Journal of Autism and Developmental Disorders
https://doi.org/10.1007/s10803-023-05990-z

ORIGINAL PAPER

Objective Measurement of Social Gaze and Smile Behaviors in Children with Suspected Autism Spectrum Disorder During Administration of the Autism Diagnostic Observation Schedule, 2nd Edition

Yeojin A. Ahn[1] · Jacquelyn M. Moffitt[1] · Yudong Tao[2] · Stephanie Custode[1] · Meaghan Parlade[1] · Amy Beaumont[1] · Sandra Cardona[1] · Melissa Hale[1] · Jennifer Durocher[1] · Michael Alessandri[1] · Mei-Ling Shyu[2] · Lynn K. Perry[1] · Daniel S. Messinger[1,2,3,4]

Accepted: 9 April 2023

Abstract
Best practice for the assessment of autism spectrum disorder (ASD) symptom severity relies on clinician ratings of the Autism Diagnostic Observation Schedule, 2nd Edition (ADOS-2), but the association of these ratings with objective measures of children's social gaze and smiling is unknown. Sixty-six preschool-age children (49 boys, $M = 39.97$ months, $SD = 10.58$) with suspected ASD (61 confirmed ASD) were administered the ADOS-2 and provided social affect calibrated severity scores (SA CSS). Children's social gaze and smiling during the ADOS-2, captured with a camera contained in eyeglasses worn by the examiner and parent, were obtained via a computer vision processing pipeline. Children who gazed more at their parents ($p = .04$) and whose gaze at their parents involved more smiling ($p = .02$) received lower social affect severity scores, indicating fewer social affect symptoms, adjusted $R^2 = .15$, $p = .003$.

Keywords Objective measurement · Autism spectrum disorder · Autism diagnostic observation schedule · Social gaze · Smiling

Autism spectrum disorder (ASD) is diagnosed on the basis of behavioral presentation and, by definition, involves persistent disturbances of social communication, as well as restricted and repetitive behaviors (APA, 2013). Current best practice for measuring children's ASD symptom severity is based on expert clinician observation during assessment, such as the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2), a semi-structured, play-based assessment that creates opportunities for observation of social communication skills and restricted repetitive behaviors or interests (Lord et al., 2012). Following the administration of the ADOS-2, clinicians provide calibrated severity scores (CSS) in the social affect (SA) and restricted and repetitive behaviors (RRB) domains. Although ADOS-2 CSS are the gold standard quantitative characterization of the ASD phenotype (Lord et al., 2012), the ADOS-2 training is time-intensive and costly (Dawson & Sapiro, 2019), and obtaining adequate interrater reliability in administration and scoring of the ADOS-2 is challenging in non-research settings (Kamp-Becker et al., 2018; Zander et al., 2016). Objective processing of children's social communication behaviors from video has the potential to provide quantifiable measures for key behavioral features of ASD during clinical assessments. In the present study, we obtained objective measurements of children's social behaviors (i.e., social gaze at adults and the social gaze involving smiles) from video recordings captured by adult-worn camera-embedded eyeglasses during the clinical assessment (ADOS-2).

ASD symptoms include reductions in spontaneously attending to others and decreased eye contact (i.e., social gaze; APA, 2013). In the social affect domain of ADOS-2

✉ Daniel S. Messinger
dmessinger@miami.edu

[1] Department of Psychology, University of Miami, Coral Gables, FL, USA

[2] Department of Electrical and Computer Engineering, University of Miami, Coral Gables, FL, USA

[3] Departments of Pediatrics and Music Engineering, University of Miami, Coral Gables, FL, USA

[4] Department of Psychology, University of Miami, 5665 Ponce de Leon Blvd., P.O. Box 248185, Coral Gables, FL 33124, USA

Published online: 27 April 2023

scoring, clinicians assess children's ability to make appropriate eye contact with them. If the child uses poorly modulated eye contact to initiate, terminate, or regulate social interaction, they receive a high symptom severity score. Early social gaze predicts the severity of ASD symptoms among infants at risk for ASD (Ibanez et al., 2013). Longitudinal studies suggest that the frequency of social gaze declines from 6 to 36 months of age in infants who were later diagnosed with ASD, with significant differences evident by 12 months when compared to low-risk, typically developing infants (Gangi et al., 2021; Ozonoff et al., 2010). Earlier eye-tracking studies also suggest that toddlers with ASD have difficulties directing attention to the face of an interactive partner who is trying to engage their attention using eye contact and child-directed speech (Chawarska et al., 2012; Jones et al., 2008; Wang et al., 2018).

A recent meta-analysis involving 91 eye-tracking studies of 4209 individuals (ASD: 2027; non-ASD: 2182) across all age ranges (i.e., infancy to adulthood) suggests that individuals with ASD attended to the eyes of a face for shorter durations than individuals without ASD, and the group differences in the time spent attending to the eyes were not moderated by age (Ma et al., 2021). In an eye-tracking study of gaze following in preschool children, forty children with autism (mean age = 5 years and 10 months, *SD* = 11 months, range = 45 months) attended to a video presentation of a woman gazing at the child for less time than 25 children with pervasive developmental disorder-not otherwise specified (mean age = 6 years and 2 months, *SD* = 8 months, range = 30 months; Falck-Ytter et al., 2012). These findings suggest decrements in social gaze for preschoolers with ASD and, in particular, autism, a particularly severe form of the disorder.

Assessments of social behaviors in young children (e.g., ADOS-2 (Lord et al., 2012), Communication and Symbolic Behavior Scales (Wetherby & Prizant, 2003), Early Social Communication Scales (Mundy et al., 2013) are administered by trained or certified examiners. However, manual coding of social gaze may yield varying detection of gaze behavior between coders and even within the same coder (Edmunds et al., 2017). This is the case even when frame-by-frame coding is employed and even when this coding is supported by a manual coding computer interface. On the other hand, objective coding of a given video recording via a given software pipeline will is designed to reliably produce the same coded output (e.g., gaze detection) on multiple occasions. Although eye tracking methods enable objective detection of social gaze, conventional screen-based eye tracking method is not suitable for measuring social gaze in naturalistic interactions. Likewise, wearable eye trackers are costly and may be burdensome to children with ASD (Chong et al., 2020).

Previous investigators have objectively measured children's social gaze at an examiner during semi-naturalistic adult–child interaction by implementing an adult-worn first-person video camera and a fully automated system for eye contact detection (Chong et al., 2017, 2020; Fathi et al., 2012; Rehg et al., 2014). Using a deep neural network model for automatic detection of gaze captured by the adult-worn first-person camera, Chong et al. (2020) found that greater frequency and duration of automatically measured direct gaze of children with ASD between the ages of 18 and 60 months during social communication assessments were associated with lower SA severity scores on the ADOS-2. The assessments involved interactions between the child and the examiner and were modified to ensure that both child and examiner were seated and facing each other at a table to better capture the child's social gaze at the examiner. In the current study, we adopted the gaze detection model developed by Chong et al. (2020) to examine the proportional duration of child's social gaze at both the examiner and parent during the ADOS-2 and its association with clinician rating of SA symptom severity on the ADOS-2. Unlike in Chong et al., children and examiners were allowed to move about the room as afforded by the ADOS-2 protocol.

Facial expressions, such as shared smiles, are another dimension of social communication associated with the heterogeneity of ASD presentation. In the ADOS-2, children are assessed on their ability to direct appropriate facial expressions (e.g., smiles) to the examiner or parent in order to communicate their affective states (Lord et al., 2012). A less appropriate and more limited range of facial expressions, as well as lack of smiling in response to the examiner and parent, are associated with higher symptom severity scores. Assessing temperament, Macari et al. (2018) found that lower intensity of joyful expression during a positive-emotion eliciting activity was associated with higher SA symptom severity among 43 children with ASD (mean age = 21.9 months, *SD* = 3.0). Further, there is evidence that the integration of gaze and positive affect (i.e., social smiling) may be especially impaired in children with ASD relative to those with developmental delays or typical development (Dawson et al., 1990; Joseph & Tager-Flusberg, 1997; Kasari et al., 1990). The current study investigated whether objectively measured proportion of social gaze involving smiling during a clinical assessment (i.e., ADOS-2) in preschool-age children at risk for ASD is associated with their ASD symptom severity scores.

