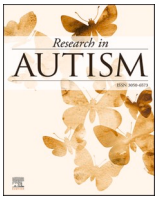
125 (2025) 202617

Contents lists available at ScienceDirect

Research in Autism

journal homepage: www.elsevier.com/locate/rasd



Is there an optimal self-report measure to investigate autism-related sex differences?

Lucy H. Waldren [a,*], Lucy A. Livingston [b,c], Punit Shah [a,*]

[a] Department of Psychology, University of Bath, Bath, United Kingdom
[b] Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, United Kingdom
[c] Neuroscience and Mental Health Innovation Institute, Cardiff University, Cardiff, United Kingdom

ARTICLE INFO

Keywords:
Autism
Autistic Traits
Gender
Sex

ABSTRACT

There is growing research interest in autism-related sex differences. Many behavioural and cognitive sex differences have been identified, with implications for research and clinical practice. Much of this research has relied on self-report autism measures, which are assumed to measure autistic traits equally in males and females. However, robust evidence for this assumption is lacking. Previous findings have not been replicated and no study has directly compared sex differences across multiple self-report autism measures in the same sample. To address this gap in research, a large sample of adults ($N = 1000$, 500 female) completed a series of self-report autism measures (AQ-50, −28, −26, −20, −10, −9, BAPQ, CATI). Following preregistered measurement invariance analyses, only the AQ-9, AQ-28, and CATI showed good-to-acceptable invariance to sex when specifying a multi-factor structure, and all 8 measures showed non-invariance to sex when capturing a general autism construct. We discuss the implications of these findings for investigating autism-related sex differences in future research.

Introduction

In recent years, autism-related sex differences have garnered considerable research interest in at least three ways. First, research continues to assess the extent of male preponderance in relation to autism and autistic traits. For example, although around three males are diagnosed for every 1 female (Loomes et al., 2017), research indicates that females may be under-diagnosed due to male-skewed diagnostic tools and/or societal and clinical biases (see Lai & Szatmari, 2020). Second, much research has focused on sex differences in specific patterns of autistic traits, both within diagnosed autistic and general population samples (see Hull et al., 2017). For example, females may show fewer or more subtle repetitive behaviours and restricted interests (e.g., Lai et al., 2015; Sutherland et al., 2017), and sometimes exhibit fewer social difficulties or appear 'superficially social' (e.g., Boorse et al., 2019; Wood-Downie et al., 2021). Third, attention has been paid to autism-related sex differences in relation to other behaviours and cognition, with implications for their expression, diagnosis, and management. For example, autism-related sex differences have been found in mental health difficulties (e.g., So et al., 2021), motivation for friendship (e.g., Sedgewick et al., 2016), and social and non-social cognition (see Hull et al., 2017 for meta-analysis). Based on this growing body of research on autism-related sex differences, the existence of a 'female autism phenotype' has even been theorised (Hull et al., 2020; Whitlock et al., 2020), which may require female-specific diagnostic tools and clinical support.

* Correspondence to: Department of Psychology, University of Bath, Bath BA2 7AY, United Kingdom.
E-mail addresses: lhw38@bath.ac.uk (L.H. Waldren), p.shah@bath.ac.uk (P. Shah).

https://doi.org/10.1016/j.reia.2025.202617
Received 23 August 2024; Received in revised form 4 December 2024; Accepted 17 January 2025
Available online 16 May 2025

Critically, however, research into autism-related sex differences requires an important assumption to be met. That is, self-report autism measures should be tapping into equivalent psychometric constructs across the sexes. If they are not, findings on sex differences using these measures are difficult or impossible to interpret. If the tools are not measuring the same constructs in males and females, meaningful comparisons cannot be made at either overarching factor or individual item levels. A widely used method to assess whether a measure is capturing the same psychometric construct across groups (e.g., sex) is measurement invariance analysis (see van de Schoot et al., 2015). Whilst such tests are common practice in several sub-domains of psychological science (e.g., social psychology; Callan et al., 2015) and clinical research (e.g., Fried et al., 2016), they are not used frequently in autism research. This is surprising, given the long-standing research and clinical interest in sex differences, and thus the need for self-report autism measures that are invariant to sex.

Some limited research has investigated measurement invariance to sex in adult self-report autism measures. The Autism-Spectrum Quotient has shown measurement invariance to sex across diagnosed autistic males and females in its 28-item form (AQ-28; Hoekstra et al., 2011; Grove et al., 2017). Surprisingly, however, to our knowledge, this has not been directly investigated in the more widely used 50-item version (AQ-50; Baron-Cohen et al., 2001). Murray and colleagues (2017, 2019) have also shown, using item response theory rather than measurement invariance analysis, that the 10-item AQ (AQ-10; Allison et al., 2012) is generally not biased across the sexes, although one item appears potentially problematic. The AQ-26 (Austin, 2005; Hurst et al., 2007) and AQ-9 (Jia et al., 2019) – other less frequently used variants of the AQ-50 – also appear to have potentially good factorial validity and invariance to sex. Additionally, the Broad Autism Phenotype Questionnaire (BAPQ; Hurley et al., 2007) has demonstrated measurement invariance to sex in one sample of parents of children with diagnosed autism (Broderick et al., 2015). And most recently, English et al. (2021) developed the Comprehensive Autistic Trait Inventory (CATI), which has demonstrated measurement invariance to sex in a large sample of ~1000 non-autistic individuals.

These studies have laid the groundwork for establishing appropriate self-report autism measures to study sex differences. Equally, however, they highlight several outstanding issues that need to be addressed and built upon. First, none of the aforementioned studies have openly accessible datasets, and there is generally a lack of replicatory research. Given how widely self-report autism measures are used in clinical and psychological science, it is imperative that we have openly accessible data in this field, to accompany robust and reproducible evidence for measurement invariance by sex (see Maassen et al., 2023). The present study will therefore, to our knowledge, represent the first study on autism-related sex differences using measurement invariance with a fully open dataset and analytic code.

Second, some previous research has not always been able to test for the most stringent levels of invariance, generally focussing on whether the measures have similar factor structures without imposing strict constraints on the models (e.g., Broderick et al., 2015). This might be due to issues with small samples or the heterogeneity in clinically diagnosed groups. By circumventing these issues in the present study, it will provide us with a better opportunity to use multigroup confirmatory factor analysis at four levels of increasing stringency (see Methods) towards an improved understanding of each measure's invariance to sex.

Third, almost all studies have selectively focussed on clinical samples, thus it is unclear whether their findings generalise to other populations where self-report autism measures are commonly used. Establishing this relationship is critical, given the increased use of self-report autism measures in non-clinical samples to inform understanding of autism and psychological science more generally (e.g., Clutterbuck et al., 2021; Happé & Frith, 2020; Riglin et al., 2021). To address this issue, we will draw on a large, non-clinical sample in the current research.

A final issue is that the aforementioned self-report autism measures have only ever been assessed in independent samples, making it difficult to compare them in terms their (in)appropriateness for studying sex differences. For example, the AQ-28 was found to be invariant to sex in Dutch cohorts (Grove et al., 2017), and the BAPQ in a highly unusual sample of people with a genetic liability for autism from across the United States (Broderick et al., 2015).

Overall, to advance understanding of autism-related sex differences, the current study aimed to test whether eight different self-report autism measures (three core measures, including six variations of the AQ) show measurement invariance to sex. Additionally, we compared autism scores by sex using conventional group comparisons. If different self-report autism measures were equally invariant to sex, the measure that produced the largest and most consistent sex difference in autism scores was taken as being the most sensitive to sex differences. Taken together, this study aims to establish which measure, if any, is optimal for investigating autism-related sex differences in future research.

Methods

Participants

One-thousand adults (500 females, 500 males) were recruited via online platform *Prolific*. Using data from the UK Office for National Statistics, we stratified participants by age to form a sample that is broadly representative of the UK (*Mean* = 45.94 years, *Range* = 18–87 years). Specific power analyses for measurement invariance analyses cannot be conducted, but larger samples with equal group sizes are known to increase analysis power and model stability (Putnick & Bornstein, 2016). Simulations suggest minimum sample sizes for measurement invariance of 150 per group (Chen, 2007), which our study exceeds. The sample size has also been determined with a view to provide 95 % power to detect small effect sizes ($d = 0.23$) in the group comparisons (where $\alpha = 0.05$, 2-tailed). Twenty-two additional participants were excluded for failing an embedded attention check ("*Please select "Definitely Disagree" to show you are reading the question.*") presented within the AQ-50, and, given the nature of the current research, two participants were excluded for specifying their sex at birth as "other".

