ORIGINAL PAPER

Are Autistic Traits Measured Equivalently in Individuals With and Without An Autism Spectrum Disorder? An Invariance Analysis of the Autism Spectrum Quotient Short Form

Aja L. Murray · Tom Booth · Karen McKenzie · Renate Kuenssberg · Michael O'Donnell

Abstract It is common to administer measures of autistic traits to those without autism spectrum disorders (ASDs) with, for example, the aim of understanding autistic personality characteristics in non-autistic individuals. Little research has examined the extent to which measures of autistic traits actually measure the same traits in the same way across those with and without an ASD. We addressed this question using a multi-group confirmatory factor invariance analysis of the Autism Quotient Short Form (AQ-S: Hoekstra et al. in J Autism Dev Disord 41(5): 589–596, 2011) across those with (n = 148) and without (n = 168) ASD. Metric variance (equality of factor loadings), but not scalar invariance (equality of thresholds), held suggesting that the AQ-S measures the same latent traits in both groups, but with a bias in the manner in which trait levels are estimated. We, therefore, argue that the AQ-S can be used to investigate possible causes and consequences of autistic traits in both groups separately, but caution is due when combining or comparing levels of autistic traits across the two groups.

A. L. Murray (✉) · T. Booth
Centre for Cognitive Ageing and Cognitive Epidemiology,
Department of Psychology, University of Edinburgh,
7 George Square, Edinburgh EH8 9JZ, UK
e-mail: s0785823@sms.ed.ac.uk

K. McKenzie
Clinical Psychology Department, University of Edinburgh,
Edinburgh, UK

R. Kuenssberg
NHS Fife, Kirkcaldy, UK

M. O'Donnell
Department of Psychology, University of Edinburgh,
Edinburgh, UK

Introduction

Formally, autism spectrum disorders (ASDs)—though considered highly heterogeneous—are defined with respect to three main characteristics: deficits in social interaction, deficits in communication, and a restricted behavioural repertoire (APA 2000). Intellectual disability, physical disorders and sensory symptoms are also commonly associated with ASDs (Matson et al. 2011; Matson and Shoemaker 2009; Rogers et al. 2003).

Increasingly, and for a number of different reasons, researchers are measuring autistic traits in those without ASD. One motivation is the argument that such research has the potential to enhance the understanding of the determinants and consequences of clinical ASD. Clinical samples can be limited in size due to the relatively low frequency of cases and including non-clinical cases can increase statistical power by capturing some sub-clinical variability in autistic characteristics (e.g. Lundström et al. 2012; Sung et al. 2005). For example, measures of the 'broader autism phenotype' or of 'autistic-like traits' aim to capture subclinical characteristics or traits which are related to a genetic liability to ASDs and may, therefore, assist in identifying the genetic determinants of ASD (Wheelwright et al. 2010).

Autistic traits in those without ASD may also be interesting in their own right. Several studies have examined the relation between autistic traits in such populations to personality traits and interests (Austin 2005; Baron-Cohen et al. 2001; Kunihira et al. 2006). One line of research in

this paradigm has led to the development of the Autism Quotient Questionnaire (AQ) and its variants (e.g. Baron-Cohen et al. 2006; Hoekstra et al. 2011) which aim to quantify an individual's position on an assumed continuum from ASD to normality (Baron-Cohen et al. 2001).

A number of authors have explicitly addressed the hypothesis that individuals with an ASD are simply those who score at the extreme end of a normal distribution of traits and abilities related to social adaptation and communication (Constantino and Todd 2003). A distinction is made between clinical autism and autistic like traits (ALTs) in the general population, but the distinction is one of degree and not kind. The key difference is that those with a clinical diagnosis of ASD cross a threshold beyond which functioning is significantly impaired, whereas those high on ALTs but without a clinical diagnosis of ASD remain able to cope with the social interaction demands of society (Lundström et al. 2012). This hypothesis has been addressed by examining the latent distribution of ASD measures across those with and without ASD using taxometric or latent class analysis methods, with the evidence at present pointing to ASD as a distinct category from ALTs (Frazier et al. 2010, 2012). Again, this line of research depends on measuring autistic traits in individuals without a clinical diagnosis of ASD.

Accurately identifying those who fall into the ASD category has also been a prominent concern among clinicians and researchers. Studies seeking to develop new psychometric tools for identifying individuals with ASD have utilised samples without ASD to evaluate the discriminative power of the instruments (e.g. Allison et al. 2011; Baron-Cohen et al. 2001). Such studies allow the utility of candidate screening tools to be assessed in non-clinical populations, where they may act as a 'red flag' to assist frontline health professionals in making the decision whether to refer an individual for full diagnostic assessment for an ASD (Allison et al. 2011).