Previous objective measurements of child facial expressions using computer vision, have been leveraged to better understand early emotional expressions and communication (Mattson et al., 2013; Messinger et al., 2009). More recently, tablet-based studies applying computer vision indicated that children with ASD more frequently displayed neutral expressions than children without ASD when watching emotion-eliciting videos (Carpenter et al., 2021; Egger et al., 2018). Although automated facial expression detection has yielded reliable measurement of child facial expressions, previous investigations have involved small samples of infants (Mattson et al., 2013; Messinger et al., 2009) or more structured protocols, such as facial expression analysis while viewing a video on a tablet (Carpenter et al., 2021) or posing an expression in front of a computer monitor (Manfredonia et al., 2019). The current study employed commercial software (iMotions, 2016), which has been used for previous studies involving individuals with ASD (Trevisan et al., 2016; Manfredonia et al., 2019) to detect preschool-age children's smiling from first-person videos captured with adult-worn glasses camera during a naturalistic interaction in the ADOS-2.

The degree to which expert clinician ADOS-2 ratings reflect objective differences in ASD-related social behaviors during the ADOS-2 and the degree to which these objective behaviors during the assessment correspond to clinician ratings of ASD symptom severity remain unclear. The ADOS-2 provides a naturalistic setting, in which the child is free to move around and spontaneously respond and initiate interactions with the examiner and the parent. The current study aimed to elucidate the degree to which objective measures of children's social gaze and smile behaviors during naturalistic interactions with both the examiner and the parent in the ADOS-2 remains consistent indices of key documented ASD deficits, such as reductions in social gaze and social smiles. We also investigated whether these objectively measured ASD-related behaviors would be evident in interactions with the examiner, parent, or both. The current study capitalized on automated measurement to investigate associations between clinical indices of ASD and objective measures of child social gaze (at both examiners and parents) and the proportional amount of smiling in the child's social gaze during the ADOS-2. Children's gaze and smile behaviors were recorded with adult-worn camera-embedded eyeglasses during the administration of the assessment. Quantification of ASD-related social communication behaviors using automatically acquired measures may potentially supplement current assessments and enhance clinical categorization and individualized referral based on *objective* information about an individual child's ASD symptoms, facilitating early detection of ASD and access to vital intervention services that can improve development.

The current study used first-person video from an adult-worn eyeglasses-embedded camera to detect social gaze and smiling in children being assessed for ASD during the ADOS-2. Child gaze and smile behaviors were measured using computer vision approaches that have previously been successfully utilized for examining gaze and smiles. We examined linear associations of objective measures of social gaze and smile behaviors (i.e., social gaze and social gaze involving smiling) with clinical indices of autism severity on the ADOS-2 (examiner ADOS-2 ratings). We hypothesized that objective measures of proportional duration of social gaze and the proportion of social gaze involving smiling would be associated with the expert ADOS-2 calibrated severity scores in the SA domain.

Methods

Participants

Sixty-six children (49 boys and 17 girls) with suspected ASD and their parents or legal guardians participated in the study. The mean age of the children was 39.97 months ($SD = 10.58$ months, range: 24–68 months). Parents indicated that 43 children were White (65.2%), 19 (28.8%) were Black, and four (6.1%) were multiracial. With respect to ethnicity, forty-eight children were identified as Hispanic or Latinx (72.7%). Participants were recruited through an autism specialty clinic housed in the Department of Psychology, which provides free ASD assessments for underserved families in the community. Based on this assessment (see "Procedure and Materials" section), 61 children (92.4%) received a diagnosis of ASD and 5 (7.6%) received a Developmental Delay (DD) diagnosis. Table 1 provides sample demographic information. All procedures were approved by the university Institutional Review Board.

Procedure and Materials

Families were invited to participate in the study during their child's clinical evaluation at a university-based autism specialty clinic. A research staff member described the study procedures and addressed questions. Written informed consent was obtained from parents or legal guardians prior to initiating the study protocol. The child, parent, and examiner were present for the Autism Diagnostic Observation Schedule, Second Edition (ADOS-2) administration, which was provided as part of the standard clinic assessment battery that also included administration of age-appropriate developmental assessments and parent interviews. The ADOS-2 is a 40–60 min semi-structured observational assessment of play, social interaction, and communicative skills that is designed as a diagnostic tool for identifying ASD (Lord et al., 2012). Children were administered one of four ADOS-2 modules—sets of developmentally appropriate activities—based on the child's age and expressive language use. Children completing module T or 1 (i.e., using no verbal language or only single words), module 2 (using phrase speech), or module 3 (fluent speech) were included in the current study. Module T is used with toddlers who are 30 months or younger and have no verbal language or only single words. Module 1 is

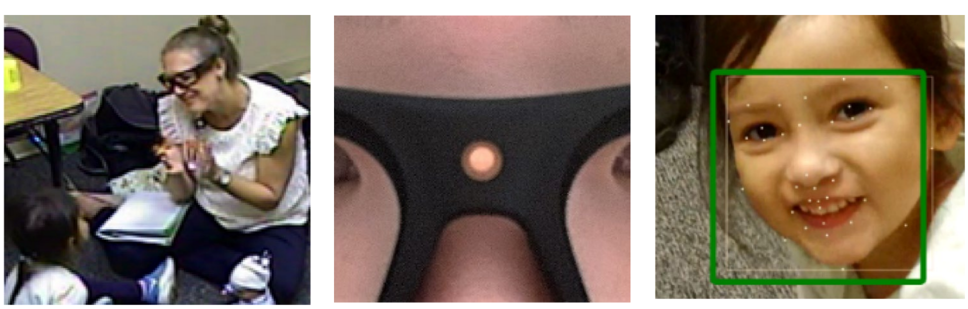
Table 1 Participant characteristics

	Total ($n=66$)	Module T ($n=14$)	Module 1 ($n=39$)	Module 2 ($n=9$)	Module 3 ($n=4$)
Mean (SD)					
Age (months)	40.0 (10.58)	28.6 (1.70)	39.4 (6.54)	53.7(10.90)	54.8 (9.43)
ADOS-2 Total CSS	6.7 (2.33)	7.57 (2.59)	6.4 (2.27)	6.4 (2.07)	6.8 (2.63)
ADOS-2 SA CSS	6.0 (2.66)	7.14 (3.21)	5.4 (2.29)	6.1 (2.80)	7.3 (2.87)
ADOS-2 RRB CSS	8.0 (2.03)	7.8 (1.42)	8.5 (1.86)	7.4 (2.88)	5.5 (1.00)
N (%)					
Sex					
Male	49 (74.2)	11 (78.6)	30 (76.9)	6 (66.7)	2 (50.0)
Female	17 (25.8)	3 (21.4)	9 (23.1)	3 (33.3)	2 (50.0)
Ethnicity					
Latinx	48 (72.7)	8 (57.1)	30 (76.9)	7 (77.8)	3 (75.0)
Non-Latinx	18 (27.3)	6 (42.9)	9 (23.1)	2 (22.2)	1 (25.0)
Race					
White	43 (65.2)	6 (42.9)	27 (69.2)	6 (66.7)	4 (100.0)
Black	19 (28.8)	6 (42.9)	11 (28.2)	2 (22.2)	0 (0.0)
Multiracial	4 (6.1)	2 (14.3)	1 (2.6)	1 (11.1)	0 (0.0)
Diagnosis					
ASD	61 (92.4)	13 (92.9)	37 (94.9)	7 (77.8)	4 (100.0)
DD	5 (7.6)	1 (7.1)	2 (51.1)	2 (22.2)	0 (0.0)

ADOS-2 Autism diagnostic observation schedule-second edition, *SA* Social affect, *RRB* Restricted and repetitive behavior, *CSS* Calibrated severity score, *ASD* Autism spectrum disorder, *DD* Developmental delay

used for children 31 months of age or older who are using no verbal language or only single words. Module 2 is used for children of any age who are using flexible phrase speech. Module 3 is used for children of any age who are using fluent speech. Of the sample, 14 children were evaluated using Module T, 39 with Module 1, 9 with Module 2, and 4 with Module 3 (see Table 1).