Measures and procedure

Three self-report autism measures were administered. The 50-item Autism-Spectrum Quotient (AQ), including several of its shorter variants, the Broad Autism Phenotype Questionnaire, and the new Comprehensive Autistic Trait Inventory. These measures were selected because they: 1) were designed for both males and females, 2) capture a wide range of autistic traits across autism diagnostic criteria, and 3) are openly accessible, enabling our materials and datasets to also be openly accessible.

The 50-Item Autism-Spectrum Quotient (AQ-50) was developed as one of the first self-report measures of autism, in line with the Diagnostic and Statistical Manual of Mental Disorders IV guidelines (DSM-IV; American Psychiatric Association, 1994; Baron-Cohen et al., 2001). It has been used extensively in diagnosed autistic (e.g., Ashwood et al., 2016) and general population samples (see Ruzich et al., 2015 for meta-analysis), including several studies on sex differences (e.g., Baron-Cohen et al., 2014; Lai et al., 2017). The AQ-50 contains 50 statements regarding diverse autistic traits, which reflect the measure's five subscales: Social Skill, Communication, Attention Switching, Attention to Detail, and Imagination. Individuals indicate how strongly each statement applies to them on a 4-point Likert scale (1 = *Definitely Agree*, 4 = *Definitely Disagree*), before their responses are binarized and one point is awarded for each item they endorse. Scores can range from 0 (low autistic traits) to 50 (high autistic traits), with moderate-to-high internal consistency found across total ($\alpha = .67$; Hurst et al., 2007) and subscale measures ($\alpha = .63$–$.77$; Baron-Cohen et al., 2001). Two cut-off values ($\geq$26 and $\geq$32) have been identified as demonstrating high sensitivity (.77–.95) and specificity (.52–.74) to identify individuals with an autism diagnosis (Woodbury-Smith et al., 2005).

The 28-Item Autism-Spectrum Quotient (AQ-28) is an abridged version of the AQ-50 that retains the measure's sensitivity and specificity (Hoekstra et al., 2011). Previous research suggests the AQ-28 shows measurement invariance to sex (Grove et al., 2017), and has thus been used across psychological research to investigate autism-related sex differences (e.g., Livingston et al., 2022; Shah et al., 2019). The 28 items reflect Social Behavioural Difficulties (Social Skills, Switching, Routine, and Imagination) and fascination of Numbers and Patterns subscales, with high internal consistency found across total ($\alpha = .77$–$.86$) and subscale measures ($\alpha = .67$–$.86$). Items are scored using a 4-point Likert scale, with total scores ranging from 28 (low autistic traits) to 112 (high autistic traits). Two cut-off values ($>$65 and $\geq$70) have been identified as demonstrating high sensitivity (.94–.97) and specificity (.82–.91) to autism.

The 26-Item Autism-Spectrum Quotient (AQ-26) was developed after the psychometric properties of the AQ-50's five-factor structure were questioned (Austin, 2005). The AQ-26 retains three subscales which align with the DSM-IV's triad of diagnostic criteria: Social Skills, Details/Patterns, and Communication/Mindreading. Twenty-six items are measured using a 4-point Likert scale, and possible scores range from 26 (low autistic traits) to 104 (high autistic traits), with no cut-off value proposed. Whilst the measure has been used in some non-clinical autism research (e.g., Camodeca et al., 2019; Ingersoll, 2010), the moderate-to-high internal consistency ($\alpha = .66$–$.85$) found by Austin (2005) has been questioned, with Hurst et al. (2007) finding moderate total ($\alpha = .65$) and low-to-moderate subscale values ($\alpha = .42$–$.75$).

The 20-Item Autism-Spectrum Quotient (AQ-20) is an abridged version of the AQ-50 that was developed as part of the Adult Psychiatric Morbidity in England survey (McManus et al., 2009). While the measure's internal consistency has not been investigated, the AQ-20 has been used across neurodevelopmental research (e.g., Chaplin et al., 2017; McCarthy et al., 2015). Retaining the AQ-50's five subscales, binarized item responses are used to create a score between 0 (low autistic traits) and 20 (high autistic traits), with a cut-off value of $\geq$ 10 reported in previous research (e.g., McCarthy et al., 2019; National Institute for Health and Clinical Excellence, 2012). The AQ-20 was developed with three items reverse-phrased from their original wording in the AQ-50 (Brugha et al., 2007), however, because the current study will compute AQ-20 scores from participant's AQ-50 responses, all items were phrased according to Baron-Cohen et al. (2001).

The 10-Item Autism-Spectrum Quotient (AQ-10), another clinically-focussed variant of the AQ-50, was developed to aid clinicians in making autism referrals by selecting the 10 items (two from each subscale) most predictive of autism diagnosis (Allison et al., 2012). The measure was found to show high internal consistency (total $\alpha = .85$; subscale α not reported) and has been used across clinical practice and research (e.g., Weir et al., 2020; Taylor et al., 2021; Livingston et al., 2020). Whilst the AQ-10 has not been assessed for measurement invariance to sex, Murray et al., (2017, 2019) used item response theory to suggest the measure does not show a strong sex bias. Binarized response scores can range from 0 (low autistic traits) to 10 (high autistic traits), and a cut-off value of $\geq$ 6 was identified as having high sensitivity (.88) and specificity (.91) to identify when a referral for a specialist autism diagnostic assessment may be required (Allison et al., 2012; National Institute for Health and Care Excellence, 2021).

The 9-Item Autism-Spectrum Quotient (AQ-9) was recently developed to provide an updated AQ measure which mirrors the DSM-5 diagnostic criteria (Jia et al., 2019; American Psychiatric Association, 2013). Whilst the AQ-9 has not been externally validated, Jia and colleagues found high internal consistency across the measure's two subscales: Attention to Detail and Social Communication ($\alpha = .80$–$.92$; total α not reported), as well as measurement invariance to sex. The AQ-9 was developed using a 7-point Likert scale, creating a total score range from 9 (low autistic traits) to 63 (high autistic traits), with no autism screening value proposed. Because current participant responses were collected using a 4-point Likert scale, current scores may not correspond to those reported in Jia et al. (2019).

The Broad Autism Phenotype Questionnaire (BAPQ) was developed to measure "milder" characteristics of autism (Hurley et al., 2007, p. 1) akin to autistic traits, known as the 'Broad Autism Phenotype' (BAP). The BAPQ has been used to investigate the differences in BAP traits across family members of autistic individuals (e.g., Sasson et al., 2013, 2014). The BAPQ contains 36 items designed to measure the characteristics of its three subscales: Aloof Personality (Social Behaviour), Rigid Personality (Stereotyped-Repetitive Behaviour), and Pragmatic Language Deficits (Communication), in line with the DSM-IV (American Psychiatric Association, 1994). Participants indicate how frequently each statement applies to them on a 6-point Likert scale (1 = *Very Rarely*, 6 = *Very Often*). Scores are then averaged across all/subscale items, creating four metrics of BAP traits ranging from 1 (low BAP traits) to 6 (high BAP traits).

Hurley et al. (2007) found the BAPQ to have high internal consistency across total ($\alpha = .95$) and subscale ($\alpha = .85$–$.94$) measures, and suggested a total score cut-off value of 3.15 due to its high sensitivity (.82) and specificity (.78).

The Comprehensive Autistic Trait Inventory (CATI) was recently developed in light of advances in autism understanding and changing DSM-5 criteria (English et al., 2021; American Psychiatric Association, 2013); most specifically, the inclusion of sensory sensitivities into diagnostic criteria. Using a 5-point Likert scale, participants indicate the extent to which they agree (1 = *Definitely Disagree*, 5 = *Definitely Agree*) with 42 items across six subscales (Social Interactions, Communication, Social Camouflage, Cognitive Rigidity, Repetitive Behaviours, and Sensory Sensitivity), with high internal consistency reported across total ($\alpha = .95$) and subscale ($\alpha = .81$–$.94$) measures. Total scores can range from 42 (low autistic traits) to 210 (high autistic traits), with a cut-off value of ≥ 134 suggested as the optimal value to discriminate between autistic and non-autistic individuals (sensitivity =.83, specificity =.79).

Overall, participants completed three core self-report autism measures, the AQ-50, BAPQ, and the CATI, which also generated overall AQ-28, −26, −20, −10, and −9 scores. Together, we have eight different self-report measures of autism. After providing informed consent, participants completed the three core measures in a counterbalanced, pseudo-randomised order before answering demographic questions regarding age (years) and sex at birth (male, female, other). Participants were debriefed following completion of the study, which was covered by the local ethics committee.

