

# Psychometric Analysis of the Social Communication Questionnaire Using an Item-Response Theory Framework: Implications for the Use of the Lifetime and Current Forms

Tianlan Wei · Steven Randall Chesnut · Lucy Barnard-Brak · David Richman

Published online: 28 October 2014  
© Springer Science+Business Media New York 2014

**Abstract** The Social Communication Questionnaire (SCQ) was developed as a screener of Autism Spectrum Disorder (ASD). To date, the majority of the SCQ utility studies focused on its external validity (e.g., ROC curve analyses), but very few have addressed the internal validity issues. With samples consisting of 2,134 individuals available from the National Database for Autism Research (NDAR), the current study examined the factor structure, item-level characteristics, and measurement equivalence of the SCQ forms (i.e., Lifetime form and Current form) using both the classical true score theory and the Item Response Theory (IRT). While our findings indicate sufficient psychometric properties of the SCQ Lifetime form, measurement issues emerged with respect to the SCQ Current form. These issues include lower internal consistencies, a weaker factor structure, lower item discriminations, significant pseudo-guessing effects, and subscale-level measurement bias. Thus, we caution researchers and clinicians about the use of the SCQ Current form. In particular, it seems inappropriate to use the Current form as an alternative to the Lifetime form among children younger than 5 years old or under other special situations (e.g., teacher-report data), although such practices were advised by the publisher of the SCQ. Instead, we recommend modifying the wording of the Lifetime form items rather than switching to the Current form where a 3-month timeframe is specified for responding to SCQ items. Future studies may consider investigating the association between the

temporality of certain behaviors and the individual's potential for being diagnosed with ASD, as well as the age neutrality of the SCQ.

**Keywords** Social Communication Questionnaire (SCQ) · Current form · Lifetime form · Autism spectrum disorder (ASD) · Item response theory (IRT) · Measurement equivalence

Individuals with autism spectrum disorder (ASD) demonstrate many challenging behavioral patterns that can interrupt day-to-day functioning and classroom instruction (Dawson, Matson, and Cherry 1998; Hill and Furniss 2006; Matson and Shoemaker 2009). The difficulties that these individuals face communicating their thoughts (Aldred, Green, and Adams 2004), interpreting the expressions and communication of others (Howlin, Mawhood, and Rutter 2000), and understanding what is happening in their environment in lieu of sensory impairments or sensitivities (Rogers, Hepburn, and Wehner 2003; Tomchek and Dunn 2007) can have a negative effect on learning new skills and adaptive functioning in their environment. These difficulties, compounded throughout several years of life, can create a skills gap so great that they lag behind their peers (e.g., learning disabilities in reading, mathematics, or overall intellectual and adaptive functioning). Early intensive behavioral interventions have shown to have a positive influence on developmental outcomes (Matson and Smith 2008).

These early interventions tend to focus on improving interpersonal social interactions (e.g., Aldred et al. 2004; Bauminger 2002) and improving responses to aversive stimuli within the environment (e.g., Baranek 2002; Pfeiffer, Koenig, Kinnealey, Sheppard, and Henderson 2011). With increased instruction and individually tailored interventions, individuals

---

T. Wei (✉) · S. R. Chesnut · L. Barnard-Brak · D. Richman  
Department of Educational Psychology and Leadership, Texas Tech University, Lubbock, TX 79409, USA  
e-mail: tw1518@msstate.edu

*Present Address:*

T. Wei  
Department of Counseling and Educational Psychology, Mississippi State University, Mississippi State, MS 39762, USA

with ASD can achieve higher levels of adaptive functioning in their daily lives, including academic settings. Diagnosing ASD and providing intensive early interventions are costly and time consuming to public schools (Hess, Morrier, Heflin, and Ivey 2008). Because of this, screeners such as the Social Communication Questionnaire (SCQ; Berument, Rutter, Lord, Pickles, and Bailey 1999) have been developed to allow parents, guardians, and teachers to help school psychologists and clinicians understand where they should focus their efforts for instruction and whether the child should be referred for a comprehensive diagnostic evaluation of ASD.

Before requesting a full clinical evaluation, screeners that caregivers (e.g., parents, guardians, teachers) can use to help determine the likelihood of a child having ASD are typically employed. In order for these screeners to be accessible to those not trained in special education or diagnostics, they are typically behavioral checklists that allow the respondents to report whether the child has demonstrated behavioral patterns that are characteristic of ASD. Each one of these observable, behavioral items contribute to an overall score that is compared with a *cutoff score* used to determine whether further diagnostic evaluation and instructional resources should be allocated to the student. There have been a few screeners for ASD developed in the last 20 years (e.g., the SCQ, Berument et al. 1999; Modified Checklist for Autism in Toddlers [M-CHAT], Robins, Fein, Barton, and Green 2001); however, the SCQ has been one of the most widely adopted (Wilkinson 2010). There are two versions of the questionnaire: SCQ Lifetime, which is completed with reference to the individual's entire developmental history, and SCQ Current, which is completed with reference to the individual's behavior during the last three months. The SCQ has found popularity not only in helping children receive the services and attention needed to intervene in skills development, but also for researchers studying a wide range of pervasive developmental disabilities.

The SCQ was developed using a sample of 200 participants ranging from 4 to 40 years of age (Berument et al. 1999). This norming sample represented a wide variety of individuals with and without pervasive developmental disabilities (e.g., ASD). With an established *cutoff score* of 15, Berument et al. (1999) reported a sensitivity of 0.96 and a specificity of 0.80 when comparing ASD to other diagnoses. However, subsequent validity studies did not confirm such psychometric properties of the SCQ when applied to different populations (e.g., individuals with intellectual disabilities, individuals with Down syndrome). At the cut-off of 15, most of these studies reported lower (< 0.80) and unsatisfactory balance of sensitivities and specificities (e.g., Allen, Silove, Williams, and Hutchins 2007; Brooks and Benson 2013; Eaves, Wingert, Ho, and Mickelson 2006; Oosterling et al. 2010; Snow and Lecavalier 2008; Wiggins, Bakeman, Adamson, and Robins 2007; Witwer and Lecavalier 2007), and quite a few of them suggested lowering the cut-off for improving the discriminant utility of

the SCQ (e.g., Corsello et al. 2007; Johnson et al. 2011; Lee, David, Rusyniak, Landa, and Newschaffer 2007; Schanding, Nowell, and Goin-Kochel 2012; Wiggins et al. 2007). Therefore, the unpredictable nature of the SCQ has caused some researchers and clinicians to question its validity. Additionally, it is not always clear when researchers should utilize the Current version over the Lifetime version or if either provides more reliable results. While the SCQ Current was designed to measure change over time for assessing the effectiveness of therapeutic or educational interventions (Rutter et al. 2003a, b), many actually rely on it for children under 5 years old (see Brooks and Benson 2013; Corsello et al. 2007; Lee et al. 2007). The few studies which explicitly evaluated the SCQ Current version (Corsello et al. 2007; Oosterling et al. 2010) found particularly lower specificities of it among very young (approximately 2 to 4 years old) children. As such, it remains questionable whether the SCQ Current might be a qualified alternative to the Lifetime form for screening ASD in young children.

To date, the majority of the studies examining the appropriateness of the SCQ as a screener have focused on its external validity through receiver operating characteristic (ROC) curve analyses (e.g., Allen et al. 2007; Corsello et al. 2007; Snow and Lecavalier 2008; Wiggins et al. 2007). There is a paucity of research that carefully analyzes the internal validity of the SCQ since its initial development (Berument et al. 1999). This gap warrants attention as understanding merely the scale-level characteristics of the SCQ may limit the practical implications for the use of the scale. The purpose of this study was to evaluate the SCQ using the Item Response Theory (IRT; Lord and Novick 1968) framework. IRT offers three primary advantages over some traditional approaches to examination of the internal validity of psychometric measures: (1) the estimate of an examinee's latent trait is independent of the particular sample of items that are administered; (2) the item characteristics are independent of the particular sample of respondents; and (3) a statistic indicating the precision with which each respondent's latent trait as estimated is provided, and this statistic is free to vary among respondents (Hambleton and Swaminathan 1985). As such, the IRT analysis is capable of yielding item-level characteristics of the SCQ, which may provide valuable insight into its psychometric properties, suggestions for revisions, and directions for future uses.