During the ADOS-2 administration, the examiner and parent wore camera-embedded eyeglasses (housing a small video camera in the nasal bridge; Fig. 1a, b), which provided continuous high resolution video recordings of children's social gaze and social gaze involving smiling (Fig. 1c). Pivothead Durango video recording camera glasses (specifications: 1080p video resolution, 30 fps, 75° field-of-view, 44.1 kHz audio recording microphone; Pivothead, Denver, Colorado, USA) were worn by 30 examiners and 30 parents, and Orca video recording camera glasses (specifications: 1080p video resolution, 30 fps, 147° field-of-view, 32 kHz audio recording microphone; Zentronix Corp, Allston, Massachusetts, USA) were worn by 36 examiners and 36 parents. Three Pivothead and two Orca parent camera did not provide recordings due to technical errors. Thus, a total of



(a) (b) (c)

Fig. 1 **a** An examiner wearing the camera-embedded eyeglasses during the ADOS-2. **b** A close-up image of the embedded video camera in the eyeglasses. **c** A video frame of child's gaze captured with the eyeglasses camera, in which FACET algorithm (green box) detects the presence of smile. Written permission was obtained from the parent of the child in the photos

66 examiner and 61 parent videos were available for video processing. The lenses were removed from the glasses to provide an unobstructed view of the adult's eyes. Parents were asked to complete a demographics questionnaire during the assessment. The recording time for each video included parent questionnaire completion, which allowed us to capture children's spontaneous social gaze and smile behaviors toward their parents.

Measures

Autism Diagnostic Observation Schedule-2 (ADOS-2)

The study examiner's administration of the ADOS-2 provided scores in the SA domain. The SA includes items that are pertinent to communication and reciprocal social interaction, such as eye contact, facial expressions directed to others, and shared enjoyment in interaction. The ADOS-2 provides CSS based on data-derived 14-item algorithms composed of the most discriminative items in each module (Esler et al., 2015; Hus et al., 2014; Janvier et al., 2021; Lord et al., 2012). The CSS ranges from 1 (*little to no evidence of ASD-related symptomology*) to 10 (*high level of symptoms*) and is designed to quantify severity that is comparable across modules, by accounting for child age and language use. We examined associations of objective measures of social gaze and smile behaviors during the ADOS-2 with SA and RRB CSS (See Supplementary Information for associations of objective measures with a specific ADOS-2 item assessing unusual eye contact).

Diagnosis

Diagnoses were provided by one of three ASD-experienced clinical psychologists based on DSM-5 criteria (APA, 2013) informed by the ADOS-2, parental interviews, and developmental assessments that varied based on the child's age. Each child was seen by only one psychologist. Sixty-one children (92.4%) received a diagnosis of ASD, and 5 children (7.6%) received a diagnosis of DD. Three children who were assigned DD diagnoses received Total (combined SA and RRB domains) CSS suggestive of ASD (≥ 4) and 5 children with ASD diagnoses received Total CSS of 3.

Objective Measurements

The first-person video approach and computer vision processing pipeline allowed for efficient objective detection of social gaze (Chong et al., 2020) and smiling (Bartlett et al., 2008; Littlewort et al., 2011). Sixty-six examiner first-person videos and 61 parent first-person videos captured with the camera-embedded eyeglasses were processed with a pipeline consisting of face detection (Zhang et al., 2016) and age estimation of the detected faces for obtaining video streams of only the child face (Rothe et al., 2018), which were then processed with a gaze estimation model (Chong et al., 2020) and the FACET facial action unit estimation algorithm (Littlewort et al., 2011; iMotions, 2016; see Fig. 2). The code for the current processing pipeline is available at https://github.com/taoyudong/ADOS_Child_Face_Analysis.

The first part of the pipeline focused on obtaining facial images of child faces from each examiner and parent video. The raw videos captured with the eyeglasses were input to a deep cascaded multi-task convolutional neural network

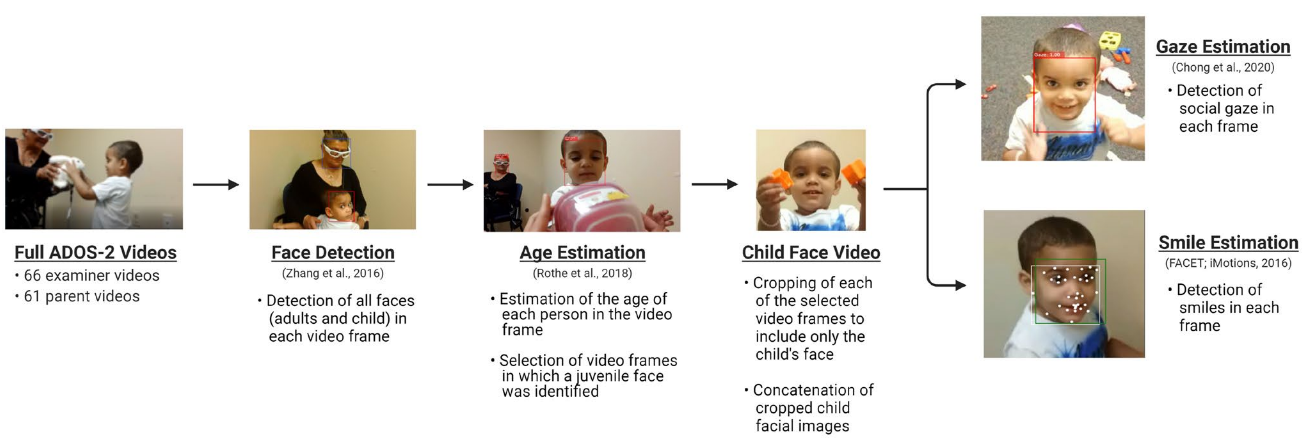


Fig. 2 The pipeline of video processing for obtaining objective measurements of social gaze and social gaze involving smiles. The overall proportion of social gaze was computed as the number of frames in which social gaze was detected divided by the total number of frames in the full ADOS-2 raw video. The proportion of social gaze involving smiles was computed as the number of frames in which both social gaze and smiling were detected divided by the number of frames in which social gaze was detected. Both social gaze and social gaze involving smiles were calculated separately for child actions directed at the parent and at the examiner

(MTCNN; Zhang et al., 2016), which is a state-of-the-art algorithm to detect faces from the images using convolutional neural network architecture. Previous assessments of the MTCNN's area under curve indicated a range of 0.61 to 0.95 on various sets of images (Jain & Learned-Miller, 2010; Yang et al., 2016, 2019), some of which were at various difficulty levels (easy, medium, and hard sets; Jain & Learned-Miller, 2010). Then, the Deep Expectation model (Rothe et al., 2018) was used to estimate the age of each person in the video frame, based on visual features of the face. The model had achieved the average error of 2.68 years and 3.09 years on two different datasets (Panis et al., 2016; Ricanek & Tesafaye, 2006). The video frames in which a juvenile face was identified were selected. Each of these video frames was cropped to include only the child's face. The cropped child facial images were normalized to the same 224×224 pixel size and concatenated into video streams – hereby referred to as the *child face videos*. A total of 66 child face videos from the examiner-worn camera and 61 child face videos from the parent-worn camera were obtained.

Social Gaze Detection Child social gaze was estimated from the child face videos using a deep convolutional neural network (CNN) classifier model trained to detect moments of eye contact in egocentric view (Chong et al., 2020). This model was trained on over 4 million facial images of more than 100 children with diverse demographic backgrounds from the examiner-worn camera-embedded eyeglasses during social communication assessments. Chong et al. (2020) chose a decision threshold of 0.9 via grid search on a held-out training sample in order to predict the presence or absence of eye contact in each frame based on the final eye contact score produced by the model. That classifier model yielded an F1 score of 0.93. The model provides a probability score of child social gaze, ranging from 0 to 1, for each video frame. In the current study, we used a decision threshold of .75 to predict the presence or absence of social gaze in each frame. That is, the presence of child social gaze in each frame was indicated if the model estimated .75 or greater for the probability of social gaze. This criterion was based on the area under the Precision/Recall curve that maximized detection accuracy while minimizing the event-level social gaze count difference between the estimate and human annotation of 53,734 facial images from 24 child face videos. The overall proportion of social gaze equaled the number of frames in which gaze was detected divided by the total number of frames in the raw video (the entire ADOS-2 duration). Thus, the denominator of the proportion included both frames in which the child's face was and was not detected. Social gaze at the parent and social gaze at the examiner were calculated separately.

