Received: 9 June 2023 | Accepted: 14 October 2023

DOI: 10.1002/aur.3046

RESEARCH ARTICLE

Development and validation of an Arabic language eye-tracking paradigm for the early screening and diagnosis of autism spectrum disorders in Qatar

Fouad A. Al-Shaban[1] | **Iman Ghazal**[1] | **I. Richard Thompson**[1] | **Eric W. Klingemier**[2] | **Mohammed Aldosari**[3] | **Hawraa Al-Shammari**[1] | **Fatema Al-Faraj**[1] | **Saba El-Hag**[4] | **Mohamed Tolefat**[5] | **Mogahed Ali**[5] | **Bisher Nasir**[6] | **Thomas W. Frazier**[7]

[1]Neurological Disorders Research Center, Qatar Biomedical Research Institute, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

[2]Center for Autism, Cleveland Clinic, Cleveland, Ohio, USA

[3]Neurological Institute, Cleveland Clinic, Cleveland, Ohio, USA

[4]Sidra Research, Qatar Foundation, Doha, Qatar

[5]Shafallah Center for Children with Disabilities, Doha, Qatar

[6]Department of Biology, Drexel University, Philadelphia, Pennsylvania, USA

[7]Department of Psychology, John Carroll University, Cleveland, Ohio, USA

Correspondence
Thomas W. Frazier, Department of Psychology, John Carroll University, Cleveland, OH, USA.
Email: tfrazier@jcu.edu

Funding information
Qatar Biomedical Research Institute, Hamad Bin Khalifa University

Abstract

Abnormal eye gaze is a hallmark characteristic of autism spectrum disorder (ASD). The primary aim of the present research was to develop an Arabic version of an objective measure of ASD, the "autism index" (AI), based on eye gaze tracking to social and nonsocial stimuli validated initially in the United States. The initial phase of this study included the translation of English language eye-tracking stimuli into stimuli appropriate for an Arabic-speaking culture. During the second phase, we tested it on a total of 144 children with ASD, and 96 controls. The AI had excellent internal consistency and test–retest reliability. Moreover, the AI showed good differentiation of ASD from control cases (AUC = 0.730, SE = 0.035). The AI was significantly positively correlated with SCQ total raw scores ($r = 0.46$, $p < 0.001$). ADOS-2 scores were only available in the ASD group and did not show a significant relationship with AI scores ($r = 0.10$, $p = 0.348$), likely due to the restricted range. The AI, when implemented using Arabic-translated stimuli in a Qatari sample, showed good diagnostic differentiation and a strong correlation with parent-reported ASD symptoms. Thus, the AI appears to have cross-cultural validity and may be useful as a diagnostic aide to inform clinical judgment and track ASD symptom levels as part of the evaluation process.

Lay Summary

This study aimed to create an Arabic version of a tool called the "autism index" (AI), which uses eye gaze tracking to assess autism spectrum disorder (ASD). The researchers translated the AI's eye-tracking tests into Arabic and tested it on

children with ASD ($n = 144$), non-autistic children ($n = 84$), and children with developmental delays ($n = 12$). The AI proved to be reliable and effectively distinguished children with ASD from the control groups and demonstrated cross-cultural validity. It also showed a strong correlation with parent-reported ASD symptoms. In conclusion, the Arabic version of the AI showed promise as a valuable cross-cultural objective tool for screening, diagnosing, and tracking ASD symptoms.

KEYWORDS
autism spectrum disorder, diagnosis, eye tracking, gaze, risk assessment

INTRODUCTION

Social attention is a key developmental parameter (Frazier et al., 2021) that likely influences a range of developmental processes, including social cognition (Frischen et al., 2007). Numerous studies have identified gaze differences between individuals with autism spectrum disorder (ASD) and controls across a wide range of ages and stimulus paradigms (Chita-Tegmark, 2016a, 2016b; Klin & Jones, 2008; Papagiannopoulou et al., 2014). Reductions in eye gaze to social stimuli and increases in attention to non-social stimuli are a replicable and relatively stable feature of ASD (Frazier et al., 2017), consistent with original conceptualizations of autism (Kanner, 1943), and is considered a red flag in all diagnostic instruments (Lord, Luyster, et al., 2012; Lord, Rutter, et al., 2012; Rapin, 1997). More than a decade of research into eye gaze abnormalities has confirmed social attention deficits as a main feature of ASD, even in very young children (Jones et al., 2008; Jones & Klin, 2013; Klin et al., 2009; Lord et al., 1994; Papagiannopoulou et al., 2014; Rice et al., 2012; Risi et al., 2006). Across studies, various stimulus patterns have induced social attention abnormalities; ranging from diminished fixation to others' eyes and social scenes as early as 6 months of age, to gaze abnormalities during dyadic or joint attention display in preschoolers and older children, to abnormal gaze toward dynamic social stimuli in older high-functioning individuals (Chawarska et al., 2013; Lord et al., 1994; Magrelli et al., 2013; Sasson & Elison, 2012; Vivanti et al., 2013).

Presently, ASD is identified in the context of a clinical evaluation that is typically based on some combination of parent-report, parent-interview, and clinical observation tools. These methods are heavily influenced by subjective perceptions and require substantial training and ongoing inter-rater reliability checks (Bishop & Seltzer, 2012; Pierce et al., 2015). Several diagnostic tools have been validated as a "gold-standard" to enable clinicians to differentiate ASD from other neurodevelopmental disorders. Specifically, the Autism Diagnostic Observation Schedule-2nd edition (ADOS-2) (Falkmer et al., 2013; Lord, Luyster, et al., 2012; Lord, Rutter, et al., 2012) is widely considered once of the most effective evaluation tools for clinical and research contexts. Nonetheless, to date, all measures proven effective in diagnosing ASD are subjective in nature and even the ADOS-2 is limited by the need for extensive training and ongoing reliability checks to maintain accuracy and fidelity. None of the present diagnostic procedures include an objective marker that provides immediate interval-scale measurements and highly reliable scores across the full range of behaviors in individuals affected with ASD. Thus, the development of objective measures of ASD has the potential to greatly enhance clinical evaluation, supplementing existing subjective assessment methods.

Recently, research has been geared toward investigations of objective markers for ASD diagnosis (Ansel et al., 2019; Frazier et al., 2016). Technological advancements such as eye gaze tracking have shown promise as a method for producing objective markers for ASD (Frazier et al., 2016; Pierce et al., 2011). Two recent studies provided some support for the potential distinctive value of eye gaze tracking. In these studies, individual social stimuli had moderate but potentially informative discriminative value (Areas Under the Curve [AUC] = 0.71–0.72) in distinguishing individuals with ASD and other developmental delays from healthy control individuals (Chevallier et al., 2015; Guillon et al., 2014).