In the context of health disparities research and individual assessment, researchers have also become increasingly concerned about measurement equivalence (Teresi 2006). An instrument is not *fair* when two groups of an equal amount of a latent trait earn different scores on the same item (Gall, Gall, and Borg 2006); such measurement bias is also termed as *differential item functioning* (DIF; Holland and Wainer 1993). While there have yet to be any published issues regarding biased outcomes of the SCQ, the fact that diagnoses of ASD are four to five times more prevalent amongst boys

than girls (Blumberg et al. 2013) warrants further examination of the inventory as a screener.

Thus, the purpose of the current study was to examine the internal validity of the SCQ as a screening instrument for ASD. To achieve this purpose, the psychometric properties of the SCQ were evaluated according to both the classical true score theory procedures through confirmatory factor analyses and item response theory techniques. IRT techniques were utilized to examine for the presence of subsequent differential item functioning according to gender with each form of the SCQ and across forms of the SCQ (e.g., Current versus Lifetime) in general.

## Method

### Participants

Our samples consisted of a total of 2,134 individuals with the SCQ item-level scores available from the National Database for Autism Research (NDAR). Among them, 769 individuals were surveyed using the SCQ Current form (Sample 1) and 1,660 individuals using the SCQ Lifetime form (Sample 2), with an overlap of 295 individuals having valid scores of both forms. Sample 1 contained 81.4 % males ( $n = 626$ ) and 18.6 % females ( $n = 143$ ), with ages ranging from 1 year and 5 months to 21 years and 8 months ( $M = 8$  years and 10 months,  $SD = 5$  years and 6 months). Sample 2 contained 80.8 % males ( $n = 1,341$ ) and 19.2 % females ( $n = 319$ ), with ages ranging from 1 year and 3 months to 57 years and 10 months ( $M = 10$  years and 9 months,  $SD = 6$  years and 11 months). Sample 1 was used for the IRT and DIF analyses of the SCQ Current, and Sample 2 was used for the analyses of the SCQ Lifetime. When analyzing the DIF of SCQ Current versus SCQ Lifetime, we combined the two samples and excluded the overlapping 295 individuals with both scores available to avoid correlated errors.

### Measures

The Social Communication Questionnaire (SCQ; Rutter et al. 2003a, b) is a 40-item, parent-report measure for screening symptomatology associated with autism spectrum disorder (ASD). All 40 items are administered in a dichotomous format (i.e., *yes/no*), with Item 1 simply documenting whether or not the child is able to speak with short phrases or sentences, and Items 2 through 40 used for scoring. For Items 2, 9, and 19 through 40, an individual scores 1 for response option *no* and 0 for *yes*; for the other items (i.e., Items 3 to 8 and 10 to 18), the individual scores 1 for *yes* and 0 for *no*. A score of 1 as opposed to 0 on an item always indicates a higher risk for ASD, and there is no need to reverse score. In the current

study, we use “response option 1” to refer to a score of 1 on the item regardless of the verbal response (*yes/no*).

The standardization study by Berument et al. (1999) took four steps to assess the validity of the SCQ. First, a factor analysis was performed to determine whether the scale reflected the differentiated conceptualizations of three main domains of abnormality found in autism, and the results indicate that all 39 items (Items 2 through 40) gauge four domains of symptomatology: Social Interaction ( $k=20$ ,  $\alpha=.91$ ), Communication ( $k=6$ ;  $\alpha=.71$ ), Abnormal Language ( $k=5$ ;  $\alpha=.79$ ), and Stereotyped Behavior ( $k=8$ ;  $\alpha=.67$ ). Second, the combination of individual items was assessed by noting the items-total correlations and how well they differentiated ASDs from other diagnoses, and 33 out of the 39 items (85 %) showed statistically significant differentiation. Third, the correlations between the SCQ and the Autism Diagnostic Interview-Revised (ADI-R; Rutter et al. 2003a, b) for the total scores as well as the ADI domain totals, and the correlation coefficients were statistically significant for all comparisons. Fourth, ROC curves were applied to determine how the SCQ differentiated ASD from other diagnoses, and a sensitivity of 0.96 and a specificity of 0.80 were reported when comparing ASD to other diagnoses at a cut-off of 15. In addition, convergent validity and discriminant validity of the SCQ were assessed using a *multitrait-multimethod matrix*, and it was reported that the validation support was moderately strong for the Reciprocal Social Interaction and Communication domains, while weaker for the Restricted, Repetitive, and Stereotyped Patterns of Behavior domain.

### Analyses

A major assumption of IRT analyses is unidimensionality, which refers to the notion that a set of items measures a single latent construct (Lord and Novick 1968). To test this assumption, confirmatory factor analyses (CFA) were performed separately on the four subscales of both forms using *Mplus* v. 6.0 (Muthén and Muthén 2010) with robust weighted least squares (WLS) estimation. Chi-Square ( $\chi^2$ ) tests of model fit were referred to along with approximate fit indices (AFIs) for determining acceptable model fit. Due to the well-known criticism that the  $\chi^2$  test is severely sensitive to sample size (Schmitt 2011), the following criteria of AFIs were consulted upon significant results of the  $\chi^2$  test: Root Mean Square Error of Approximation (RMSEA)  $\leq .06$ , Comparative Fit Index (CFI)  $\geq .95$ , and Weighted Root Mean Square Residual (WRMR)  $\leq 1.0$  (Bentler 2007; Hu and Bentler 1999; Yu, 2002). A CFA model with fit indices close to these cutoff values is considered to fit the data well, while a model with fit indices well above or below these values (e.g., CFI  $< 0.90$ , RMSEA  $> .10$ ) is considered to fit the data poorly.

Given the dichotomous response format of the SCQ items, Birnbaum's (1968) three-parameter (3PL) logistic model was employed for the IRT estimates in IRTPRO v. 2.1 (Cai, Thissen, and du Toit 2011). The mathematical formula of the item characteristic curve (ICC) with an item is written as:

$$P_i = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \quad (1)$$

where  $P_i$  is the probability of a person's endorsement of response option 1 for item  $i$ , the mathematical constant  $e$  is the base of the natural logarithm,  $\theta$  (theta) is the latent trait,  $a_i$  is the discrimination parameter for item  $i$ , and  $b_i$  is the threshold parameter (de Ayala 2009). The threshold parameter  $b_i$  represents the value of  $\theta$  where an individual randomly selected from all persons with this level of latent trait has the probability of 0.5 to endorse option 1, as well as the inflection point of the ICC. The discrimination parameter  $a_i$  is proportional to the slope of the ICC at its inflection point, the slope at  $b_i = 0.25 \times a_i$ . According to Baker (1985, 2001), item discrimination is considered to be "very low" for  $a < .34$ , "low" for  $.35 \leq a \leq .64$ , "moderate" for  $.65 \leq a \leq 1.34$ , "high" for  $1.35 \leq a \leq 1.69$ , and "very high" for  $a > 1.70$ . For item discrimination to be reasonably good, it should also not exceed 2.50 (de Ayala 2009). Finally, the pseudo guessing parameter of item  $i$  was denoted by  $c_i$ , indicating the probability of a response of 1 when  $\theta$  approaches negative infinite ( $-\infty$ ). In other words, it reflects that some respondents with infinitely low thetas may endorse 1 when they should not (de Ayala 2009).