Smile Detection Smiling (Facial Action Coding System Action Unit 12) was estimated for each frame of the child face videos, based on the face detection function of the Facial FACET module in the iMotions Biometric Research Platform (iMotions, 2016; Manfredonia et al., 2019), which is a commercial tool for automated facial expression analysis. Using computer vision algorithms, FACET (formerly the Computer Expression Recognition Toolbox; CERT; Littlewort et al., 2011) provides automatic identification of the probability of the activity of specific facial actions (i.e., Action Units; AUs) from the Facial Action Coding System (FACS), by estimating the locations of 20 facial features. FACET provided an 'Evidence Value' for the activated facial action unit 12 (AU12; indexing smile), estimated as the odds, on a logarithmic (base 10) scale, of AU12 being present in each frame. For example, the Evidence Value 0 indicates that there is an equal chance that the expression is to be categorized by an expert human coder as expressing AU12 than not. The negative Evidence Values indicated the AU12 being less likely to be categorized by an expert human coder as not expressing the AU12 than expressing it. We considered frames in which the odds were greater than 0 to indicate AU12 being present. The proportion of social gaze involving smiling was computed as the number of frames in which both social gaze and AU12 (i.e., smiling) were detected divided by the number of frames in which social gaze was detected. Social gaze at the parent that involved smiling was calculated independently from social gaze at the examiner that involved smiling.

Manual Annotation Twenty-four child face videos (12 examiner videos and 12 parent videos; a total of 53,734 facial images) were randomly selected for frame-by-frame manual coding of social gaze and smiling to assess the concordance between objective measurement and manual human annotation in the current dataset. Moderately high concordance between objective measurement of child social gaze and manual coding was indicated by the F1 score of .78 (92.6% agreement, $\kappa = .78$; McHugh, 2012). Lower concordance (F1 score = .66, 69% agreement, $\kappa = .38$) was obtained for smile between objective measurement (obtained with FACET) and manual coding.

Results

IBM SPSS Statistics version 26.0.0 for Windows was used for analyses. We first describe objective social gaze and smile behavior measures during the ADOS-2 and then examine their associations with children's ADOS-2 scores.

We obtained an average of 77,377.29 video frames ($SD = 27,974.50$), approximately 43 min, of the ADOS-2 from the examiner-worn camera and 72,538.15 frames

($SD = 25{,}123.45$; approximately 40 min) from the parent-worn camera. The examiner-worn camera detected an average of 5660.58 frames ($SD = 5425.19$) of the child's face, and the parent worn-camera detected an average of 3716.34 frames ($SD = 3942.56$) of the child's face. The mean proportion of frames in which a child's face was detected by the examiner-worn camera was .07 ($SD = .05$). The mean proportion of frames in which a child's face was detected by the parent-worn camera was .05 ($SD = .04$). The proportion of video frames in which a child's face was detected by the examiner-worn camera was associated with child ADOS-2 SA CSS, $r(64) = .29$, $p = .02$. That is, children whose faces were detected at higher rates in video from the examiner-worn camera exhibited higher social affect symptomatology. There was no association between the proportion of frames in which a child's face was detected in video from the parent-worn camera and the SA CSS, $r(59) = .10$, $p = .44$.

Distributions of the proportion of child social gaze and social gaze involving smiling (at both parent and examiner) are shown in Fig. S1. Analyses included all data and were robust to outlier removal (see Supplementary Information). We examined the distribution of objectively measured social gaze and smiling involved in social gaze by demographic variables (Table 2). Non-parametric Welch's F-tests were performed for all comparisons to account for non-equivalent sample sizes. These objective measures did not differ by child sex (female versus male), ethnicity (Latinx versus non-Latinx), or the ADOS-2 modules (T versus 1 versus 2 versus 3; see Table 2). The proportion of social gaze at examiner differed by child race (White versus Black versus Multiracial), $F(2, 11.34) = 5.47$, $p = .02$. A contrast analysis indicated that White children engaged in a greater proportion of social gaze at the examiner than non-White (Black and Multiracial combined; $M = .004$, $SD = .003$, $n = 23$) children, $t(46.46) = 2.93$, $p = .01$, $d = .72$. No difference was observed for child social gaze at the parent based on child race. Although overall the proportions of social gaze at the parent that involved smiling varied by race, $F(2, 9.18) = 5.17$, $p = .03$, Games-Howell post-hoc pairwise comparisons indicated no significant differences between White and Black ($p = .96$), White and Multiracial ($p = .05$), or Black and Multiracial children ($p = .09$). Further, a contrast analysis revealed no difference between White ($n = 38$) and Black/Multiracial ($n = 21$) children, $t(56) = 1.33$, $p = .19$, $d = .19$.

Table 3 shows overall pairwise univariate correlations between objective measurements of children's gaze and smile behaviors, ADOS-2 severity scores, and child age. We are alert to the possibility of inflation of alpha resulting from multiple correlations. Consequently, we applied a Benjamin-Hochberg correction with a false discovery rate of 5% to correlations between objective measurements of children's gaze and smile behaviors and ADOS-2 severity scores to account for Type 1 error. There were no significant associations between objective measures and the ADOS-2 RRB CSS (see Table 3). The RRB CSS was positively associated with the total CSS, $r = 0.47$, $p < .001$, but not with the SA CSS, $r(63) = .19$, $p = .21$. The proportion of social gaze at the examiner was not significantly associated with the SA CSS, $r(64) = -.06$, $p = .72$. The proportion of social gaze at the examiner that involved smiling was inversely associated with the SA CSS, $r(63) = -.28$, $p = .04$. The proportion of social gaze at the examiner was not associated with the proportion of social gaze at the examiner that involved smiling, $r(63) = .01$, $p = .97$. The proportion of social gaze at the parent was also inversely associated with the SA CSS, $r(59) = -.32$, $p = .04$. Likewise, the proportion of social gaze at the parent involving smiling was inversely associated with the SA CSS, $r(57) = -.34$, $p = .03$. The proportion of social gaze at the parent was not associated with the proportion of social gaze at the parent that involved smiling, $r(57) = .22$, $p = .20$. Partial correlations controlling for child race did not affect the significance of these associations (see Table S1). Uncorrected univariate correlations are reported in Table S2.

Stepwise multiple regressions were performed to establish the best predictors of the SA CSS. Initial regression models included the objective measures that had a significant univariate association ($p < .05$) with the SA CSS (i.e., proportion of social gaze at parent and the proportions of social gaze at parent and examiner involving smiling). Both forward and backward stepwise regressions yielded the same final model. In the final model, the proportion of social gaze at the parent and the proportion of social gaze at the parent involving smiling were unique predictors of the SA CSS, adjusted $R^2 = .15$, $F(2, 56) = 6.29$, $p = .003$ (Table 4, Fig. 3). We found no collinearity between these variables (Tolerance = .95, VIF = 1.05). A greater proportion of social gaze at parent, $\beta = -141.38$, $t(56) = -2.12$, $p = .04$, and a greater proportion of social gaze at parent that involved smiling, $\beta = -4.15$, $t(56) = -2.32$, $p = .02$, were associated with lower SA CSS.

Sensitivity analyses that included only children with ASD diagnoses ($n = 61$) and only children who received ADOS module T or 1 ($n = 53$) yielded results equivalent to those including the entire sample (see Supplementary Information).