An important concept when utilising a single inventory across different groups, e.g. individuals with and without ASD, is measurement equivalence (Kim and Yoon 2011). Briefly, tests of measurement equivalence assess the hypothesis that the constructs being measured by an inventory have the same meaning and measurement properties across the groups. There are several available methods for testing the measurement equivalence of an instrument across groups in both confirmatory factor analytic (CFA) and item response theory (IRT) frameworks (Kim and Yoon 2011). When several layers of constructs are of interest i.e. the instrument has a higher-order structure, then multi-group CFA approaches provide an efficient means of gauging measurement invariance at all levels of this higher-order structure. This has the advantage of allowing specific loci of a lack of invariance to be identified e.g. in first or second-order traits. This is of relevance when assessing invariance of a measure of autistic traits because it is thought that these traits form a hierarchy including at least two levels of generality (Hoekstra et al. 2011). Work by Hoekstra et al. (2011) has supported the existence of both first-order autistic traits and a second-order 'Social Behaviour' construct.

Measurement equivalence can be assessed at a number of levels. Depending on the purpose of measuring autistic traits in individuals without ASD, different levels of measurement equivalence are required in order for the conclusions drawn to be valid. The importance of achieving different levels of measurement equivalence for different types of analyses has in general been widely discussed (e.g. see Borsboom 2006), however, the implications for ASD research has received less attention. In the present study, we outline the levels of equivalence required for the above outlined applications and then proceed to test this in a measure of ASD commonly used in both those with and without ASD.

Although stronger forms of equivalence may be needed for some applications of ASD measures, their most common uses rely only on metric invariance. Metric invariance implies that the same latent traits underlie the observed variables in both groups. Metric invariance is, therefore, important if the aim is to make inferences about the causes and consequences of autistic traits from autistic traits in the general population to ASDs in the clinical ASD population. In addition, in combined analyses of two groups measured by a test that does not exhibit metric invariance across these groups, differences in factor loadings can result in the appearance of additional non-substantive factors (Meredith and Teresi 2006). This has the potential to confuse attempts to identify the dimensionality of autistic traits and any putative correlates of identified dimensions.

Scalar invariance may also be desirable in some cases. Scalar invariance is required whenever group means are to be compared, otherwise observed mean differences or lack thereof may not represent the extent of difference in the latent trait of interest (Borsboom 2006). For example, group differences in scores on an ASD measures have been used as an index of the discriminative power of that test, however, without assessing scalar invariance, it is not known how this observed difference corresponds to mean differences on the latent traits measured by such test (e.g. Baron-Cohen et al. 2001). Similarly, it is possible that a test of ASD traits used to screen for or diagnose ASD could systematically under- or over-estimate the scores of individuals with ASD relative to individuals without ASD if the test is biased, thus reducing its discriminative power. Assessing scalar invariance can evaluate this possibility. Finally, any aggregated factor analyses of scores of those with and without ASD in which scalar invariance does not hold can result in additional factors appearing due to the

differences in thresholds across the groups (Meredith and Teresi 2006). Again, where the dimensionality of ASD traits may be of interest, this could produce misleading results.

It is important to test empirically, and not simply assume, measurement equivalence across groups of people with and without ASD, as there are potentially a number of reasons to think that it could be violated. For example, particular items may be less relevant, or have a different meaning or interpretation to individuals without, as compared to individuals with, a clinical diagnosis of ASD. In addition, ASD is associated with difficulties in self-perception, such that there may be systematic under-reporting or possible increased inaccuracies in any self-report measure of autistic traits in individuals with ASD (Johnson et al. 2009). To our knowledge, no study has previously examined measurement invariance of a test of autistic traits across individuals with and without ASD. It was, therefore, the aim of the present study to test for invariance of a popular measure of autistic traits across individuals with a clinical diagnosis of ASD and control individuals without a diagnosis of ASD.

Methods

Participants

ASD Group

We utilised archival data from case notes on a total of 148 participants with a diagnosis of Asperger syndrome (AS) or high functioning autism (HFA), who had previously completed the full AQ (Baron-Cohen et al. 2001). HFA was defined as meeting the criteria for Autism but having normal intellectual function while AS was defined as meeting the criteria for HFA but with no history of language delay (Baron-Cohen and Wheelwright 2004). The sample included 107 males and 41 females. The age range of the sample was 17–62 with a mean of 33.3 (SD = 10.7). For a comprehensive description of this sample, including information on data collection and the diagnostic process please see Kuenssberg et al. (2012).

Non-ASD Control Group

We recruited a control group of 166 participants from a large university community and from social networking sites. Information was collected via an online questionnaire, which included both the full AQ (Baron-Cohen et al. 2001) and questions on the demographic characteristics of participants. The control sample included 40 male and 126 female participants. The age of the control sample ranged from 17 to 65 with a mean of 30.1 (SD = 11.30).

Measures

The AQ-S is a short form of the full AQ (Baron-Cohen et al. 2001) motivated by the benefits of a shorter form for use in large scale studies where the full AQ may be too lengthy. Hoekstra et al. (2011) arrived at the structure and content of the AQ-S using item selection and validation analyses from the full AQ in a sample of individuals with ASD and 2 control populations. The AQ was designed to capture the core dimensions of autistic traits in adults with normal intelligence. The AQ-S includes 28 of the 50 original AQ items and retains its broad dimensionality. Thus, the items measure the domains of: social skills, routine, switching, imagination, and numbers/patterns and a higher-order social behaviour factor defined by the first four of these five factors. In ASD and control samples, CFA analyses have generally found reasonable fit of this structure for the AQ-S (Kuenssberg et al. 2012; Hoekstra et al. 2011).