The development of a quantitative, interval-scale measure of autism symptoms, including measures of the core symptom domains, would represent a major step forward in the technology used to capture autism symptom levels and identify cases that are at risk for ASD diagnosis. Most recently, researchers showed that eye gaze to social and non-social stimuli can be aggregated into an autism risk index with high validity for the identification of ASD (Frazier et al., 2018). This work has also shown that social attention processes are a distinct behavioral dimension that shows cross-cultural consistency and that remains stable throughout various age groups, with females exhibiting a slightly stronger preference, on average, for social attention (Frazier et al., 2021). Thus, it remains possible that recent advances in the development of social attention-based metrics for assessing the likelihood of ASD might be transportable across cultural contexts after appropriate adaptation.

Following this logic, the primary purpose of the present research was to build on prior research in a US population (Frazier et al., 2018) to create a highly similar Arabic-language stimulus battery for collecting and

scoring the autism risk index. The stimuli needed to be appropriate for the Middle Eastern culture, well accepted by parents, and maintain the child's attention during data collection. The secondary purpose was to compute the autism risk (AI) by aggregating gaze metrics across the new Arabic-language stimuli in an identical fashion to the prior US work, and validate the Arabic AI within the Qatari population through recruitment and collection of eye tracking data of a large sample of ASD affected individuals and non-autistic (NA) children and children with other developmental delays (DD) between the ages of 3 and 15 years. All children referred to this study were either already diagnosed or suspected to be on the autism spectrum. We hypothesized that the AI would show substantial diagnostic validity—the ability to differentiate ASD and control cases—with Area Under the Curve (AUC) >0.70. Second, we hypothesized that the clinical autism symptom measurements would be significantly associated with the social and non-social attention indices comprising the AI.

METHODS

Participants

All research participants were between the age of 3 and 15 years. The study sample consisted of two groups: ASD ($n = 144$) and NA ($n = 84$) and DD controls ($n = 12$). The ASD group was recruited from local special needs clinics and centers in which a prior ASD diagnosis was done. We also recruited cases suspected to be on the autism spectrum, which we confirmed using the tools in our research protocol described in the diagnosis section. Meanwhile, the control group participants were either siblings of ASD participants, recruited through primary care clinics, or through research contacts. The total control sample consisted of 39% siblings of participants with ASD, and 61% were not related to any case of ASD. Given the small sample of DD, all control group participants were grouped into a single comparison sample. The exclusion criteria were for children with certain disabilities that will pose difficulties in performing the eye tracking testing (i.e., severe physical disabilities, visual and auditory impairments, etc.).

Diagnosis

The diagnostic assessment was conducted for all ASD participants who never received a diagnosis using the ADOS-2, and for all participants suspected to be on the autism spectrum. A consensus diagnosis was based on a parent interview, psychosocial, developmental, and clinical history conducted by our research team whose members are all licensed administrators of the ADOS-2. After conducting the assessments, the team met to confirm the presence/absence of ASD using DSM-5 criteria (American Psychiatric Association, 2013). Eligibility for participation in the DD group required any other neurodevelopmental or neuropsychiatric disorder diagnosis other than ASD. Eligibility in the NA group required no past or current developmental or psychiatric difficulties. All group participants were screened using the SCQ and seen by our neurodevelopmental physician with years of experience in ASD phenotyping and diagnosis, to ascertain or rule out ASD in each sample.

Clinical assessments

The clinical assessments in our research protocol included the ADOS-2, Arabic, and English versions of the SCQ (Aldosari et al., 2019; Rutter et al., 2003; Chandler et al., 2007), and DSM-5 criteria. As the gold-standard clinical observation measure used to assess autism symptoms severity, ADOS-2 was used. For the present study, the ADOS-2 total, social affect sub-scale, and restricted/repetitive behavior sub-scale raw scores were converted to calibrated severity scores based on the ADOS-2 module used and the comparison scores for each participant. During the parent interview, the research team completed the Arabic or English version of the SCQ, and a clinical and developmental history form. As part of clinical history, results from previous assessments such as IQ, language, and other developmental concerns, were requested from the families of consented participants and retrieved from clinical history records of record-review participants.

Eye tracking stimuli creation

The creation of the Arabic version of the eye-tracking stimuli was completed in three stages. The first stage was obtaining the English version (Cleveland Clinic) of the eye-tracking stimuli. The second stage included the translation and back translation (from English to Arabic) of the content by the QBRI team. Finally, a local Qatari production company was hired for the creation of the Arabic version of the eye-tracking stimuli. The research team replicated the setup and setting of the English version at Hamad Bin Khalifa University (HBKU). The team then reviewed the content, compared them to match the Cleveland Clinic (CC) stimuli as much as possible, and recommended adjustments as needed. ROIs were identified in the Qatari Stimuli to match the English Stimuli (Figure 1).

Eye tracking data acquisition and processing

Eye tracking data was collected in a quiet room, using the SMI RED250 remote eye tracker system attached to a 19-inch LCD stimulus presentation monitor. Binocular gaze, 3D eye position, pupil, and timestamp data were collected at a sampling rate of 250 Hz. Gaze capture was

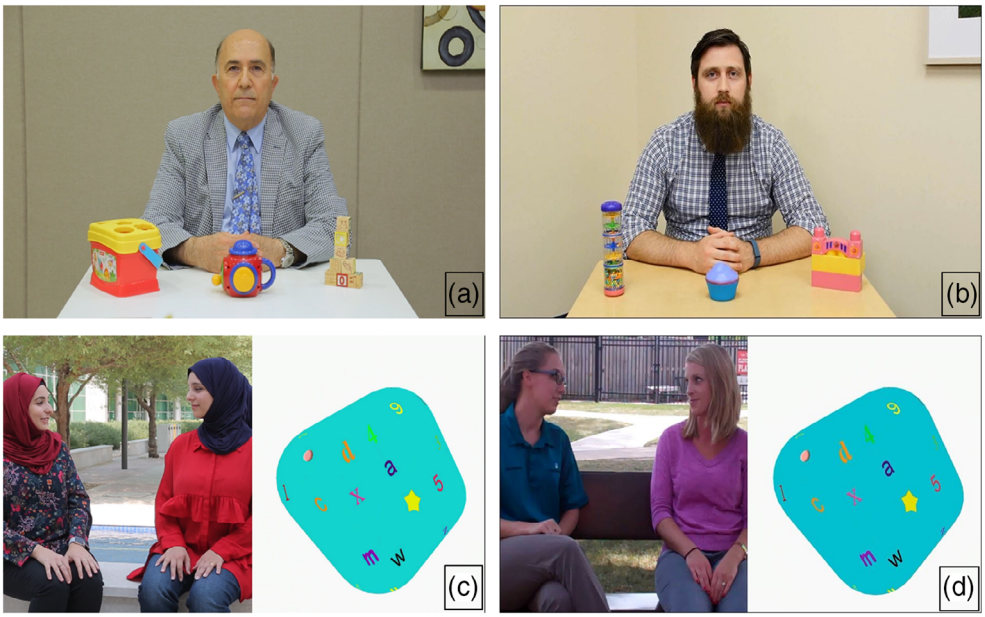

FIGURE 1 Sample Qatar versus US stimuli. (a and c) Qatar version, (b and d) Cleveland Clinic version.

