



HHS Public Access

Author manuscript

J Am Acad Child Adolesc Psychiatry. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

J Am Acad Child Adolesc Psychiatry. 2016 April ; 55(4): 301–309. doi:10.1016/j.jaac.2016.01.011.

Development of an Objective Autism Risk Index Using Remote Eye Tracking

Dr. Thomas W. Frazier, PhD,

Mr. Eric W. Klingemier, BA,

Ms. Mary Beukemann [Research Coordinator],

Dr. Leslie Speer, PhD, NCSP,

Dr. Leslie Markowitz, PsyD,

Dr. Sumit Parikh, MD,

Dr. Steven Wexberg, MD,

Dr. Kimberly Giuliano, MD,

Dr. Elaine Schulte, MD, MPH,

Dr. Carol Delahunty, MD,

Correspondence to: Thomas W. Frazier, PhD, Center for Autism, Cleveland Clinic, 2801 Martin Luther King Jr. Drive CRS10, Cleveland, OH 44104; fraziet2@ccf.org.

This article is discussed in an editorial by Dr. Frederick Shic on page xx.

Clinical guidance is available at the end of this article.

Supplemental material cited in this article is available online.

Drs. Frazier and Youngstrom served as the statistical experts for this research.

Supplemental material cited in this article is available online.

The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Authors' contributions statement: T.W.F. and E.W.K. designed the present study. T.W.F. obtained funding to support data collection and analyses. T.W.F. supervised data collection but was not directly involved with patients during the collection process. T.W.F., E.A.Y., A.Y.H., and M.S.S. supervised interpretation of the study. T.W.F. conducted data management and data analyses. All authors contributed to writing and revision.

Disclosure: Dr. Frazier has received federal funding or research support from, acted as a consultant to, received travel support from, and/or received a speaker's honorarium from the Simons Foundation, the Ingalls Foundation, Forest Laboratories, Ecoeos, IntegraGen, Kugona LLC, Shire Development, Bristol-Myers Squibb, the National Institutes of Health, and the Brain and Behavior Research Foundation. Dr. Parikh has received research funding from Edison Pharmaceuticals and the National Institutes of Health and has served on the advisory board without funding for Stealth Pharmaceuticals, the United Mitochondrial Disease Foundation, the Cyclic Vomiting Syndrome Association, and the International Foundation for CDKL5 Research. Dr. Eng has served as a member of the external advisory boards of N-of-One, the Center for Personalized Medicine, Mission Health, Asheville, NC, and CareSource, and as an unpaid member of the external advisory boards of EcoEos and Medical Mutual of Ohio. Dr. Manos has received research support and served in an advisory role for Shire Development Inc. Dr. Hardan has received research funding from Forest Pharmaceuticals and Bristol-Myers Squibb and has served as a consultant to IntegraGen. Dr. Youngstrom has served as a consultant to Otsuka, Lundbeck, and Pearson Publishing. He is an author of a measure under development with Western Psychological Services. He has received travel support from BristolMyers Squibb. Mr. Klingemier has received support from Kugona LLC. Ms. Beukemann has received support from Kugona LLC. Drs. Speer, Markowitz, Wexberg, Giuliano, Schulte, Delahunty, Ahuja, and Strauss report no biomedical financial interests or potential conflicts of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Dr. Veena Ahuja, MD,
Dr. Charis Eng, MD, PhD,
Dr. Michael J. Manos, PhD,
Dr. Antonio Y. Hardan, MD,
Dr. Eric A. Youngstrom, PhD, and
Dr. Mark S. Strauss, PhD

Abstract

Objective—Abnormal eye gaze is a hallmark characteristic of autism spectrum disorder (ASD), and numerous studies have identified abnormal attention patterns in ASD. The primary aim of the present study was to create an objective, eye tracking-based autism risk index.

Method—In initial and replication studies, children were recruited after referral for comprehensive multidisciplinary evaluation of ASD and subsequently grouped by clinical consensus diagnosis (ASD n=25/15, non-ASD n=20/19 for initial/replication samples). Remote eye tracking was blinded to diagnosis and included multiple stimuli. Dwell times were recorded to each a priori-defined region-of-interest (ROI) and averaged across ROIs to create an autism risk index. Receiver operating characteristic curve analyses examined classification accuracy. Correlations with clinical measures evaluated whether the autism risk index was associated with autism symptom severity independent of language ability.

Results—In both samples, the autism risk index had high diagnostic accuracy (area under the curve [AUC]=.91 and .85, 95% CIs=.81–.98 and .71–.96), was strongly associated with Autism Diagnostic Observation Schedule–Second Edition (ADOS-2) severity scores ($r=.58$ and $.59$, $p<.001$), and not significantly correlated with language ability ($r |-.28|$, $p>.095$).

Conclusion—The autism risk index may be a useful quantitative and objective measure of risk for autism in at-risk settings. Future research in larger samples is needed to cross-validate these findings. If a validated scale for clinical use, this measure could inform clinical judgment regarding ASD diagnosis and track symptom improvements.

Keywords

autism spectrum disorder; remote eye tracking; objective measure; autism symptoms; risk

INTRODUCTION

Deficits in eye gaze are a hallmark feature of autism spectrum disorder (ASD)^{1, 2} and are included in gold-standard diagnostic instruments.^{3, 4} More than a decade of research into abnormalities of eye gaze has confirmed social attention deficits as a key feature of ASD.^{5–9} Across studies, diverse stimulus paradigms have elicited social attention abnormalities, ranging from decreased fixation to others' eyes⁵ and social scenes¹⁰ as early as 6 months of age, to gaze abnormalities during dyadic or joint attention bids in preschoolers¹¹ and older children,¹² to aberrant gaze toward dynamic social stimuli in older high-functioning individuals.¹³ Subtler, but identifiable, gaze abnormalities have also been seen in family

members with the broad autism phenotype.¹⁴ This implies that eye gaze patterns, particularly those based on dynamic temporal analysis,¹⁵ may be a promising objective risk marker of ASD as well as a quantitative measure of autism symptoms spanning the full continuum of behavior. Two recent studies provided some support for the potential discriminative value of eye gaze tracking.^{16, 17} In these studies, individual stimulus paradigms had modest but potentially informative discriminative value (areas under the curve [AUC] = .71–.72) in separating ASD and developmental delay¹⁷ or healthy control cases.¹⁶ However, no published studies have evaluated whether aggregating eye tracking metrics across stimulus paradigms might show sufficient validity (AUC .80) to inform clinical judgment by accurately discriminating ASD from a clinically realistic comparison group.