The analysis of DIF was performed using DIFAS v. 5.0 (Penfield 2012). The DIF analysis is conducted by comparing a focal group (usually the minority) against a reference group (usually the majority). The Mantel-Haenszel Chi-Square (MH  $\chi^2$ ) procedure was used for detecting DIF in polytomous items (Penfield 2012) with  $\alpha = .05$ . The MH  $\chi^2$  statistic is distributed with one degree of freedom, with a higher value of a particular item indicating a higher probability for this item to display DIF (Mantel 1963). To reduce Type 1 error, we used the Benjamini and Hochberg false discovery rate (BH-FDR; Benjamini and Hochberg 1995) to adjust each item's  $p$  value of MH  $\chi^2$ . According to Kim and Oshima (2012), the BH-FDR is the most balanced adjustment in lowering the Type 1 error rate as compared with Bonferroni correction and Holm's procedure. In terms of effect size of differences, DIFAS v. 5.0 provides Mantel-Haenszel Common Log-Odds Ratio (MH LOR; Mantel and Haenszel 1959) statistics for estimating magnitude and direction of DIF. Positive MH LOR values indicate DIF in favor of the reference group, while negative values indicate DIF in favor of the focal group (Penfield 2012). In addition, Breslow-Day Chi-Square (BD; Breslow and Day 1980) statistics were

also consulted for detecting nonuniform DIF. Finally, the impact of DIF at the subscale level was examined by Differential Test Functioning (DTF) analyses. DTF represents the aggregated DIF across the items of a test or subscale and is essentially the variance estimator of item-level DIF effects. Thus, it can provide information concerning the overall impact of DIF effects. DIF effect variance is considered to be small for  $\tau^2$  (Tau-square)  $< 0.07$ , medium for  $0.07 \leq \tau^2 \leq 0.14$ , and large for  $\tau^2 > 0.14$  (Penfield and Algina 2006).

## Results

### Unidimensionality

Table 1 summarizes the fit indices of the one-factor CFA models along with the internal consistency reliabilities of each subscale: Social Interaction (SI), Abnormal Language (AL), Communication (COMM), and Stereotyped Behavior (SB). While results indicated acceptable to excellent reliabilities for SCQ Lifetime, the subscales of SCQ Current appeared to have lower internal consistencies. In addition, internal consistencies were shown to be similar within each pair of comparison groups for subsequent DIF analyses (see Table 1). In terms of unidimensionality, three subscales (i.e., Abnormal Language, Communication and Stereotyped Behavior) showed sufficient unidimensionality for both SCQ Lifetime and Current, while Social Interaction of the Current form demonstrated particularly poorer fit.

### Item Response Theory Analyses

Table 2 summarizes IRT parameter estimates for the items of all four subscales for both SCQ versions.

Based on Baker's (1985, 2001) guidelines, 18 out of the 39 items (46 %) demonstrated high to very high discrimination ( $a \geq 1.35$ ) in the Current form, while 33 items (85 %) demonstrated high to very high discrimination in the Lifetime form. Some items warranted particular attention because they demonstrated low to very low discrimination ( $a \leq 0.64$ ) in either SCQ version (i.e., Items 4 [inappropriate questions], 9 [inappropriate facial expression], 15 [hand and finger mannerisms], 17 [deliberate self-injury], 19 [friends], 23 [gestures], and 39 [imaginative play with peers]). In general, items of the Lifetime form showed higher and more consistent discriminations within subscales than those of the Current form. For example,  $a$  ranged from 0.99 to 2.96 ( $M = 1.57$ ,  $SD = 0.70$ ) for SB items in the Current form, while from 1.69 to 2.98 ( $M = 2.30$ ,  $SD = 0.49$ ) in the Lifetime form. Such comparisons are illustrated by the ICCs in Figs. 1 and 2.

In terms of item thresholds, the Lifetime form subscales demonstrated higher average thresholds than the Current form ones. The average threshold (difficulty) was  $-0.09$  ( $SD = 0.46$ )



**Table 1** CFA fit indices for subscales of the SCQ current and lifetime versions

Factor	$\chi^2$ (df)	CFI	RMSEA	WRMR	Cronbach's $\alpha^a$		
					T	F	M
<b>Current</b>							
Abnormal language	37.894 (5)*	0.963	.103	1.311	.61	.60	.61
Communication	80.666 (9)*	0.964	.102	1.670	.65	.70	.64
Social interaction	1,483.139 (170)*	0.762	.100	2.579	.81	.84	.80
Stereotyped behavior	112.655 (20)*	0.935	.078	1.488	.72	.71	.72
<b>Lifetime</b>							
Abnormal language	27.812 (5)*	0.996	.061	1.008	.79	.80	.79
Communication	112.506 (9)*	0.988	.084	1.930	.70	.75	.68
Social Interaction	1,210.454 (170)*	0.960	.061	2.058	.91	.92	.90
Stereotyped behavior	23.315 (20)	1.000	.010	0.601	.86	.86	.85

\*  $p < .0001$

<sup>a</sup>T = Total sample, F = Female, M = Male

for AL (Current) and 0.19 ( $SD=0.22$ ) for AL (Lifetime), 0.00 ( $SD=0.48$ ) for COMM (Current) and 0.69 ( $SD=0.44$ ) for COMM (Lifetime), 0.11 ( $SD=0.58$ ) for SB (Current) and 0.10 ( $SD=0.25$ ) for SB (Lifetime), and  $-0.04$  ( $SD=0.58$ ) for SI (Current) and 0.26 ( $SD=0.37$ ) for SI (Lifetime). Finally, the Lifetime form also appeared to be less influenced by the pseudo guessing effects because only two items (5 %) had significant nonzero  $c$  values, as compared with the Current form in which seven items (18 %) had significant  $c$  values.

We also compared the test information (Hambleton and Swaminathan 1985) curves among the subscales. The test information function is defined for a particular set of items (e.g., a subscale) at each point along with the continuum of the latent factor (Studts 2008). The contribution of each item to the total test information function is additive. As shown in Fig. 3, items of the Lifetime form altogether provided more information for measurement of children's social communication skills than those of the Current form.

### Differential Item Functioning

The analyses of differential item functioning (DIF) were first conducted separately on the two SCQ versions, both with male as the reference group and female as the focal group. Next, the two samples were combined for DIF to contrast the Current form and the Lifetime form. Table 3 presents results of the three sets of DIF analyses. For SCQ Lifetime, six out of the 39 items (15.4 %) demonstrated significant DIF across gender: Items 19 (friends) and 34 (imitative social play) of SI, Item 9 (inappropriate facial expression) of COMM, Item 6 (neologism) of AL, and Items 10 (use of other's body to communicate) and 13 (circumscribed interests) of SB. The Differential Test Functioning (DTF) effect variances were small for SI ( $\tau^2=0.03$ ) and AL ( $\tau^2=0.06$ ), medium for COMM ( $\tau^2=0.14$ ), and large for SB ( $\tau^2=0.19$ ).