Analyses

All analyses were conducted in R (R Core Team, 2021) and pre-registered (see https://psyarxiv.com/r4t9v/ for our Stage 1 manuscript with In Principle Acceptance; accepted September 2022). Measurement invariance analysis utilised the *lavaan* package (Rosseel, 2012) and the annotated analysis script, including all used packages, is openly accessible along with the dataset, and can be found in the Supplementary Materials. Relevant items for the AQ-28, −26, −20, −10, and −9 were extracted from the AQ-50 and total scores computed for each measure following reverse scoring. To inspect and visualize the data, descriptive statistics (e.g., mean, *SD*, range) and internal consistency measures (e.g., α, ω) are reported for all eight self-report autism measures. The correlations between the measures are also reported. These analyses were conducted across the whole sample, as well as separately in males and females.

To test for measurement invariance across sex, multigroup confirmatory factor analyses – using robust diagonally weighted least square estimates (WLSMV) to account for ordered-categorical data (DiStefano & Morgan, 2014; Flora & Curran, 2004; Li, 2016a, 2016b; Sass, 2011; Williams et al., 2018) – were conducted at four levels of increasing stringency (Putnick & Bornstein, 2016; van de Schoot et al., 2012). *Configural* invariance measures if the same latent factor structure is found between groups when fitting a model without any constraints. *Metric* invariance measures if each item shows a similar contribution to factors between males and females by constraining the configural models with equal factor loadings. *Scalar* invariance measures if the mean differences in the latent factor can explain all mean differences in the shared variance of items by constraining the metric models with equal intercepts. *Residual* invariance measures if the explained variance of each item is equal between groups by constraining the scalar models with equal variance of residuals.

Table 1
Descriptive statistics.

Measure	Group	Descriptives			Reliability		Sex comparison
		M	*SD*	*Range*	*α*	ω	
AQ50	Total	115.27	18.16	66–181	.90	.91	$t(997.71) = 4.17$, $p < .001$, $d = .26$ [.14, .39]
	Male	117.64	18.17	68–181	.90	.91	
	Female	112.89	17.86	66–170	.90	.91	
AQ28	Total	65.16	11.88	34–102	.87	.88	$t(996.19) = 4.33$, $p < .001$, $d = .27$ [.15, .40]
	Male	66.77	12.02	37–102	.88	.88	
	Female	63.55	11.52	34–97	.87	.87	
AQ26	Total	61.07	11.60	31–98	.88	.88	$t(998.00) = 3.83$, $p < .001$, $d = .24$ [.12, .37]
	Male	62.46	11.52	33–98	.87	.88	
	Female	59.67	11.52	31–93	.88	.88	
AQ20	Total	45.52	7.55	25–71	.76	.78	$t(996.16) = 4.15$, $p < .001$, $d = .26$ [.14, .39]
	Male	46.51	7.65	25–71	.76	.77	
	Female	44.54	7.33	26–70	.76	.78	
AQ10	Total	21.26	4.29	10–38	.71	.75	$t(997.75) = 5.01$, $p < .001$, $d = .32$ [.19, .44]
	Male	21.93	4.27	11–38	.70	.75	
	Female	20.59	4.20	10–35	.71	.75	
AQ9	Total	22.46	5.01	9–36	.78	.80	$t(998.00) = 5.31$, $p < .001$, $d = .34$ [.21, .46]
	Male	23.29	4.94	11–36	.77	.80	
	Female	21.63	4.94	9–34	.77	.80	
CATI	Total	112.50	29.06	43–197	.95	.95	$t(997.45) = 1.23$, $p = .22$, $d = .08$ [−.05, .20]
	Male	113.64	28.70	48–197	.95	.95	
	Female	111.37	29.39	43–196	.95	.95	
BAPQ	Total	112.71	25.95	52–192	.94	.94	$t(997.35) = 2.81$, $p = .005$, $d = .18$ [.05, .30]
	Male	115.00	26.19	52–192	.94	.94	
	Female	110.41	25.53	53–186	.94	.94	

Note. 95 % Confidence Intervals of effect size are shown in brackets [].

At each level, model fit was assessed across several fit indices (e.g., CFI and TLI > 0.95, RMSEA < 0.06, and SRMR < 0.08, are indicative of good model fit; Hu & Bentler, 1999; Chen, 2007). If model fit was found to worsen, it was taken as indication of measurement non-invariance (following Chen, 2007; Meade et al., 2008; Putnick & Bornstein, 2016). Because of the large sample size, change in χ^2 was not used to test for measurement invariance, following recommendations from Cheung and Rensvold (2002) and Meade et al. (2008). Partial invariance (Jung & Yoon, 2016) was also not investigated as the current study aimed to test full invariance of the different measures in their current form. Nonetheless, as the dataset is openly accessible, it will enable future analyses on partial invariance and other psychometric analyses (e.g., MIMIC modelling). Irrespective of invariance to sex, we compared males and females on all self-report autism measures using independent samples t-tests (where $\alpha = 0.05$, 2-tailed), accompanied with estimated effect sizes (Cohens *d*) with 95 % CIs.

Exploratory section

Outside of autism research, 1-item measures have been widely used to capture psychological constructs (see Postmes et al., 2013 for a meta-analysis of frequently cited 1-item measures). They often highly correlate with larger measurement scales and may be more equipped to detect individual differences as they are not biased by or towards specific traits (Nagy, 2002). We consequently included an additional 1-item Overall Autistic Trait Scale (OATS) in this study, which asked participants to rate "*On a scale of 0–100, how many autistic personality traits do you think you have?*", to ascertain how this approach performs within the context of autistic traits.

Results

At the observational level, all measures showed excellent to acceptable internal consistency for measuring autistic traits. As shown in Table 1, males reported significantly greater autistic traits than females in all measures except the CATI, and a moderate to strong correlation was found between total scores on all eight measures (Fig. 1).

When considering latent constructs, all measures showed consistent composite reliability between males and females (see Table 3). When modelling the factor structures unique to each autistic trait measure, the AQ-50, −26, and −20 produced wide-ranging coefficient omegas covering poor to excellent reliability. In these instances – in addition to all other measures – factors relating to social skills appeared to have the greatest reliability (see Supplementary Materials S.M.3.3). All AQ-related measures showed good-to-excellent composite reliability using a single-factor structure that captures autism as a general construct. However, reliability values for the CATI and BAPQ exceeded 1.0, suggesting that these measures may not work effectively as a unidimensional scale.

When modelling the unique multi-factor structures, the AQ-9 showed good measurement invariance to sex at the strictest level of constraints, whilst the AQ-28 and the CATI showed acceptable measurement invariance. The AQ-50, −26, −20, and BAPQ showed poor model fit and measurement non-invariance to sex at even the least constrained, configural level.[1] Applying a single-factor structure produced poor model fit and measurement non-invariance to sex in all measures at the configural level (see Table 2).

Chi-square statistics suggested a significant difference between males and females in autistic traits across all measures ($p < .001$).[2] The AQ-50 and AQ-26 showed the greatest difference in chi-square contribution (> 1000), whilst the AQ-20 and AQ-9 showed the smallest difference (< 100). Females contributed significantly more to the chi-square statistics in the AQ-50, −28, −26, and −20, suggesting that female scores were further from the expected score distribution than males (Table 3). The same pattern of results was found across the AQ-9, CATI, and BAPQ when specifying multi-factor structures, but reversed when specifying a single-factor model.

Latent variable means were higher for males than females in all measures (see Table 3), except for the sensory sensitivity subscale of the CATI, where the latent mean of males was 0.22 *SDs* below that of females (see S.M.3.3 for all subscale results). There was moderate variation in latent mean differences across all multi-factor models (*SD* = 0.00–0.37), with lower deviations observed in social skill related factors. Latent mean differences showed greater consistency across single-factor models (*SD* = 0.02–0.31), with the smallest sex differences observed in the CATI and AQ-9.

Exploratory results

When using the 1-item measure (OATS), there was no significant difference ($t(997.58) = .58$, $p = .56$, $d = .04$, 95 % CI[−.09, .16]) in the number of autistic traits reported between males (*M* = 23.79, *SD* = 22.04, *Range* = 0–90) and females (*M* = 22.99, *SD* = 21.59, *Range* = 0–90). Significant correlations between participant scores on the OATS and all eight established measures were observed (Fig. 1).

[1] Ordinal categorical data can also be investigated using an alternative constraint approach (Svetina et al., 2020), which produces a very similar pattern of results to those reported in Table 2. We report these alternative findings, in addition to non-scaled values, in the Supplementary Materials for the interested reader (S.M.1 – S.M.2.2). The AQ-10 could not be investigated using a multi-factor structure as only two items load onto each factor.

[2] Chi-square statistics have been included at the request of a reviewer. These values have not been used as indicators of model fit, but to provide greater insight into the source of measurement non-invariance.

