Original Research

Test-Retest Reliability and Validity of the Autism Symptoms Questionnaire

Richard E. Frye, MD, PhD;* Marie Tippett, BA; Leanna Delhey, BA; John Slattery, BA

Arkansas Children's Hospital Research Institute, Little Rock, AR

Although many instruments can measure symptoms in Autism Spectrum Disorder (ASD), only a few are directly based on the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR) criterion. One of these tools called the Autism Symptoms Questionnaire (ASQ) is directly based on the DSM-IV-TR criteria for ASD and has been used in several clinical studies. However, its basic psychometric characteristics such as reliability and validity have not been studied. In this study we analyze the data from 165 individuals that were seen in an autism multispecialty clinic whose caretakers completed the ASQ at least twice. In addition, concurrent to completing the ASQ, most caretakers also completed two other commonly used validated questionnaires used in autism research: the Social Responsiveness Scale (SRS) and the Aberrant Behavior Checklist (ABC). The data was analyzed using a mixed-model to control for within-subject variation and time between assessments. First, we found excellent test-retest reliability for the ASQ with statistically significant r values for the mixed-model. Reliabilities of 0.93, 0.94, 0.91 and 0.99 were found for the social, communication and stereotyped behavior subscales and total score, respectively. The Minimal Clinically Important Difference (MCID) was calculated as 0.11, 0.09, 0.12 and 0.04 points for the social, communication and stereotyped behavior subscales and total score, respectively. The r values for the mixed-model associating the ASQ with the SRS and ABC varied between subscales but were moderate to good in magnitude and statistically significant, demonstrating that the ASQ measures a valid psychometric construct. Overall, this study suggests that the ASQ has acceptable reliability. The ASQ's close correspondence to the DSM criteria is a strength of this instrument. With further research, it may be possible to develop a similar tool based on the DSM-V criteria. It appears that a Minimal Clinically Important Difference is approximately 0.1 point on any subscale, providing validation that changes on the ASQ can be considered important.

[N A J Med Sci. 2015;8(4):149-153. DOI: *10.7156/najms.2015.0804149]*

Key Words: *autism spectrum disorder, reliability, behavior*

INTRODUCTION

Many instruments have been developed to measure symptoms of Autism Spectrum Disorder (ASD), but only a few are directly based on the Diagnostic and Statistical Manual of Mental Disorders, 4th Edition, Text Revision (DSM-IV-TR) criterion. One of these tools called the Autism Symptoms Questionnaire (ASQ; The Center for Autism and Related Disorders, Inc., Tarzana, CA) is a checklist directly based on the DSM-IV-TR criteria for ASD and has been used in several clinical studies.[1,2]

The ASQ measures symptoms of (a) atypical social interactions using 12 questions, resulting in a score ranging from 0 to 4, (b) atypical communication using 15 questions, resulting in a score ranging from 0 to 5 and (c) stereotyped behavior using 7 questions, resulting in a score ranging from 0 to 4. In addition, all of the subscales can be added to calculate a total symptom composite score which ranges from 0 to13. Intervention autism studies suggest a 1.1 point change as clinically meaningful.[2] The ASQ has the advantage of having face validity since the questions included closely match the criteria outlined in the DSM-IV-TR. The ASQ is quick and easy to fill out as questions are answered as 'yes' or 'no' and can generate a suggested diagnosis of Pervasive Developmental Disorder - Not Otherwise Specified (PDD-NOS), Autistic Disorder, or Asperger Syndrome based on the pattern of answers.

Many other standard instruments for assessing ASD symptoms have undergone assessments of their reliability and validity. For example, the Aberrant Behavior Checklist (ABC) was designed to measure disruptive behaviors in individuals with developmental disabilities[3] and has been shown to have convergent and divergent validity in ASD[4] and has been used in multiple autism clinical trials.[5] The Social Responsiveness Scale (SRS) measures the severity of social skill deficits[6] which has been validated and shown to be reliable in several independent ASD samples across several cultures, including Chinese,[7] United Kingdom,[8]

Received: 09/03/2015; Revised: 10/09/2015; Accepted: 10/11/2015
***Corresponding Author:** Slot 512-41B, Room R4025, 13 Children's Way, Little Rock, AR 72202. Tel: 501-364-4662. Fax: 501-364-1648. (Email: REFrye@uams.edu)

