Supplement 1

Measurement Invariance Analyses.

Measurement equivalence of autism symptoms between females and males with autism spectrum disorder (ASD) was evaluated using a series of multigroup confirmatory factor analyses. For each analysis, indicators were scales from each of the major instruments (ADI-R social, communication, behavior; ADOS reciprocal social, communication, and restricted/repetitive; SRS total raw score, RBS-R total raw score). The model chosen to evaluate measurement invariance was a two-factor model based on *DSM-5* domains (Table S1 provides indicator-factor correspondence). Preliminary analyses indicated fit better to the two-factor DSM-5 model than single factor and three factor models (See Supplementary Table 1 footnote). The model included covariances between error terms of measures with the same rater (Subjective Report: SRS with RBS-R, Interview: ADI-R domain scores, Observation: ADOS domain scores).[50] Additional single-factor models were computed separately for the ADI-R domain scores and ADOS scale scores to test whether measurement invariance held for specific diagnostic measures. The purpose of single-factor measurement invariance analyses of the ADI-R and ADOS was to investigate sex differences in measurement across the major instruments used for diagnostic evaluations of ASD, not to investigate the specific factor structure of these measures.

Measurement invariance analyses were conducted in a stepwise fashion to examine decreases in model fit when increasing restrictions were placed on model parameters. The baseline model for these analyses estimated separate factor loadings, item intercepts, and residual variances and evaluated *configural invariance*—whether factors are represented by the same sets of items across groups. This model served as a comparison to more restrictive models. The next model fixed factor loadings across groups to examine weak or *metric invariance*—whether groups show a similar pattern of relationships

between indicators and the latent constructs they measure. The next model fixed both factor loadings and item intercepts across groups to examine strong or *scalar invariance*. This model tested whether the constructs were measured on the same scale across groups. The final model fixed factor loadings, item intercepts, and residual variances across groups to examine *strict invariance*. Scalar measurement invariance is preferred prior to making group comparisons to ensure that any observed sex differences are not an artifact of measurement differences. Model comparisons for measurement invariance analyses were based on empirical work indicating that a drop in CFI or TLI >.01 or an increase in RMSEA >.01 with a non-overlapping 95% CI imply measurement non-equivalence.[51-53] For the present study, if any indices fell beyond this cutoff, the more restrictive model was considered to have reduced fit.

Table S1. Factor-Indicator Correspondence for Multigroup Confirmatory Factor Measurement Invariance Analyses.

Indicator	Rater source	*DSM-5* two-factor	*DSM-IV-TR* three-factor
ADI-R social interaction	Interview	SCI	Social
ADI-R communication	Interview	SCI	Communication
ADI-R restricted/repetitive behavior	Interview	RRB	RRB
ADOS reciprocal social	Clinician observation	SCI	Social
ADOS communication	Clinician observation	SCI	Communication
ADOS restricted/repetitive	Clinician observation	RRB	RRB
SRS total raw score	Subjective report	SCI	Social
RBS-R total raw score	Subjective report	RRB	RRB

Note: The DSM-5 model was chosen for measurement invariance analyses based on superior fit ($X^2(36)=154.14$, CFI=.979, TLI=.968, RMSEA=.053, 95%CI=.044–.062, Bayesian Information Criterion= 106976) to a single factor solution ($X^2(46)=313.18$, CFI=.952, TLI=.933, RMSEA=.076, 95%CI=.068–.084, Bayesian Information Criterion 107104) and slightly better fit with a more parsimonious model to the three-factor solution ($X^2(40)=135.10$, CFI=.981, TLI=.965, RMSEA=.055, 95%CI=.045–.064, Bayesian Information Criterion= 107004). ADI-R=Autism Diagnostic Interview–Revised; ADOS=Autism Diagnostic Observation Schedule; RBS-R=Repetitive Behavior Scale–Revised; RRB=Restricted/Repetitive Behavior; SCI=Social Communication/Interaction; SRS=Social Responsiveness Scale.

Table S2. Measurement Invariance of Autism Symptom Indicators Across Females and Males With Autism Spectrum Disorder (ASD).