Beyond accurate discrimination, objective measures of autism symptom severity are needed to provide quantitative assessments for tracking intervention effectiveness. At present, autism symptoms are measured using direct clinical observation, parent interview, and/or parent-report.^{18, 19} These methods are heavily influenced by subjective perceptions, and both parent-interview and clinician observation measures also require substantial training with ongoing inter-rater reliability checks. Parent-report questionnaires are easier to obtain and have shown validity for separating ASD and non-ASD,^{20–23} but they are heavily influenced by rater biases (e.g. halo or contrast effects), measurement context, and are often conflated with other psychopathology symptoms,^{24, 25} reducing their effectiveness in clinically challenging samples. Lastly, none of the current diagnostic approaches readily produce interval-scale measurements that yield high reliability across the full range of behavior in neurotypical and ASD-affected individuals. This is true even for the Social Responsiveness Scale (SRS), where a floor effect is observed when converting low raw scores to standard scores.²⁶ Development of quantitative, interval-scale measures of autism symptoms, including measures of the core symptom domains of social communication/interaction (SCI) and restricted/repetitive behavior (RRB), would represent a major step forward in the technology used to capture autism symptom levels and risk for categorical ASD diagnosis.

Remote eye tracking is a promising technology for development as an objective measure of autism. In addition to literature support for gaze abnormalities in ASD, remote eye tracking is easier to calibrate and collect in young or severely impaired children relative to traditional headgear-based eye tracking methods and other methods that require significant preparation (e.g. electroencephalogram [EEG]/event-related potential [ERP]), physical restrictions (e.g. magnetic resonance imaging [MRI]/magnetoencephalography [MEG]), and possibly sedation (e.g. MRI). Young children are familiar with watching TV, and their attention is often sufficiently captured over short intervals. When attention capture is more challenging, visual stimuli can be repeated and interspersed around short breaks, with multiple stimuli used to further enhance novelty and attention. Use of multiple stimulus paradigms also permits capture of different aspects of attention, including attention to socially appropriate targets and non-social/distractor targets. Relative to the massive literature examining social attention in ASD, very little research has focused on abnormalities of attention to nonsocial/distractor stimuli.²⁷ Inappropriate attention to nonsocial stimuli is frequently observed clinically and is an important part of the description of RRB symptoms.²⁸ Supporting this

observation, Sasson et al.²⁷ recently demonstrated that attention to objects vs. people and attention to high vs. low autism-interest items were associated with greater RRB symptoms. This study suggests that it may be possible to identify characteristic gaze patterns to nonsocial stimuli that more strongly associate with RRB than SCI symptoms. Visual attention paradigms can readily integrate both social and nonsocial/distractor targets without adding time or reducing participant engagement.

The primary aim of the present research was to develop and replicate an objective measure of autism symptom level based on eye gaze tracking to social and nonsocial stimuli, hereafter referred to as the “autism risk index” (ARI). We hypothesized that ASD-affected children would show less attention to social and greater attention to nonsocial targets than children without an ASD diagnosis but with other developmental neuropsychiatric concerns (non-ASD). Based on this expectation, the ARI was created by averaging dwell times to a priori social and nonsocial target regions of interest. In initial and replication samples, the ARI was expected to show strong discrimination (AUC .80) of ASD and non-ASD cases and be significantly related to overall autism symptom severity but not language measures.

METHOD

Samples

Participants were children, ages 3.0 to 8.11, referred to a tertiary care multidisciplinary ASD specialty clinic. Referrals were made by local pediatricians, following autism screening, if there was clinical concern of social deficits or ASD, or if parents or teachers had concerns. Patients were consecutively recruited at the time of the diagnostic evaluation visit (initial study - July 2014 and June 2015; replication study - August 2015–November 2015). Gaze data were collected prior to the consensus diagnosis team meeting, and the research team was blinded to participant diagnosis. Procedures of this research were reviewed and approved by the Cleveland Clinic institutional review board (IRB).

Eye Tracking

Eye tracking data were collected in a quiet room adjacent to the diagnostic clinic. Data were recorded at using an SMI remote eye tracker (initial study: Red-m at 120Hz, replication study: Red250 at 60Hz) attached to the frame of a 1280 horizontal X 1024 vertical 19-inch LCD stimulus presentation monitor. Spatial resolution of these systems was 0.1°, and average gaze position accuracies were 0.5°. The system allows for head movement (32 × 21 × 25cm for Red-m and 32 × 21 × 30 for Red250) at a maximum distance of 75cm. In the initial study, a 3- or 5-point calibration was obtained prior to the experiment. In the replication study, an initial and four additional 5-point calibrations were obtained at fixed times throughout the experiment (see Supplement 1, available online). Proportion net dwell time to each ROI was derived using SMI BeGaze software. Dwell time was defined as the sum of all sample durations (all fixations and saccades) falling within the ROI divided by the total stimulus time.

Visual Stimulus Battery

Stimuli were presented using SMI Experiment Center, and stimuli for the initial study were selected to represent multiple distinct types previously used in the eye gaze literature, including static facial affect, biological vs. non-biological pairings, and dynamic/naturalistic scenes. Figures S1–S2 (available online) present example stimuli created for the initial and replication studies, and Tables S1–S2 (available online) list all stimuli and ROIs. Stimuli were presented in a single order, intermixed with attention-grabbing stimuli, gaze recalibration, and receptive language stimuli, and other stimuli not considered for the present paper. Total experiment time was approximately 7 minutes for both the initial and replication studies.

For the initial study, a priori ROIs were identified by the first author, who did not participate in data collection or diagnostic evaluations. ROIs were drawn to capture important social (faces, key body movements) and nonsocial target stimulus elements (distractors). A priori ROIs were further restricted to key time points within each stimulus based on a socially relevant action. When relevant, a priori ROIs were also designated across the total stimulus period to capture basic attention to social versus nonsocial elements. For example, in dynamic joint attention stimuli, a temporal ROI evaluated gaze to the most relevant social action (e.g., gaze-and-point to a target), but the total stimulus period was also examined to capture overall attention to the social (e.g., face) and nonsocial information (e.g., clock). A total of 68 a priori ROIs were identified, including 51 social and 17 nonsocial ROIs.

Replication study stimuli were chosen based on results from the initial study. Specifically, the replication study focused on stimuli showing the strongest validity in the initial study—joint attention and child joke stimuli—and enriched for nonsocial ROIs. New stimuli were created to mimic videos used in the initial study. A total of 42 a priori ROIs were identified, including 19 social and 23 non-social ROIs.