Japanese,[9] Taiwanese,[10] and German[11,12] populations as well as American preschool children[13] and adolescents.[14] In addition, the SRS has been shown to have good correspondence to the gold-standard Autism Diagnostic Interview-Revised (ADI-R) while being much more time efficient and cost effective.[15]

Although the ASQ has been used in research and clinical settings, its basic psychometric characteristics such as reliability and validity have not been studied. This is important as such information is necessary to interpret the significance of previous research studies. In addition, since the DSM-V has been developed, it would be advantageous to understand whether the ASQ is a reliable and valid instrument in order to guide the possibility of developing a similar checklist based on the DSM-V. In this study we examine the test-retest reliability and validity of the ASQ in an American clinical sample of children with ASD and examine the changes in the ASQ that would represent a measurable and meaningful clinical change.

METHODS

Data used in this study was collected as part of our routine care for autism patients seen in the Arkansas Children's Hospital Autism Multispecialty Clinic between September 2012 and September 2015. Patients seen in the clinic are routinely requested to consent to allow their medical information to be anonymously abstracted into a clinical database that contained medical history, pertinent physical examination findings and the results of neurological and metabolic testing. Approximately 98% of parents who were approached signed the consent. This design is similar to our previous studies.[16,17]

As part of the intake process at each clinic appointment, caretakers of patients with ASD are asked to complete several standardized questionnaires including the ASQ, the SRS and the ABC as well as others, to monitor their on-going ASD symptoms. Compliance is high but not perfect. From our database we found 165 individuals who completed the ASQ at least twice, so test-retest reliability could be calculated. Of these 165 individuals, they attended from 2 to 4 clinic visits, with 163 completing the ASQ on the first clinic visit, 158 completing the ASQ on the 2nd clinic visit, 67 completing the ASQ on the 3rd clinic visit and 23 completing the ASQ on a 4th clinic visit.

In order to obtain baseline characteristics of the ASQ we calculated the mean and standard deviation (SD) for the subscale and total score across each individual and all visits. In order to calculate the correspondence between repeated ASQ scores and ASQ scores with other instruments a mixed-model was used to account for individual subject variation as well as the variation in time between completing the ASQ repeatedly. The 'glimmix' procedure of SAS 9.1 (SAS Institute Inc., Cary, NC) was used to analyze the data. The general mixed model is in matrix form

$$y = X\beta + Z\gamma + \varepsilon \quad (1)$$

where y is the dependent variable, which in this case is the ASQ score, X is the design matrix for the fixed effects, β is a vector containing the parameters of the fixed effects, Z is the design matrix for the random effects, γ contains the parameters of the random effects and ε is the variance-covariance matrix of the model error. The key assumption of the mixed model are that both γ and ε have the expected value of zero (i.e., $E(\gamma) = 0$ and $E(\varepsilon) = 0$) and known covariance structure given by the matrices Var (γ) and Var(ε).

For the test-retest analysis the values for each row of the fixed-effects design matrix X are given by

$$x(t, A, t) = [c\ A\ t] \quad (2)$$

where c is the constant with value 1, A is the repeated ASQ score and t is the time (in days) between the original and repeated score.

For the validity analysis examining the association between the ASQ score and the ABC and SRS scores, the values for each row of the fixed-effects design matrix X are given by

$$x(t, A) = [c\ S] \quad (3)$$

where c is the constant with value 1, S is the score from the ABC or SRS scales.

The values for each row of the random-effects design matrix Z are given by

$$z(p, t) = [c_1 \ldots c_p] \quad (4)$$

where c_i is 1 for participant i and 0 otherwise where i is the participant index going from 1 to p and p is the maximum number of participants.

To provide an index of the association between variables in the model we calculated the correlation coefficient, r, for the model as the square-root of the r-squared define as:

$$r^2 = 1 - (\text{sum of squares residual/sum of squares total}) \quad (5)$$

We calculated with test-retest reliability as the correlation coefficient derived from the mixed-model using the X matrix derived from Equation (2). Each possible pair of clinic visits was entered as independent values, although the subject variable in the mixed-model accounted for the repeated use of subjects. We interpreted the correlation coefficient by convention for test-retest reliability such that r >= 0.9 is considered excellent, r >= 0.80 and < 0.9 is considered good, r >= 0.7 and < 0.8 is considered acceptable, r >= 0.6 and < 0.7 questionable, r >= 0.5 and < 0.6 is poor and an r < 0.50 is considered unacceptable.