For the SCQ Current, while only five out of the 39 items (12.8 %) demonstrated significant DIF across gender, four of

them (i.e., Items 11 [unusual preoccupations], 12 [repetitive use of objects], 13 [circumscribed interests], 18 [carry objects around]) were of the SB subscale, meaning that 50 % of the SB items were showing measurement bias. In other words, even given the same severity of stereotyped behaviors, parents of boys were more likely than those of girls to endorse option 1 (*yes*) of the statements “preoccupied with odd interests (e.g., traffic lights, drainpipes, or timetables)” (MH  $\chi^2=7.10$ , BH FDR corrected  $p=.02$ , MH LOR = +0.73), “odd interest in an object (e.g., spinning the wheels of a car)” (MH  $\chi^2=10.52$ ,  $p=.005$ , MH LOR = +0.83), and “interests that are *unusual* in intensity (e.g., trains or dinosaurs)” (MH  $\chi^2=6.16$ ,  $p=.03$ , MH LOR = +0.57). On the other hand, parents of girls were more likely to endorse option 1 (*yes*) of the statement “objects that she/he has to carry around (*other* than a soft toy or comfort blanket)” (MH  $\chi^2=23.70$ ,  $p<.0001$ , MH LOR =  $-1.17$ ). The DTF effect variances were small for SI ( $\tau^2=0.01$ ), medium for AL ( $\tau^2=0.08$ ), and large for COMM ( $\tau^2=0.15$ ) and SB ( $\tau^2=0.35$ ). Finally, only one item (i.e., Item 13 [circumscribed interests] of SCQ Lifetime) showed significant nonuniform DIF (BD=8.93, BH FDR corrected  $p=.02$ ).

In terms of the DIF results contrasting the SCQ Current and Lifetime, 16 out of the 39 items (41 %) demonstrated significant DIF and 14 of them were of SI (i.e., Items 21, 26 to 30, 32 to 35, and 37 to 40), meaning that parents responded differently to 70 % of the SI items according to whether they were administered in a Lifetime or Current version. Moreover, 50 % of the COMM items ( $k=3$ ) had significant nonuniform DIF as indicated by the BD statistics. The DTF effect variances were small for AL ( $\tau^2=-0.01$ ) and SB ( $\tau^2=0.03$ ) and large for COMM ( $\tau^2=0.17$ ) and SI ( $\tau^2=0.61$ ).

### Discussion

Whereas the majority of studies in the literature examined the external, utility validity (e.g., ROC curve analyses) of the

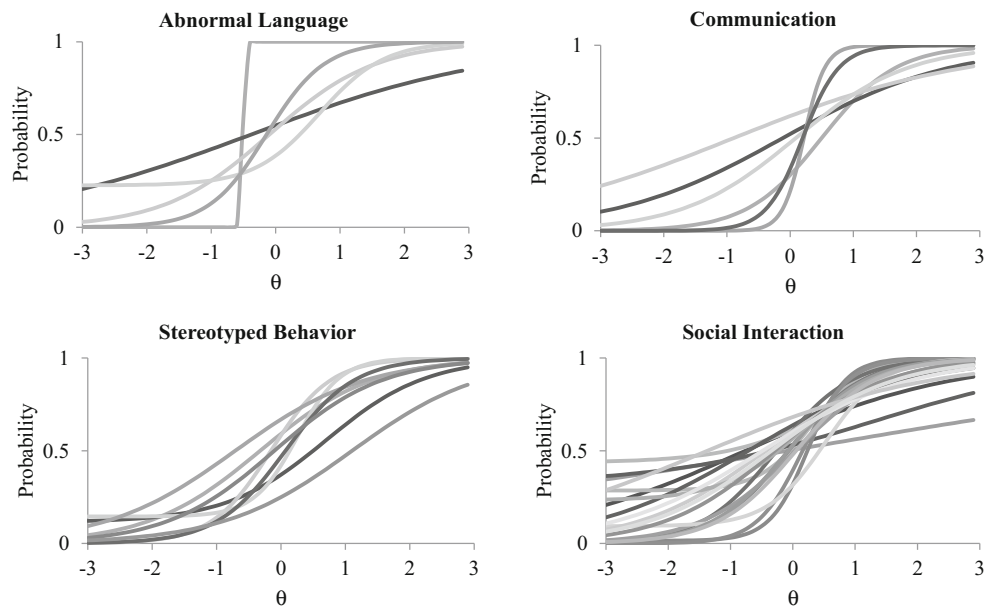
**Table 2** IRT parameter estimates (a, b, c) by subscale for the SCQ current and lifetime forms

Item#	SCQ current			SCQ lifetime		
	a <sub>i</sub> (se)	b <sub>i</sub> (se)	c <sub>i</sub> (se)	a <sub>i</sub> (se)	b <sub>i</sub> (se)	c <sub>i</sub> (se)
<b>Social interaction</b>						
17	0.24 (21.94)	0.52 (54.05)	0.07 (2.79)	0.49 (8.8E +12)	−0.11 (1.1E +13)	0 (0.02)
19	0.60 (1.88E +04)	−0.76 (3.48E +04)	0 (0.08)	1.06 (0.07)	0.02 (0.07)	0 (0)
21	1.24 (0.48)	0.62 (0.32)	0.44 (0.08)	1.54 (0.26)	0.34 (0.11)	0.09 (0.05)
22	2.36 (0.6)	0.31 (0.14)	0.29 (0.06)	2.44 (0.31)	0.47 (0.06)	0.08 (0.02)
23	0.60 (0.23)	1.29 (0.56)	0.32 (0.1)	1.26 (0.36)	0.76 (0.16)	0.26 (0.06)
26	0.79 (5.76)	−0.72 (4.67)	0 (0.89)	2.03 (0.12)	−0.07 (0.05)	0 (0)
27	1.69 (0.34)	−0.07 (0.15)	0.01 (0.07)	2.06 (0.14)	0.61 (0.05)	0 (0)
28	3.29 (0.54)	0.24 (0.05)	0.01 (0.02)	2.94 (0.25)	0.71 (0.05)	0 (0)
29	1.67 (0.18)	−0.26 (0.06)	0 (0)	2.86 (0.19)	0.13 (0.04)	0 (0)
30	2.72 (0.31)	0.16 (0.05)	0 (0)	3.00 (0.23)	0.56 (0.04)	0 (0)
31	1.33 (0.15)	0.04 (0.07)	0 (0)	2.12 (0.14)	0.18 (0.05)	0 (0)
32	2.13 (0.43)	0.52 (0.09)	0.09 (0.04)	1.79 (0.27)	1.12 (0.09)	0.02 (0.02)
33	1.63 (0.17)	−0.07 (0.06)	0 (0)	2.25 (0.16)	0.49 (0.04)	0 (0)
34	0.98 (0.12)	−0.44 (0.09)	0 (0)	2.06 (0.14)	0.11 (0.05)	0 (0)
35	1.56 (0.48)	0.17 (0.25)	0.23 (0.1)	1.94 (0.12)	0.15 (0.05)	0 (0)
36	1.13 (0.13)	−0.29 (0.08)	0 (0)	2.16 (0.14)	0.07 (0.05)	0 (0)
37	1.58 (0.18)	0.05 (0.06)	0 (0)	2.42 (0.18)	0.32 (0.04)	0 (0)
38	0.84 (0.11)	−0.49 (0.11)	0 (0)	2.23 (0.15)	−0.03 (0.05)	0 (0)
39	0.56 (0.87)	−1.37 (13.81)	0 (0)	2.68 (0.17)	−0.41 (0.05)	0 (0)
40	1.01 (1.12)	−0.34 (4.22)	0 (0.01)	2.45 (0.15)	−0.18 (0.05)	0 (0)
<b>Communication</b>						
2	1.69 (0.27)	0.50 (0.11)	0 (0)	1.43 (0.18)	1.05 (0.11)	0 (0)
9	0.75 (46.16)	−0.12 (33.93)	0 (2.19)	0.56 (2.02)	1.35 (5.3)	0 (0.01)
15	0.54 (118.65)	−0.89 (273.86)	0 (0.03)	1.08 (0.1)	0.19 (0.06)	0 (0)
20	1.12 (0.13)	0.09 (0.08)	0 (0)	1.45 (0.12)	0.34 (0.05)	0 (0)
24	6.01 (1.76)	0.22 (0.05)	0 (0)	47.60 (0.29)	0.50 (0.01)	0.01 (0.01)
25	3.46 (0.46)	0.20 (0.05)	0 (0)	5.27 (0.8)	0.68 (0.04)	0 (0.01)
<b>Abnormal language</b>						
3	47.25 (1.36)	−0.51 (0.01)	0 (0)	7.17 (1.51)	−0.13 (0.03)	0 (0)
4	0.52 (7.19)	−0.37 (7.88)	0 (1.41)	1.33 (0.11)	0.39 (0.06)	0 (0)
5	1.21 (0.15)	−0.10 (0.08)	0 (0)	1.90 (0.15)	0.14 (0.05)	0 (0)
6	1.99 (0.89)	0.68 (0.15)	0.23 (0.08)	1.89 (0.15)	0.40 (0.05)	0 (0)
7	2.23 (0.27)	−0.14 (0.06)	0 (0)	3.88 (0.39)	0.13 (0.04)	0 (0)
<b>Stereotyped behavior</b>						
8	1.14 (1.14E +05)	−0.32 (1.18E +04)	0 (0)	2.33 (0.16)	−0.09 (0.04)	0 (0)
10	1.28 (0.48)	0.70 (0.25)	0.11 (0.11)	1.72 (0.12)	0.26 (0.05)	0 (0)
11	2.15 (0.25)	−0.16 (0.06)	0 (0)	2.98 (0.22)	0.01 (0.04)	0 (0)
12	2.96 (1)	0.25 (0.11)	0.15 (0.06)	2.97 (0.22)	−0.02 (0.04)	0 (0)
13	0.99 (0.13)	−0.71 (0.11)	0 (0)	2.09 (0.14)	−0.27 (0.04)	0 (0)
14	1.87 (0.2)	0.07 (0.06)	0 (0)	2.43 (0.18)	0.19 (0.04)	0 (0)
16	1.20 (0.14)	−0.10 (0.08)	0 (0)	2.16 (0.16)	0.17 (0.04)	0 (0)
18	1.00 (0.13)	1.11 (0.14)	0 (0)	1.69 (0.12)	0.56 (0.05)	0 (0)