In the present study, both our ASD sample and control sample completed the full AQ, but we selected only the AQ-S items for analysis. Each item of the AQ-S has four response options from 'definitely agree to definitely disagree'. Half of these items are reverse keyed and we rescored these items so, for all items, higher scores indicate a higher degree of autistic traits. Although previous analyses of the full AQ have tended to use a dichotomous scoring system, we elected to use the full four-point scale for our analyses because multi-group CFA with dichotomous items has lower power and requires larger sample sizes to detect non-invariance (Kim and Yoon 2011).

Statistical Analysis

Data Screening

As a check on the appropriateness of the data for our planned statistical analysis, we examined data skewness, missingness and communalities.

Measurement Invariance Analysis

We estimated a series of CFA models in order to first assess the fit of the proposed structure of the AQ-S in both the ASD and control groups. We also fit these same CFA models to only the males and only the females in the sample. This was to assess the possibility that sex differences in autistic traits contributed to any non-invariance observed across the ASD and control samples (e.g. see Rivet and Matson 2011).

Next, we estimated multi-group CFA models to assess measurement equivalence. Multi-group CFA for the analysis of measurement equivalence can be implemented using either a forwards or backwards approach. In the forwards approach,

the baseline model is the least constrained model and cross-group equality constraints are added successively. In the backwards approach, the baseline model is fully constrained and mis-specified constraints successively released. Here we employ a forward selection method for our analyses. In assessing the equivalence of a measure of autism, the forwards multi-group CFA approach is most appropriate because it allows the less strict assumption of factorial or metric invariance (equality of factor loadings across groups) to be tested before proceeding to a test of scalar invariance (equality of factor loadings and thresholds across groups) and the even stronger assumption of strict invariance (equality of factor loadings, thresholds and item residual variances). Forward selection has been shown to detect differential item functioning with greater accuracy and to be less prone to the substantial type I error produced by backward selection (Stark et al. 2006).

Here we followed the sequence of analyses suggested by Chen et al. (2005) for high-order structures. We first tested for configural equivalence across the ASD and control groups for the complete second-order structure. Next, metric equivalence constraints were added to the first, then second order factor loadings. Finally, scalar equivalence constraints were added to the first, then second order thresholds and intercepts.

Model Estimation

All models were estimated in Mplus 6.11 (Muthén and Muthén 2010) using weighted least squares means and variances estimation (theta parameterization) to account for the categorical measurement scale of the questionnaire (Rhemtulla et al. 2012).

Model Specification

Single Group CFAs As a preliminary test of the appropriateness of invariance analysis in our data, we individually fit single group CFA models to both the first and second order proposed structure of the AQ-S in the ASD and control groups (Table 3). We fit the model suggested by Hoekstra et al. (2011) in the original validation study for the AQ-S and depicted in Fig. 1. For the purposes of scaling and identification, we fixed the variance of the latent factors to be 1.0.

MG-CFA Configural Invariance The configural invariance model (Table 4, M1) provides the baseline model for all subsequent analyses (see section "Model Evaluation"). Here, factor loadings and item thresholds are free to vary across both groups. Factor variances are again fixed to 1.0 across both groups for identification. Latent factor means were fixed to zero in both groups, while item residuals were fixed to 1.0 in both groups.
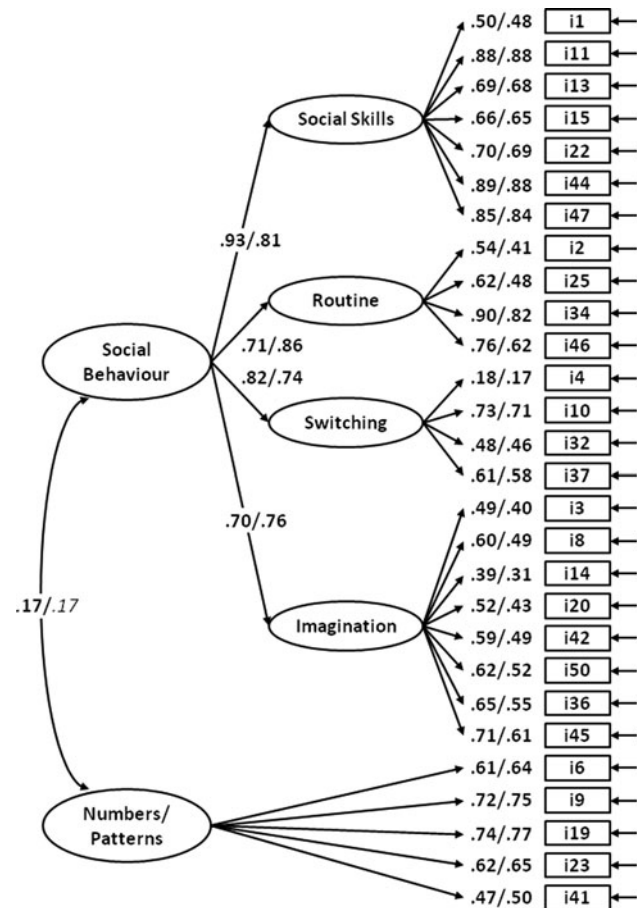