From this information the Minimal Clinically Important Difference (MCID) was calculated using the standard error of measurement (SEM) method given as

$$\text{SEM} = \text{Standard Deviation} * \text{sqrt}(1 - \text{r}) \quad (6)$$

For 406 visits, information was available for the SRS and/or ABC as well as the ASQ. The ABC is a 58-item validated questionnaire that measures disruptive behaviors in individuals with developmental disabilities[3] across five dimensions: Irritability (15 items, range 0-45); Social Withdrawal (16 items, range 0-48); Stereotypy (7 items, range 0-21); Hyperactivity (16 items, range 0-48) and Inappropriate Speech (4 items, range 0-12). Each item is rated 0 to 3 with higher scores indicating greater severity. The SRS is a 65-item questionnaire that measures the severity of social skill deficits across five domains: Social Awareness (8 items), Social Cognition (12 items), Social Communication (22 items), Social Motivation (11 items), Autistic Mannerisms (12 items) and total (65 items).[6] Each item is rated 0 to 3 with higher scores indicating greater severity. Standardized T-scores (mean 50, standard deviation 10) range 30-90. To help determine validity of the ASQ, the mixed-model calculated the association between the ASQ subscales and total score with the scales of the other instruments using the X matrix in Equation (3). The correlation coefficient for the association was calculated as the square root of Equation (5).

Table 1. Average and standard deviations (SD) of Autism Symptoms Questionnaire (ASQ) values for each patient visit and overall of all ASQ questionnaires completed.

	Visit 1 (n=163)		Visit 2 (n=158)		Visit 3 (n=67)		Visit 4 (n=23)		All Visits (n=411)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Social	1.9	1.4	2.7	1.4	3.1	1.1	3.5	1.0	2.5	1.5
Communication	3.5	1.6	3.8	1.4	4.0	1.2	4.3	1.2	3.8	1.5
Stereotyped Behavior	2.3	1.3	2.8	1.2	3.2	1.0	3.3	1.2	2.7	1.3
Total	7.7	3.6	9.4	3.4	10.2	2.9	11.2	2.8	9.0	3.5

Table 2. Results from the mixed model. Results of model F-tests for individual factors (second rating, time). Test-Retest reliability as determined by the r statistic of the mixed-model examining the association between Autism Symptoms Questionnaire (ASQ) scales taking into account time between test-retest responses and individual subject variation. Using the r statistic and the standard deviation in Table 1, the Minimal Clinically Important Difference is calculated using the standard error of measurement method.

	Repeat Rating		Time		Reliability (r statistic)	Minimally Clinically Important Difference
	F-test	p	F-test	P	r	
Social	17.11	<0.0001	17.15	<0.0001	0.93	0.11
Communication	8.96	0.003	8.91	0.003	0.94	0.09
Stereotyped Behavior	21.82	<0.0001	16.60	<0.0001	0.91	0.12
Total	20.28	<0.0001	21.99	<0.0001	0.99	0.04

RESULTS

Table 1 outlines the means and standard deviations for each individual visit as well as the overall means and standard deviations for all ASQs completed. **Table 2** outlines the test-retest reliability correlations between ASQs completed on the various visits. The table outlines the significance of the two fixed effects, the repeated rating and time, both of which were significant in all cases. The correlation coefficient was large in all cases, suggesting excellent test-retest reliability. On average there was 412 (SD = 219) days between clinical visits.

From the test-retest reliability we were able to calculate the MCID, which was approximately 0.1 points for social, communication and stereotyped behavior subscales and approximately 0.04 points for the total score.

In order to investigate construct validity, we examined the correlation between the ASQ and the SRS and ABC. Although the majority of the correlations were highly significant, the correlations were small to moderate with particularly large correlations only between the ASQ Communication subscale and the subscales of the ABC. Overall the ASQ correlated best with the ABC (average correlation 0.72, SD 0.08) as compared to the SRS (average correlation 0.65, SD 0.08). The ASQ Communication subscale correlated best with ABC (average correlation 0.80, SD 0.01) and SRS (average correlation 0.77, SD 0.02) subscales, showing a good association. The ASQ Stereotyped Behavior demonstrated slightly less strong correlation with the ABC (average correlation 0.74, SD 0.01) and SRS (average correlation 0.67, SD 0.02) subscales. Of the ASQ subscales, the ASQ Social demonstrated that lowest association overall with the ABC (average correlation 0.59, SD 0.05) and SRS (average correlation 0.54, SD 0.03) subscales. The Total ASQ score correlated moderately with the ABC (average correlation 0.71, SD 0.03) and SRS (average correlation 0.62, SD 0.04) overall.