Discussion

The current study examined associations between the objective measures of social gaze and smile behaviors in children and ASD symptom severity in a sample of children being assessed for ASD, the overwhelming majority of whom received an ASD diagnosis. Objective quantification of ASD-related characteristics of social gaze and smiles occurring during the ADOS-2 assessment indicate the potential of

Table 2 Group comparisons of objective measurements of social gaze and smile behaviors

Objective measures		*n*	M	SD	F	df1	df2	*p*
Proportion of social gaze at examiner								
Child sex								
	Male	49	.009	.013	.064	1	42.168	.802
	Female	17	.009	.009				
Child ethnicity								
	Latinx	48	.009	.008	.193	1	19.412	.665
	Non-Latinx	18	.007	.019				
Child race								
	White	43	.011	.014	5.473	2	11.344	.022*
	Black	19	.004	.004				
	Multiracial	4	.006	.003				
ADOS-2 module								
	Mod T	14	.007	.007	1.016	3	11.048	.422
	Mod 1	39	.007	.007				
	Mod 2	9	.006	.005				
	Mod 3	4	.036	.033				
Proportion of social gaze involving smiles at examiner								
Child sex								
	Male	48	.449	.170	1.337	1	25.522	.258
	Female	17	.388	.191				
Child ethnicity								
	Latinx	47	.435	.176	.023	1	30.056	.879
	Non-Latinx	18	.428	.181				
Child race								
	White	42	.418	.182	.498	2	8.088	.625
	Black	19	.451	.162				
	Multiracial	4	.506	.200				
ADOS-2 module								
	Mod T	14	.404	.152	.609	3	11.521	.622
	Mod 1	38	.432	.187				
	Mod 2	9	.498	.167				
	Mod 3	4	.396	.186				
Proportion of social gaze at parent								
Child sex								
	Male	45	.004	.006	2.029	1	58.518	.160
	Female	16	.002	.002				
Child ethnicity								
	Latinx	43	.004	.004	.125	1	22.139	.727
	Non-Latinx	18	.003	.007				
Child race								
	White	39	.004	.006	2.544	2	10.593	.125
	Black	18	.002	.003				
	Multiracial	4	.002	.002				
ADOS-2 module								
	Mod T	13	.004	.004	.550	3	10.346	.659
	Mod 1	36	.003	.004				
	Mod 2	8	.002	.003				
	Mod 3	4	.010	.013				

Table 2 (continued)

Objective measures		*n*	M	SD	F	df1	df2	*p*
Proportion of social gaze involving smiles at parent								
Child sex								
	Male	44	.394	.180	.080	1	22.305	.780
	Female	15	.377	.199				
Child ethnicity								
	Latinx	41	.411	.180	1.775	1	31.578	.192
	Non-Latinx	18	.341	.186				
Child race								
	White	38	.402	.187	5.165	2	9.185	.031*
	Black	18	.387	.186				
	Multiracial	3	.246	.066				
ADOS-2 module								
	Mod T	13	.351	.127	.555	3	10.931	.655
	Mod 1	34	.398	.200				
	Mod 2	8	.438	.173				
	Mod 3	4	.349	.242				

F F statistics of Welch's test; *df*1 between-groups degrees of freedom; *df*2 within-groups degrees of freedom

$^*p<.05$

Table 3 Associations among objective measurements of child's social gaze and smile behaviors during the ADOS-2, ADOS-2 severity scores, and child age

	1	2	3	4	5	6	7	8
1. Child age in months	–							
2. Total CSS	.02	–						
3. SA CSS	.02	.92**	–					
4. RRB CSS	.02	.47**	.19	–				
5. Proportion of social gaze at examiner	.27*	− .12	− .06	− .15	–			
6. Proportion of social gaze at examiner involving smiles	.06	− .23	− .28*	.14	.005	–		
7. Proportion of social gaze at parent	.06	− .34*	− .32*	− .11	.55**	.04	–	
8. Proportion of social gaze at parent involving smiling	.06	− .23	− .34*	.15	.09	.59*	.22	–

Pearson's correlation coefficients (two-tailed) are shown in the table

The *p*-values are adjusted using the Benjamini–Hochberg correction

*Adjusted $p<.05$; **adjusted $p<.01$

using first-person video for capturing social behaviors relevant to the ASD phenotype. In accordance with the research hypothesis, objective measurement of children's social gaze and smile behaviors during a clinical assessment were associated with clinician ratings of SA symptom severity. Children who demonstrated higher levels of social gazing at the parent and whose social gaze at the parent involved more smiling during the ADOS-2 received lower ADOS-2 scores in the SA domain. Our results are congruent with other reports of associations between objective measurement of children's social gaze and smile behaviors and ASD symptom severity, underscoring the potential of digital measures for better understanding the ASD phenotype (Chong et al., 2020; Dawson & Sapiro, 2019; Hashemi et al., 2018).

Chong et al. (2020) reported that, among 25 children with ASD, those who demonstrated more frequent and longer duration of social gaze at the examiner during the Brief Observation of Social Communication Change (BOSCC; an assessment of change in ASD symptoms) received lower examiner-rated BOSCC SA symptom severity scores. In the current study, adopting the first-person video approach that Chong et al. (2020) pioneered, objective measures of social gaze at the examiner in the ADOS-2 were not associated with the examiner ratings of ADOS-2 SA symptom severity scores. In Chong et al. (2020), the social communication assessments were modified in order to elicit more instances of child's direct gaze. Examiner-child interaction during the assessments took place at a table with the child sitting directly across from the examiner, allowing for more conducive setting for collecting gaze data. Other prior work on automated measurement of "response to name" behaviors (Campbell et al., 2019; Hashemi et al., 2018), which included both ASD and typically developing children, also involved a structured setting with both the child and examiner stationary, as well as the use of video recording of the child behavior captured with a stationary tablet.

The strength of associations between objective measures of social gaze and SA severity scores in the ADOS-2 were lower than the association between objectively measured social gaze and SA symptom severity score in the BOSCC in Chong et al., (2020; $r = -.78$ for duration). Discrepancies in the strength of association may be due to protocol differences. The ADOS-2 administration in the current study afforded the majority of children greater freedom of movement. Especially during ADOS-2 Modules T and 1, and to some degree during Module 2, children could spontaneously explore the room, and examiners adapted to children's movement as they conducted the administration. The current results indicate that objective measurement of social gaze shows some concordance with clinician ratings of ASD symptom severity. While rates of concordance or levels of association were not high, this is the first study in which social gaze was measured in a context in which children were free to locomote.

We found associations between ADOS-2 SA CSS and the objective measure of social gaze at the parent, but not with social gaze at the *examiner.* One possibility is that parents were generally seated during the ADOS-2. Thus, gaze at the parent required the child to initiate the interaction. By contrast, the examiner was required to frequently initiate contact with the child to complete activities in the ADOS-2. As the examiner's role included facilitating interaction, the examiner gazed frequently at the child. This may have created a baseline level of child social gaze at the examiner, rendering variance in social gaze at the examiner less meaningful with respect to symptom severity.

The current smiling results were consistent with previous studies documenting reduced social smiling in ASD. There is a robust evidence for impairment in the integration of gaze and smiling in children with ASD relative to those without ASD (Dawson et al., 1990; Joseph & Tager-Flusberg, 1997; Kasari et al., 1990; Ozonoff et al., 2010). Studies of infants at risk for ASD indicated lower rates of social smiling for at-risk infants who are later diagnosed with ASD compared to their typically developing peers (Brian et al., 2008; Landa et al., 2007; Nichols et al., 2014; Ozonoff et al., 2010). Similarly, using automated analysis of facial

Table 4 Final regression model: Objective measures of child social gaze and smile behaviors predicting ADOS-2 social affect calibrated severity score ($n = 66$)

Model	B	Std. Error	β	*t*	*p*	Zero-order	Partial	Tolerance	VIF	*F*	R^2	Adjusted R^2	*p*
(Constant)	8.11	.76		10.73	.000					6.29	.43	.15	.003**
Proportion of social gaze at parent	− 141.38	66.71	− .26	− 2.12	.04*	− .32	− .27	.95	1.05				
Proportion of social gaze at parent involving smiling	− 4.15	1.79	− .29	− 2.32	.02*	− .34	− .30	.95	1.05				