Fig. 1 Structural diagram and standardized parameter estimates for the second order metrically invariant model (M3, Table 5). *i* item. Item numbers correspond to the numbers from the original AQ. Parameter estimates are presented for both the ASD group (*on the left hand side of the forward slash*) and the non-ASD group (*right hand side of the forward slash*)

MG-CFA Metric Invariance Metric invariance was assessed in both the 1st-order factors (Table 4, M2) and 2nd-order factor (Table 4, M3), by constraining the loadings to equivalence across groups.

MG-CFA Item Threshold Invariance Next, equivalence constraints were placed on the thresholds of the indicator items across groups (Table 4, M4). The AQ-S has a four point response scale resulting in three thresholds per item. In this model, the factor means of the first order factors were freely estimated in the second group (here the control group), but remained fixed at 0 and 1.0 respectively in the reference group (here the ASD group).

Model Evaluation

Model fit was evaluated using comparative fit index (CFI), Tucker-Lewis Index (TLI), Root Mean Square Error of

Approximation (RMSEA) and Weighted Root Mean Square Residual (WRMSR). We judged fit to be good when CFI and TLI values were >.90–.95, WRMSR values were <.90, and RMSEA values were <.08 (Hu and Bentler 1999; Schermelleh-Engel et al. 2003; Yu 2002). Though it is often desirable to compare models, especially non-nested models, based on the Akaike and Bayesian Information Criteria, these are not available in the current analyses as we utilise a item level estimation technique (WLSMV), which does not maximize the log-likelihood required to calculate AIC and BIC.

In testing measurement invariance, the assumptions of invariance are assumed to hold if the additional constraints placed on models yield negligible changes in model fit. There is no uniformly agreed upon criteria for judging change in model fit within invariance models. Here we rely on the studies of Vandenberg and Lance (2000) and Chen (2007) and suggest drop in fit of −.010 for TLI, −.005 for the CFI, alongside a change in RMSEA of .010, to be indicative of substantive decline in model fit, and thus, that the assumptions of the invariance constraints are violated.

Results

Descriptive Statistics

Substantial skew and kurtosis were evident in both the ASD (±2.43 and ±7.22 respectively) and control (±1.27 and ±1.24 respectively) groups. However, WSLMV is robust to non-normality, and as such, the levels of skew and kurtosis were not deemed problematic. Item communalities in both the ASD (range = .24–.60, mean = .40) and control (range = .21–.75, mean = .44) suggested the data were suitable for factor analysis.

In both samples, the percentage of missing data was low (ASD = .27 %; control: .84 %). Less than 5 % missingness is generally considered unproblematic (Tabachnick and Fidell 2007).

Subscale Means

Subscale means and standard deviations first split by ASD versus control and then by males and females are provided in Table 1. For the social skills, routine, and switching factors means were higher in ASD relative to controls and in males relative to females. For the imagination and numbers/patterns subscale, however, there was essentially no observed mean difference between ASD and control and between males and females (see section "Discussion").

Table 2 reports the correlations of the five subscales scores of the AQ-S. Given the same sizes in the current study, differences in Pearson's correlations of approximately .20

would be indicative of a significant difference. Consideration of the estimates in Table 2 suggests that although correlations between subscales show some variation across groups, these variations do not reach a threshold for being statistically significantly different. However, it is important to note that this may be a reflection of the size of the current samples.

Single Group CFAs

Fit indexes for the single group CFA analyses are provided in Table 3. Model fit for the first-order and second-order models in the ASD and control were acceptable based on the lower limits of model fit, and in line with model fits reported in previous studies. Fit in the ASD group was slightly better than that of the control group. The inclusion of the second-order factor in both the ASD and control groups resulted in a small decrease in model fit at the third decimal place in the CFI, TLI and RMSEA. However, given the magnitude of the factor correlations (ranging from $r = .46$ to .77; see also the sum score correlations in Table 1), which indicate the plausibility of a second-order factor, and past research on both the AQ and AQ-S suggesting its inclusion, the small differences in fit were not deemed substantive enough to support rejecting the second-order model. Modification indices (MI) and expected parameter changes (EPC) suggested the presence of some cross-loadings and correlated residuals in both samples and pointed to item complexity as an important source of mis-fit in the model. In the ASD group, the largest MIs and EPCs were associated with cross loadings of item 41 on social behaviour (MI = 16.13, EPC = .26) and item 2 (MI = 15.71, EPC = .43) on numbers and patterns. In the control group the largest MIs and EPCs were again associated with cross loadings of item 41 on social behaviour (MI = 20.65, EPC = .36) as well as item 46 on Social Skills (MI = 16.80, EPC = .72). Given that fit was well within acceptable levels and because we wished to avoid capitalising on chance (e.g. McDonald and Ho 2002) we did not make any post hoc modifications to our models based on these MIs and EPCs.