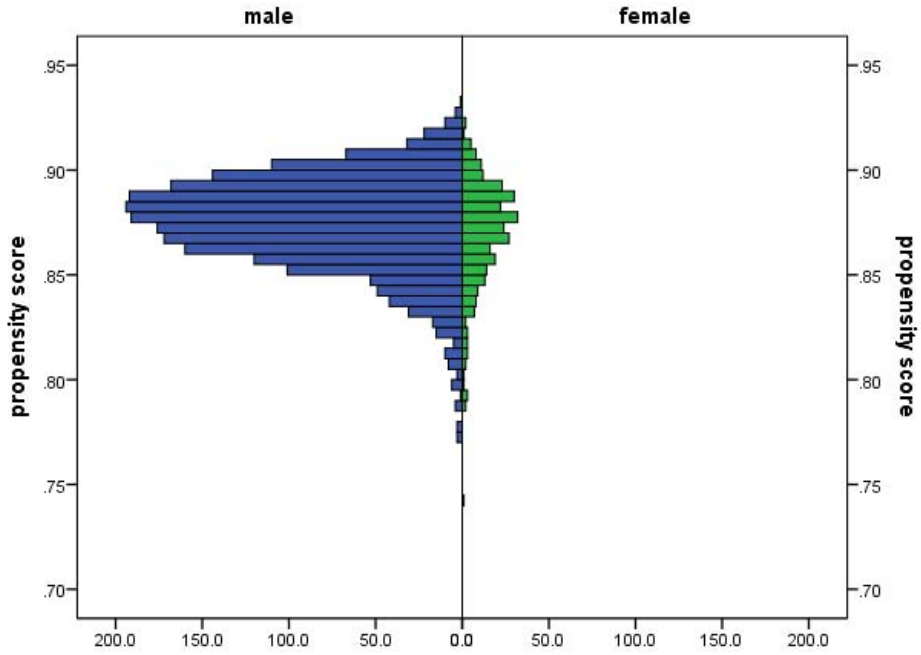
	X^2	Par	CFI / Δ	TLI / Δ	RMSEA / Δ	RMSEA 95% CI	Invariance Rule Accepted
***DSM-5* model**							
Configural Invariance	154.46	62	.977	.951	.065	.055-.075	-
Metric Invariance	146.68	58	.979 / .002	.962 / .011	.057 / .008	.048-.067	Yes
Scalar Invariance	165.41	50	.978 / -.001	.967 / .003	.053 / .005	.045-.062	Yes
Strict Invariance	178.54	42	.977 / -.001	.972 / .005	.049 / .004	.042-.057	Yes
ADI-R							
Configural Invariance	17.29	17	.990	.937	.118	.073-.169	-
Metric Invariance	7.56	14	.998 / .008	.997 / .060	.027 / .091	.000-.057	Yes
Scalar Invariance	14.91	11	.995 / -.003	.996 / -.001	.031 / -.004	.007-.053	Yes
Strict Invariance	15.07	9	.996 / .001	.997 / .002	.024 / .007	.000-.044	Yes
ADOS							
Configural Invariance	0.04	17	1.00	1.00	.001	.000-.045	-
Metric Invariance	0.28	16	1.00 / <.001	1.00 / <.001	.001 / <.001	.000-.030	Yes
Scalar Invariance	8.06	13	.999 / -.001	.998 / -.002	.023 / -.022	.000-.050	Yes
Strict Invariance	9.54	10	.999 / <.001	.999 / .001	.013 / .010	.000-.050	Yes

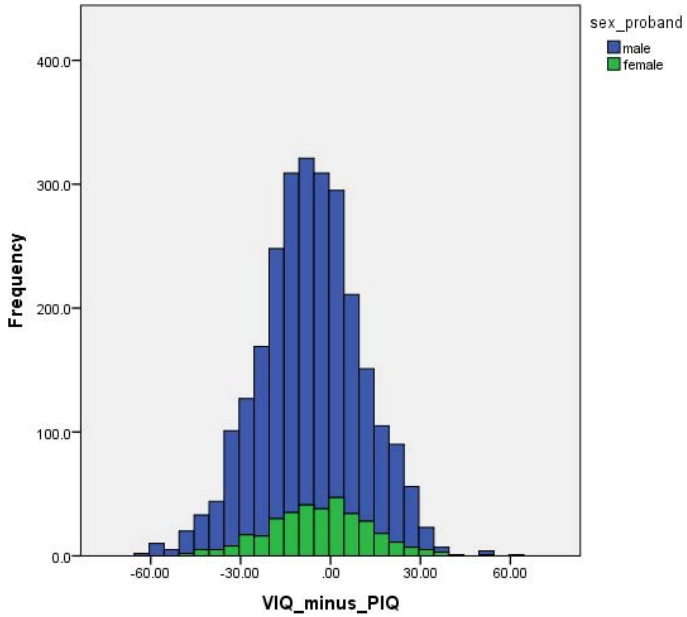
Note: Positive Δ values indicate improved fit, negative Δ values indicate decreased fit. ADI-R=Autism Diagnostic Observation Schedule; ADOS=Autism Diagnostic Observation Schedule; CFI=Comparative Fit Index; *DSM-5=Diagnostic and Statistical Manual of Mental Disorders–Fifth Edition*; Par=number of free parameters; RMSEA=Root Mean Square Error of Approximation; TLI=Tucker-Lewis Index.

Figure S1. Linear propensity score distributions for males and females. Note: Strong overlap in the propensity scores for males and females with autism spectrum disorder was observed.

Figure S2. Distribution of verbal IQ minus performance IQ discrepancies in females and males with autism spectrum disorder.

Figure S3. Q-Q plot of observed and expected *p*-values for social communication/interaction indicators. Note: A significant deviation from expected *p*-values is observed.


male
female
propensity score
propensity score
.95
.90
.85
.80
.75
.70
200.0
150.0
100.0
50.0
0.0
50.0
100.0
150.0
200.0


sex_proband
male
female
400.0
300.0
200.0
100.0
0.0
Frequency
-60.00
-30.00
.00
30.00
60.00
VIQ_minus_PIQ