Results of stepwise regression employing a $p < .05$ inclusion criterion

$^*p < .05$; $^{**}p < .01$

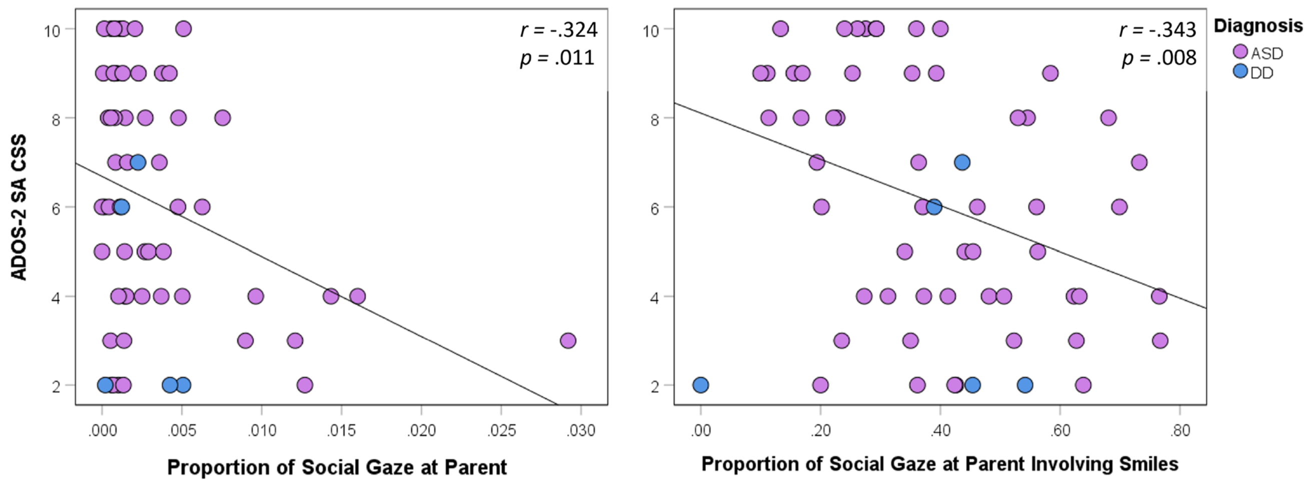


Fig. 3 Proportions of social gaze at parent and social gaze at parent involving smiling versus ADOS-2 social affect calibrated severity score. r = Pearson correlation coefficient (two-tailed). The x-axes show objective measures of social gaze and smile behaviors at parent. The y-axes show the ADOS-2 social affect calibrated severity score (SA CSS). The purple dots represent children who received an ASD diagnosis. The blue dots represent children who received a non-ASD developmental delay (DD) diagnosis. Removing the outlier (purple) in the bottom right portion of the left graph did not appreciably affect the correlation between the proportion of social gaze at the parent and SA CSS, $r(58) = -.31$, $p = .02$ (see Supplementary Information)

expressions, Manfredonia et al. (2019) found less evidence of objectively detected smiling (via FACET) in older children with ASD (mean age = 9.2 years), when asked to show happiness, was associated with greater social communication difficulties reported by parents. Likewise, in the current study, zero-order correlations indicate that higher levels of objectively measured smiling when gazing at the examiner and when gazing at the parent were associated with lower ASD symptom severity in preschoolers. Current findings based on objective measures corroborate earlier evidence of reduced social smiling in children with ASD and suggest the potential of objective measures of preschool-age children's social gaze and smile behaviors that are key ASD symptom domains (Lord et al., 2012). Further, social gaze and smile behaviors at the parent were the only unique predictors of SA symptom severity in multiple regression models. These results suggest the clinical importance of children's social gaze and smile behaviors at parents even while being assessed by an examiner.

Although the regression model indicated that social gaze at the examiner that involved smiling was not a unique predictor of SA CSS, zero-order correlations indicated that the objective measure of social gaze at the examiner that involved smiling was associated with the SA CSS. Thus, sharing positive emotion with the examiner remains a potentially useful correlate of social affect symptom severity, supporting the potential utility of examiner as well as parent-worn video cameras in future research.

This study appears to be the first to examine associations between concurrent objective measures of social gaze and smiling during the ADOS-2 and examiner-rated autism symptom severity. Like Chong et al., we adopted adult-worn first-person video camera that avoided intrusiveness and burden on children with ASD. In addition, the approach utilized readily available software to detect social gaze and smiling. Although preprocessing of the video stream (identification of child faces) was used, we have made software code for all processing steps available (see "Methods" section).

The current study has several noteworthy limitations. The study design did not afford prediction of future symptoms or associations with other ASD symptom severity measures. In addition, the limited sample size of children in a relatively large age range (24–68 months) necessitated the use of four ADOS-2 modules (T, 1, 2, and 3). Consequently, we examined associations of objective measurements with calibrated severity scores that allowed for comparison across ADOS-2 modules. Moreover, the use of a post hoc criterion of .75 for identifying social gaze (75% or higher probability of social gaze indicated the presence of social gaze) in each frame of the videos maximized reliability in the current sample but may not be optimally generalizable to other samples. Likewise, low reliability for the detection of smiling between FACET and human coders suggests limitations of the FACET algorithm. The FACET algorithm was trained on adult facial expressions (iMotions, 2016; Littlewort, 2011), and there is an urgent need for training on child faces that can be extended to naturalistic settings (Rehg, 2013).

Previous validation studies of automated facial expression analysis reported that facial expression analysis algorithms (e.g., FACET) appear to be most accurate for White faces, as the databases used in training FACET algorithms included only White faces (Phillips et al., 2011; Stöckli et al., 2018).

Our results indicated that White children engaged in a greater proportion of social gaze at the examiner than Black and Multiracial children. This may suggest that our automated processing pipeline for child social gaze and smile behaviors may be biased to identify White faces more accurately. Alternately, White children may have gazed more at the examiner than Black and Multiracial children. It is also noteworthy that the three clinicians (all self-identified as White, of whom one identified as Latina) were a different race than the Black children. This may have contributed to lower levels of social gaze at the examiner for these children. Crucially, the strength of association between the proportion of social gaze at the examiner and social affect scores was comparable between White ($r = -.12$), Black ($r = -.12$), and Black/Multiracial ($r = -.08$) children. Future investigations should continue to scrutinize how racial and cultural factors may influence objective behavioral phenotyping in children affected by ASD.

It is also possible that the use of the eyeglasses during the ADOS-2 altered the assessments. For example, a child may gaze more at a parent who typically does not wear glasses, because they were intrigued by the novelty of the eyeglasses during the ADOS-2. However, clinicians' anecdotal feedback indicated that they did not feel that wearing the eyeglasses during the ADOS-2 was disruptive, nor that it altered assessment outcomes. Nevertheless, future research could empirically compare assessment outcomes of children who were administered the ADOS-2 with adults who did and did not wear camera-embedded eyeglasses.

Conclusions

The current study provides an initial investigation into the objective detection of social gaze and smile behaviors during the gold-standard ASD assessment. During the ADOS-2, higher levels of objectively measured child social gaze at the parent and a higher proportion of social gaze at the parent that involved smiling were uniquely associated with lower examiner ADOS-2 SA severity scores. These findings contribute to the cross-validation of examiner ratings of SA domain symptoms and objective measurements of social behavior.

Further development of the technology and computational methods (e.g., machine learning) used to identify social communication behaviors, including social gaze and smiling, may allow for more efficient indexing of ASD symptom severity in larger samples (Bone et al., 2015; Dawson & Sapiro, 2019; de Belen et al., 2020; Moffitt et al., 2022; Rehg et al., 2014). Nevertheless, the current results suggest that objective measures of children's social gaze and smile behaviors reflect underlying ASD-related characteristics. If replicated, these objective measurement approaches have the potential to supplement current assessments and inform clinician severity ratings of young children being assessed for ASD, which may increase the capacity of clinicians to observe children in multiple contexts (e.g., home, school) to inform diagnostic decisions. At the most basic level, clinicians might consider the use of objective measures of a child's level of social gaze (and social gaze accompanied by smiling), calibrated in the context of clinic-specific levels of these behaviors, in adjudicating a child's level of social affective communication disturbances. Through such initiatives and continued investigation, the current project may enhance early detection and treatment of ASD and contribute to improving the lives of individuals with ASD and other neurodevelopmental disorders.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10803-023-05990-z.

Acknowledgments We thank the families who participated. The project described was supported in part by Grant Number UL1TR002736, Miami Clinical and Translational Science Institute, from the National Center for Advancing Translational Sciences and the National Institute on Minority Health and Health Disparities. The project was also supported in part by an Autism Science Foundation Predoctoral Fellowship (18-001) entitled "Automating and accelerating the autism diagnostic process," awarded to Y. A. Ahn.