Social Communication Questionnaire, the current study is a comprehensive review of the interval validity issues of both

SCQ forms (i.e., Lifetime and Current). We could only locate one other study (Magyar, Pandolfi, and Dill 2012) that

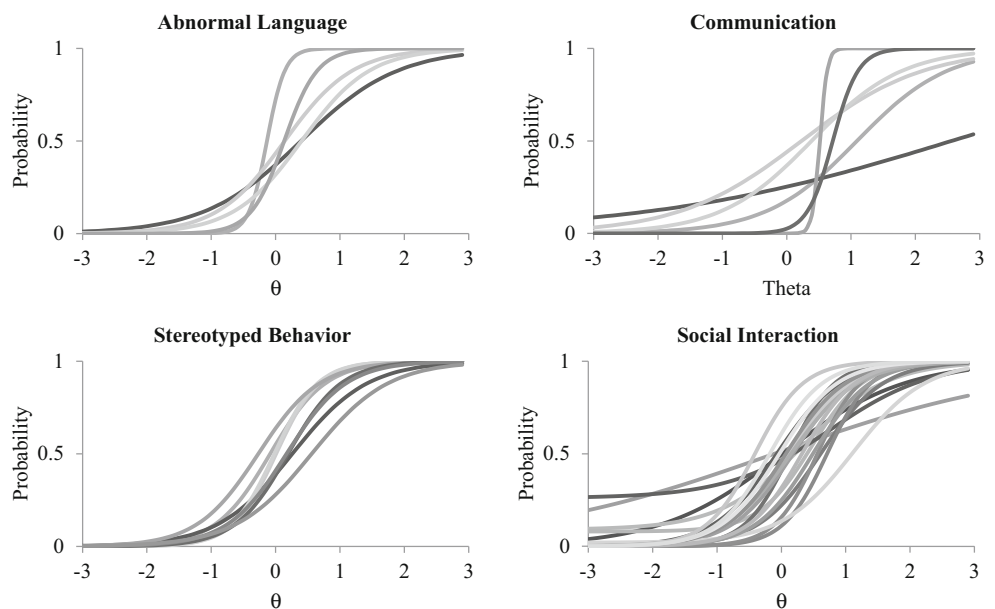
**Fig. 1** Item characteristic curves (ICCs) of SCQ current by subscale



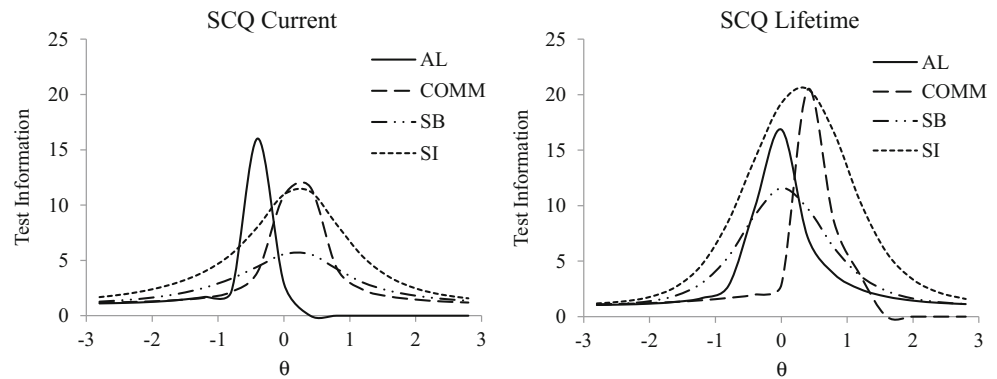
examined the internal validity of the SCQ. In Magyar et al. (2012), the sample consisted of approximately 70 participants and was limited to the classical true score theory and factor analysis. In the current study, classical true score theory techniques were first utilized via the examination of the internal consistencies. Then, IRT techniques were utilized to examine item-level characteristics and measurement equivalence across items. In view of extant literature, the current study represents the largest sample among the validity studies on SCQ. With this large sample available from NDAR, we consider our findings to possess valuable information for research and clinical practice on Autism Spectrum Disorder.

Our findings indicate sufficient psychometric properties of the SCQ Lifetime form, but several measurement issues emerged with respect to the Current form. First, the SCQ Current subscales demonstrated lower internal consistencies and a weaker factor structure. In particular, the unidimensionality of the Current form subscales does not appear to be strong as indicated by the fit indices (Bentler 2007; Hu and Bentler 1999; Yu, 2002). We need to note that when Berument et al. (1999) established the four-factor structure of the SCQ, no data were available for the evaluation of the Current form. Therefore, researchers need to carefully examine the factor structure of the SCQ when necessary in their studies. Second, the IRT-based examination indicated that the

**Fig. 2** Item characteristic curves (ICCs) of SCQ lifetime by subscale



**Fig. 3** Test information functions of SCQ current and lifetime



same items in the SCQ Lifetime may not perform as well when utilized in the Current form. That is, certain behaviors observed within a shorter time frame (i.e., the past 3 months for the Current form), rather than in one's lifetime, may not be sufficient indicators of an individual's potential for being diagnosed with ASD. In terms of the IRT parameters, the superiority of the Lifetime form mainly concerns the item discrimination ( $a$ ) that the majority of the 39 items (92 %) demonstrated higher discrimination in the Lifetime form (see Table 2 and Figs. 1 and 2). When administered in the Current form, less than 50 % of the items showed reasonably high discrimination, and some (e.g., Items 17 [deliberate self-injury] and 19 [friends]) were low in discrimination. Lastly, there appears to be more significant  $c$  (pseudo guessing) values in the Current form as compared with the Lifetime form, suggesting that parents are more likely to endorse option 1 on certain items in the Current form (e.g., Items 21 [imitation], 22 [pointing to express interest], and 23 [gestures]) when they would not be expected to given their otherwise low trait values. In addition, the higher incidence of the pseudo-guessing effect in the Current form items (18 % as compared with 5 % in the Lifetime items) also indicates that the reliability of some measures may be impaired when a narrow time frame is specified in the Current form.