Table 3. Correlation coefficients derived from the mixed-model regression examining the association between Autism Symptoms Questionnaire (ASQ) subscales and total score (columns) and T Scores from the Social Responsiveness Scale (SRS) subscales and total (rows) from 406 cases. Both scales were completed at the same clinical visit. All correlations are significant $p < 0.0001$.

	ASQ Social	ASQ Communication	ASQ Stereotyped Behavior	ASQ Total
SRS Awareness	0.54	0.77	0.72	0.65
SRS Cognition	0.57	0.77	0.69	0.63
SRS Communication	0.50	0.75	0.69	0.58
SRS Motivation	0.55	0.79	0.69	0.65
SRS Mannerisms	0.58	0.78	0.67	0.64
SRS Total	0.51	0.75	0.66	0.57

Table 4. Correlation coefficients derived from the mixed-model regression examining the association between Autism Symptoms Questionnaire (ASQ) subscales and total score (columns) and Aberrant Behavior Checklist subscales (rows) from 405 cases. Both scales completed for the same clinical visit. All correlations are significant $p < 0.0001$.

	ASQ Social	ASQ Communication	ASQ Stereotyped Behavior	ASQ Total
ABC Irritability	0.65	0.81	0.74	0.74
ABC Social Withdrawal	0.53	0.80	0.74	0.69
ABC Stereotyped Behavior	0.58	0.79	0.75	0.69
ABC Hyperactivity	0.62	0.81	0.75	0.73
ABC Inappropriate Speech	0.65	0.81	0.76	0.74
ABC Total	0.59	0.80	0.72	0.69

DISCUSSION

Overall this study demonstrates that the ASQ has excellent test-retest reliability when the time between the testing and intrasubject variation is taken into account. Children with ASD can be highly variable in their behavior, so it is important to use an instrument that can repeatedly measure the symptoms reliably. Despite these positive results further research should explore more detailed measures of reliability and validity such as examining Cronbach's alpha internal consistency and examining the diagnostic association of this test with gold standards such as the ADI-R or the Autism Diagnostic Observation Scale (ADOS).

We also examined construct validity by examining the correlation of the ASQ with other validated ASD scales. ASQ Subscales correlated with similar subscales on the ABC and SRS, suggesting that they are measuring similar, but potentially distinct, variations of the same construct. Further examination of the ASQ subscales with other methods, such as factor analysis, may help reveal distinct constructs that they are measuring.

Interestingly, the ASQ Communication subscale was found to correlate best with scores on the ABC and SRS instruments. This is important as development of language and communication skills in childhood is associated with favorable long-term outcomes in children with ASD[18-20] and the development of verbal communication skills is closely linked to the quality of life of the parent[21] and is closely associated with ASD severity.[22] Thus, communication is an important core ASD symptom that probably has wide effects on many other core and associated ASD symptoms.

The ASQ is designed to provide a suggested diagnosis based on the pattern of answers and is not particularly designed to examine severity. This may have led to some of its limitations psychometrically speaking. Reassuringly, the MCID was found to be a change of approximately 0.1 point on an ASQ subscale, suggesting that changes in the ASQ on repeated measurements may have some clinical importance.

Further research into the factors which result in variation in the ASQ as well as the distinct constructs that it may be measuring may lead to the development of an improved instrument which could be designed around the new DSM-V criteria for ASD.

ABBREVIATIONS

ABC: Aberrant Behavior Questionnaire; ASD: autistic spectrum disorder; ASQ: autism symptoms questionnaire; MCID: Minimally Clinically Important Difference; SRS: Social Responsiveness Scale

CONFLICT OF INTEREST

The author has no conflicts of interest to declare.

FINANCIAL DISCLOSURES

The author has no financial disclosures to declare.