automatically calibrated using 2-, 5-, or 9-point calibrations (starting with 5 or 9-point depending on the functional level of the child) and provided position accuracy to 0.4°. Dwell proportion (percentage on target region relative to total time on screen) was the primary measurement of interest, although additional measures were also collected, including visits to regions of interest, saccade average length, saccade peak velocity, and blink frequency.

For the eye-tracking assessment, we followed recommendations from Sasson and Elison (2012), in which the child was seated alone or in his/her parent's lap approximately 60–65 cm from the LCD display and viewed the stimuli subtending a visual angle of approximately 18.8 degrees. Standard room lighting was used, and the room was sparse, with visual barriers to reduce distraction. After calibration, children were told, "You will see some videos, pay attention, but look however you want." Stimuli was presented using the SMI Experiment Center. Gaze data was captured during viewing of a 10 min battery consisting of initial and recurring calibration and multiple stimuli from each of the following paradigms: dynamic individual faces, static side-by-side faces, joint attention bids, gaze following, reciprocal interactions, dynamic social versus dynamic geometric images, and passive viewing of social/object arrays. The visual paradigm ended with a short gaze accuracy validation step. The researchers used the same measurement techniques in the Arabic-language study as in the English-language study but calculated z-scores using the mean and standard deviation of the original study. The researchers then combined indicators that showed less attention to social cues in children with ASD (which were given a negative value) with indicators that showed more attention to non-social cues in ASD-affected individuals. Test–retest comparisons were done for 28 participants from all groups, the time between administrations of the test and retest averaged 8 months (Table 1).

Statistical analyses

The study identified outliers and high-leverage cases using univariate and bivariate distributions. Analyses were performed with and without these cases, but there were no significant differences, so all available data was included. Descriptive statistics were used to characterize the sample and comparisons were made between baseline and retest participants using appropriate statistical tests. Receiver operating characteristic (ROC) curve analysis assessed the validity of the AI, SCQ, and ADOS-2 severity scores. Concurrent validity for autism severity was evaluated using Spearman's rank-order correlations between these measures. However, the correlation between ADOS-2 scores and other measures was anticipated to be lower due to the restricted range of these scores. The internal consistency reliability of the AI and its attention indicators was estimated using Cronbach's α, while test–retest reliability was evaluated using Pearson's r. Area under the ROC curve was used to quantify validity. To establish diagnostic validity, a 95% confidence interval of the ROC curve should be larger than 0.80.

To validate the Arabic language AI, we computed Spearman rank-order correlation coefficients examining the relationships between the social vs. non-social attention indices, the ADOS-2 total, severity scores, and SCQ raw scores. ADOS-2 calibrated severity scores were computed using existing norms. A type 1 error rate of 0.05 was used for each analysis. In addition, to avoid inflation of Type 1 error for these correlations, we used a Benjamini-Hochberg false discovery rate correction, and only correlations exceeding $r = 0.40$ were considered clinically meaningful. The validation sample was intentionally over-powered for detection of a significant area under the ROC curve (AUC of >0.70). Descriptive statistics were used to characterize the sample and

TABLE 1 Test–retest samples, details of period between tests.

Test	Retest	Number of days	Number of months
5/14/2019	10/14/2019	153	5
3/24/2019	12/9/2019	260	8.15
3/25/2019	12/9/2019	259	8.14
3/25/2019	12/5/2019	255	8.10
3/25/2019	12/10/2019	260	8.15
3/25/2019	12/11/2019	261	8.16
3/27/2019	12/8/2019	256	8.11
4/4/2019	12/5/2019	245	8.1
3/28/2019	12/10/2019	257	8.12
3/28/2019	12/10/2019	257	8.12
3/28/2019	12/11/2019	258	8.13
3/28/2019	12/9/2019	256	8.11
4/1/2019	12/9/2019	252	8.8
4/1/2019	12/8/2019	251	8.7
4/2/2019	12/8/2019	250	8.6
4/3/2019	12/11/2019	252	8.8
4/3/2019	12/12/2019	253	8.9
4/3/2019	12/11/2019	252	8.8
4/3/2019	12/9/2019	250	8.6
4/3/2019	12/5/2019	246	8.2
8/4/2019	1/13/2020	162	5.9
8/4/2019	1/13/2020	162	5.9
8/4/2019	1/13/2020	162	5.9
8/4/2019	1/13/2020	162	5.9
2/25/2019	1/12/2020	321	10.18
2/25/2019	1/12/2020	321	10.18
2/25/2019	1/12/2020	321	10.18
3/7/2019	1/12/2020	311	10.5
3/7/2019	1/12/2020	311	10.5
4/13/2019	1/12/2020	274	8.30
4/13/2019	1/12/2020	274	8.30
4/17/2019	1/13/2020	271	8.27
7/7/2019	1/12/2020	189	6.5

comparisons were made between baseline and retest participants using Chi-square (categorical variables) or *t*-tests (continuous variables) and Cohen's d was presented to evaluate the magnitude of baseline and retest sample differences.

RESULTS

Sample description

Baseline sample accounting has been described previously (Frazier et al., 2021). The final baseline sample included 240 participants (144 ASD, 84 NA, and 12 DD) and the retest sample included 28 participants (16 ASD, 9 NA, and 3 DD) (Table 2). The distribution of diagnoses, age, sex, ASD sex ratio, SCQ total raw scores, tracking ratio, and number of valid stimuli were consistent between the baseline sample and the retest sample. However, the retest sample had significantly higher proportions of participants with language/communication disorders and global developmental delay/intellectual disability. Overall, both the baseline and retest samples consisted of individuals with wide ranges of ages, neuropsychiatric diagnoses, and autism symptoms, making the present sample a relatively challenging case for diagnostic differentiation.