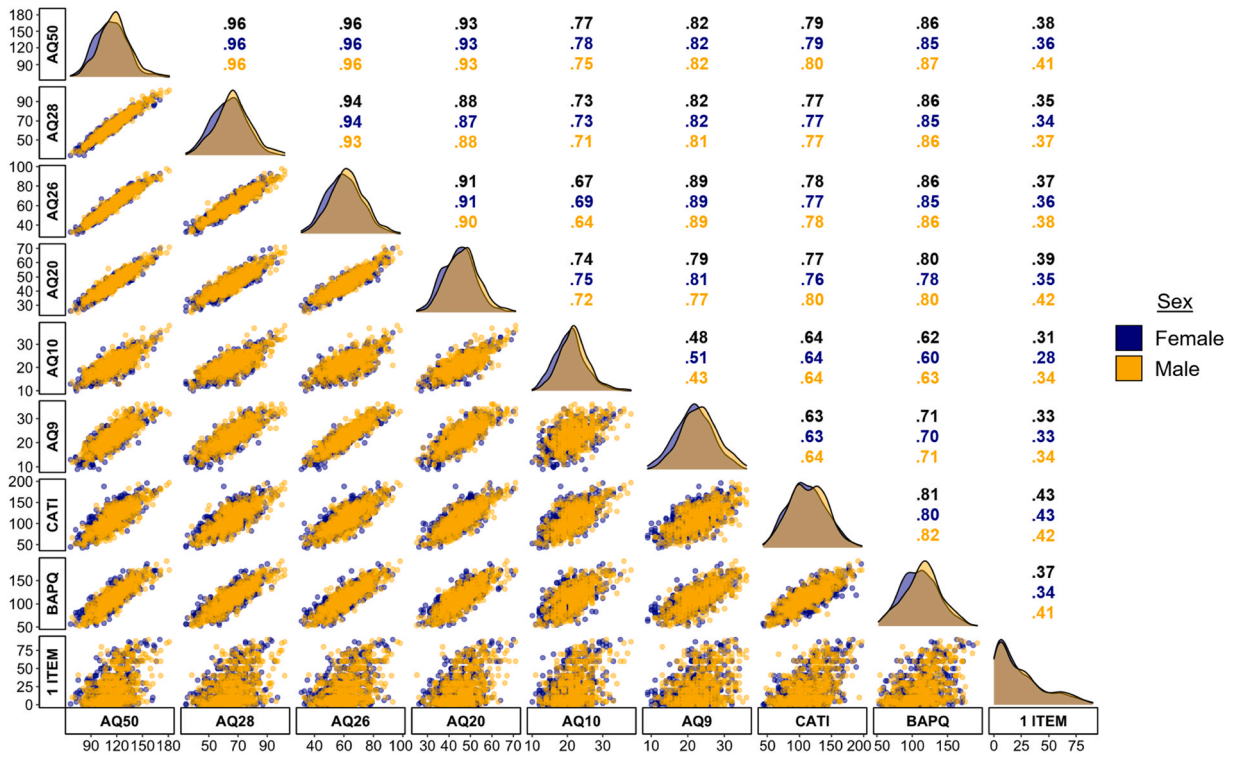


Fig. 1. Distribution of autistic traits and the relationships between trait measures. *Note*. Scatter plots and spearman correlation coefficients between the self-report autism measures split by sex. Males are shown in yellow, females are shown in blue, and correlation coefficients for the total sample are shown in black. All correlations are significant at $p < .001$. Density plots show the distribution of autistic trait scores within each measure, overlaying male and female distributions separately.

Discussion

This study investigated measurement invariance to sex in eight self-report autism measures that are widely used in research. Using a large, sex-balanced sample broadly representative of the UK population, a striking pattern of results was observed. Only three measures – the AQ-9, AQ-28, and CATI – showed measurement invariance to sex when a factor structure was specified, and all eight measures showed measurement non-invariance to sex when capturing a general autism construct. These results suggest that many self-report autism measures may not capture the same constructs in males and females in general population samples, implicating the interpretation of previous results and the best practice for research moving forward.

Finding good-to-acceptable model fit across increasingly constrained models in the AQ-9, AQ-28, and CATI, replicates previous research testing these measures for measurement invariance to sex (English et al., 2021; Jia et al., 2019). These findings bolster evidence and support the use of these tools for investigating sex differences in autism in non-clinical populations. Equally, the current results raise caution over the continued use of the AQ-50, −26, −20, −10, and BAPQ for investigating sex differences in autism, given their current measurement non-invariance to sex.

The psychometric explanation for this pattern of results appears two-fold. First, the AQ-9, −28, and CATI showed consistently good-to-excellent composite reliability, suggesting all factors within their model structure can accurately capture an autism construct. The remaining AQ measures showed poor composite reliability across many of their factors, suggesting the opposite, except for social skill related constructs. Finding such a large discrepancy in reliability within multiple measures is arguably surprising, and may reflect that social skill factors capture a narrower array of traits and linguistics (e.g., "*I find social situations easy*" and "*I enjoy social situations*") compared to other autistic trait factors. Second, differences in chi-square contribution were smaller for measures that showed more invariance, and larger in non-invariant measures. Although all measures showed larger chi-square contributions for females, reinforcing the potential male-biases within these tools, those that minimised this discrepancy were more likely to show measurement-invariance to sex.

Interestingly, this pattern was not so stark when considering differences in latent means, although discrepancies were smallest in social skill related factors, again suggesting greater consistency in scoring across more reliable constructs. The sensory sensitivity subscale in the CATI was the only factor to show a preponderance towards females. This is in line with current literature on sensory sensitives in autism (e.g., Lane et al., 2022; Osório et al., 2021), suggesting that measures aligning with more up-to-date theory may be more adept at capturing autistic constructs across sex.

Indeed, of the eight measures currently tested, only the AQ-9 and CATI were developed in-line with the current DSM-5 clinical

Table 2
Measurement invariance fit statistics.

Measure	Analysis	Muli-factors				One factor			
		CFI_S	TLI_S	$RMSEA_S$	SRMR	CFI_S	TLI_S	$RMSEA_S$	SRMR
AQ50	Configural	0.81	0.80	0.08	0.11	0.74	0.73	0.10	0.12
	Metric	0.84	0.84	0.07	0.11	0.81	0.81	0.08	0.12
	Scalar	0.82	0.82	0.08	0.11	0.77	0.77	0.09	0.12
	Residual	0.82	0.82	0.08	0.11	0.77	0.77	0.09	0.12
AQ28	Configural	0.93	0.92	0.07	0.08	0.72	0.70	0.14	0.13
	Metric	0.94	0.93	0.07	0.08	0.81	0.81	0.11	0.13
	Scalar	0.92	0.92	0.07	0.08	0.76	0.77	0.12	0.13
	Residual	0.92	0.92	0.07	0.08	0.76	0.77	0.12	0.13
AQ26	Configural	0.91	0.90	0.10	0.11	0.82	0.81	0.14	0.12
	Metric	0.92	0.92	0.09	0.11	0.89	0.89	0.11	0.12
	Scalar	0.91	0.91	0.10	0.11	0.86	0.86	0.12	0.12
	Residual	0.91	0.91	0.10	0.11	0.86	0.86	0.12	0.12
AQ20	Configural	0.85	0.82	0.11	0.10	0.78	0.75	0.12	0.12
	Metric	0.87	0.85	0.10	0.10	0.83	0.82	0.10	0.12
	Scalar	0.85	0.84	0.10	0.10	0.80	0.81	0.11	0.12
	Residual	0.85	0.84	0.10	0.10	0.80	0.81	0.11	0.12
AQ10	Configural	–	–	–	–	0.90	0.87	0.12	0.08
	Metric	–	–	–	–	0.93	0.92	0.09	0.08
	Scalar	–	–	–	–	0.91	0.91	0.10	0.08
	Residual	–	–	–	–	0.91	0.91	0.10	0.08
AQ9	Configural	0.99	0.99	0.06	0.05	0.81	0.74	0.30	0.18
	Metric	0.99	0.99	0.06	0.05	0.88	0.87	0.22	0.18
	Scalar	0.99	0.99	0.07	0.05	0.85	0.87	0.21	0.18
	Residual	0.99	0.99	0.07	0.05	0.85	0.87	0.21	0.18
CATI	Configural	0.92	0.91	0.07	0.07	0.76	0.75	0.12	0.12
	Metric	0.93	0.93	0.06	0.07	0.84	0.83	0.10	0.12
	Scalar	0.92	0.92	0.07	0.07	0.77	0.78	0.12	0.12
	Residual	0.92	0.92	0.07	0.07	0.77	0.78	0.12	0.12
BAPQ	Configural	0.89	0.88	0.10	0.09	0.81	0.80	0.12	0.11
	Metric	0.92	0.92	0.08	0.09	0.88	0.87	0.10	0.11
	Scalar	0.90	0.91	0.09	0.09	0.84	0.85	0.11	0.11
	Residual	0.90	0.91	0.09	0.09	0.84	0.85	0.11	0.11

Note. Scaled model fit indices are reported. Criteria for good model fit were CFI/TLI >0.95 and RMSEA <0.06/SRMR <0.08. Model fit was considered to have significantly worsened when ΔCFI/ ΔTLI >−0.01 and ΔRMSEA >0.015/ΔSRMR >0.3 (metric) or >−0.15 (scalar; Chen, 2007; Putnick & Bornstein, 2016). The AQ10 could not be investigated under a multi-factor structure as only two items load onto each factor.

criteria (APA, 2013). Knowledge of sex differences in autism has advanced significantly between and beyond the publication of the DSM-IV (APA, 1994) and the DSM-5 (APA, 2013), and thus the grounding of these measures in more current understanding may be driving their propensity to accurately capture autism constructs across sex. These results echo a broader call amongst research to re-assess how older – and somewhat theoretically outdated – self-report autism measures are used in research, and potentially supersede them with newer measures that can accurately capture autism constructs across sex and thus be more useful for examining sex differences (e.g., Ashwood et al., 2016; Bertrams & Shah, 2021; Taylor et al., 2021).