For both studies, all ROIs were truly a priori. No preliminary analyses of ROI validity were conducted to choose ROIs and no post hoc modifications were made to ROIs to enhance validity.

Procedure

Eye tracking data collection followed recommendations from Sasson and Elison.⁹ Children were seated alone or in their parent's lap approximately 65cm from the LCD display and viewed stimuli subtending a visual angle of 18.8°. Standard room lighting was used, and the room was sparse, with visual barriers used to reduce distraction. After calibration, children were told, "You will see some pictures and videos, pay attention, but look however you want." Data for individual ROIs were excluded if proportion dwell to any location on the stimulus was <40%. Gaze needed to be detected on-screen 40% of the time and participants had to have at least 20 ROIs available dwell data to consider the eye tracking evaluation valid. A minimum tracking ratio of 40% was selected based on recommendations to exclude children with very low overall attention to stimuli,⁹ and this ratio was combined with a minimum number of ROI to ensure that sufficient data contributed to computation of the ARI.

The same procedures were followed for the replication study with a few notable exceptions. First, an anonymous reviewer of an early draft of this manuscript suggested using healthy control participants to select ROIs. Following this suggestion, we recruited 12 healthy control children (ages 2–15; 4 females), and their gaze data were viewed in SMI Be Gaze software to identify a priori social targets (see Supplement 2, available online). Second, results of the initial study identified trends toward lower scores on the ARI in older individuals and individuals with higher tracking ratios. Additionally, inspection of healthy control data identified longer dwell times to social ROIs in older children and children with higher tracking ratios. Preliminary analyses confirmed this observation, identifying consistently positive small-to-medium-sized relationships between social ROI dwell times and age/tracking ratios. For this reason, social ROIs were residualized after regressing age and tracking ratio. Due to these methodological differences, the replication study should be considered a partial replication and extension that generalizes the approach rather than an exact replication.

Consensus Diagnosis

Consensus diagnosis was based on diagnostic interviews conducted by a psychologist, developmental and psychosocial history confirmed by the psychologist, medical history confirmed by a physician, cognitive testing administered by a speech language pathologist, and the Autism Diagnostic Observation Schedule-Second Edition (ADOS-2) administered by a reliable administrator. Within two weeks of the initial visit, a multidisciplinary team meeting was conducted to confirm the presence/absence of *DSM-5* criteria for ASD and document any other psychiatric diagnoses.

Clinical Measures

ADOS-2—The ADOS-2 is the gold-standard clinical observation measure used to assess autism symptom severity. For the present study, the ADOS-2 total, social affect subscale, and restricted/repetitive behavior subscale raw scores were converted to calibrated severity scores.^{29, 30}

SRS-2—Parents completed the SRS-2 as part of the clinical evaluation. The SRS-2 is a 65-item, ordinally scaled (1= “not true” to 4= “almost always true”) quantitative assessment of the level of autism traits. The SRS sex-adjusted total T-score has been extensively validated and distinguishes youth with autism from other psychiatric conditions.^{31, 32}

Language—Receptive and expressive language was collected as part of the clinical evaluation using the Mullen Scales of Early Learning,³³ the Clinical Evaluation of Language Fundamentals – Fourth Edition³⁴ or Preschool Version – Second Edition,³⁵ or the Preschool Language Scales – Fifth Edition.³⁶ For Mullen subscales, T-scores were converted to standard scores (M=100, SD=15).

Child Behavior Checklist (CBCL)—Other psychopathological symptoms were collected using the CBCL – ages 1.5 to 5 and 6–18 parent-report versions.³⁷ Total problems T-score was used to describe the sample and examine whether other psychopathology influences the ARI.

Statistical Analyses

The study design and analyses followed recommendations for evaluating test validity (See Table S3, available online).^{38, 39} Univariate and bivariate distributions were examined to identify outliers and ensure that high leverage cases did not unduly influence relationships. Descriptive statistics were presented separately for patients with ASD and non-ASD patients to characterize the sample. Comparisons between ASD and non-ASD groups were made across demographic and clinical measures using independent samples t-tests or Chi-square statistics.

To develop the ARI, the directionality and discriminative strength of individual ROIs was evaluated by computing independent samples t-tests and associated effect sizes (Cohen's *d*). The dependent variable was proportion dwell time to each ROI. ASD-affected children were predicted to look less at social targets and more at non-social targets. Effect sizes were transformed so that positive values represent differences in the predicted direction. The number of ROIs in the predicted direction and the number of ROIs with statistically significant differences in the predicted direction were compared to expected proportions (.50 and .05, respectively) using a one-sample proportion test. After establishing the expected directionality for the majority of ROIs, dwell times to social and nonsocial ROIs were standardized using the non-ASD means and standard deviations. These standardized ROI scores were separately averaged to create social and nonsocial attention measures, respectively. Finally, the social and nonsocial attention measures were averaged (after reflection of the social attention index - multiplying each score by -1) to form the ARI. Internal consistency reliability of the social and nonsocial attention measures was computed using Cronbach's α , using packets of nonsocial ROIs as items (4–7 items per packet).

Validity of the ARI was estimated using receiver operating characteristic curve analyses. Area under the curve (AUC) was calculated using consensus diagnosis (ASD vs. non-ASD) or ADOS classifications (non-spectrum vs. autism spectrum) as the state variable. To determine the potential incremental validity of social and non-social attention measures in predicting ASD, hierarchical logistic regressions were computed with social and non-social attention measures alternating as the predictors in steps 1 and 2. Incremental validity of the ARI over the SRS-2 was made using hierarchical logistic regression, with the SRS-2 entered in step 1 and the ARI in step 2. A significant increase in R^2 from step 1 to 2 would indicate that the ARI improves detection of ASD diagnoses.⁴⁰

Convergent validity of the ARI with clinical measures of autism symptoms was evaluated by computing using Spearman's rank-order correlations between eye tracking measures and autism symptoms (ADOS-2 and SRS-2 scores), demographics (age and sex), and tracking ratio (total time-on-screen). Non-parametric (Spearman's) partial correlations were computed between the ARI and language measures and CBCL behavior scores after accounting for ADOS-2 total calibrated severity scores.