In terms of DIF, our analyses reveal that the Lifetime form was established with sufficient measurement equivalence across gender because the percentage of items showing significant DIF were limited to 25 % or below within each subscale (Teresi 2006). However, we must note that the DIF in some Lifetime items (i.e., Items 6, 9, 10, 19, 13, and 34) will likely confound any item-level comparisons even though the subscale-level scores can be trusted. For instance, we do not recommend using the raw scores on Item 9 to examine gender differences in facial expression because the DIF results show that a parent is more likely to indicate that the child's facial expression has been inappropriate if the child is a girl rather than a boy. On the other hand, the Current form warrants some concern with regard to the measure of stereotyped behaviors because 50 % of the SB items demonstrated significant DIF across gender. It is possible that the wording or examples of

specific items to be gender-stereotypical (e.g., “drainpipes,” “spinning the wheels of a car,” “soft toy,” “comfort blanket”) and may interfere with a parent's or caregiver's responding according to the child's gender. While this observation is preliminary and certainly requires further investigations, there is empirical evidence that children's gender-role development begins early at 3- or 4-years old (Bussey and Bandura 1992, 2004). We recommend using more gender-neutral activities as examples in future revisions of the SCQ items.

It is worth particular attention that the DIF analysis contrasting the Lifetime and Current forms yielded significant and substantial DIF, particularly in the Social Interaction subscale. That is, a parent will likely respond differently according to which form is administered, even given the same level of social interaction skills of the child. Although the SI items were phrased differently in SCQ Current and Lifetime that an age limit was specified in the Lifetime version (e.g., “does she/he ever . . .” in the Current form versus “when she/he was 4 to 5, did she/he ever . . .” in the Lifetime form), the DIF is unlikely a function of this difference in wording given that six of the 14 items had significant DIF in favor of the reference group (MH LOR > 0) while the other eight in favor of the focal group (MH LOR < 0). Along with our finding that the unidimensionality of the SI subscale is particularly weak in the Current form, the measure of individuals' social interaction skills for the purpose of ASD screening using the Current form may be problematic.

In summary, the IRT-based examination and analyses of measurement equivalence do not indicate good psychometric properties of the Current form, which aligns with the findings from the external validity studies of the SCQ. Our findings, however, indicate that simply lowering the cut-off may not solve the measurement issues because the instrument suffers from an insufficient and unbalanced distribution of item discriminations, as well as some degree of pseudo guessing effects. We caution researchers and clinicians about the administration and interpretation of the SCQ Current form. In particular, it may not be appropriate for some studies to use the Current form among children below 5 years old (e.g., Corsello et al. 2007; Lee et al. 2007; Oosterling et al. 2010) or



**Table 3** DIF results by subscale and item

Item#	DIF by gender-SCQ Current			DIF by gender-SCQ Lifetime			DIF by version (Current vs. Lifetime)		
	MH $\chi^2$	MH LOR	BD	MH $\chi^2$	MH LOR	BD	MH $\chi^2$	MH LOR	BD
<b>Social interaction</b>									
17	0.00	0.01	0.04	0.30	-0.12	0.08	2.36	0.20	18.94***
19	1.32	0.27	2.35	10.49*	-0.55	0.79	1.17	-0.14	0.20
21	2.50	0.35	0.02	1.69	0.26	3.92	9.42**	-0.39	1.07
22	0.00	0.02	1.33	0.15	-0.10	0.35	0.02	-0.03	0.77
23	0.01	0.00	0.24	0.02	0.04	2.91	4.04	-0.24	0.54
26	5.50	-0.56	0.08	0.57	0.16	0.38	29.86***	-0.73	0.00
27	0.42	0.18	0.14	0.20	-0.12	0.68	22.07***	0.70	0.91
28	0.00	0.02	0.13	0.04	-0.08	0.56	17.68***	1.73	0.02
29	2.52	-0.43	0.05	1.37	0.31	1.14	5.48*	-0.36	0.91
30	0.07	-0.11	0.97	0.97	-0.26	0.27	73.84***	1.29	3.70
31	0.02	0.06	1.89	0.67	-0.19	0.80	1.08	0.15	0.23
32	0.51	0.19	1.17	0.09	-0.09	0.06	48.28***	1.68	0.51**
33	0.03	0.07	0.16	0.09	0.09	2.05	14.63***	0.55	2.34
34	1.54	-0.31	0.20	8.13*	0.55	6.07	7.11*	-0.37	0.02
35	4.68	-0.52	0.02	1.26	0.26	1.25	6.83*	-0.36	2.14
36	0.58	0.21	0.77	5.23	0.47	5.00	2.78	-0.24	0.02
37	0.11	-0.11	0.03	0.03	-0.06	0.45	8.12**	0.42	0.01
38	0.00	0.04	0.90	0.08	-0.08	0.75	34.82***	-0.79	1.67
39	1.61	0.29	2.78	0.17	-0.12	1.13	10.80**	-1.32	11.94
40	0.24	0.14	0.35	2.43	-0.36	0.14	23.41***	-0.67	0.21
<b>Communication</b>									
2	3.07	-0.63	0.08	0.00	-0.04	1.29	12.41**	0.63	1.46
9	2.58	-0.43	2.05	18.52***	-0.86	2.57	0.49	0.12	0.50
15	11.42**	0.84	0.12	3.46	0.40	0.83	3.88	-0.29	6.78*
20	0.38	0.21	0.76	0.86	0.25	3.82	19.50***	-0.74	3.40
24	0.93	-0.37	0.00	0.28	0.18	3.48	0.89	0.19	12.29**
25	0.01	-0.09	0.04	2.44	0.47	0.77	4.17	0.39	11.71**
<b>Abnormal language</b>									
3	0.02	0.13	0.51	0.46	0.24	0.80	0.05	-0.08	1.47
4	1.73	0.35	0.11	0.06	0.07	1.01	3.02	0.28	0.22
5	6.55	-0.72	0.25	2.28	0.34	1.50	0.36	-0.11	0.26
6	0.04	-0.09	0.11	8.96*	-0.65	2.21	0.43	-0.12	0.07
7	1.52	0.39	0.67	0.06	0.10	1.85	0.05	-0.06	0.04
<b>Stereotyped behavior</b>									
8	0.05	0.08	0.00	3.44	-0.39	0.74	0.21	0.07	1.09
10	2.50	-0.37	0.92	6.42*	-0.50	0.79	1.98	-0.19	3.95
11	7.10*	0.73	0.00	0.16	0.11	0.01	1.10	-0.18	2.27
12	10.52**	0.83	4.61	4.79	0.48	1.10	0.00	0.01	0.04
13	6.16*	0.57	0.39	23.19***	1.02	8.93*	7.47	0.40	3.72
14	0.65	-0.24	0.40	2.41	-0.35	0.01	0.23	0.08	0.36
16	2.02	-0.37	1.30	0.09	-0.08	0.02	5.11	-0.33	0.00
18	23.70***	-1.17	0.02	0.50	-0.16	0.57	0.93	0.15	1.11

(BH FDR corrected) \* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$

other special situations (e.g., teacher-report; Schanding et al. 2012). First, the Current form was not intended as an

alternative to the Lifetime form (Rutter et al. 2003a, b). Second, the SCQ was developed for individuals from 4 to

40 years old, and the downward extension to 2 years old needs to be undertaken with caution (Rutter et al. 2003a, b).