All eight measures showed measurement non-invariance to sex when measuring autism as a general construct. This was even found at the configural level before any constraints had been imposed, suggesting that overarching autism is not being captured equally between males and females. This is arguably concerning given that many of the studies that use these measures rely on their one-factor structure and total trait scores to infer the general level of autistic traits within a sample (e.g., Hargitai et al., 2023; Livingston et al., 2022; Taylor et al., 2023). Although this limitation is widely discussed across other psychological measures (e.g., Neff et al., 2017; Reise et al., 2013), the implications within autism research have not been readily considered. For example, using total scores on a measure with a non-invariant 1-factor structure may reduce the reliability of results that focus on autism related sex-differences. How research continues to use self-report measures to study autism, therefore, may require further reflection and discussion. This may lead to the application of more nuanced approaches that facilitate the accurate capture of sex differences.

Indeed, many studies have raised concern over the psychometric properties of older measures, specifically concerning their ability to accurately capture autism constructs (e.g., Ashwood et al., 2016; Bertrams & Shah, 2021; Taylor et al., 2021). This is compounded by the often-poor implementation of these tools, which may stem from erroneous or misleading guidelines (e.g., Waldren et al., 2025). When coupled with the current results, it becomes apparent that a stark review into how (and which) self-report autism measures should be used in research is required, with the aim of ensuring that autism constructs are accurately captured moving forward. For example, the AQ-9 showed excellent measurement invariance to sex in our study, suggesting it would be a suitable tool for investigating sex differences in autism. However, the AQ-9 has been used sparingly in autism research, with, to our knowledge, only three published studies to date utilising the measure (Belcher et al., 2023; Francis et al., 2024; Jia et al., 2022).

Furthermore, our exploratory 1-item questionnaire (OATS) significantly correlated with all other measures, in keeping with previous research that has used a 1-item tool to capture attitudes towards autistic people (Hanel & Shah, 2020). This suggests this

Table 3
Composite reliability, difference in latent means, and chi-square contributions.

Measure	Analysis	Multiple Factors					One Factor				
		Female ω	Male ω	Δ Mean	Female χ^2	Male χ^2	Female ω	Male ω	Δ Mean	Female χ^2	Male χ^2
AQ50	Configural	0.77 - 0.93	0.66–0.95	-	5201.27	4925.20	0.90	0.89	-	7080.36	5978.73
	Metric	0.76 - 0.94	0.69–0.93	-	4552.20	4323.47	0.90	0.90	-	5434.05	4647.80
	Scalar	0.75 - 0.93	0.66–0.95	0.05 - 0.25	5131.07	4864.51	0.91	0.90	0.09	6454.90	5522.73
	Strict	0.75 - 0.93	0.66–0.95	0.05 - 0.25	5131.07	4864.51	0.91	0.90	0.09	6454.90	5522.73
AQ28	Configural	0.72 - 0.88	0.71 - 0.89	-	1355.30	1094.22	0.91	0.92	-	4061.59	3226.36
	Metric	0.71 - 0.90	0.72 - 0.88	-	1221.87	999.16	0.93	0.91	-	2860.40	2293.36
	Scalar	0.72 - 0.88	0.71 - 0.89	0.00 - 0.32	1415.72	1159.38	0.93	0.92	0.08	3528.89	2859.99
	Strict	0.72 - 0.88	0.71 - 0.89	0.00 - 0.32	1415.72	1159.38	0.93	0.92	0.08	3528.89	2859.99
AQ26	Configural	0.69 - 0.94	0.69 - 0.94	-	2051.10	1720.76	0.90	0.88	-	3866.67	2753.91
	Metric	0.68 - 0.93	0.7 - 0.95	-	1733.03	1482.28	0.89	0.90	-	2472.89	1807.13
	Scalar	0.69 - 0.94	0.69 - 0.95	0.04 - 0.27	2042.65	1746.52	0.89	0.89	0.15	3165.65	2321.07
	Strict	0.69 - 0.94	0.69 - 0.95	0.04 - 0.27	2042.65	1746.52	0.89	0.89	0.15	3165.65	2321.07
AQ20	Configural	0.46 - 0.88	0.32 - 0.89	-	1062.96	1024.49	0.76	0.75	-	1588.51	1335.47
	Metric	0.46 - 0.88	0.31 - 0.89	-	951.27	919.14	0.74	0.77	-	1236.77	1047.41
	Scalar	0.45 - 0.88	0.29 - 0.89	0.06 - 0.37	1085.89	1051.08	0.77	0.74	0.19	1429.88	1248.81
	Strict	0.45 - 0.88	0.29 - 0.89	0.06 - 0.37	1085.89	1051.08	0.77	0.74	0.19	1429.88	1248.81
AQ10	Configural	-	-	-	-	-	0.75	0.72	-	296.08	254.29
	Metric	-	-	-	-	-	0.72	0.76	-	226.38	196.45
	Scalar	-	-	-	-	-	0.73	0.75	0.31	278.53	261.13
	Strict	-	-	-	-	-	0.73	0.75	0.31	278.53	261.13
AQ9	Configural	0.77 - 0.89	0.72 - 0.90	-	66.30	83.29	0.82	0.79	-	1394.32	1041.88
	Metric	0.76 - 0.90	0.73 - 0.90	-	68.67	85.39	0.79	0.81	-	854.14	642.61
	Scalar	0.77 - 0.89	0.73 - 0.90	0.10 - 0.34	120.88	132.92	0.81	0.81	0.05	1066.81	829.35
	Strict	0.77 - 0.89	0.73 - 0.90	0.10 - 0.34	120.88	132.92	0.81	0.81	0.05	1066.81	829.35
CATI	Configural	0.89 - 0.95	0.84 - 0.94	-	2921.43	2959.14	1.04	1.03	-	7394.75	6572.46
	Metric	0.89 - 0.95	0.85 - 0.94	-	2515.02	2550.33	1.04	1.03	-	5292.92	4749.09
	Scalar	0.89 - 0.95	0.84 - 0.94	−0.22 - 0.27	2989.92	2999.94	1.04	1.03	0.04	7288.02	6489.96
	Strict	0.89 - 0.95	0.84 - 0.94	−0.22 - 0.27	2989.92	2999.94	1.04	1.03	0.04	7288.02	6489.96
BAPQ	Configural	0.77 - 0.98	0.82 - 0.97	-	3206.39	3385.00	1.01	1.00	-	5286.76	5037.75
	Metric	0.79 - 0.97	0.80 - 0.97	-	2489.63	2613.83	1.02	0.99	-	3713.61	3551.04
	Scalar	0.78 - 0.98	0.81 - 0.97	0.02 - 0.16	3073.11	3231.59	1.02	1.00	0.13	4797.20	4583.77
	Strict	0.78 - 0.98	0.81 - 0.97	0.02 - 0.16	3073.11	3231.59	1.02	1.00	0.13	4797.20	4583.77

Note. Female/Male ω represent the range (across factors, Multiple Factors) and total (One Factor) composite reliability for each measure at each level of stringency. Δ *Mean* represents the magnitude of standard deviation for which the male latent mean is higher than female latent mean. Latent means can only be calculated for scalar and strict levels of stringency where intercepts have been constrained. *Female/Male* χ^2 represent the contribution of each sex to scaled chi-square statistics using traditional measurement invariance loadings, with higher values reflecting greater deviation from expected outcomes. For unstandardised and alternative invariance loadings, please see Supplementary Materials S.M.3.1 and S.M.3.2. The AQ10 could not be investigated under a multi-factor structure as only two items load onto each factor.

approach may be a suitable alternative to measure sex-differences in autism, given its consistency with other measures, innate measurement invariance to sex, and the resource constraints (e.g., financial, time) that often accompany research, especially within low and middle income countries (Bauer et al., 2022). Improving autism science through the accurate measurement of autistic traits may require a shift towards using existing measures that are theoretically current and psychometrically robust, like the AQ-9, or towards developing new measures, like the 1-item questionnaire, in collaboration with the neurodivergent community, that can accurately capture autistic constructs across sex (see Hobson et al., 2023). Nevertheless, for any such changes in research practice to occur it is evident that further research and discussion is needed. We hope that our study adds valuable insight on this issue and, by making our dataset and analysis code openly accessible, can provide an impetus for further research.