Autism index score range and reliability

Autism index scores were fairly normally distributed in the ASD (skew = 0.21, kurtosis = 0.08) and control (skew = 0.34, kurtosis = −0.40) participants showed a wide range of scores on the autism index ($z = -2.28$ to $+4.26$) with a pronounced shift upward in the ASD group (Cohen's $d = 0.86$) (Figures 1 and 2, and Table 2 and 3). When considered separately, the social and non-social attention indicators had very good to excellent internal consistency and test–retest reliability. Internal consistency and test–retest reliability were excellent (Table 4 and 5).

Diagnostic and concurrent validity

Autism index scores showed good differentiation of ASD from control cases (AUC = 0.730, SE = 0.035) (Figure 3 and Table 6). Evaluating subsets of cases with more stringent validity (Tracking ratio ≥0.80, number of valid stimuli ≥35) did not improve diagnostic accuracy (AUC = 0.689, SE = 0.050) likely due to the loss of more significantly affected ASD cases. In the full sample, autism index scores were significantly positively correlated with SCQ total raw scores ($r = 0.46$, $p < 0.001$) (Figure 4). ADOS-2 scores were only available in the ASD group and these scores did not show a significant relationship with autism index scores ($r = 0.10$, $p = 0.348$), likely due to the somewhat restricted range of scores in this group (82.6% of scores fell between 3 and 8) (Tables 4–6).

Exploring an alternative algorithm – machine learning

The observed autism index diagnostic accuracy is impressive given that the original index selected a subset of social and non-social indicators showing significant but modest validity for ASD diagnosis and conducted linear averaging across these indicators. For the present study, this average was rigidly applied to examine strict replicability and extension to a Qatar population. Thus, the observed diagnostic accuracy is likely conservative, as one would expect that a more optimal set of indicators

TABLE 2 Participant characteristics in the full baseline sample and the retest sub-sample.

	Baseline	Retest		
	M (SD)	*M* (SD)	X^2/t (*p*)	Cohen's *d*
Total *N*	240	28	2.15 (0.342)	0.19
Non-autistic (*n*, %)	84 (35.0%)	9 (32.1%)		
Developmental delay (*n*, %)	12 (5.0%)	3 (10.7%)		
Autism spectrum disorder (*n*, %)	144 (60.0%)	16 (57.2%)		
Age (range)	7.7 (3.6, 1.4–16)	7.7 (3.1, 1.4–13)	−0.80 (0.935)	0.10
Female (*n*, %)	75 (31.3%)	8 (28.6%)	0.13 (0.721)	0.05
ASD sex ratio (female: male)	1: 3.7	1:4.3	0.10 (0.751)	0.04
Other diagnoses (*n*, %)			19.6 (0.001)	0.60
Language or communication disorder	51 (21.2%)	16 (57%)		−4.18
GDD/ID	19 (7.9%)	8 (28.6%)		−3.45
Anxiety disorder	28 (11.7)	1 (3.6%)		1.30
Attention-deficit/hyperactivity disorder	86 (35.8%)	6 (21.4%)		1.52
Other	19 (7.9%)	1 (3.6%)		0.82
SCQ total raw score				
Non-autistic	1.4 (2.2)	0.8 (0.4)	0.92 (0.363)	0.12
Developmental delay	4.0 (4.9)	-	-	
Autism spectrum disorder	16.4 (7.2)	20.1 (5.1)	−1.97 (0.051)	0.26
ADOS-2 total severity (Autism cases only)	6.1 (2.1)	-	-	
Overall tracking ratio (%)	79.1% (14.3%)	77.6% (15.2%)	0.64 (0.526)	0.08
Number of valid stimuli (out of 44)	35.3 (7.9)	34.3 (8.1)	0.75 (0.457)	0.10

Note: SCQ = Social communication total raw score Chi-square statistics were converted to Cramer's V. As an effect size metric, Cramer's V is roughly equivalent to r and, therefore, to provide a common metric, Cramer's V was converted to Cohen's d via r. Only one ADOS-2 total severity score was available in the retest sample and no SCQ total raw scores were available in the developmental delay group for the retest sample.
Abbreviation: DD, developmental delay.

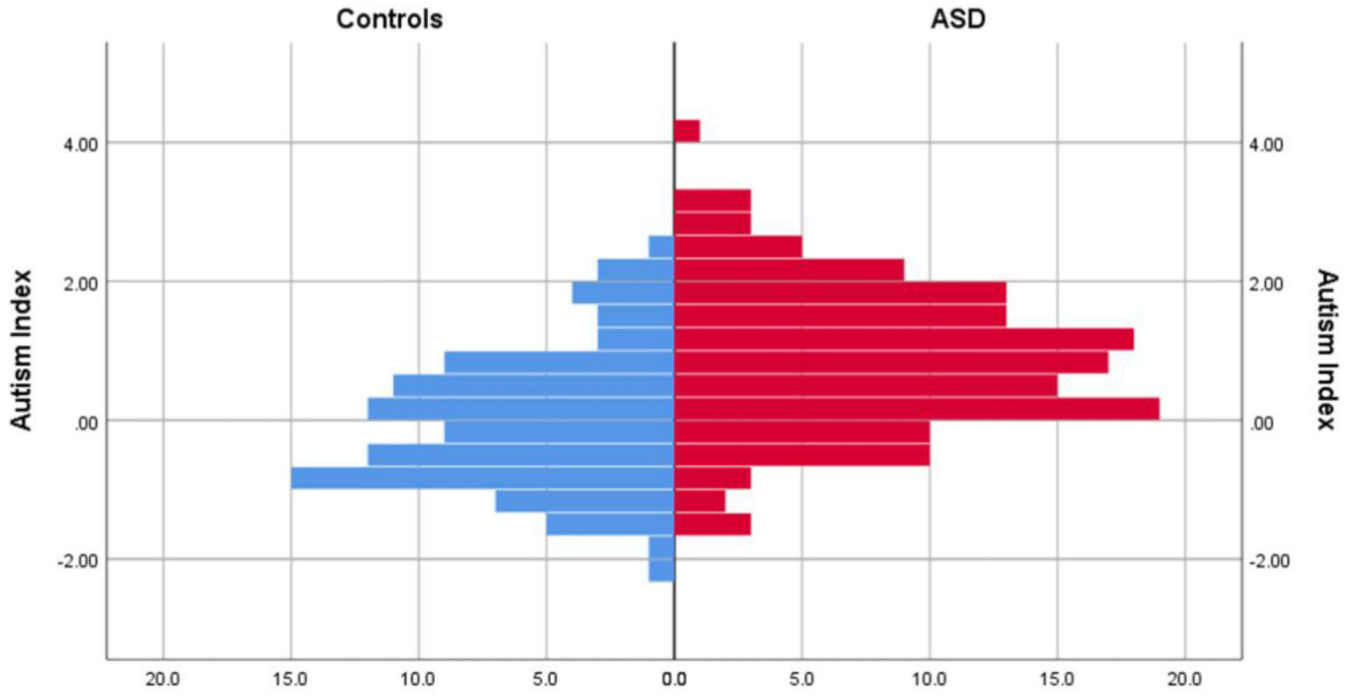

FIGURE 2 Side-by-side histograms of autism index scores for ASD-diagnosed and non-autistic control participants.