Power to detect a significant AUC ($p < .05$) was computed in the initial and replication samples using pROC in R.^{41, 42} Minimum AUC of .73 and .77 were detectable with power (1-B) = .80 in the initial and replication samples. Group comparisons between ASD and non-ASD cases for individual ROIs had adequate power (.73) for detecting large differences

(Cohen's $d = .80$; $p < .05$). Both samples also had good power ($.77$) for detecting large positive relationships ($r = .40$) between eye tracking measures and clinical measures of autism symptom severity.

ROC analyses were computed using pROC,⁴¹ and non-parametric Spearman's rank-order bivariate and partial correlations were implemented using the ppcor program⁴³ in R. Individual ROI analyses were computed using IBM SPSS v23.0.⁴⁴ In spite of strong directional predictions for all group differences and bivariate relationships, $\alpha = .05$ two-tailed was used. Emphasis was placed on effect magnitude and confidence intervals, as these are the most crucial features for test validity. Effect sizes conventions were small ($d = .20$), medium ($d = .50$), and large ($d = .80$)⁴⁵ for group comparisons. A rough guideline for evaluating AUC values is: $< .70$ = poor; $.70$ – $.79$ = fair; $.80$ – $.89$ = good; and $.90$ – 1.00 = excellent.⁴⁶

RESULTS

Participant Accounting

Figure S3 (available online) describes participant inclusion/exclusion. Of the individuals who consented, 6 children from the initial study and 3 children from the replication study could not adequately attend to the stimuli at least 40% of the time (ASD $n = 1/2$, non-ASD $n = 5/1$). All individuals who could not achieve a valid administration had low language scores ($SS < 74$) and/or severe autism symptom levels (ADOS-2 calibrated severity score > 7).

Sample Descriptions

Table 1 presents sample characteristics for the initial and replication studies. As expected, the group with ASD had higher autism symptom severity scores on the ADOS-2 and lower language scores. The non-ASD group had a range of psychiatric diagnoses, with one non-ASD participant receiving no clinical diagnosis in each sample. Consistent with the at-risk, referred nature of these samples, there were no significant differences in SRS-2 or CBCL total problems scores, and high scores did not discriminate cases of ASD and non-ASD cases in either sample (SRS-2: AUC = $.58$ and $.43$, 95% CI = $.39$ – $.77$ and $.22$ – $.66$; CBCL: AUC = $.36$ and $.41$, 95% CI = $.18$ – $.54$ and $.19$ – $.64$). Broad ranges of ADOS-2 total severity scores, SRS Total T-scores, and CBCL scores were observed, with near complete overlap in parent-reported autism traits and slightly more behavior problems in non-ASD cases (See Figures S4 and S5, available online). Importantly, the tracking ratio (total time-on-screen) did not significantly differ between patients with ASD and non-ASD patients.

Individual ROIs

Dwell time differences between participants with ASD and non-ASD participants were in the expected direction for the majority of ROIs (initial study: 57 of 68 [$z = 4.54$, $p < .001$]; replication study: 37 of 42 [$z = 4.43$, $p < .001$]; Figure 1a and 1b) and a substantial minority were statistically significant in the expected direction in both samples (initial study: 19 of 68 [$z = 7.05$, $p < .001$] and replication study: 11 of 42 [$z = 5.67$, $p < .001$]). Effect sizes were highly variable across ROIs ($d = -.41$ – 1.35). In the initial sample, 9 of the 10 largest effect sizes (Cohen's $d > .75$) were from joint attention or child joke stimuli, supporting focus on these

stimuli in the replication study. Internal consistency reliability was adequate for the social attention measure ($\alpha=.73$ and $.76$) and marginal to adequate for the nonsocial attention measure ($\alpha=.53$ and $.76$) across the two samples.

Identification of Cases of ASD

ARI scores had limited overlap (Cohen's $d=1.15$ and 1.41) and showed very good discrimination (AUC=.91 and $.85$, 95%CI=.81–.98 and $.71$ –.96; Figure 2a and 2b) between non-ASD and ASD cases in initial and replication samples, respectively. Discrimination was also excellent for cases meeting threshold on the ADOS-2 (AUC=.90 and $.86$, 95%CI=.80–.98 and $.69$ –.98). The ARI showed substantial incremental validity over the SRS-2 Total T-score (smallest $R^2=.46$, $X^2[1]=12.16$, $p<.001$).

Correlations With Autism Symptom Severity

The ARI was strongly correlated with ADOS-2 total ($r=.58$ and $.59$, $p=.001$; Figure 3a and b) and domain severity scores (See Table S4, available online), but was not significantly related to age, sex, or tracking ratio ($|r| .28$, $p .060$; See Table S5, available online). After accounting for total autism symptom severity, ARI scores were not significantly related to language ability ($|r| .28$, $p .096$). A modest significant relationship with CBCL Total Problems scores was observed in the replication sample. However, in both samples, the relationship between ARI and ADOS-2 total scores remained strong after accounting for total behavior problems (See Table S4, available online). SRS-2 scores were not related to eye tracking measures. Supplement 3, Table S6 and Figures S6–S7 (available online) provide additional information on the discriminative value and autism symptom correlations for the social and nonsocial attention measures.

Combined Sample

The ARI was computed in a highly similar fashion across the initial and replication studies. For this reason, we also examined discrimination after combining the samples (total $N=79$, ASD $n=40$, non-ASD $n=39$). The ARI had modest overlap (Figure 4) and very good discrimination between patients with ASD and non-ASD (AUC=.89, 95%CI=.81–.95). The optimal cutpoint based on Youden's statistic was $z=.104$, with sensitivity of $.80$ (95%CI=.68–.93) and specificity of $.82$ (95%CI=.69–.92). False negative ASD-diagnosed children had lower ADOS-2 total severity scores than correctly identified patients with ASD (4.5 vs. 6.1) and false positive non-ASD diagnosed children had higher ADOS-2 total severity scores than correctly identified patients with non-ASD (4.0 vs. 2.6; $F(3,72)=27.50$, $p<.001$). There were no differences between identified and missed cases for receptive language or CBCL Total Problems ($p>.190$). Using the combined sample, the correlation between ARI scores and ADOS-2 total severity scores was strong ($r=.59$, $p<.001$) and remained strong after adjusting for receptive language ($r=.57$, $p<.001$) or CBCL Total Problems ($r=.66$, $p<.001$).