Research may wish to extend this work to consider clinical and non-clinical participants in addition to sex differences. Whilst this type of measurement invariance has been tested in previous studies (e.g., Murray et al., 2014), it has not been examined across numerous measures in the same dataset, nor has it readily been combined with sex differences (see Edwards, Wright, Sargeant, Cortese, & Wood-Downie, 2023; Tsirgiotis, Young, & Weber, 2023). Generating a more complete picture of self-report autism measures in this manner may further elucidate how to advance best practice in autism research. Furthermore, future research could examine a broader scope of self-report measures, given the arguably limited selection that were currently investigated. This was in-part due to many measures not being openly accessible (e.g., Social Responsiveness Scale; Constantino et al., 2000), highlighting the importance of incorporating open-science into future research and measure development to facilitate high quality autism research.

One may further argue that the current study is limited for applying traditional measurement invariance constraints to ordinal categorical data, when this can be less meaningful at more stringent levels of invariance (Svetina et al., 2020). We find the same pattern of results, however, when using alternative model constraints (see Supplementary Materials), suggesting the current pattern of results is robust. Research has also debated whether the traditional model fit criteria can over inflate good model fit in Likert data (Xia & Yang, 2019). If this were the case, however, it arguably makes the current findings – and non-invariance to sex – even more striking, and underscores that further consideration is needed over how (and which) self-report measures are continually used in research.

One such consideration would be to examine the item-level measurement-invariance. Whilst beyond the scope of the current study, investigating partial-invariance may shed important light onto the specific items that are failing to capture the same construct across sex and driving the current pattern of results (e.g., Belcher et al., 2023). By making the data set from this study openly accessible, we encourage future research to investigate these parameters, in addition to implementing wider invariance techniques.

In summary, the current study set out to investigate, which, if any, self-report measure(s) were optimal for investigating autism-related sex differences. Of the eight measures investigated (BAPQ, CATI, and six variants of the AQ50), the AQ-9, AQ-28, and CATI showed good-to-acceptable measurement invariance to sex when specifying a multi-factor structure, whilst all measures showed measurement non-invariance to sex when measuring a general autism construct. These findings paint an alarming picture for the propensity of existing self-report measures to capture autism traits equally across sex, with implications for how these measures should be used and interpreted within autism research. Further discussion and research in this area is now required, which we hope to facilitate through our open-access analysis code and dataset.

CRediT authorship contribution statement

Waldren Lucy H: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Visualization, Writing – original draft, Writing – review & editing. **Livingston Lucy A:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing. **Shah Punit:** Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Funding

LHW was supported by a doctoral studentship from the Economic and Social Research Council. LAL was supported by a fellowship from The Waterloo Foundation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank Sam Taylor for their thought-provoking discussions on the exploratory 1-item measure, and Luca D. Hargitai for their feedback on the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.reia.2025.202617.

Data availability

Data and code are openly accessible and are available in the Supplementary Material and in an online repository.

References

Allison, C., Auyeung, B., & Baron-Cohen, S. (2012). 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 & Adolescent Psychiatry, 51*(2), 202–212. https://doi.org/10.1016/j.jaac.2011.11.003

American Psychiatric Association. (1994). Diagnostic and Statistical Manual Of Mental Disorders (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). Diagnostic and Statistical Manual of Mental Disorders (5th ed.). Washington, DC: Author.

Ashwood, K. L., Gillan, N., Horder, J., Hayward, H., Woodhouse, E., McEwen, F. S., Findon, J., Eklund, H., Spain, D., Wilson, C. E., Cadman, T., Young, S., Stoencheva, V., Murphy, C. M., Robertson, D., Charman, T., Bolton, P., Glaser, K., Asherson, P., Simonoff, E., & Murphy, D. G. (2016). Predicting the diagnosis of autism in adults using the Autism-Spectrum Quotient (AQ) questionnaire. *Psychological Medicine, 46*(12), 2595–2604. https://doi.org/10.1017/S0033291716001082

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. https://doi.org/10.1016/j.paid.2004.04.022

Baron-Cohen, S., Cassidy, S., Auyeung, B., Allison, C., Achoukhi, M., Robertson, S., Pohl, A., & Lai, M.-C. (2014). Attenuation of typical sex differences in 800 adults with autism vs. 3,900 controls. *PLOS ONE, 9*(7), Article e102251. https://doi.org/10.1371/journal.pone.0102251

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. https://doi.org/10.1023/a:1005653411471

Bauer, K., Morin, K. L., Renz, T. E., & Zungu, S. (2022). Autism assessment in low- and middle-income countries: Feasibility and usability of western tools. *Focus on Autism and Other Developmental Disabilities, 37*(3), 179–188. https://doi.org/10.1177/10883576211073691

Belcher, H. L., Uglik-Marucha, N., Vitoratou, S., Ford, R. M., & Morein-Zamir, S. (2023). Gender bias in autism screening: Measurement invariance of different model frameworks of the Autism Spectrum Quotient. *BJPsych Open, 9*(5), Article e173. https://doi.org/10.1192/bjo.2023.562

Bertrams, A., & Shah, P. (2021). Internal reliability, homogeneity, and factor structure of the ten-item Autism-Spectrum Quotient (AQ-10) with two additional response categories. *Experimental Results, 2*, Article e3. https://doi.org/10.1017/exp.2020.70

Boorse, J., Cola, M., Plate, S., Yankowitz, L., Pandey, J., Schultz, R. T., & Parish-Morris, J. (2019). Linguistic markers of autism in girls: Evidence of a "blended phenotype" during storytelling. *Molecular Autism, 10*(1), 14. https://doi.org/10.1186/s13229-019-0268-2

Broderick, N., Wade, J. L., Meyer, J. P., Hull, M., & Reeve, R. E. (2015). Model invariance across genders of the Broad Autism Phenotype Questionnaire. *Journal of Autism and Developmental Disorders, 45*(10), 3133–3147. https://doi.org/10.1007/s10803-015-2472-z

Brugha, T., McManus, S., Meltzer, H., Smith, J., Scott, F. J., Purdon, S., Harris, J., & Bankart, J. (2007). Autism spectrum disorders in adults living in households throughout England: Report from the Adult Psychiatric Morbidity Survey. *National Centre for Social Research, 2009*. ⟨https://files.digital.nhs.uk/publicationimport/pub01xxx/pub01131/aut-sp-dis-adu-liv-ho-a-p-m-sur-eng-2007-rep.pdf⟩.

Callan, M. J., Kim, H., & Matthews, W. J. (2015). Age differences in social comparison tendency and personal relative deprivation. *Personality and Individual Differences, 87*, 196–199. https://doi.org/10.1016/j.paid.2015.08.003

Camodeca, A., Hosack, A., & Todd, K. Q. (2019). Investigation of broad autism phenotype traits as measured by the 26-Item Autism Quotient. *Journal of Psychoeducational Assessment, 37*(3), 338–357. https://doi.org/10.1177/0734282918768706

Chaplin, E., McCarthy, J., Underwood, L., Forrester, A., Hayward, H., Sabet, J., Mills, R., Young, S., Asherson, P., & Murphy, D. (2017). Characteristics of prisoners with intellectual disabilities. *Journal of Intellectual Disability Research, 61*(12), 1185–1195. https://doi.org/10.1111/jir.12441

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Clutterbuck, R. A., Callan, M. J., Taylor, E. C., Livingston, L. A., & Shah, P. (2021). Development and validation of the Four-Item Mentalising Index. *Psychological Assessment, 33*(7). https://doi.org/10.1037/pas0001004

Constantino, J. N., Przybeck, T., Friesen, D., & Todd, R. D. (2000). Reciprocal social behavior in children with and without pervasive developmental disorders. *Journal of Developmental & Behavioral Pediatrics, 21*(1), 2. doi:10.1097/00004703-200002000-00002.