TABLE 3 Sample demographics and eye tracking quality for valid and invalid participants.

	Valid	Invalid	
	M (SD)	*M* (SD)	X^2, *t* (*p*)
N	136	19	
Age	7.6 (3.4)	6.5 (4.0)	1.34 (0.182)
Male (*N*, %)	92 (67.6%)	8 (42.1%)	4.75 (0.029)
SCQ	9.9 (9.2)	18.9 (7.8)	2.81 (0.006)
ADOS raw	16.2 (5.1)	19.8 (3.6)	2.03 (0.046)
ADOS total calibrated severity score	6.1 (1.9)	6.7 (1.4)	0.89 (0.378)
ASD (*n*, %)	85 (62.5%)	14 (73.7%)	0.90 (0.342)
Number of valid stimuli	31.5 (7.2)	10.5 (2.9)	12.53 (<0.001)
Tracking ratio (%)	80.1% (13.2%)	48.0% (6.3%)	10.38 (<0.001)

Note: Invalid cases were individuals with fewer than 15 valid stimuli out of a possible 40. All participants (valid and invalid) had tracking ratios above 40%. The ADOS-2 was only administered to ASD-affected individuals and a small number of NA controls where concern of ASD was identified but ruled out. SCQ scores were only available for 107 of 155 participants.

TABLE 4 Internal consistency and test–retest reliability coefficients for social and non-social attention indicators and the autism index.

	Internal consistency α	Test–retest *r*
Social attention ($k = 177$)	0.90	0.73
Non-social attention ($k = 195$)	0.80	0.44
Autism index ($k = 372$)	0.91	0.73

Note: k = number of gaze indicators.

TABLE 5 Summary statistics of the test–retest samples.

Mean (average)	8.1921875
Median	8.18
Range	5.5
Geometric mean	8.069696
Standard deviation	1.3561866
Variance	1.83924209
Sample standard deviation	1.37788696
Sample variance	1.89857248

and algorithms might be obtained. To explore this possibility, the available indicators used to compute the autism index were input to support vector machine and Random Forest analyses to predict ASD diagnosis. Results indicated a potential improvement in validity, using the same indicator set, by applying a support vector machine (radial basis kernel, 230 support vectors, cost = 1, gamma = 0.003, epsilon = 0.1, 10-fold cross-validation) or a random forest algorithm (500 trees, variables tried at each split = 123, leave-one-out cross-validation) (SVM – $R^2 = 0.26$, AUC = 0.799; RF – $R^2 = 0.21$, AUC = 0.768). However, these results require replication in a new sample and are still based on using the English version study's Means and SDs to standardize the indicators. Thus, even these values may be conservative relative to what is possible by computing a fully Qatari-specific algorithm (Figure 5).

DISCUSSION

The present study aimed to validate the Arabic version of the eye-tracking AI and assess its diagnostic utility in a sample of children with and without ASD. Results indicated that the Arabic version of the AI showed good diagnostic differentiation, which supports its utility for screening and as a diagnostic aide to inform clinical judgment. These findings are consistent with previous studies that have reported the efficacy of the AI or similar algorithms in identifying children at risk for ASD (Frazier et al., 2016; Zwaigenbaum & Penner, 2018). The current study also found a strong correlation between the Arabic version of the AI and the SCQ, which suggests its ability to discriminate between symptom levels in children with ASD (Frazier et al., 2018). These findings are in line with previous studies that have demonstrated the validity of the SCQ as a screening tool for ASD (Aldosari et al., 2019; Berument et al., 1999; Corsello et al., 2007).

Moreover, the high internal consistency and test–retest reliability of the Arabic AI in the current study supports its potential clinical value as an assessment tool for ASD. Prior studies on gaze-based measures have suggested moderate test–retest stability (Farzin et al., 2011). However, the current study found strong test–retest reliability for the social attention indicators and moderate reliability for the non-social indicators, which suggest that the Arabic AI is measuring a stable trait consistent with the diagnosis of ASD. These findings are consistent

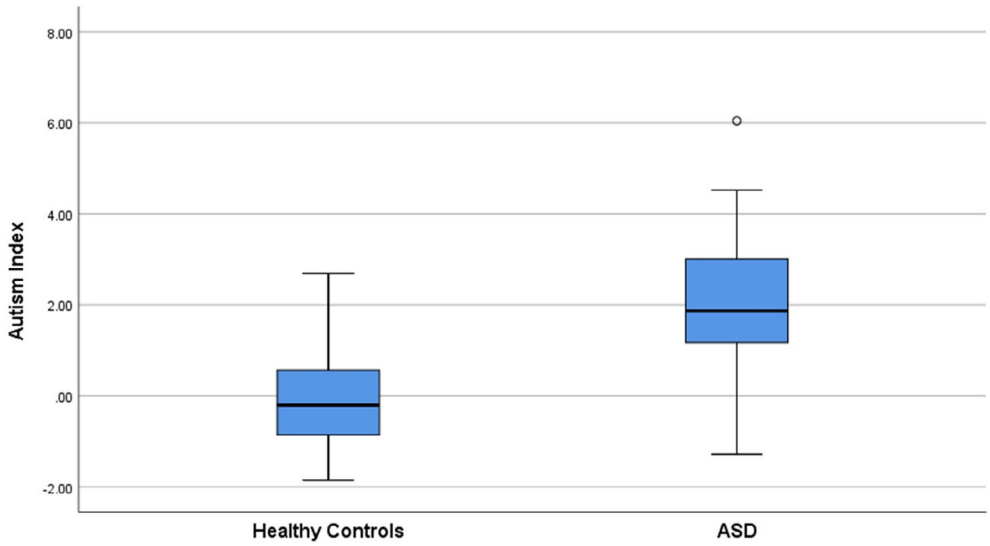

FIGURE 3 Boxplot (±95% CI) of autism index scores for ASD diagnosed and non-autistic control participants.

TABLE 6 Sample demographics and eye tracking quality for ASD diagnosed and non-autistic control participants.