Author Contributions Conceptualization: YAA, JMM, DSM; Methodology: YAA, JMM, YT, SC, LKP, DSM; Investigation: YAA, JMM, YT, SC, MP, AB, SC, DSM; Formal analysis: YAA, JMM, YT, LKP, DSM; Writing – original draft preparation: YAA, DSM; Writing – review and editing: all authors; Funding acquisition: YAA, LKP, DSM; Supervision: MH, JD, MA, MS, LKP, DSM.

Code Availability The code for the current processing pipeline is available at https://github.com/taoyudong/ADOS_Child_Face_Analysis.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The University of Miami Institutional Review Board (IRB) approved this study.

Consent to Participate Written informed consent was obtained from the parents.

Consent to Publish The authors affirm that human research participants provided informed consent for publication of the images in Figs. 1a, c, and 2.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders: DSM-5.

Bartlett, M., Littlewort, G., Wu, T., & Movellan, J. (2008). Computer expression recognition toolbox. In *Proceedings of 2008 8th IEEE International Conference on Automatic Face & Gesture Recognition*, Amsterdam, Netherlands. https://doi.org/10.1109/AFGR.2008.4813406.

Bone, D., Goodwin, M. S., Black, M. P., Lee, C., Audhkhasi, K., & Narayanan, S. (2015). Applying machine learning to facilitate autism diagnostics: Pitfalls and promises. *Journal of Autism and Developmental Disorders, 45*, 1121–1136. https://doi.org/10.1007/s10803-014-2268-6

Brian, J., Bryson, S., Garon, N., Roberts, W., Smith, I. M., Szatmari, P., & Zwaigenbaum, L. (2008). Clinical assessment of autism in high-risk 18-month-olds. *Autism, 12*(5), 433–456. https://doi.org/10.1177/1362361308094500

Campbell, K., Carpenter, K. L., Hashemi, J., Espinosa, S., Marsan, S., Borg, J. S., Chang, Z., Qiu, Q., Vermeer, S., Adler, E., Tepper, M., Egger, H. L., Baker, J. P., Sapiro, G., & Dawson, G. (2019). Computer vision analysis captures atypical attention in toddlers with autism. *Autism, 23*(3), 619–628. https://doi.org/10.1177/1362361318766247

Carpenter, K. L. H., Hahemi, J., Campbell, K., Lippmann, S. J., Baker, J. P., Egger, H. L., Espinosa, S., Vermeer, S., Sapiro, G., & Dawson, G. (2021). Digital behavioral phenotyping detects atypical pattern of facial expression in toddlers with autism. *Autism Research, 14*(3), 488–499. https://doi.org/10.1002/aur.2391

Chawarska, K., Macari, S., & Shic, F. (2012). Context modulates attention to social scenes in toddlers with autism. *Journal of Child Psychology and Psychiatry, 53*(8), 903–913. https://doi.org/10.1111/j.1469-7610.2012.02538.x

Chong, E., Chanda, K., Ye, Z., Southerland, A., Ruiz, N., Jones, R. M., Rozga, A., & Rehg, J. M. (2017). Detecting gaze towards eyes in natural social interactions and its use in child assessment. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 1*(3), 1–20. https://doi.org/10.1145/3131902

Chong, E., Clark-Whitney, E., Southerland, A., Stubbs, E., Miller, C., Ajodan, E. L., Silverman, M. R., Lord, C., Rozga, A., Jones, R. M., & Rehg, J. M. (2020). Detection of eye contact with deep neural networks is as accurate as human experts. *Nature Communications, 11*(1), 6386. https://doi.org/10.1038/s41467-020-19712-x

Dawson, G., Hill, D., Spencer, A., Galpert, L., & Watson, L. (1990). Affective exchanges between young autistic children and their mothers. *Journal of Abnormal Child Psychology, 18*(3), 335–345. https://doi.org/10.1007/BF00916569

Dawson, G., & Sapiro, G. (2019). Potential for digital behavioral measurement tools to transform the detection and diagnosis of autism spectrum disorder. *JAMA Pediatrics, 173*(4), 305–306. https://doi.org/10.1001/jamapediatrics.2018.5269

de Belen, R. A. J., Bednarz, T., Sowmya, A., & Del Favero, D. (2020). Computer vision in autism spectrum disorder research: a systematic review of published studies from 2009 to 2019. *Translational Psychiatry*. https://doi.org/10.1038/s41398-020-01015-w

Edmunds, S. R., Rozga, A., Li, Y., Karp, E. A., Ibanez, L. V., Rehg, J. M., & Stone, W. (2017). Brief report: Using a point-of-view camera to measure eye gaze in young children with autism spectrum disorder during naturalistic social interactions: A pilot study. *Journal of Autism and Developmental Disorders, 47*, 898–904. https://doi.org/10.1007/s10803-016-3002-3

Egger, H. L., Dawson, G., Hashemi, J., Carpenter, K. L. H., Espinosa, S., Campbell, K., Brotkin, S., Schaich-Borg, J., Qiu, Q., Tepper, M., Baker, J. P., Bloomfield, R. A., & Sapiro, G. (2018). Automatic emotion and attention analysis of young children at home: a ResearchKit autism feasibility study. *npj Digital Medicine, 1*(1), 20. https://doi.org/10.1038/s41746-018-0024-6

Esler, A. N., Bal, V. H., Guthrie, W., Wetherby, A., Weismer, S. E., & Lord, C. (2015). The autism diagnostic observation schedule, toddler module: Standardized severity scores. *Journal of Autism and Developmental Disorders, 45*(9), 2704–2720. https://doi.org/10.1007/s10803-015-2432-7

Falck-Ytter, T., Fernell, E., Hedvall, Å. L., von Hofsten, C., & Gillberg, C. (2012). Gaze performance in children with autism spectrum disorder when observing communicative actions. *Journal of Autism and Developmental Disorders, 42*, 2236–2245. https://doi.org/10.1007/s10803-012-1471-6

Fathi, A., Li, Y., & Rehg, J. M. (2012). Learning to recognize daily actions using gaze. *European Conference on Computer Vision*. https://doi.org/10.1007/978-3-642-33718-5_23

Gangi, D. N., Boterberg, S., Schwichtenberg, A. J., Solis, E., Young, G. S., Iosif, A.-M., & Ozonoff, S. (2021). Declining gaze to faces in infants developing autism spectrum disorder: Evidence from two independent cohorts. *Child Development, 92*, e285–e295. https://doi.org/10.1111/cdev.13471

Hashemi, J., Dawson, G., Carpenter, K. L., Campbell, K., Qiu, Q., Espinosa, S., Marsan, S., Baker, J. P., Egger, H. L., & Sapiro, G. (2018). Computer vision analysis for quantification of autism risk behaviors. *IEEE Transactions on Affective Computing, 12*(1), 215–226. https://doi.org/10.1109/TAFFC.2018.2868196

Hus, V., Gotham, K., & Lord, C. (2014). Standardizing ADOS domain scores: Separating severity of social affect and restricted and repetitive behaviors. *Journal of Autism and Developmental Disorders, 44*(10), 2400–2412. https://doi.org/10.1007/s10803-012-1719-1

Ibanez, L. V., Grantz, C. J., & Messinger, D. S. (2013). The development of referential communication and autism symptomatology in high-risk infants. *Infancy, 18*(5), 687–707. https://doi.org/10.1111/j.1532-7078.2012.00142.x

iMotions. (2016). *Facial Expression Analysis: The definitive guide*. https://imotions.com/facialexpression-guide-ebook/

Jain, V., & Learned-Miller, E. (2010). Fddb: A benchmark for face detection in unconstrained settings. *UMass Amherst Technical Report*. http://works.bepress.com/erik_learned_miller/55/

Janvier, D., Choi, Y. B., Klein, C., Lord, C., & Kim, S. H. (2021). Brief report: Examining test-retest reliability of the autism diagnostic observation schedule (ADOS-2) calibrated severity scores (CSS). *Journal of Autism and Developmental Disorders, 52*, 1388–1394. https://doi.org/10.1007/s10803-021-04952-7

Jones, W., Carr, K., & Klin, A. (2008). Absence of preferential looking to the eyes of approaching adults predicts level of social disability in 2-year-old toddlers with autism spectrum disorder. *Archives of General Psychiatry, 65*(8), 946–954. https://doi.org/10.1001/archpsyc.65.8.946