DISCUSSION

The present investigation demonstrates the strong potential for remote eye tracking as an objective tool for quantifying autism risk and estimating autism symptom severity. As expected, individual ROIs had variable, but generally modest, levels of discrimination of

patients with ASD and non-ASD. In contrast, by measuring a core cognitive feature of autism—dysfunctional attention to social and nonsocial information—the composite ARI had substantial diagnostic accuracy, dramatically outperformed the SRS-2, and showed strong relationships with a gold-standard measure of autism symptom severity. These results are very promising, particularly because the current version of the ARI was conservative: all a priori ROIs were included regardless of direction and validity level. Enhancements to the present approach are possible and may increase the validity and utility of the ARI. Potential improvements include: continuing to focus on the highest-validity stimuli (as was done in the replication study), differentially weighting stimuli, adding non-redundant and more dynamic gaze metrics (e.g. time-to-first-fixation to target ROIs, revisits to target, saccades, etc.),¹⁵ and including a larger number of nonsocial target ROIs. Adding nonsocial targets may be particularly important for increasing the relationship between nonsocial attention and RRB symptoms. Even if the ARI is unchanged in future validation studies, confidence intervals from the combined sample suggested that diagnostic accuracy should remain in the good-to-excellent range. Replication of these findings is warranted and would represent a major step forward, as objective markers of autism are sorely needed.

The present data suggested that the ARI may have incremental validity for ASD identification when used in conjunction with other clinical measures. The ARI accounted for substantial predictive variance after accounting for the SRS-2, and relationships between the ARI and ADOS-2 overall severity scores were high but did not suggest redundancy. Additionally, missed cases had a different pattern of ADOS-2 scores than correctly identified cases. Future research is needed to establish precise estimates of stand-alone and incremental validity of the ARI for categorical ASD diagnosis.

Beyond enhancing clinical diagnosis, objective measures are needed that grade autism symptom severity and track symptom changes with treatment. At present, treatment-mediated changes in autism symptoms are gauged using subjective measures, decreasing the reliability of treatment effect estimates. The result is inefficient evaluation of promising treatments. Objectively and quantitatively measuring symptom severity would facilitate more accurate characterization of treatment response and enhance our ability to assess developmental trajectories in autism symptoms. Longitudinal studies of the ARI and the associated social and non-social attention measures, including collection during treatment studies, will be needed to determine if these measures have adequate test-retest reliability and are sensitive to change.

Remote eye tracking-based measures have several desirable features beyond the objective and quantitative nature of these measurements. Administration of visual attention paradigms is rapid (<10 minutes), can be largely automated, requires limited technical expertise, and does not involve ongoing inter-rater reliability checks. Parents are also likely to have high acceptance of remote eye tracking as part of the clinical evaluation. The data are easily acquired in most children, and reduced or altered eye contact is well understood as a symptom of autism. In some parents, lack of an objective measure can lead to delayed or diminished acceptance of the clinical diagnosis. Scalable eye-tracking solutions for other neuropsychiatric disorders are in development or early adoption,^{47, 48} further supporting the

viability of the approach. Thus, development of an eye tracking system implementing the ARI described here appears feasible and potentially informative to clinical practice.

Even with these advantages, there are technical limitations that will delay immediate clinical adoption of remote eye-tracking as an objective measure. Hardware and software costs are substantial and scoring is labor intensive. Both of these limitations can be remedied in future work. Remote eye tracking hardware continues to decrease in cost, with less sophisticated models that are sufficient for collecting viewing time at \$100. Validation studies will be needed to demonstrate that the present findings can be obtained with less sophisticated hardware. Similarly, clinical viability of a remote eye tracking-based measure can be improved by automating scoring. Automation is highly feasible and would only require add-on software that could be updated as the paradigm is enhanced and scoring algorithms are refined. Future work is also needed to extend the present findings to younger patients. Additional breaks in testing, re-testing at a later date, and liberal use of attention-grabbing stimuli may be useful methods for improving the number of young and low-functioning children who achieve a valid evaluation.

The primary limitations of the present study were modest sample sizes for the initial and replication studies, lack of calibration quality measures for the initial study, and evaluations of test performance under high prevalence conditions. The combined sample size of the present study was larger than most of the recently reviewed eye tracking studies in children and included a developmental disability control group larger than almost all previous studies.⁸ The group with non-ASD represented a challenging comparison cohort of children referred for clinical evaluation of ASD, with a wide range of clinician-observed and parent-reported autism symptoms, high levels of other behavior problems, and highly overlapping levels of receptive language relative to the group with ASD. In spite of its use in the diagnostic evaluation, several children in the group with non-ASD had ADOS-2 scores overlapping the group with ASD, and all but two children with non-ASD had some form of developmental neuropsychiatric diagnosis. The clinically realistic nature of the group with non-ASD indicates that diagnostic discrimination values are not likely to be inflated, an important consideration for test evaluation studies that has not been addressed in any previous eye tracking investigations. Comparisons that use healthy controls produce much larger effect sizes that are prone to greater shrinkage when the same test is used in clinical settings where other developmental conditions that could have high false positive rates are common.⁴⁹

When considering these limitations, it is important to keep in mind that the overarching objective was to demonstrate the potential clinical value of combining eye tracking measurements into a risk index. In this light, the present results suggest substantial potential for the ARI to inform clinical practice. These studies should include additional clinical measures with better coverage of SCI and RRB domains, examine test-retest stability, estimate sensitivity to change, and include a healthy comparison cohort that is not used to identify social ROIs. A healthy comparison group will help with understanding the relative impairment in social and nonsocial attention in the group with non-ASD, tracking expected developmental changes in social and nonsocial attention, and identifying whether an enhanced ARI would improve screening in low base rate settings. If the above

recommendations are undertaken and validation of the present approach is achieved, remote eye tracking would be the first clinically scalable quantitative and objective measure of autism.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was made possible by a generous donation from the Stephan and Allison Cole Family Research Fund. The work was also supported by funding from the Case Western Reserve University International Center for Autism Research and Education (ICARE) and funding for the Developmental Synaptopathies Consortium (U54NS092090). The Developmental Synaptopathies Consortium is part of NCATS Rare Disease Clinical Research Network (RDCRN), an initiative of the Office of Rare Disease Research (ORDR). This consortium is funded through collaboration between NCATS, and the National Institute of Neurological Disorders and Stroke (NINDS) of the National Institutes of Health.

The authors would like to thank the children and parents/caregivers who participated in this study.