DiStefano, C., & Morgan, G. B. (2014). A comparison of diagonal weighted least squares robust estimation techniques for ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 425–438. https://doi.org/10.1080/10705511.2014.915373

Edwards, H., Wright, S., Sargeant, C., Cortese, S., & Wood-Downie, H. (2023). Research Review: A systematic review and meta-analysis of sex differences in narrow constructs of restricted and repetitive behaviours and interests in autistic children, adolescents, and adults. *Journal of Child Psychology and Psychiatry*. https://doi.org/10.1111/jcpp.13855

English, M. C. W., Gignac, G. E., Visser, T. A. W., Whitehouse, A. J. O., Enns, J. T., & Maybery, M. T. (2021). The comprehensive autistic trait inventory (CATI): Development and validation of a new measure of autistic traits in the general population. *Molecular Autism, 12*(1), 37. https://doi.org/10.1186/s13229-021-00445-7

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466

Francis, L. J., Sailer, A. B., Payne, V. J., & McKenna, U. (2024). Engaging the AQ10 to predict professional burnout or poor work-related psychological wellbeing among anglican clergy in wales. *Journal of Religion and Health, 63*(2), 1661–1676. https://doi.org/10.1007/s10943-024-02006-7

Fried, E. I., van Borkulo, C. D., Epskamp, S., Schoevers, R. A., Tuerlinckx, F., & Borsboom, D. (2016). Measuring depression over time … Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment, 28*(11), 1354–1367. https://doi.org/10.1037/pas0000275

Grove, R., Hoekstra, R. A., Wierda, M., & Begeer, S. (2017). Exploring sex differences in autistic traits: A factor analytic study of adults with autism. *Autism, 21*(6), 760–768. https://doi.org/10.1177/1362361316667283

Hanel, P. H. P., & Shah, P. (2020). Simplifying the measurement of attitudes towards autistic people. *Experimental Results, 1*, Article e29. https://doi.org/10.1017/exp.2020.36

Happé, F., & Frith, U. (2020). Annual research review: Looking back to look forward – changes in the concept of autism and implications for future research. *Journal of Child Psychology and Psychiatry, 61*(3), 218–232. https://doi.org/10.1111/jcpp.13176

Hargitai, L. D., Livingston, L. A., Waldren, L. H., Robinson, R., Jarrold, C., & Shah, P. (2023). Attention-deficit hyperactivity disorder traits are a more important predictor of internalising problems than autistic traits. *Scientific Reports, 13*(1). https://doi.org/10.1038/s41598-022-26350-4

Hobson, H., Linden, A., Crane, L., & Kalandadze, T. (2023). Towards reproducible and respectful autism research: Combining open and participatory autism research practices. *Research in Autism Spectrum Disorders, 106*, Article 102196. https://doi.org/10.1016/j.rasd.2023.102196

Hoekstra, R. A., Vinkhuyzen, A. A. E., Wheelwright, S., Bartels, M., Boomsma, D. I., Baron-Cohen, S., Posthuma, D., & van der Sluis, S. (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. https://doi.org/10.1007/s10803-010-1073-0

Hu, L., & 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. https://doi.org/10.1080/10705519909540118

Hull, L., Petrides, K. V., & Mandy, W. (2020). The female autism phenotype and camouflaging: A narrative review. *Review Journal of Autism and Developmental Disorders, 7*(4), 306–317. https://doi.org/10.1007/s40489-020-00197-9

Hull, L., Petrides, K. V., Allison, C., Smith, P., Baron-Cohen, S., Lai, M.-C., & Mandy, W. (2017). "Putting on my best normal": Social camouflaging in adults with autism spectrum conditions. *Journal of Autism and Developmental Disorders, 47*(8), 2519–2534. https://doi.org/10.1007/s10803-017-3166-5

Hurley, R. S. E., Losh, M., Parlier, M., Reznick, J. S., & Piven, J. (2007). The broad autism phenotype questionnaire. *Journal of Autism and Developmental Disorders, 37*(9), 1679–1690. https://doi.org/10.1007/s10803-006-0299-3

Hurst, R. M., Mitchell, J. T., Kimbrel, N. A., Kwapil, T. K., & Nelson-Gray, R. O. (2007). Examination of the reliability and factor structure of the Autism-Spectrum Quotient (AQ) in a non-clinical sample. *Personality and Individual Differences, 43*(7), 1938–1949. https://doi.org/10.1016/j.paid.2007.06.012

Ingersoll, B. (2010). Broader autism phenotype and nonverbal sensitivity: Evidence for an association in the general population. *Journal of Autism and Developmental Disorders, 40*(5), 590–598. https://doi.org/10.1007/s10803-009-0907-0

Jia, R., Steelman, Z. R., & Jia, H. H. (2019). Psychometric assessments of three self-report autism scales (AQ, RBQ-2A, and SQ) for general adult populations. *Journal of Autism and Developmental Disorders, 49*(5), 1949–1965. https://doi.org/10.1007/s10803-019-03880-x

Jia, R., Steelman, Z., & Jia, H. (2022). What Makes One Intrinsically Interested in IT? An Exploratory Study on Influences of Autistic Tendency and Gender in the U.S. and India. *Management Information Systems Quarterly, 46*(3), 1603–1634.

Jung, E., & Yoon, M. (2016). Comparisons of three empirical methods for partial factorial invariance: Forward, backward, and factor-ratio tests. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(4), 567–584. https://doi.org/10.1080/10705511.2015.1138092

Lai, M.-C., & Szatmari, P. (2020). Sex and gender impacts on the behavioural presentation and recognition of autism. *Current Opinion in Psychiatry, 33*(2), 117–123. https://doi.org/10.1097/YCO.0000000000000575

Lai, M.-C., Lombardo, M. V., Auyeung, B., Chakrabarti, B., & Baron-Cohen, S. (2015). Sex/gender differences and autism: Setting the scene for future research. *Journal of the American Academy of Child and Adolescent Psychiatry, 54*(1), 11–24. https://doi.org/10.1016/j.jaac.2014.10.003

Lai, M.-C., Lombardo, M. V., Ruigrok, A. N., Chakrabarti, B., Auyeung, B., Szatmari, P., Happé, F., & Baron-Cohen, S. (2017). Quantifying and exploring camouflaging in men and women with autism. *Autism, 21*(6), 690–702. https://doi.org/10.1177/1362361316671012

Lane, A. E., Simpson, K., Masi, A., Grove, R., Moni, M. A., Montgomery, A., Roberts, J., Silove, N., Whalen, O., Whitehouse, A. J. O., & Eapen, V. (2022). Patterns of sensory modulation by age and sex in young people on the autism spectrum. *Autism Research, 15*(10), 1840–1854. https://doi.org/10.1002/aur.2762

Li, C.-H. (2016a). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychological Methods, 21*(3), 369–387. https://doi.org/10.1037/met0000093

Li, C.-H. (2016b). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Livingston, L. A., Shah, P., Milner, V., & Happé, F. (2020). Quantifying compensatory strategies in adults with and without diagnosed autism. *Molecular Autism, 11*(1), 15. https://doi.org/10.1186/s13229-019-0308-y

Livingston, L. A., Waldren, L. H., Walton, E., & Shah, P. (2022). Emotion processing differences mediate the link between sex and autistic traits in young adulthood. *JCPP Advances, 2*(3), Article e12096. https://doi.org/10.1002/jcv2.12096

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child and Adolescent Psychiatry, 56*(6), 466–474. https://doi.org/10.1016/j.jaac.2017.03.013

Maassen, E., D'Urso, E. D., van Assen, Nuijten, M. B., De Roover, K., & Wicherts, J. M. (2023). The dire disregard of measurement invariance testing in psychological science.Psychological Methods. Advance online publication. doi:10.1037/met0000624.

