 (74.6%)		84 (66.7%)		11 (64.7%)		2 (50%)
Age in years	8.8 ± 2.3		8.5 ± 2.2		7.5 ± 2.4		8.3 ± 3.4	
Ethnicity								
Caucasian		**109 (76.7%)**		**71 (56.3%)**		12 (70.6%)		4 (100%)
African American		**13 (9.2%)**		**38 (30.2%)**		1 (5.9%)		0 (0%)
More than one Race		12 (8.5%)		9 (7.1%)		3 (17.6%)		0 (0%)
Other or Unknown		8 (5.6%)		8 (6.4%)		1 (5.9%)		0 (0%)
Non-ASD Diagnosis								
Language Disorders		–		23 (18.3%)		–		0 (0%)
ADHD		–		56 (44.4%)		–		0 (0%)
Mood/Anxiety Disorder		–		24 (19%)		–		0 (0%)
Intellectual Disability		–		20 (15.9%)		–		4 (100%)
ADOS								
Module 1		**15 (10.6%)**		**0 (0%)**		16 (94%)		3 (75%)
Module 2		17 (12%)		16 (13%)		1 (6%)		1 (25%)
Module 3		110 (77.4%)		110 (87%)		0 (0%)		0 (0%)
Mean VIQ	*85.9 ± 24.8*		*93.2 ± 19.3*		20.3 ± 14.3		33.5 ± 23.9	
Mean NVIQ	90.7 ± 23.7		93.3 ± 18.1		32.6 ± 17.2		36.0 ± 14.0	

Note. **Bold**=$p<.001$; *Italics*=$p<.01$; Verbal level of validation groups across the top row was based on parent report during the ASI.

ASD=Autism Spectrum Disorder. ADHD=Attention-Deficit/Hyperactive Disorder. ADOS=Autism Diagnostic Observation Schedule. VIQ=Verbal Intelligence Quotient. NVIQ=Nonverbal Intelligence Quotient

Table 3

Final ASI School-Age Verbal Algorithm Items and Nonverbal "Best Differentiating" Items (algorithm not yet available)

Verbal Algorithm	Best Differentiating Nonverbal Items
Social Communication	
- Direct Gaze - Pointing (to express interest) - Gesture - Nodding - Sharing - Response to Distress - Appropriate Social Responses - Inappropriate Social Responses - Asking for Information - Starting Conversations - Responding to Conversational Leads - Responding to Questions - Social Chat - Preservation - Odd Phrases - Intonation - Inappropriate Facial Expressions	- Direct Gaze - Pointing (to express interest) - Gesture - Nodding - Sharing - Response to Distress - Appropriate Social Responses - Inappropriate Social Responses - Response to Name
Peer Interaction	
- Time Alone - Group Play with Peers - Approaching Children - Response to Children - Awkard Interactions - Socializing with Peers - Maintaining Friendships	- Time Alone - Group Play with Peers
Restricted and Repetitive Behaviors	
- Circumscribed Interests OR Unusual Preoccupations (higher score) - Sensory Aversions - Unusual Sensory Interests - Initiation of Appropriate Activities - Routines	- Circumscribed Interests OR Unusual Preoccupations (higher score) - Hand/Finger OR Other Complex Mannerisms (higher score)

Table 4

Validation Sample: Sensitivity and Specificity of Individual Instruments and Instrument Combinations

	ASD n=142			Non-ASD n=126			ASD vs Non-ASD		
	Sensitivity	95% CI		Specificity	95% CI		PPV	95% CI	
ASI Verbal School-age (Cutoff=38)	87	81	92	62	53	70	72	65	79
ADI-R *ASD* * cutoff	89	83	94	67	58	75	75	68	82
ADI-R *Autism* cutoff	76	68	83	87	79	92	86	79	92
ADOS *ASD* cutoff	94	89	98	79	70	85	83	77	89
ADOS *Autism* cutoff	82	74	88	87	80	93	88	81	93
ASI Verbal School-age (Cutoff= 38) + ADOS (*ASD* Cut-off)	82	74	88	92	86	96	92	86	96
ADI-R (*Autism* cut-off) + ADOS (*ASD* cut-off)	71	63	78	94	89	98	94	87	97
ADI-R (*ASD* cut-off) * + ADOS (*ASD* cut-off)	84	77	89	90	83	94	90	84	95

Note.

*
from Hus et al. (2013), following CPEA guidelines. ASD=Autism Spectrum Disorder. ASI=Autism Screening Interview. ADI-R=Autism Diagnostic Interview-Revised. ADOS=Autism Diagnostic Observation Schedule