References

1. Kanner L. Autistic disturbances of affective contact. *Nervous Child*. 1943; 2:217–250.
2. Rapin I. Autism. *N Engl J Med*. 1997; 337:97–104. [PubMed: 9211680]
3. Lord, C.; Rutter, M.; DiLavore, PC.; Risi, S.; Gotham, K.; Bishop, SL. Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) Manual (Part 1): Modules 1–4. Torrance, CA: Western Psychological Services; 2012.
4. Lord C, Rutter M, LeCouteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *J Autism Dev Disord*. 1994; 24:569–685.
5. Jones W, Klin A. Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*. 2013; 504:427–431. [PubMed: 24196715]
6. Jones W, Carr K, Klin A. Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Arch Gen Psychiatry*. 2008; 65:946–954. [PubMed: 18678799]
7. Klin A, Lin DJ, Gorrindo P, Ramsay G, Jones W. Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*. 2009; 459:257–261. [PubMed: 19329996]
8. Papagiannopoulou EA, Chitty KM, Hermens DF, Hickie IB, Lagopoulos J. A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Soc Neurosci*. 2014; 9:610–632. [PubMed: 24988218]
9. Sasson NJ, Elison JT. Eye tracking young children with autism. *J Vis Exp*. 2012; (61):3675. [PubMed: 22491039]
10. Chawarska K, Macari S, Shic F. Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biol Psychiatry*. 2013; 74:195–203. [PubMed: 23313640]
11. Vivanti G, Trembath D, Dissanayake C. Atypical monitoring and responsiveness to goal-directed gaze in autism spectrum disorder. *Exp Brain Res*. 2014; 232:695–701. [PubMed: 24292493]
12. Magrelli S, Jermann P, Noris B, et al. Social orienting of children with autism to facial expressions and speech: a study with a wearable eye-tracker in naturalistic settings. *Front Psychol*. 2013; 4:840. [PubMed: 24312064]
13. Rice K, Moriuchi JM, Jones W, Klin A. Parsing heterogeneity in autism spectrum disorders: visual scanning of dynamic social scenes in school-aged children. *J Am Acad Child Adolesc Psychiatry*. 2012; 51:238–248. [PubMed: 22365460]

14. Dalton KM, Nacewicz BM, Alexander AL, Davidson RJ. Gaze-fixation, brain activation, and amygdala volume in unaffected siblings of individuals with autism. *Biol Psychiatry*. 2007; 61:512–20. [PubMed: 17069771]
15. Guillon Q, Hadjikhani N, Baduel S, Roge B. Visual social attention in autism spectrum disorder: insights from eye tracking studies. *Neurosci Biobehav Rev*. 2014; 42:279–297. [PubMed: 24694721]
16. Chevallier C, Parish-Morris J, McVey A, et al. Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Res*. 2015; 8:620–8. [PubMed: 26069030]
17. Pierce K, Marinero S, Hazin R, McKenna B, Barnes CC, Malige A. Eye Tracking Reveals Abnormal Visual Preference for Geometric Images as an Early Biomarker of an Autism Spectrum Disorder Subtype Associated with Increased Symptom Severity [Epub ahead of print]. *Biol Psychiatry*. 2015 Apr 11.10.1016/j.biopsych.2015.03.032
18. Bishop SL, Seltzer MM. Self-reported autism symptoms in adults with autism spectrum disorders. *J Autism Dev Disord*. 2012; 42:2354–2363. [PubMed: 22361924]
19. Risi S, Lord C, Gotham K, et al. Combining information from multiple sources in the diagnosis of autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. 2006; 45:1094–1103. [PubMed: 16926617]
20. Corsello C, Hus V, Pickles A, et al. Between a ROC and a hard place: decision making and making decisions about using the SCQ. *J Child Psychol Psychiatry*. 2007; 48:932–940. [PubMed: 17714378]
21. Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: Diagnostic validity. *Br J Psychiatry*. 1999; 175:444–451. [PubMed: 10789276]
22. Chandler S, Charman T, Baird G, et al. Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. *J Am Acad Child Adolesc Psychiatry*. 2007; 46:1324–1332. [PubMed: 17885574]
23. Constantino, JN.; Gruber, CP. *The social responsiveness scale manual, second edition (SRS-2)*. Los Angeles, CA: Western Psychological Services; 2012.
24. Frazier TW, Youngstrom EA, Embacher R, et al. Demographic and clinical correlates of autism symptom domains and autism spectrum diagnosis. *Autism*. 2014; 18(5):571–582. [PubMed: 24104512]
25. Hus V, Bishop S, Gotham K, Huerta M, Lord C. Factors influencing scores on the social responsiveness scale. *J Child Psychol Psychiatry*. 2013; 54:216–224. [PubMed: 22823182]
26. Frazier TW, Ratliff KR, Gruber C, Zhang Y, Law PA, Constantino JN. Confirmatory factor analytic structure and measurement invariance of quantitative autistic traits measured by the Social Responsiveness Scale-2. *Autism*. 2014; 18:31–44. [PubMed: 24019124]
27. Sasson NJ, Turner-Brown LM, Holtzclaw TN, Lam KS, Bodfish JW. Children with autism demonstrate circumscribed attention during passive viewing of complex social and nonsocial picture arrays. *Autism Res*. 2008; 1(1):31–42. [PubMed: 19360648]
28. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5. Arlington, VA: American Psychiatric Association; 2013.
29. Gotham K, Pickles A, Lord C. Standardizing ADOS scores for a measure of severity in autism spectrum disorders. *J Autism Dev Disord*. 2009; 39:693–705. [PubMed: 19082876]
30. Hus V, Gotham K, Lord C. Standardizing ADOS domain scores: separating severity of social affect and restricted and repetitive behaviors. *J Autism Dev Disord*. 2014; 44:2400–2412.
31. Virkud YV, Todd RD, Abbacchi AM, Zhang Y, Constantino JN. Familial aggregation of quantitative autistic traits in multiplex versus simplex autism. *Am J Med Genet B Neuropsychiatr Genet*. 2009; 150B:328–334. [PubMed: 18618672]
32. Constantino, JN.; Gruber, CP. *Social Responsiveness Scale: Manual*. Los Angeles, CA: Western Psychological Services; 2005.
33. Mullen, EM. *Mullen Scales of Early Learning*. Circle Pines, MN: American Guidance Service Inc; 1995.
34. Semel, E.; Wiig, EH.; Secord, WA. *Clinical Evaluation of Language Fundamentals, fourth edition (CELF-4)*. Toronto: The Psychological Corporation; 2003.