In Lee et al. (2007), researchers were advised via personal communication with SCQ publishers to use the current form of the SCQ over the lifetime form for individuals younger than 4 years old. For future research where the SCQ is to be used among children below 4 years old, we recommend modifying the wording of the Lifetime items (i.e., removing the timeframe “when she/he was 4 to 5”) rather than switching to the Current form where a 3-month time frame is specified. Given the psychometric properties revealed by this study, it is difficult to rely on the Current form to accurately measure change over time for evaluating the effectiveness of interventions, as intended by the test developers (Rutter et al. 2003a, b). We consider it necessary to administer the SCQ Current form along with the modified (i.e., time frame removed) Lifetime form for an examination of concurrent validity in such program evaluations. It is recommended that future studies investigate the association between the temporality (e.g., is the past 3 months an appropriate timeframe for the measure?) of certain behaviors and the potential of being diagnosed with ASD.

Finally, an investigation of *age neutrality* may be another promising direction for future research considering the wide age range (e.g., 4 to 40 years) that the SCQ purports to accurately measure. The age neutrality of the SCQ is questionable based upon previous research (e.g., Barnard-Brak et al., in progress, Corsello et al. 2007; Allen et al. 2007; Wiggins et al. 2007) given this broad range of ages. We suggest that for some items such as those concerning pronoun reversal would not be developmentally appropriate at some ages but would be considered so at older ages. Other items would appear to more resistant to item parameter drift (Wells, Subkoviak, and Serlin 2002) or the presence of longitudinal DIF such as items concerning self-injurious behavior, which ostensibly be infrequent at any age in the typically developing population. Future research should examine SCQ scores from a longitudinal perspective to discern the presence of DIF across time. The lack of non-uniform DIF would appear to support the ability of researchers to examine for the presence of DIF across time without having to be concerned about the interaction of variables indicating the presence of cross-sectional DIF.

Limitations that emerged as part of conducting this study need to be noted. First, while our analyses benefit from large samples available from National Database for Autism Research (NDAR), limited demographic information were available for the data set to be considered nationally-representative of individuals who may receive the SCQ. Second, the 295 participants we excluded from the DIF analysis contrasting Lifetime and Current forms may contain valuable information because they have valid data of both forms. Future studies may examine, for example, how a

caregiver of the same individual may respond differently to the Lifetime and Current forms. Interestingly, caregivers could rate an individual as having had a behavior on the Lifetime form but not endorse this same behavior on the Current form. This discrepancy could indicate that certain behaviors are a function of the developmental course of autism spectrum disorder, suggesting a lack of age neutrality of the SCQ. Finally, because our data were drawn from a set of studies from NDAR (e.g., “Computer Adaptive Testing of Adaptive Behavior of Children and Youth with Autism,” “Biological and Information Processing Mechanisms Underlying Autism”), there may be study-level characteristics which influence the data analysis and inferences. Future studies utilizing this data set may consider adopting multilevel analytic techniques if the nested nature of the data would appear to be influential. In conclusion, the current study provides a comprehensive examination of the internal validity of the SCQ according to both classical true score and item response theory techniques using the largest sample of SCQ responses in extant, peer-reviewed literature. Item response theory techniques were emphasized in the discussion of implications as item-level characteristics and their interactions with respondents were able to be examined using these techniques. We conclude that the Lifetime form of the SCQ appears to have sufficient psychometric properties, while the Current form of the SCQ appears to have more questionable qualities. We suggest that the Current form of the SCQ be used with particular caution given the evidence of its questionable internal validity.

**Acknowledgments** Data and/or research tools used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the NIH-supported National Database for Autism Research (NDAR). The unique ID for the current study is [NDAR Study 358](#), DOI: [10.15154/1158938](#). NDAR is a collaborative biomedical informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in autism. The study included the following collection names (along with submitters): Autism Risk, Prenatal Environmental Exposures, and Pathophysiologic Markers (Irva Hertz-Picciotto); Biological and Information Processing Mechanisms Underlying Autism (Nancy Minshew); CBT for Anxiety in Adolescents with Autism (Eric A. Storch); Computer Adaptive Testing of Adaptive Behavior of Children and Youth with Autism (Wendy Coster); Decoding 'what' and 'who' in the auditory system of children with autism spectrum disorder (Vinod Menon); Dissecting Epistasis and Pleiotropy in Autism towards Personalized Medicine (Lauren Weiss); Early Autism Risk Longitudinal Investigation (Craig Newschaffer); Electrophysiologic Studies of Language Impairment in ASD (Timothy Roberts); Functional Neuroimaging of Attention in Autism (Benjamin Yerys); Identifying Brain-Based Biomarkers for ASD and their Biological Subtypes (Bradley Peterson); Longitudinal MRI Study of Infants at Risk for Autism (Bradley Peterson); Longitudinal MRI Study of Infants at Risk for Autism (Joseph Piven); Restricted Interests and Repetitive Behaviors in Autism (Gabriel Dichter); Neural Dissection of Hyperactivity/Inattention in Autism (Francisco Xavier Castellanos); PDN Screening

Data (Audrey Thurm); Sex specific dissection of autism genetics (Lauren Weiss); Studies to Advance Autism Research and Treatment (STAART) (Joseph Buxbaum); UIC ACE: Translational Studies of Insistence on Sameness in Autism (Stephen Guter). This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or of the researchers submitting original data to NDAR.

**Conflict of Interest** Tianlan Wei, Steven Randall Chesnut, Lucy Barnard-Brak, and David Richman declared no potential conflicts of interest with respect to the research, authorship, funding, and/or publication of this article.

**Experiment Participants** The authors declared that data and/or research tools used in the preparation of this manuscript were obtained and analyzed from the controlled access datasets distributed from the NIH-supported National Database for Autism Research (NDAR). The NDAR Policy addresses informed consent concerns for both prospective studies and retrospective studies using existing data. In addition, NIH has obtained a Certificate of Confidentiality and developed sample informed consent language. NIH has also prepared a brochure for research participants.