Joseph, R. M., & Tager-Flusberg, H. (1997). An investigation of attention and affect in children with autism and Down Syndrome. *Journal of Autism and Developmental Disorders, 27*(4), 385–396. https://doi.org/10.1023/A:1025853321118

Kamp-Becker, I., Albertowski, K., Becker, J., Ghahreman, M., Langmann, A., Mingebach, T., Poustka, L., Weber, L., Schmidt, H., Smidt, J., Stehr, T., Roessner, V., Kucharczyk, K., Wolff, N., & Stroth, S. (2018). Diagnostic accuracy of the ADOS and ADOS-2 in clinical practice. *European Child and Adolescent Psychiatry, 27*(9), 1193–1207. https://doi.org/10.1007/s00787-018-1143-y

Kasari, C., Sigman, M., Mundy, P., & Yirmiya, N. (1990). Affective sharing in the context of joint attention interactions of normal, autistic, and mentally retarded children. *Journal of Autism and Developmental Disorders, 20*(1), 87–100. https://doi.org/10.1007/BF02206859

Landa, R. J., Holman, K. C., & Garrett-Mayer, E. (2007). Social and communication development in toddlers with early and later diagnosis of autism spectrum disorders. *Archives of General*

Psychiatry, 64(7), 853–864. https://doi.org/10.1001/archpsyc.64.7.853

Littlewort, G., Whitehill, J., Wu, T., Fasel, I. R., Frank, M. G., Movellan, J. R., & Bartlett, M. S. (2011). The computer expression recognition toolbox (CERT). *Face and Gesture, 2011*, 298–305. https://doi.org/10.1109/FG.2011.5771414

Lord, C., Rutter, M., DiLavore, P., Risi, S., Gotham, K., & Bishop, S. (2012). *Autism diagnostic observation schedule–2nd edition (ADOS-2)* (p. 284). Western Psychological Corporation.

Ma, X., Gu, H., & Zhao, J. (2021). Atypical gaze patterns to facial feature areas in autism spectrum disorders reveal age and culture effects: A meta-analysis of eye-tracking studies. *Autism Research, 14*(12), 2625–2639. https://doi.org/10.1002/aur.2607

Macari, S., DiNicola, L., Kane-Grade, F., Prince, E., Vernetti, A., Powell, K., Fontenelle, S., IV., & Chawarska, K. (2018). Emotional expressivity in toddlers with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 57*(11), 828–836. https://doi.org/10.1016/j.jaac.2018.07.872

Manfredonia, J., Bangerter, A., Manyakov, N. V., Ness, S., Lewin, D., Skalkin, A., Boice, M., Goodwin, M. S., Dawson, G., Hendren, R., Leventhal, B., Shic, F., & Pandina, G. (2019). Automatic recognition of posed facial expression of emotion in individuals with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 49*(1), 279–293. https://doi.org/10.1007/s10803-018-3757-9

Mattson, W. I., Cohn, J. F., Mahoor, M. H., Gangi, D. N., & Messinger, D. S. (2013). Darwin's Duchenne: Eye constriction during infant joy and distress. *PLoS One, 8*(11), e80161. https://doi.org/10.1371/journal.pone.0080161

McHugh, M. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282. https://doi.org/10.11613/BM.2012.031

Messinger, D. S., Mahoor, M. H., Chow, S. M., & Cohn, J. F. (2009). Automated measurement of facial expression in infant-mother interaction: A pilot study. *Infancy, 14*(3), 285–305. https://doi.org/10.1080/15250000902839963

Moffitt, J. M., Ahn, Y. A., Custode, S., Tao, Y., Mathew, E., Parlade, M., Hale, M., Durocher, J., Alessandri, M., Perry, L. K., & Messinger, D. S. (2022). Objective measurement of vocalizations in the assessment of autism spectrum disorder symptoms in preschool age children. *Autism Research*. https://doi.org/10.1002/aur.2731

Mundy, P.C., Delgado, C.F., Block, J.J., Venezia, M., Hogan, A.E., & Seibert, J.M. (2013). Early social communication scales (ESCS). https://doi.org/10.1007/978-1-4419-1698-3_287

Nichols, C. M., Ibañez, L. V., Foss-Feig, J. H., & Stone, W. L. (2014). Social smiling and its components in high-risk infant siblings without later ASD symptomatology. *Journal of Autism and Developmental Disorders, 44*(4), 894–902. https://doi.org/10.1007/s10803-013-1944-2

Ozonoff, S., Iosif, A.-M., Baguio, F., Cook, I. C., Hill, M. M., Hutman, T., Rogers, S. J., Rozga, A., Sangha, S., & Sigman, M. (2010). A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(3), 256–266.

Panis, G., Lanitis, A., Tsapatsoulis, N., & Cootes, T. F. (2016). Overview of research on facial aging using the FG-NET ageing database. *IET Biometrics, 5*(2), 37–46. https://doi.org/10.1049/iet-bmt.2014.0053

Phillips, P. J., Jiang, F., Narvekar, A., Ayyad, J., & O' Toole, A. J. (2011). An other-race effect for face recognition algorithms. *ACM Transactions on Applied Perception, 8*(2), 1–11. https://doi.org/10.1145/1870076.1870082

Rehg, J. M. (2013). Behavior imaging and the study of autism. *Proceedings of the 15th ACM on International Conference on Multimodal Interaction, Association for Computing Machinery.* pp. 1–2. https://doi.org/10.1145/2522848.2532203

Rehg, J. M., Rozga, A., Abowd, G. D., & Goodwin, M. S. (2014). Behavioral imaging and autism. *IEEE Pervasive Computing, 13*(2), 84–87.

Ricanek, K., & Tesafaye, T. (2006). Morph: A longitudinal image database of normal adult age-progression. *7th International Conference on Automatic Face and Gesture Recognition (FGR06)*, pp. 341–345 https://doi.org/10.1109/FGR.2006.78

Rothe, R., Timofte, R., & Van Gool, L. (2018). Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision, 126*(2), 144–157. https://doi.org/10.1007/s11263-016-0940-3

Stöckli, S., Schulte-Mecklenbeck, M., Borer, S., & Samson, A. C. (2018). Facial expression analysis with AFFDEX and FACET: A validation study. *Behavior Research Methods, 50*, 1446–1460. https://doi.org/10.3758/s13428-017-0996-1

Trevisan, D. A., Bowering, M., & Birmingham, E. (2016). Alexithymia, but not autism spectrum disorder, may be related to the production of emotional facial expressions. *Molecular Autism, 7*(46). https://doi.org/10.1186/s13229-016-0108-6

Wang, Q., Campbell, D. J., Macari, S. L., Chawarska, K., & Shic, F. (2018). Operationalizing atypical gaze in toddlers with autism spectrum disorders: A cohesion-based approach. *Molecular Autism*. https://doi.org/10.1186/s13229-018-0211-y

Wetherby, A. M., & Prizant, B. M. (2003). *Communication and symbolic behavior scale (CSBS)*. Paul H. Brookes Publishing Company.

Yang, L., Ertugrul, I. O., Cohn, J. F., Hammal, Z., Jiang, D., & Sahli, H. (2019). FACS3D-Net: 3D convolution based spatiotemporal representation for action unit detection. *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 538–544. https://doi.org/10.1109/ACII.2019.8925514

Yang, S., Luo, P., Loy, C.-C., & Tang, X. (2016). Wider face: A face detection benchmark. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. https://doi.org/10.1109/CVPR.2016.596

Zander, E., Willfors, C., Berggren, S., Choque-Olsson, N., Coco, C., Elmund, A., Hedfors Moretti, Å., Holm, A., Jifält, I., Kosieradzki, R., Linder, J., Nordin, V., Olafsdottir, K., Poltrago, L., & Bölte, S. (2016). The objectivity of the autism diagnostic observation schedule (ADOS) in naturalistic clinical settings. *European Child and Adolescent Psychiatry, 25*, 769–780. https://doi.org/10.1007/s00787-015-0793-2

Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters, 23*(10), 1499–1503. https://doi.org/10.1109/LSP.2016.2603342