	ASD	Non-autistic controls	
	M (SD)	*M* (SD)	X^2, *t* (*p*)
N	85	51	
Age	8.5 (3.1)	6.1 (3.3)	4.32 (<0.001)
Male (*N*, %)	70 (82.4%)	22 (43.1%)	22.40 (<0.001)
SCQ	15.9 (7.7)	1.7 (2.3)	11.41 (<0.001)
ADOS raw	16.2 (5.1)		
ADOS total calibrated severity score	6.1 (1.9)		
Number of valid stimuli	29.0 (7.0)	35.7 (5.5)	5.82 (<0.001)
Tracking ratio (%)	75.1% (12.5%)	88.4% (10.0%)	6.47 (<0.001)

Note: Invalid cases were individuals with fewer than 15 valid stimuli out of a possible 40. All participants (valid and invalid) had tracking ratios above 40%. The ADOS was only administered to ASD-affected individuals and a small number of non-autistic controls where concern of ASD was identified but ruled out. SCQ scores were only available for 107 of 155 participants.

with previous studies that have reported the stability of gaze-based measures in children with ASD (Chawarska et al., 2013). Nonetheless, it may be useful to explore additional non-social indicators or other ways to enhance reliability of the assessment of non-social attention in future studies.

Objective measures are necessary to grade the severity of ASD symptoms and monitor changes over time, as subjective measures may be inconsistent and often require substantial training and ongoing reliability checks to maintain accuracy (Bishop & Seltzer, 2012; Pierce et al., 2015). The results of the current study suggest that AI may be a useful tool in conjunction with other commonly used clinical measures for identifying ASD. Interestingly, supporting the contention of incremental validity with other clinical measures, missed cases (ASD diagnosis but low AI index scores) also had a different pattern of ADOS-2 scores compared to correctly identified cases. However, more research is needed to determine the precise level of stand-alone and incremental validity of the AI for categorical ASD diagnosis.

LIMITATIONS

The current study describes the development and initial validation of an Arabic version of the gaze-based AI as a screening and diagnostic tool for ASD in the Arabic-speaking population. However, the study was limited by its small sample size overall, especially for the DD group. Therefore, it is essential to conduct further research with a larger sample size, particularly for establishing the validity of the AI in distinguishing between ASD and non-ASD developmentally delayed cases. It is important to note that this limitation is somewhat offset by the conservative nature of applying the original English-language AI algorithm to a Qatari sample without developing local norms or an Arabic-specific algorithm.

Nonetheless, future work is required to replicate these findings with larger, more representative samples that include a wide range of ASD and non-ASD developmental disability cases. Although generalizability was obtained between US-English and Qatar-Arabic stimuli, additional generalizability studies should be considered to other populations and cultural contexts. Moreover, the absence of IQ test results for some of the participants was a limitation and should be addressed in future similar studies. Finally, despite the majority of the control group participants in our sample not being siblings of ASD nor considered to be from the at-risk category, it is still worth noting that including siblings of the ASD samples was a limitation that should be avoided in future research.

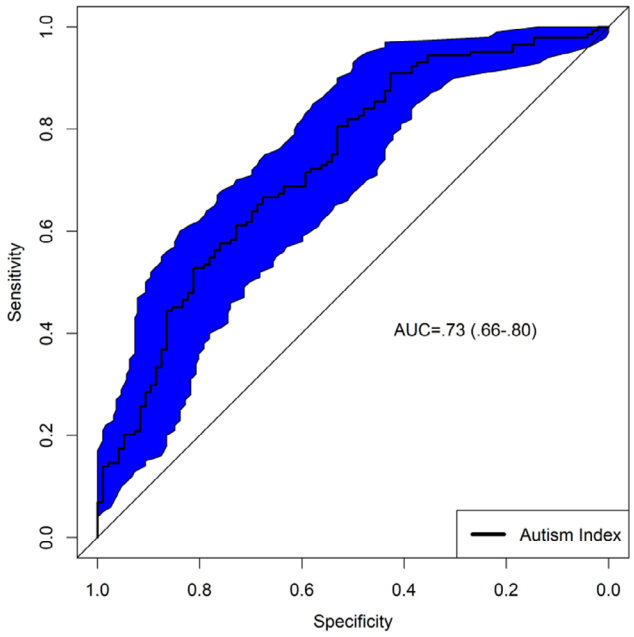

FIGURE 4 Receiver operating characteristics curve for the autism index predicting clinical ASD diagnosis ($n = 240$).

CONCLUSION

While social attention has been found to be consistent across cultures (Frazier et al., 2021), and AI has shown promise in identifying ASD in English-speaking populations (Frazier et al., 2018), it is crucial to determine whether the tool can be successfully implemented in other contexts through further cross-cultural validation studies. This will expand the diagnostic tools available for ASD screening and improve access to accurate diagnosis for individuals from diverse backgrounds.

Although the current study supports the validity of the AI as a diagnostic tool for ASD in the context of Qatar, it is necessary to evaluate its clinical implementation to determine its diagnostic accuracy as part of the typical workflow. This involves testing the AI in clinical settings, comparing its accuracy to existing diagnostic tools such as the ADOS-2, and examining the feasibility of implementing the AI in clinical practice. Guidelines for the use of the AI in clinical settings are also needed to ensure its proper utilization and interpretation.

Future research should involve testing the AI in a larger and more diverse sample set and examining its predictive value in comparison to other established diagnostic tools.

AUTHOR CONTRIBUTIONS

Dr. Alshaban and Dr. Thomas W. Frazier had full access to all the data in the study and take responsibility for the

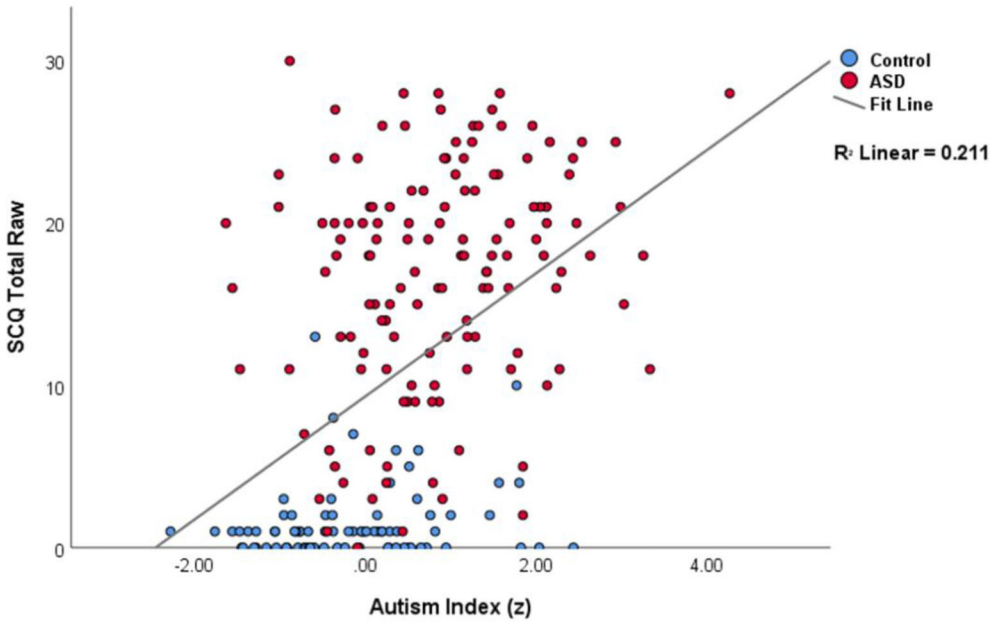

FIGURE 5 Bivariate correlation between the autism index and SCQ total raw scores.