REFERENCES

1. Frye RE, Delatorre R, Taylor H, et al. Redox metabolism abnormalities in autistic children associated with mitochondrial disease. Transl Psychiatry. 2013;3:e273.
2. Frye RE, DeLatorre R, Taylor HB, et al. Metabolic effects of sapropterin treatment in autism spectrum disorder: a preliminary study. Transl Psychiatry. 2013;3:e237.
3. Aman MG, Singh NN, Stewart AW, Field CJ. The aberrant behavior checklist: a behavior rating scale for the assessment of treatment effects. Am J Ment Defic. 1985;89:485-491.
4. Kaat AJ, Lecavalier L, Aman MG. Validity of the aberrant behavior checklist in children with autism spectrum disorder. J Autism Dev Disord. 2014;44:1103-1116.
5. Frye RE, Slattery J, MacFabe DF, et al. Approaches to studying and manipulating the enteric microbiome to improve autism symptoms. Microb Ecol Health Dis. 2015;26:26878.

6. Constantino JN. The Social Responsiveness Scale. Los Angeles: Western Psychological Services; 2002.
7. Zhou H, Zhang L, Wu L, et al. Validity and reliability analysis of the Chinese parent version of the Autism Spectrum Rating Scale (6-18 years). Psychiatry Res. 2015. [epub]
8. Wigham S, McConachie H, Tandos J, Le Couteur AS. The reliability and validity of the Social Responsiveness Scale in a UK general child population. Res Dev Disabil. 2012;33:944-950.
9. Takei R, Matsuo J, Takahashi H, Uchiyama T, Kunugi H, Kamio Y. Verification of the utility of the social responsiveness scale for adults in non-clinical and clinical adult populations in Japan. BMC Psychiatry. 2014;14:302.
10. Wang J, Lee LC, Chen YS, Hsu JW. Assessing autistic traits in a Taiwan preschool population: cross-cultural validation of the Social Responsiveness Scale (SRS). J Autism Dev Disord. 2012;42:2450-2459.
11. Bolte S. Brief Report: the Social Responsiveness Scale for Adults (SRS-A): initial results in a German cohort. J Autism Dev Disord. 2012;42:1998-1999.
12. Bolte S, Poustka F, Constantino JN. Assessing autistic traits: cross-cultural validation of the social responsiveness scale (SRS). Autism Res. 2008;1:354-363.
13. Duku E, Vaillancourt T, Szatmari P, et al. Investigating the measurement properties of the social responsiveness scale in preschool children with autism spectrum disorders. J Autism Dev Disord. 2013;43:860-868.
14. Pearl AM, Murray MJ, Smith LA, Arnold M. Assessing adolescent social competence using the Social Responsiveness Scale: should we ask both parents or will just one do? Autism. 2013;17:736-742.
15. Murray MJ, Mayes SD, Smith LA. Brief report: excellent agreement between two brief autism scales (Checklist for Autism Spectrum Disorder and Social Responsiveness Scale) completed independently by parents and the Autism Diagnostic Interview-Revised. J Autism Dev Disord. 2011;41:1586-1590.
16. Frye RE, Melnyk S, Macfabe DF. Unique acyl-carnitine profiles are potential biomarkers for acquired mitochondrial disease in autism spectrum disorder. Transl Psychiatry. 2013;3:e220.
17. Frye RE. Biomarkers of Abnormal Energy Metabolism in Children with Autism Spectrum Disorder. N A J Med Sci. 2012;5:141-147.
18. Mukaddes NM, Tutkunkardas MD, Sari O, Aydin A, Kozanoglu P. Characteristics of children who lost the diagnosis of autism: a sample from istanbul, Turkey. Autism Res Treat. 2014;2014:472120.
19. Tager-Flusberg H, Rogers S, Cooper J, et al. Defining spoken language benchmarks and selecting measures of expressive language development for young children with autism spectrum disorders. J Speech Lang Hear Res. 2009;52:643-652.
20. Luyster R, Qiu S, Lopez K, Lord C. Predicting outcomes of children referred for autism using the MacArthur-Bates Communicative Development Inventory. J Speech Lang Hear Res. 2007;5:667-681.
21. Tilford JM, Payakachat N, Kovacs E, et al. Preference-based health-related quality-of-life outcomes in children with autism spectrum disorders: a comparison of generic instruments. PharmacoEconomics. 2012;30:661-679.
22. Bavin EL, Kidd E, Prendergast L, Baker E, Dissanayake C, Prior M. Severity of autism is related to children's language processing. Autism Res. 2014;7:687-694.