McCarthy, J., Chaplin, E., Forrester, A., Underwood, L., Hayward, H., Sabet, J., Young, S., Mills, R., Asherson, P., & Murphy, D. (2019). Prisoners with neurodevelopmental difficulties: Vulnerabilities for mental illness and self-harm. *Criminal Behaviour and Mental Health, 29*(5–6), 308–320. https://doi.org/10.1002/cbm.2132

McCarthy, J., Chaplin, E., Underwood, L., Forrester, A., Hayward, H., Sabet, J., Young, S., Asherson, P., Mills, R., & Murphy, D. (2015). Screening and diagnostic assessment of neurodevelopmental disorders in a male prison. *Journal of Intellectual Disabilities and Offending Behaviour, 6*(2), 102–111. https://doi.org/10.1108/JIDOB-08-2015-0018

McManus, S., Meltzer, H., Brugha, T., Bebbington, P., & Jenkins, R. (2009). Adult psychiatric morbidity in England, 2007: Results of a household survey. *National Centre for Social Research.* ⟨https://files.digital.nhs.uk/publicationimport/pub02xxx/pub02931/adul-psyc-morb-res-hou-sur-eng-2007-rep.pdf⟩.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology, 93*(3), 568. https://doi.org/10.1037/0021-9010.93.3.568

Murray, A. L., Booth, T., Auyeung, B., McKenzie, K., & Kuenssberg, R. (2019). Investigating sex bias in the AQ-10: A replication study. *Assessment, 26*(8), 1474–1479. https://doi.org/10.1177/1073191117733548

Murray, A. L., Booth, T., McKenzie, K., Kuenssberg, R., & O'Donnell, M. (2014). Are autistic traits measured equivalently in individuals with and without an autism spectrum disorder? An invariance analysis of the autism spectrum quotient short form. *Journal of Autism and Developmental Disorders, 44*(1), 55–64. https://doi.org/10.1007/s10803-013-1851-6

Murray, A. L., McKenzie, K., Kuenssberg, R., & Booth, T. (2017). Do the Autism-Spectrum Quotient (AQ) and Autism-Spectrum Quotient Short Form (AQ-S) primarily reflect general ASD traits or specific ASD traits? A bi-factor analysis. *Assessment, 24*(4), 444–457. https://doi.org/10.1177/1073191115611230

Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology, 75*(1), 77–86. https://doi.org/10.1348/096317902167658

National Institute for Health and Care Excellence (2021). 2021 Surveillance of Autism (NICE guidelines CG128, CG142 and CG170). ⟨https://www.nice.org.uk/guidance/cg142/resources/2021-surveillance-of-autism-nice-guidelines-cg128-cg142-and-cg170–9140525965/chapter/Surveillance-decision?tab=evidence⟩.

National Institute for Health and Clinical Excellence. (2012). Appendix 14B: Clinical Evidence – Study Characteristics. (NICE guideline CG142). ⟨https://www.nice.org.uk/guidance/cg142/evidence/appendix-14b-case-identification-instruments–clinical-study-characteristics-pdf-186587681⟩.

Neff, K. D., Whittaker, T. A., & Karl, A. (2017). Examining the factor structure of the self-compassion scale in four distinct populations: Is the use of a total scale score justified? *Journal of Personality Assessment, 99*(6), 596–607. https://doi.org/10.1080/00223891.2016.1269334

Osório, J. M. A., Rodríguez-Herreros, B., Richetin, S., Junod, V., Romascano, D., Pittet, V., Chabane, N., Jequier Gygax, M., & Maillard, A. M. (2021). Sex differences in sensory processing in children with autism spectrum disorder. *Autism Research, 14*(11), 2412–2423. https://doi.org/10.1002/aur.2580

Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology, 52*(4), 597–617. https://doi.org/10.1111/bjso.12006

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ⟨https://www.R-project.org/⟩.

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment, 95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Riglin, L., Wootton, R. E., Thapar, A. K., Livingston, L. A., Langley, K., Collishaw, S., Tagg, J., Smith, G. D., Stergiakouli, E., Tilling, K., & Thapar, A. (2021). Variable emergence of autism spectrum disorder symptoms from childhood to early adulthood. *American Journal of Psychiatry, 178*(8), 752–760. https://doi.org/10.1176/appi.ajp.2020.20071119

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1–36. ⟨https://www.jstatsoft.org/v48/i02/⟩.

Ruzich, E., Allison, C., Smith, P., Watson, P., Auyeung, B., Ring, H., & Baron-Cohen, S. (2015). Measuring autistic traits in the general population: A systematic review of the Autism-Spectrum Quotient (AQ) in a nonclinical population sample of 6,900 typical adult males and females. *Molecular Autism, 6*, 2. https://doi.org/10.1186/2040-2392-6-2

Sass, D. A. (2011). Testing measurement invariance and comparing latent factor means within a confirmatory factor analysis framework. *Journal of Psychoeducational Assessment, 29*(4), 347–363. https://doi.org/10.1177/0734282911406661

Sasson, N. J., Faso, D. J., Parlier, M., Daniels, J. L., & Piven, J. (2014). When father doesn't know best: Selective disagreement between self-report and informant report of the broad autism phenotype in parents of a child with autism. *Autism Research: Official Journal of the International Society for Autism Research, 7*(6), 731–739. https://doi.org/10.1002/aur.1425

Sasson, N. J., Lam, K. S. L., Childress, D., Parlier, M., Daniels, J. L., & Piven, J. (2013). The broad autism phenotype questionnaire: Prevalence and diagnostic classification. *Autism Research: Official Journal of the International Society for Autism Research, 6*(2), 134–143. https://doi.org/10.1002/aur.1272

Sedgewick, F., Hill, V., Yates, R., Pickering, L., & Pellicano, E. (2016). Gender differences in the social motivation and friendship experiences of autistic and non-autistic adolescents. *Journal of Autism and Developmental Disorders, 46*(4), 1297–1306. https://doi.org/10.1007/s10803-015-2669-1

Shah, P., Livingston, L. A., Callan, M. J., & Player, L. (2019). Trait autism is a better predictor of empathy than alexithymia. *Journal of Autism and Developmental Disorders, 49*(10), 3956–3964. https://doi.org/10.1007/s10803-019-04080-3

So, P., Wierdsma, A. I., van Boeijen, C., Vermeiren, R. R., & Mulder, N. C. (2021). Gender differences between adolescents with autism in emergency psychiatry. *Autism, 25*(8), 2331–2340. https://doi.org/10.1177/13623613211019855

Sutherland, R., Hodge, A., Bruck, S., Costley, D., & Klieve, H. (2017). Parent-reported differences between school-aged girls and boys on the autism spectrum. *Autism, 21*(6), 785–794. https://doi.org/10.1177/1362361316668653

Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: an illustration using mplus and the lavaan/semtools packages. *Structural Equation Modeling: A Multidisciplinary Journal, 27*(1), 111–130. https://doi.org/10.1080/10705511.2019.1602776

Taylor, E. C., Livingston, L. A., Callan, M. J., Hanel, P. H. P., & Shah, P. (2021). Do autistic traits predict pro-environmental attitudes and behaviors, and climate change belief? *Journal of Environmental Psychology, 76*, Article 101648. https://doi.org/10.1016/j.jenvp.2021.101648

Taylor, E. C., Livingston, L. A., Clutterbuck, R. A., Callan, M. J., & Shah, P. (2023). Psychological strengths and well-being: Strengths use predicts quality of life, well-being and mental health in autism. *Autism, 27*(6), 1826–1839. https://doi.org/10.1177/13623613221146440

Tsirgiotis, J. M., Young, R. L., & Weber, N. (2023). A comparison of the presentations of males and females with autism spectrum disorder and those narrowly below the diagnostic threshold. *Autism.* , Article 13623613231190682. https://doi.org/10.1177/13623613231190682

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492. https://doi.org/10.1080/1740

van de Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement invariance. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.01064

Waldren, L. H., Livingston, L. A., Clutterbuck, R. A., Walton, E., Callan, M. J., & Shah, P. (2025). The importance of using an optimal cutoff value for the 10-item Autism-Spectrum Quotient (AQ10). *Psychological Assessment, 37*(4), 180–186. https://doi.org/10.1037/pas0001339

Weir, E., Allison, C., & Baron-Cohen, S. (2020). Identifying and managing autism in adults. *Prescriber, 31*(2), 12–16. https://doi.org/10.1002/psb.1822

Whitlock, A., Fulton, K., Lai, M.-C., Pellicano, E., & Mandy, W. (2020). Recognition of girls on the autism spectrum by primary school educators: An experimental study. *Autism Research, 13*(8), 1358–1372. https://doi.org/10.1002/aur.2316

Williams, Z. J., Failla, M. D., Gotham, K. O., Woynaroski, T. G., & Cascio, C. (2018). Psychometric evaluation of the Short Sensory Profile in youth with autism spectrum disorder. *Journal of Autism and Developmental Disorders, 48*(12), 4231–4249. https://doi.org/10.1007/s10803-018-3678-7

Woodbury-Smith, M. R., Robinson, J., Wheelwright, S., & Baron-Cohen, S. (2005). Screening adults for asperger syndrome using the AQ: A preliminary study of its diagnostic validity in clinical practice. *Journal of Autism and Developmental Disorders, 35*(3), 331–335. https://doi.org/10.1007/s10803-005-3300-7

Wood-Downie, H., Wong, B., Kovshoff, H., Mandy, W., Hull, L., & Hadwin, J. A. (2021). Sex/Gender differences in camouflaging in children and adolescents with autism. *Journal of Autism and Developmental Disorders, 51*(4), 1353–1364. https://doi.org/10.1007/s10803-020-04615-z

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modelling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods, 51*(1), 409–428. https://doi.org/10.3758/s13428-018-1055-2