integrity of the data and the accuracy of the data analysis. Concept and design: Alshaban and Frazier. Acquisition, analysis, or interpretation of data: Alshaban, Frazier, Klingemier, and Thompson. Drafting of the manuscript: Alshaban, Frazier, Ghazal, & Thompson. Critical revision of the manuscript for important intellectual content: All Authors. Statistical analysis: Frazier, Klingemier, and Thompson. Administrative, technical, or material support: Alshaban. Supervision: Alshaban.

ACKNOWLEDGMENTS

This study was funded by Qatar Biomedical Research Institute. We are immensely grateful to the families and their children for their time and participation in the research. We thank Qatar ASD Family Association, Shafallah Center for Persons with Disabilities, and all other private centers that participated in the study.

CONFLICT OF INTEREST STATEMENT

Dr. Frazier has received funding or research support from, acted as a consultant to, received travel support from, and/or received a speaker's honorarium from the PTEN Research Foundation, SYNGAP Research Fund, Malan Syndrome Foundation, ADNP Kids Research Foundation, Quadrant Biosciences, Autism Speaks, Impel NeuroPharma, F. Hoffmann-La Roche AG Pharmaceuticals, the Cole Family Research Fund, Simons Foundation, Ingalls Foundation, Forest Laboratories, Ecoeos, IntegraGen, Kugona LLC, Shire Development, Bristol-Myers Squibb, Roche Pharma, MaraBio, Scioto Biosciences, National Institutes of Health, and the Brain and Behavior Research Foundation, is employed by and has equity options in Quadrant Biosciences/Autism Analytica, has equity options in MaraBio and Springtide, and has an investor stake in Autism EYES LLC and iSCAN-R.

DATA AVAILABILITY STATEMENT

Research data are not shared.

ETHICS STATEMENT

Qatar Biomedical Research Institute Institutional Review Board approved the research.

ORCID

Fouad A. Al-Shaban https://orcid.org/0000-0002-0529-9721
Iman Ghazal https://orcid.org/0000-0002-6263-766X
I. Richard Thompson https://orcid.org/0000-0002-9554-550X
Thomas W. Frazier https://orcid.org/0000-0002-6951-2667

REFERENCES

Aldosari, M., Fombonne, E., Aldhalaan, H., Ouda, M., Elhag, S., Alshammari, H., Ghazal, I., Alsaleh, A., Alqadoumi, T., Thomson, R., Al Khasawneh, M., Tolefat, M., & Alshaban, F. (2019). Validation of the Arabic version of the Social Communication Questionnaire. *Autism*, *23*(5), 1255–1263. https://doi.org/10.1177/1362361318816065

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Association https://doi.org/10.1176/appi.books.9780890425596

Ansel, A., Posen, Y., Ellis, R., Deutsch, L., Zisman, P. D., & Gesundheit, B. (2019). Biomarkers for Autism Spectrum Disorders (ASD): A meta-analysis. *Rambam Maimonides Medical Journal*, *10*(4), e0021. https://doi.org/10.5041/RMMJ.10375

Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry*, *175*, 444–451. https://doi.org/10.1192/bjp.175.5.444

Bishop, S. L., & Seltzer, M. M. (2012). Self-reported autism symptoms in adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*(11), 2354–2363. https://doi.org/10.1007/s10803-012-1483-2

Chandler, S., Charman, T., Baird, G., Simonoff, E., Loucas, T., Meldrum, D., & Pickles, A. (2007). Validation of the Social Communication Questionnaire in a population cohort of children with autism spectrum disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *46*(10), 1324–1332. https://doi.org/10.1097/chi.0b013e31812f7d8d

Chawarska, K., Macari, S., & Shic, F. (2013). Decreased spontaneous attention to social scenes in 6-month-old infants later diagnosed with autism spectrum disorders. *Biological Psychiatry*, *74*(3), 195–203. https://doi.org/10.1016/j.biopsych.2012.11.022

Chevallier, C., Parish-Morris, J., McVey, A., Rump, K. M., Sasson, N. J., Herrington, J. D., & Schultz, R. T. (2015). Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Research*, *8*(5), 620–628. https://doi.org/10.1002/aur.1479

Chita-Tegmark, M. (2016a). Attention allocation in ASD: A review and meta-analysis of eye-tracking studies. *Review Journal of Autism and Developmental Disorders*, *3*, 209–223. https://doi.org/10.1007/s40489-016-0070-6

Chita-Tegmark, M. (2016b). Social attention in ASD: A review and meta-analysis of eye-tracking studies. *Research in Developmental Disabilities*, *48*, 79–93. https://doi.org/10.1016/j.ridd.2015.11.007

Corsello, C., Akshoomoff, N., & Stahmer, A. C. (2007). Between a ROC and a hard place: Decision making and making decisions about using the SCQ. *Journal of Child Psychology and Psychiatry*, *48*, 932–940. https://doi.org/10.1111/j.1469-7610.2007.01787.x

Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: A systematic literature review. *European Child & Adolescent Psychiatry*, *22*, 329–340. https://doi.org/10.1007/s00787-013-0375-0

Farzin, F., Scaggs, F., Hervey, C., Berry-Kravis, E., & Hessl, D. (2011). Reliability of eye tracking and pupillometry measures in individuals with fragile X syndrome. *Journal of Autism and Developmental Disorders*, *41*, 1515–1522. https://doi.org/10.1007/s10803-011-1176-2

Frazier, T. W., Klingemier, E. W., Beukemann, M., Speer, L., Markowitz, L., Parikh, S., Wexberg, S., Giuliano, K., Schulte, E., Delahunty, C., Ahuja, V., Eng, C., Manos, M. J., Hardan, A. Y., Youngstrom, E. A., & Strauss, M. S. (2016). Development of an objective autism risk index using remote eye tracking. *Journal of the American Academy of Child and Adolescent Psychiatry*, *55*(4), 301–309. https://doi.org/10.1016/j.jaac.2016.01.011