35. Wiig, EH.; Secord, WA.; Semel, E. Clinical evaluation of language fundamentals - Preschool, second edition (CELF-2). Toronto: The Psychological Corporation; 2004.
36. Zimmerman, IL.; Steiner, VG.; Pond, E. Preschool Language Scales-Fifth Edition (PLS-5). San Antonio, TX: Pearson; 2011.
37. Achenbach, TM.; Rescorla, LA. Manual for the ASEBA school-age forms and profiles. Burlington, VT: University of Vermont, Department of Psychiatry; 2001.
38. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Ann Intern Med.* 2003; 138:40–44. [PubMed: 12513043]
39. Irwig L, Bossuyt P, Glasziou P, Gatsonis C, Lijmer J. Evidence base of clinical diagnosis: Designing studies to ensure that estimates of test accuracy are transferable. *Br Med J.* 2002; 324:669–671. [PubMed: 11895830]
40. Youngstrom EA. A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *J Pediatr Psychol.* 2014; 39:204–221. [PubMed: 23965298]
41. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics.* 2011; 12:77. [PubMed: 21414208]
42. Obuchowski NA, McClish DK. Sample size determination for diagnostic accuracy studies involving binormal ROC curve indices. *Stat Med.* 1997; 16(13):1529–1542. [PubMed: 9249923]
43. Kim S. ppcor: An R Package for a Fast Calculation to Semi-partial Correlation Coefficients. *Communications for Statistical Applications and Methods.* 2015; 22(6):665–674. [PubMed: 26688802]
44. IBM Statistics for Windows [computer program]. Version 23.0. Armonk, NY: IBM Corp; 2015.
45. Cohen, J. Statistical power analysis for the behavioral sciences. 2. Hillsdale, NJ: Erlbaum; 1987.
46. Swets JA. Measuring the accuracy of diagnostic systems. *Science.* 1988; 240:1285–1293. [PubMed: 3287615]
47. Zola SM, Manzanares CM, Clopton P, Lah JJ, Levey AI. A behavioral task predicts conversion to mild cognitive impairment and Alzheimer’s disease. *Am J Alzheimers Dis Other Demen.* 2013; 28:179–184. [PubMed: 23271330]
48. Neurotrack. [Accessed June 1, 2015] 2015. <https://www.neurotrack.com/>
49. Youngstrom EA, Genzlinger J, Egerton G, Van Meter AR. Multivariate meta-analysis of the discriminative validity of caregiver, youth, and teacher rating scales for pediatric bipolar disorder: mother knows best about mania. *Arch Sci Psychol.* 3(1):112–137. in press.

Clinical Guidance

- The ARI, based on eye gaze to social and nonsocial information, may be a useful quantitative and objective measure, informing clinical judgment regarding the presence of an ASD diagnosis in at-risk settings.
- The ARI, and by extension eye tracking-based attention measures, may supplement existing clinical observation measures for grading the severity of autism symptom levels.
- Pending acquisition of data establishing sensitivity to change, the ARI may have potential as an objective measure of response to pharmacological and behavioral intervention programs.

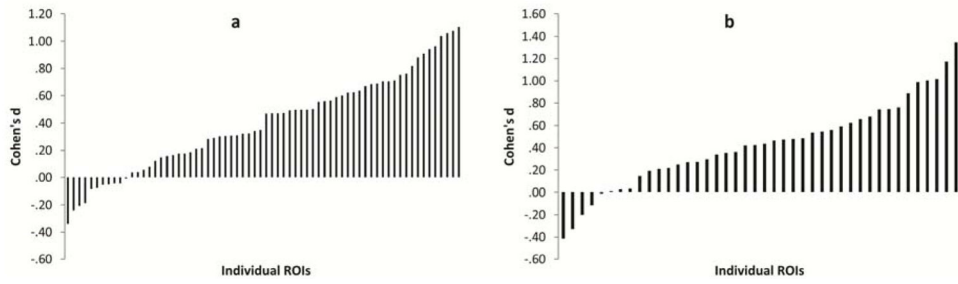


Figure 1. Effect sizes (Cohen's d) for individual regions-of-interest (ROI) representing the magnitude of group differences between cases of autism spectrum disorder (ASD) and non-ASD cases, separately for the initial (a) and replication studies (b).

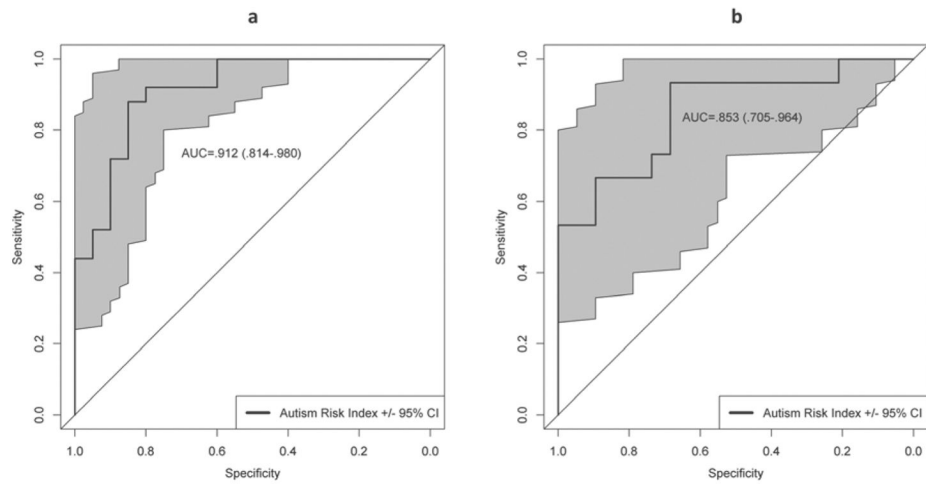


Figure 2. Areas under the curve from receiver operating characteristic curve (AUC) analysis for the autism risk index predicting consensus clinical autism spectrum disorder (ASD) diagnosis, separately for the initial (a) and replication (b) samples.