In the male, female and whole sample groups, however, the magnitude of the fit statistics for the first-order models was poor (and we, therefore, did not attempt to fit a second-order model). A possible explanation for this was that differences between the ASD and control groups (i.e. non-invariance) meant that whenever a group contained both ASD and control individuals, there was a break down in factor structure (e.g. see Meredith and Teresi 2006). We tested this possibility by fitting single group CFAs to two samples to which participants were randomly assigned and which were, therefore, heterogeneous with respect to both sex and ASD versus control diagnosis. Table 4 provides details of the sex and ASD

Table 1 AQ-S subscale means and standard deviations for sample sub-groups

	ASD	Control	Male			Female		
			All	ASD	Control	All	ASD	Control
Social skills	3.25 (.63)	2.70 (.69)	3.12 (.66)	3.24 (.63)	2.78 (.62)	2.83 (.74)	3.29 (.62)	2.67 (.71)
Routine	3.36 (.57)	2.55 (.65)	3.15 (.68)	3.34 (.57)	2.63 (.67)	2.75 (.72)	3.41 (.52)	2.53 (.65)
Switching	3.42 (.53)	2.58 (.55)	3.13 (.69)	3.39 (.54)	2.44 (.58)	2.83 (.64)	3.48 (.50)	2.65 (.53)
Imagination	2.93 (.63)	3.02 (.45)	2.91 (.59)	2.91 (.63)	2.95 (.45)	3.03 (.50)	3.00 (.64)	3.04 (.45)
Numbers and patterns	2.75 (.76)	2.73 (.70)	2.71 (.73)	2.74 (.75)	2.63 (.68)	2.78 (.73)	2.78 (.79)	2.77 (.71)

Means and standard deviations are based on complete cases

Table 2 Subscale inter-correlations for ASD, control, males and females

	1	2	3	4	5
ASD/control					
1. Social skills	–	.56**	.30**	.41**	.16
2. Routine	.43**	–	.31**	.28**	.13
3. Switching	.42**	.26**	–	.29**	−.04
4. Imagination	.51**	.37**	.41**	–	.10
5. Numbers and patterns	.09	.30**	.14	.16	–
Males/females					
1. Social Skills	–	.61**	.48**	.39**	.10
2. Routine	.54**	–	.51**	.24**	.19*
3. Switching	.49**	.49**	–	.30**	.02
4. Imagination	.46**	.29**	.25**	–	.10
5. Numbers and patterns	.17*	.22**	.08	.16	–

Note In the upper panel, ASD group correlations are shown below the diagonal and control group correlations are above the diagonal. In the lower panel, male correlations are shown below the diagonal and female correlations are shown above the diagonal

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

status of the random groups as compared to the whole sample. In addition, we fit single group CFAs to the sample of females without ASD and the sample of males with ASD in order to test the model on samples which were homogeneous with respect to both sex and ASD diagnosis. The purpose of these additional analyses was to help us to identify the source of mis-fit in the sex homogeneous, ASD heterogeneous samples.

The model fit well in the male group excluding those without an ASD diagnosis and in the female group excluding those with an ASD diagnosis. In the groups mixed with respect to sex and ASD diagnosis, fit was poor and similar to that in the single sex groups which were heterogeneous with respect to ASD diagnosis. This pointed to differences between individuals with ASD and controls as the source of mis-fit in the mixed groups and suggested

that we would observe non-invariance between ASD and control groups in subsequent stages of analysis.

Multi-Group CFA Measurement Invariance

Next, we sequentially tested for measurement invariance across ASD and control groups. The multi-group configurally invariant model (Table 5, M1) showed reasonable fit to the data and provided the initial baseline model for model fit comparisons. The assumption of factor loading or metric invariance was supported in the first-order model (Table 5, M2). Additional constraints on the second order factor loadings (Table 5, M3; see also Fig. 1) yielded a difference in CFI of −.011, above the suggested difference value of −.005. However, considered alongside the changes in TLI and RMSEA, invariance of the second order factor loadings was considered to hold. Therefore, the assumptions of factor loading, or metric invariance, were deemed to be reasonable for the AQ-S in both the first and second order factors across ASD and control groups.

We then proceeded to assess scalar invariance. The difference in model fit statistics between models M3 and M4 suggested that the assumption of scalar invariance of the indicator variables did not hold (ΔCFI = −.18; ΔTLI = −.16; ΔRMSEA = .032). We investigated the source of misfit based on the MI. However, after twelve individual item thresholds had been released, model fit was still poor, and so model modification was terminated. All thresholds for items 3, 4, 23 and 46 were released suggesting these items performed very differently across groups. Considering the response counts for these items confirmed this. For example, for item 4 (I frequently get so strongly absorbed in one thing that I lose sight of other things), the distribution of responses across the four points of the response scale were 2, 2, 30 and 113 in the ASD group, and 60, 66, 30, 10 in the control group.

As the focus of the current study was not on mean differences in latent factor scores, and the fact that we were unable to achieve first-order scalar invariance, we did not continue to fix the intercepts of the second order factors.