Frazier, T. W., Klingemier, E. W., Parikh, S., Speer, L., Strauss, M. S., Eng, C., Hardan, A. Y., & Youngstrom, E. A. (2018). Development and validation of objective and quantitative eye tracking-based measures of autism risk and symptom levels. *Journal of the American Academy of Child and Adolescent Psychiatry*, *57*(11), 858–866. https://doi.org/10.1016/j.jaac.2018.06.023

Frazier, T. W., Strauss, M., Klingemier, E. W., Zetzer, E. E., Hardan, A. Y., Eng, C., & Youngstrom, E. A. (2017). A meta-

analysis of gaze differences to social and nonsocial information between individuals with and without autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, *56*(7), 546–555. https://doi.org/10.1016/j.jaac.2017.05.005

Frazier, T. W., Uljarevic, M., Ghazal, I., Klingemier, E. W., Langfus, J., Youngstrom, E. A., Aldosari, M., Al-Shammari, H., El-Hag, S., Tolefat, M., Ali, M., & Al-Shaban, F. A. (2021). Social attention as a cross-cultural transdiagnostic neurodevelopmental risk marker. *Autism Research*, *14*(9), 1873–1885. https://doi.org/10.1002/aur.2532

Frischen, A., Bayliss, A., Tipper, A. P., & Stevens, P. (2007). Gaze cueing of attention: Visual attention, social cognition, and individual differences. *Psychological Bulletin*, *133*(4), 694–724. https://doi.org/10.1037/0033-2909.133.4.694

Guillon, Q., Hadjikhani, N., Baduel, S., & Rogé, B. (2014). Visual social attention in autism spectrum disorder: Insights from eye tracking studies. *Neuroscience & Biobehavioral Reviews*, *42*, 279–297. https://doi.org/10.1016/j.neubiorev.2014.03.013

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

Jones, W., & Klin, A. (2013). Attention to eyes is present but in decline in 2–6-month-old infants later diagnosed with autism. *Nature*, *504*(7480), 427–431. https://doi.org/10.1038/nature12715

Kanner, L. (1943). Autistic disturbances of affective contact. *Nervous Child*, *2*, 217–250.

Klin, A., & Jones, W. (2008). Altered face scanning and impaired recognition of biological motion in a 15-month-old infant with autism. *Developmental Science*, *11*(1), 40–46. https://doi.org/10.1111/j.1467-7687.2007.00640.x

Klin, A., Lin, D. J., Gorrindo, P., Ramsay, G., & Jones, W. (2009). Two-year-olds with autism orient to non-social contingencies rather than biological motion. *Nature*, *459*(7244), 257–261. https://doi.org/10.1038/nature07868

Lord, C., Luyster, R. J., Gotham, K., & Guthrie, W. (2012). (ADOS-2) manual (Part II): Toddler module. In *Autism diagnostic observation schedule* (2nd ed.). Western Psychological Services.

Lord, C., Rutter, M., DiLavore, P. C., Risi, S., Gotham, K., & Bishop, S. L. (2012). (ADOS-2) manual (Part 1): Modules 1-4. In *Autism diagnostic observation schedule* (2nd ed.). Western Psychological Services.

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*(5), 659–685. https://doi.org/10.1007/BF02172145

Magrelli, S., Jermann, P., Noris, B., Ansermet, F., Hentsch, F., Nadel, J., & Billard, A. (2013). Social orienting of children with autism to facial expressions and speech: A study with a wearable eye-tracker in naturalistic settings. *Frontiers in Psychology*, *4*, 840. https://doi.org/10.3389/fpsyg.2013.00840

Papagiannopoulou, E. A., Chitty, K. M., Hermens, D. F., Hickie, I. B., & Lagopoulos, J. (2014). A systematic review and meta-analysis of eye-tracking studies in children with autism spectrum disorders. *Social Neuroscience*, *9*(6), 610–632. https://doi.org/10.1080/17470919.2014.934966

Pierce, K., Conant, D., Hazin, R., Stoner, R., & Desmond, J. (2011). Preference for geometric patterns early in life as a risk factor for autism. *Archives of General Psychiatry*, *68*(1), 101–109. https://doi.org/10.1001/archgenpsychiatry.2010.113

Pierce, K., Marinero, S., Hazin, R., McKenna, B., Barnes, C. C., & Malige, A. (2015). Eye tracking reveals abnormal visual preference for geometric images as an early biomarker of an autism spectrum disorder subtype associated with increased symptom severity. *Biological Psychiatry*, *79*(8), 657–666. https://doi.org/10.1016/j.biopsych.2015.03.032

Rapin, I. (1997). Autism. *New England Journal of Medicine*, *337*, 97–104.

Rice, K., Moriuchi, J. M., Jones, W., & Klin, A. (2012). Parsing heterogeneity in autism spectrum disorders: Visual scanning of dynamic social scenes in school-aged children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(3), 238–248. https://doi.org/10.1016/j.jaac.2011.12.017

Risi, S., Lord, C., Gotham, K., Corsello, C., Chrysler, C., Szatmari, P., Cook, E. H., Leventhal, B. L., & Pickles, A. (2006). Combining information from multiple sources in the diagnosis of autism spectrum disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *45*(9), 1094–1103. https://doi.org/10.1097/01.chi.0000227880.42780.0e

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire manual*. Western Psychological Services.

Sasson, N. J., & Elison, J. T. (2012). Eye tracking young children with autism. *Journal of Visualized Experiments*, *61*, e3675. https://doi.org/10.3791/3675

Vivanti, G., Trembath, D., & Dissanayake, C. (2013). Atypical monitoring and responsiveness to goal-directed gaze in autism spectrum disorder. *Experimental Brain Research*, *232*(2), 695–701. https://doi.org/10.1007/s00221-013-3777-9

Zwaigenbaum, L., & Penner, M. (2018). Autism spectrum disorder: Advances in diagnosis and evaluation. *BMJ (Clinical Research Ed)*, *361*, k1674. https://doi.org/10.1136/bmj.k1674

How to cite this article: Al-Shaban, F. A., Ghazal, I., Thompson, I. R., Klingemier, E. W., Aldosari, M., Al-Shammari, H., Al-Faraj, F., El-Hag, S., Tolefat, M., Ali, M., Nasir, B., & Frazier, T. W. (2023). Development and validation of an Arabic language eye-tracking paradigm for the early screening and diagnosis of autism spectrum disorders in Qatar. *Autism Research*, 1–11. https://doi.org/10.1002/aur.3046