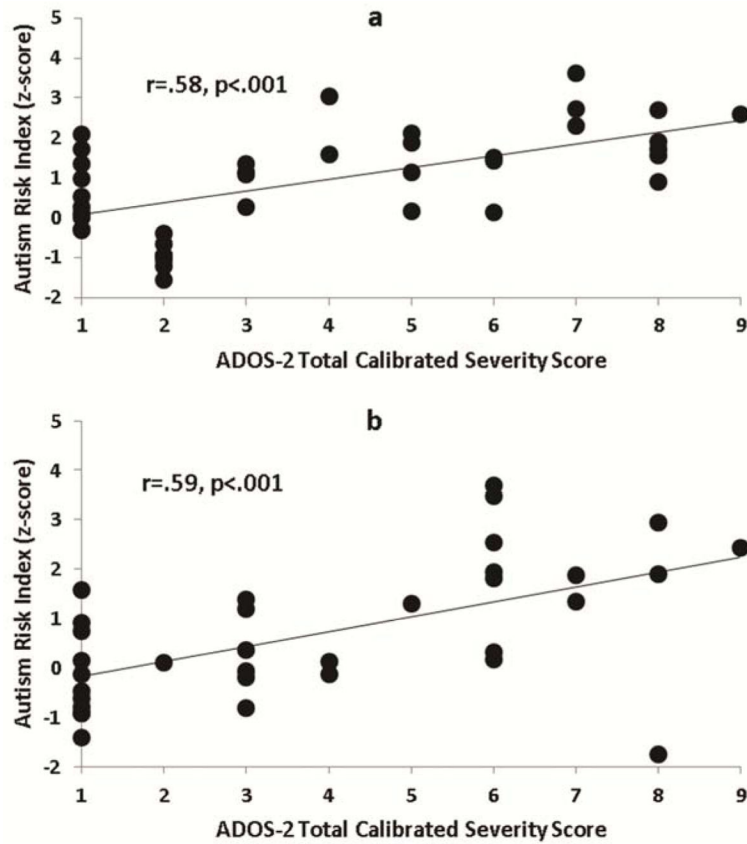


Figure 3. Relationship between autism risk index scores and Autism Diagnostic Observation Schedule – Second Edition (ADOS-2) total calibrated severity scores, separately in the initial (a) and replication (b) samples.

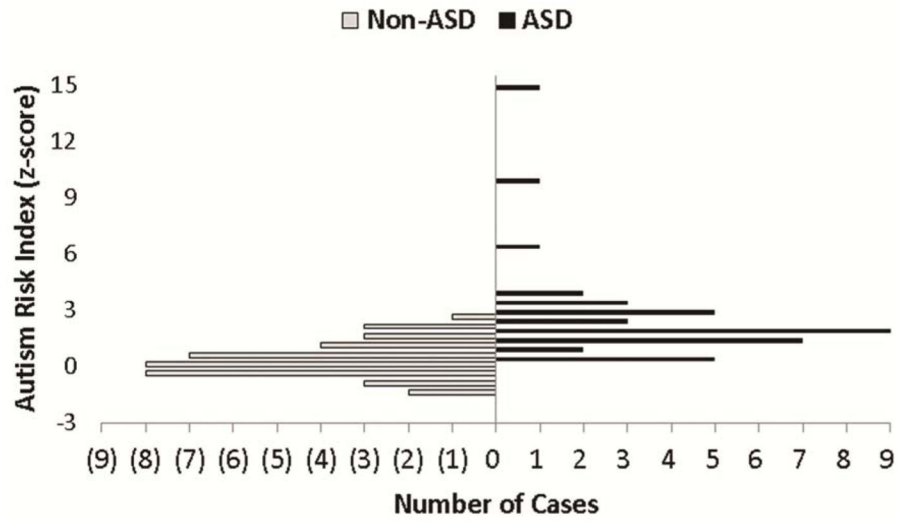


Figure 4. Autism risk index (ARI) score distributions for autism spectrum disorder (ASD) and non-ASD cases in the combined sample.

Table 1

Sample Demographic and Clinical Characteristics.

	Initial Study			Replication Study		
	Non-ASD M (SD)	ASD M (SD)	X ² /t (p)	Non-ASD M (SD)	ASD M (SD)	X ² /t (p)
N	20	25		19	15	
Age	5.9 (1.4)	5.4 (1.8)	1.13 (.263)	5.8 (1.3)	4.9 (1.6)	1.79 (.083)
Male (%)	17 (85%)	20 (80%)	0.19 (.663)	15 (79%)	11 (73%)	0.15 (.702)
Total Language	101.3 (20.8)	81.6 (21.1)	2.78 (.009)	86.3 (20.6)	81.9 (25.1)	0.53 (.599)
Receptive Language	100.8 (19.7)	80.9 (22.9)	2.70 (.011)	87.1 (23.2)	82.2 (22.4)	0.58 (.566)
Expressive Language	98.8 (20.8)	82.5 (21.4)	2.28 (.029)	87.5 (19.5)	78.8 (23.4)	1.12 (.271)
ADOS-2 Total Calibrated Severity	1.4 (0.7)	5.9 (1.9)	-10.67 (<.001)	2.6 (2.0)	5.5 (2.4)	-3.78 (.001)
ADOS-2 Social Affect Severity	1.7 (1.2)	5.9 (2.0)	-8.3 (<.001)	3.2 (2.5)	6.0 (2.4)	-3.20 (.003)
ADOS-2 Restricted/Repetitive Severity	2.7 (2.6)	6.0 (3.1)	-3.66 (.001)	2.2 (2.4)	5.4 (2.6)	-3.77 (.001)
SRS-2 Total T-score	67.3 (12.9)	70.7 (11.8)	-0.86 (.397)	66.6 (12.7)	64.6 (10.0)	0.45 (.656)
Child Behavior Checklist – Total Problems	66.3 (7.3)	62.8 (8.7)	1.35 (.184)	66.1 (13.4)	63.4 (13.3)	0.55 (.589)
Clinical Diagnoses						
Anxiety Disorder	11 (55%)	-	-	8 (42%)	-	-
Mood Disorder	1 (5%)	-	-	0 (0%)	-	-
ADHD	8 (40%)	-	-	6 (32%)	-	-
Speech/Language Disorder	7 (35%)	-	-	10 (53%)	-	-
Unspecified Behavior Disorder	5 (25%)	-	-	7 (37%)	-	-
Adjustment Disorder	1 (5%)	-	-	0 (0%)	-	-
No Diagnosis	1 (5%)	-	-	1 (5%)	-	-
Tracking Ratio (%)	74.6% (12.7%)	68.6% (14.6%)	1.45 (.153)	76.0% (16.5%)	70.9% (13.3%)	0.98 (.335)

Note. Several non-ASD children (patients without autism spectrum disorder consensus diagnosis) had multiple diagnoses. Three participants (2 non-ASD and 1 ASD) did not complete the Autism Diagnostic Observation Schedule–Second Edition (ADOS-2) due to clinical judgment or challenging behavior in the initial sample. ADHD=attention-deficit/hyperactivity disorder; SRS-2=Social Responsiveness Scale – Second Edition.