Table 3 Model fit statistics for single group models

Model	Adjusted χ^2	df	Signif.	CFI	TLI	RMSEA	WRMR
Single ASD status groups							
1st-order model—ASD group	448.239	340	$p < .001$	.920	.912	.046	.939
2nd-order model—ASD group	464.257	345	$p < .001$	.912	.904	.048	.976
1st-order model—control group	545.769	340	$p < .001$	.902	.891	.060	1.108
2nd-order model—control group	565.154	345	$p < .001$	.895	.885	.062	1.149
Single sex groups							
1st-order model—males (n = 147)	693.808	340	$p < .001$	.789	.766	.084	1.307
1st-order model—females (n = 167)	804.252	340	$p < .001$	.809	.788	.090	1.469
Single sex, single ASD status groups							
1st-order model—male with ASD (n = 107)	426.199	340	$p < .001$	.913	.903	.049	.908
1st-order model—female without ASD (n = 126)	493.769	340	$p < .001$	.913	.903	.060	1.042
Random groups mixed sex and ASD status							
1st-order model—random group 1 (n = 151)	916.713	340	$p < .001$	.711	.679	.106	1.586
1st-order model—random group 2 (n = 163)	781.930	340	$p < .001$	.822	.803	.089	1.408
Full sample							
1st-order model (n = 314)	1,490.782	340	$p < .001$	.759	.732	.104	1.948

Chi-square from WLSMV estimation is means- and variances-adjusted using the full weight matrix. The difference in the adjusted chi-square is not interpretable based on a standard chi-square distribution (Muthén and Muthén 2010)

Discussion

In the present study, we found that the higher-order structure of the AQ-S suggested by previous studies exhibited configural and metric invariance across an ASD and control group in a series of multi-group CFA models. We were not, however, able to achieve scalar invariance even in the first-order model. After releasing equality constraints on 12 item thresholds, the model fit remained outside of the bounds that would indicate acceptable fit.

Implications

The importance of assessing measurement equivalence of a measure used across two groups is widely acknowledged, however, the present study was, to our knowledge, the first to examine the measurement equivalence of a question-naire used across both ASD and control groups. This is important because measures of autistic traits are frequently administered to both groups, with varying research goals.

Which level of invariance is required and how prob-lematic violations of invariance are depend on the specific purpose of using the measure in the two groups (Borsboom 2006). Our finding that the measure exhibited invariance at the metric level suggests that the manifest variables of the AQ-S measure the same latent traits across ASD and control group. From a substantive point of view, this is interesting because it implies that the same constructs can describe autistic traits in ASD and control groups. This

does not imply that these traits are necessarily continuously distributed across these groups (e.g. see Frazier et al. 2010), but it does suggest that even if the ASD and non-ASD represent distinct latent categories, then the two cat-egories may still be definable in terms of the same latent autistic traits. To formally test this hypothesis, it would be interesting to conduct latent class or taxometric analyses including tests of invariance across categories with the AQ-S, however, we did not have the requisite sample size in the present study.

We did not observe complete scalar invariance, even in our first-order model. A lack of scalar invariance suggests test bias and means that equal observed scores on the AQ-S does not necessarily imply equal levels of autistic traits in an individual drawn from an ASD versus a non-ASD population. This may makes sense in the context of the ALT-ASD continuity hypothesis which proposes that those who score high on ALTs may not receive a diagnosis of ASD if they are able to compensate and achieve adaptive levels of social functioning in spite of being predisposed to ASD. This could result in non-diagnosed individuals hav-ing higher levels of autistic traits for the same AQ-S score an individual with a clinical diagnosis of ASD. It also is consistent with a systematic under-reporting of symptoms in individuals with ASD due to deficits in self-perception (Johnson et al. 2009). This latter hypothesis would predict that scalar invariance could be observed if data were obtained from informant raters rather than via self-report and would be interesting to test in future research.

Table 4 Cross tabulation of sex and ASD status in the whole and random samples

	Male	Female	Total
Whole			
ASD	107	41	148
% within ASD	72.3	27.7	
% within Sex	72.8	24.6	
Control	40	126	166
% within ASD	24.1	75.9	
% within Sex	27.2	75.4	
Total	147	167	314
Group 1			
ASD	58	17	75
% within ASD	77.3	22.7	
% within Sex	78.4	22.1	
Control	16	60	76
% within ASD	21.1	78.9	
% within Sex	21.6	77.9	
Total	74	77	151
Group 2			
ASD	49	23	72
% within ASD	68.1	31.9	
% within Sex	67.1	25.6	
Control	24	67	91
% within ASD	26.4	73.6	
% within Sex	32.9	74.4	
Total	73	90	163

Strengths and Limitations

In the present study we examined the measurement equivalence of the AQ-S because it is a popular measure of autistic traits used frequently in individuals both with and without ASD. Historically, defining the core features of ASD in light of its complexity and heterogeneity has proven a challenge (Rajendran and Mitchell 2007). The AQ and its derivative, the AQ-S, represented an attempt to tackle this

issue and distil ASD into its core trait dimensions using psychometric tools of analysis. The AQ has undergone extensive psychometric evaluation and the majority of studies support its psychometric utility. Wheelwright et al. (2010) note that it shows consistent results over time, produces highly heritable scores, predicts clinical diagnosis of ASD and correlates with brain function, genetic polymorphisms in ASD candidate genes, social attention and prenatal testosterone levels. The more recently developed AQ-S yielded scores that correlated strongly with those of the full AQ in both a control samples and an ASD sample, suggesting minimal loss of information moving to a briefer inventory (Hoekstra et al. 2011). They also demonstrated good sensitivity and specificity in classifying individuals as having or not having a clinical diagnosis of ASD (ibid).