## References

- Aldred, C., Green, J., & Adams, C. (2004). A new social communication intervention for children with autism: Pilot randomised controlled treatment study suggesting effectiveness. *Journal of Child Psychology and Psychiatry*, *45*(8), 1420–1430.
- Allen, C. W., Silove, N., Williams, K., & Hutchins, P. (2007). Validity of the social communication questionnaire in assessing risk of autism in preschool children with developmental problems. *Journal of Autism and Developmental Disorders*, *37*(7), 1272–1278.
- Baker, F. B. (1985). *The basics of item response theory*. Portsmouth: Heinemann.
- Baker, F. B. (2001). *The basics of Item Response Theory* (2nd ed.). College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland.
- Baranek, G. T. (2002). Efficacy of sensory and motor interventions for children with autism. *Journal of Autism and Developmental Disorders*, *5*, 397–422.
- Bauminger, N. (2002). The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. *Journal of Autism and Developmental Disorders*, *32*(4), 283–298.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B Methodological*, *57*, 289–300.
- Bentler, P. M. (2007). On tests and indices for evaluating structural models. *Personality and Individual Differences*, *42*, 825–829.
- Berument, S. K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry*, *175*, 444–451.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 392–479). Reading: Addison-Wesley.
- Blumberg, S. J., Bramlett, M. D., Kogan, M. D., Schieve, L. A., Jones, J. R., & Lu, M. C. (2013). Changes in prevalence of parent-reported autism spectrum disorder in school-aged U.S. children: 2007 to 2011–2012. *National health statistics reports; no 65*.
- Breslow, N. E., & Day, N. E. (1980). *Statistical methods in cancer research: Volume 1 the analysis of case-control studies*. Lyon: International Agency for Research on Cancer.
- Brooks, W. T., & Benson, B. A. (2013). The validity of the social communication questionnaire in adults with intellectual disability. *Research in Autism Spectrum Disorders*, *7*, 247–255.
- Bussey, K., & Bandura, A. (1992). Self-regulatory mechanisms governing gender development. *Child Development*, *63*, 1236–1250.
- Bussey, K., & Bandura, A. (2004). Social cognitive theory of gender development and differentiation. In A. H. Eagly, A. E. Beall, & R. J. Sternberg (Eds.), *The psychology of gender* (2nd ed., pp. 92–119). New York: Guilford Press.
- Cai, L., Thissen, D., & du Toit, S. H. (2011). IRTPRO: Flexible, multi-dimensional, multiple categorical IRT modeling [Computer software]. Lincolnwood, IL: Scientific Software International.
- Corsello, C., Hus, V., Pickles, A., Risi, S., Cook, E. H., Leventhal, B. L., & Lord, 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*(9), 932–940.
- Dawson, J. E., Matson, J. L., & Cherry, K. E. (1998). An analysis of maladaptive behaviors in persons with autism, PDD-NOS, and mental retardation. *Research in Developmental Disabilities*, *19*, 439–448.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- Eaves, L. C., Wingert, H. D., Ho, H. H., & Mickelson, E. C. (2006). Screening for autism spectrum disorders with the social communication questionnaire. *Developmental and Behavioral Pediatrics*, *27*(2), 94–103.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2006). *Educational research: An introduction* (8th ed.). Boston: Pearson.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Norwell: Kluwer Academic Publishers.
- Hess, K. L., Morrier, M. J., Heflin, L. J., & Ivey, M. L. (2008). Autism treatment survey: Services received by children with autism spectrum disorders in public school classrooms. *Journal of Autism and Developmental Disorders*, *38*(5), 961–971.
- Hill, J., & Furniss, F. (2006). Patterns of emotional and behavioural disturbance associated with autistic traits in young people with severe intellectual disabilities and challenging behaviours. *Research in Developmental Disabilities*, *27*(5), 517–528.
- Holland, P. W., & Wainer, H. (1993). *Differential item functioning*. Hillsdale: Lawrence Erlbaum Associates.
- Howlin, P., Mawhood, L., & Rutter, M. (2000). Autism and developmental receptive language disorder—a follow-up comparison in early adult life. II: Social, behavioural, and psychiatric outcomes. *Journal of Child Psychology and Psychiatry*, *41*(5), 561–578.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
- Johnson, S., Hollis, C., Hennessy, E., Kochhar, P., Wolke, D., & Marlow, N. (2011). Screening for autism in preterm children: Diagnostic utility of the social communication questionnaire. *Archives of Disease in Childhood*, *96*, 73–77.
- Kim, J., & Oshima, T. C. (2012). Effect of multiple testing adjustment in differential item functioning detection. *Educational and Psychological Measurement*. doi:10.1177/0013164412467033. Advance online publication.
- Lee, L., David, A. B., Rusyniak, J., Landa, R., & Newschaffer, C. J. (2007). Performance of the social communication questionnaire in children receiving preschool special education services. *Research in Autism Spectrum Disorders*, *1*(2), 126–138.
- Lord, F., & Novick, M. R. (Eds.). (1968). *Statistical theories of mental test scores*. Reading: Addison-Wesley.
- Magyar, C. I., Pandolfi, V., & Dill, C. A. (2012). An initial evaluation of the social communication questionnaire for the assessment of autism spectrum disorder in children with down syndrome. *Journal of Developmental and Behavioral Pediatrics*, *33*(2), 134–145.

- Mantel, N. (1963). Chi-square tests with one degree of freedom; Extension of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, *58*, 690–700.
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, *22*, 719–748.
- Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities*, *30*(6), 1107–1114.
- Matson, J. L., & Smith, K. R. (2008). Current status of intensive behavioral interventions for young children with autism and PDD-NOS. *Research in Autism Spectrum Disorders*, *2*(1), 60–74.
- Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles: Muthén & Muthén.
- Oosterling, I., Rommelse, N., de Jonge, M., van der Gaag, R., Swinkels, S., Roos, S., & Buitelaar, J. (2010). How useful is the social communication questionnaire in toddlers at risk of autism spectrum disorder? *Journal of Child Psychology and Psychiatry*, *51*(11), 1260–1268.
- Penfield, R. D. (2012). *DIFAS 5.0. Differential item functioning analysis system: User's manual*. Unpublished manuscript.
- Penfield, R. D., & Algina, J. (2006). A generalized DIF effect variance estimator for measuring unsigned differential test functioning in mixed format tests. *Journal of Educational Measurement*, *43*(3), 295–312.
- Pfeiffer, B. A., Koenig, K., Kinnealey, M., Sheppard, M., & Henderson, L. (2011). Effectiveness of sensory integration interventions in children with autism spectrum disorders: A pilot study. *The American Journal of Occupational Therapy*, *65*(1), 76–85.
- Robins, D. L., Fein, D., Barton, M., & Green, J. A. (2001). The modified checklist for autism in toddlers: An initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *31*, 131–144.
- Rogers, S. J., Hepburn, S., & Wehner, E. (2003). Parent reports of sensory symptoms in toddlers with autism and those with other developmental disorders. *Journal of Autism and Developmental Disorders*, *33*(6), 631–642.
- Rutter, M., Bailey, A., & Lord, C. (2003a). *The social communication questionnaire manual*. Los Angeles: Western Psychological Services.
- Rutter, M., Le Couteur, A., & Lord, C. (2003b). *Autism diagnostic interview—revised manual*. Los Angeles: Western Psychological Services.
- Schanding, G. T., Nowell, K. P., & Goin-Kochel, R. P. (2012). Utility of the social communication questionnaire-current and social responsiveness scale as teacher-report screening tools for autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *42*, 1705–1716.
- Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, *29*(4), 304–321.
- Snow, A. V., & Lecavalier, L. (2008). Sensitivity and specificity of the modified checklist for autism in toddlers and the social communication questionnaire in preschoolers suspected of having pervasive developmental disorders. *Autism: The International Journal of Research and Practice*, *12*(6), 627–644.
- Studts, C. R. (2008). *Improving screening for externalizing behavior problems in very young children: applications of item response theory to evaluate instruments in pediatric primary care* (Doctoral dissertation). From <http://digital.library.louisville.edu/utl/getfile/collection/etd/id/450/filename/451.pdf>.
- Teresi, J. A. (2006). Overview of quantitative measurement methods - equivalence, invariance, and differential item functioning in health applications. *Medical Care*, *44*(11), S39–S49.
- Tomchek, S. D., & Dunn, W. (2007). Sensory processing in children with and without autism: A comparative study using the short sensory profile. *The American Journal of Occupational Therapy*, *61*(2), 190–200.
- Wells, C. S., Subkoviak, M. J., & Serlin, R. C. (2002). The effect of item parameter drift on examinee ability estimates. *Applied Psychological Measurement*, *26*, 77–87.
- Wiggins, L. D., Bakeman, R., Adamson, L. B., & Robins, D. L. (2007). The utility of the social communication questionnaire in screening for autism in children referred for early intervention. *Focus on Autism and Other Developmental Disabilities*, *22*(1), 33–38.
- Wilkinson, L. A. (2010). *A best practice guide to assessment and intervention for autism and Asperger syndrome in schools*. London: Jessica Kingsley Publishers.
- Witwer, A. N., & Lecavalier, L. (2007). Autism screening tools: An evaluation of the social communication questionnaire and the developmental behaviour checklist-autism screening algorithm. *Journal of Intellectual and Developmental Disability*, *32*(3), 179–187.
- Yu, C. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. (Doctoral dissertation, University of California, Los Angeles). Retrieved from <http://statmodel2.com/download/Yudissertation.pdf>.