One issue, however, is that the AQ was designed for the express purpose of measuring autistic traits in adults with normal intellectual functioning (Baron-Cohen et al. 2001). For example, behaviours that are less common in high functioning ASD than low functioning ASD such as repetitive sensori-motor behaviours are not represented in the AQ-S. In fact, estimates suggest that approximately 50–70 % of individuals with ASD have an intellectual disability (Matson and Shoemaker 2009). The AQ-S is, therefore, not applicable to a large proportion of individuals with ASD and also may be less likely to show invariance across a low functioning ASD group and control group. Nevertheless, this does not undermine the validity of the current results with respect to individuals of normal intellectual functioning.

It is also important to consider the effect of our particular sample on the generalisability of results. We used a clinical sample of individuals with an ASD diagnosis, as well as a sample drawn from the general population. The relative rarity of clinical disorders can limit the sample sizes attainable for studies on clinical disorders. In the present study our sample size was small for an application of multi-group CFA, however, simulation studies have suggested that a lack of invariance can be detected for group sizes of down to 100 if the effect size is not small

Table 5 Model fit statistics for multi-group invariance models

Model	Adjusted χ^2	df	Signif.	CFI	TLI	RMSEA	WRMR
M1: configural invariance	1,029.686	690	$p < .001$	.901	.892	.056	1.508
M2: 1st-order factor loadings	1,031.686	713	$p < .001$	.907	.901	.054	1.594
Difference M2 versus M1	37.888	23	$p = .026$	.006	.009	−.002	.086
M3: 2nd-order factor loadings	1,071.350	716	$p < .001$	.895	.889	.057	1.671
Difference M3 versus M2	23.887	3	$p < .001$	−.012	−.012	.003	.077
M4: 1st-order thresholds	1,776.643	795	$p < .001$	.715	.729	.089	2.374
Difference M4 versus M3	1,246.168	79	$p < .001$	−.180	−.160	.032	.703

Chi-square from WLSMV estimation is means- and variances-adjusted using the full weight matrix. The difference in the adjusted chi-square is not interpretable based on a standard chi-square distribution. The corrected chi-square difference estimates were calculated using the DIFTEST option in Mplus (Muthén and Muthén 2010)

and once the sample size reaches 200 per group, a lack of invariance can generally be detected well, irrespective of the effect size (Kim and Yoon 2011). Our group sample sizes of 148 for the ASD group and 166 for the control group fell between these two values, suggesting that our study was sufficiently powered to detect a lack of invariance unless it was of only small effect size. Thus, while we can rule out moderate to large differences in the factor structure of the AQ-S across individuals with and without ASD in the current sample, replication of these findings and larger samples may be required to formally test for small differences in factor loadings across groups. We were also more likely to observe non-invariance by using a general population sample as our comparison control group. Had we used a sample comprising, for example, relatives of individuals with ASD, who would have likely had more similar levels of autistic traits to a clinical ASD sample we may not have observed such a large decrease in fit when adding scalar constraints across the ASD and control group.

Another factor which could affect the generalisability of the present results is the differing sex ratios of the control and ASD group. ASD is more prevalent in males than females, with estimates suggesting an average sex ratio of 4.3:1 in the disorder and with an even higher male: female ratio in samples of individuals with higher IQs, such as that utilised in the present study (Rivet and Matson 2011). Thus, in the present study the sex ratios favoured males in the ASD group and females in the control group. It is possible, given that there has been much discussion of possible sex differences in ASD, that our failure to find threshold invariance partly reflects these possible sex differences (Rivet and Matson 2011). The results from our single group CFAs using both sex/ASD homogenous versus heterogeneous samples suggested that the source of invariance had its origin in differences between ASD and control individuals rather than in sex differences. However, although these models are suggestive, they did not provide a formal test of differences across ASD status and sex. Such tests would be possible by applying a four group invariance analysis, with homogenous sex/ASD groups. Again, the current sample was not large enough for such an analysis. Sex differences in ASD are, nevertheless, not yet well understood and further research would be required to examine possible sex differences in the constructs measured by the AQ-S.

References

Allison, C., Auyeung, B., & Baron-Cohen, S. (2011). Toward brief "red flags" for autism screening: The short autism spectrum quotient and the short quantitative checklist in 1,000 cases and 3,000 controls. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(2), 202–212.e7.

American Psychiatric Association (APA). (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., p. 75). Washington, DC: American Psychiatric Association.

Austin, E. J. (2005). Personality correlates of the broader autism phenotype as assessed by the autism spectrum quotient (AQ). *Personality and Individual Differences, 38*(2), 451–460.

Baron-Cohen, S., Hoekstra, R. A., Knickmeyer, R., & Wheelwright, S. (2006). The autism-spectrum quotient (AQ)—adolescent version. *Journal of Autism and Developmental Disorders, 36*(3), 343–350.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with Asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders, 34*(2), 163–175.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., & Clubley, E. (2001). The autism-spectrum quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *Journal of Autism and Developmental Disorders, 31*(1), 5–17.

Borsboom, D. (2006). When does measurement invariance matter? *Medical Care, 44*(11), S176–S181.

Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464–504.

Chen, F. F., Sousa, K. H., & West, S. G. (2005). Teacher's corner: Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*(3), 471–492.

Constantino, J. N., & Todd, R. D. (2003). Autistic traits in the general population: A twin study. *Archives of General Psychiatry, 60*(5), 524.

Frazier, T. W., Youngstrom, E. A., Sinclair, L., Kubu, C. S., Law, P., Rezai, A., et al. (2010). Autism spectrum disorders as a qualitatively distinct category from typical behavior in a large, clinically ascertained sample. *Assessment, 17*(3), 308–320.

Frazier, T. W., Youngstrom, E. A., Speer, L., Embacher, R., Law, P., Constantino, J., et al. (2012). Validation of proposed *DSM-5* criteria for autism spectrum disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*(1), 28–40.e3.

Hoekstra, R. A., Vinkhuyzen, A. A., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., et al. (2011). The construction and validation of an abridged version of the autism-spectrum quotient (AQ-short). *Journal of Autism and Developmental Disorders, 41*(5), 589–596.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

Johnson, S. A., Filliter, J. H., & Murphy, R. R. (2009). Discrepancies between self-and parent-perceptions of autistic traits and empathy in high functioning children and adolescents on the autism spectrum. *Journal of Autism and Developmental Disorders, 39*(12), 1706–1714.

Kim, E. S., & Yoon, M. (2011). Testing measurement invariance: A comparison of multiple-group categorical CFA and IRT. *Structural Equation Modeling, 18*(2), 212–228.

Kuenssberg, R., Murray, A. L., Booth, T., & McKenzie, K. (2012). Structural validation of the abridged autism spectrum quotient—short form in a clinical sample of people with autism spectrum disorders. *Autism*. doi:10.1177/1362361312467708.

Kunihira, Y., Senju, A., Dairoku, H., Wakabayashi, A., & Hasegawa, T. (2006). Autistic traits in non-autistic Japanese populations: Relationships with personality traits and cognitive ability. *Journal of Autism and Developmental Disorders, 36*, 553–566.

Lundström, S., Chang, Z., Rastam, M., Gillberg, C., Larsson, H., Anckarsater, H., et al. (2012). Autism spectrum disorders and

autisticlike traits: Similar etiology in the extreme end and the normal variation. *Archives of General Psychiatry, 69*(1), 46.

Matson, M. L., Matson, J. L., & Beighley, J. S. (2011). Comorbidity of physical and motor problems in children with autism. *Research in Developmental Disabilities, 32*(6), 2304–2308.

Matson, J. L., & Shoemaker, M. (2009). Intellectual disability and its relationship to autism spectrum disorders. *Research in Developmental Disabilities, 30*(6), 1107–1114.

McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods, 7*, 64–82.

Meredith, W., & Teresi, J. A. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(11), S69–S77.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthen & Muthen.

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review, 27*(2), 224–260.

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354.

Rivet, T. T., & Matson, J. L. (2011). Review of gender differences in core symptomatology in autism spectrum disorders. *Research in Autism Spectrum Disorders, 5*(3), 957–976.

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.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online, 8*(2), 23–74.

Stark, S., Chernyshenko, O. S., & Drasgow, F. (2006). Detecting differential item functioning with confirmatory factor analysis and item response theory: Toward a unified strategy. *Journal of Applied Psychology, 91*(6), 1292.

Sung, Y. J., Dawson, G., Munson, J., Estes, A., Schellenberg, G. D., & Wijsman, E. M. (2005). Genetic investigation of quantitative traits related to autism: Use of multivariate polygenic models with ascertainment adjustment. *The American Journal of Human Genetics, 76*(1), 68–81.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (p. 63). Boston, MA: Pearson Education Inc.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods, 3*(1), 4–70.

Wheelwright, S., Auyeung, B., Allison, C., & Baron-Cohen, S. (2010). Defining the broader, medium and narrow autism phenotype among parents using the autism spectrum quotient (AQ). *Molecular Autism, 1*, 1–10.

Yu, C. Y. (2002). Evaluating cut-off criteria of model fit indices for latent variable models with binary and continuous outcomes. Unpublished Doctoral Dissertation, University of California, Los Angeles, CA. Accessed January 5, 2011 from http://www.statmodel.com/download/Yudissertation.pdf.