Special Issue Article

autism

Autism
1–9

DOI: 10.1177/1362361316667283
aut.sagepub.com
SAGE

Exploring sex differences in autistic traits: A factor analytic study of adults with autism

Rachel Grove[1], Rosa A Hoekstra[2], Marlies Wierda[3,4] and Sander Begeer[3,4]

Abstract

Research has highlighted potential differences in the phenotypic and clinical presentation of autism spectrum conditions across sex. Furthermore, the measures utilised to evaluate autism spectrum conditions may be biased towards the male autism phenotype. It is important to determine whether these instruments measure the autism phenotype consistently in autistic men and women. This study evaluated the factor structure of the Autism Spectrum Quotient Short Form in a large sample of autistic adults. It also systematically explored specific sex differences at the item level, to determine whether the scale assesses the autism phenotype equivalently across males and females. Factor analyses were conducted among 265 males and 285 females. A two-factor structure consisting of a social behaviour and numbers and patterns factor was consistent across groups, indicating that the latent autism phenotype is similar among both autistic men and women. Subtle differences were observed on two social behaviour item thresholds of the Autism Spectrum Quotient Short Form, with women reporting scores more in line with the scores expected in autism on these items than men. However, these differences were not substantial. This study showed that the Autism Spectrum Quotient Short Form detects autistic traits equivalently in males and females and is not biased towards the male autism phenotype.

Keywords

adults, autism spectrum disorder, autism spectrum quotient, females, gender, sex differences

Autism spectrum conditions (ASC) are characterised by social and communication difficulties alongside repetitive behaviours or restricted interests (American Psychiatric Association, 2013). ASC occur in approximately 1 to 2% of the population and are more frequently diagnosed in males, with a sex ratio of around 4:1 (Baird et al., 2006; Centers for Disease Control and Prevention, 2014). Considering individuals without an intellectual disability only, the sex ratio increases to 8-9:1 (Mandy et al., 2011; Scott et al., 2002), indicating that sex differences in the vulnerability for autism change as a function of IQ. However, others suggest this discrepancy in autism diagnosis may be overestimated and report overall sex ratios between 2.0 and 2.6:1 (Kim et al., 2011; Mattila et al., 2011). This differential diagnosis has led to an expanding body of research evaluating sex differences in autism. Understanding sex differences in more detail is important for unpacking the complex aetiology of autism (Rutter et al., 2003). This study aimed to explore differences between males and females, using standardised measures of autistic traits, in a large sample of autistic[1] adults.

An evaluation of the research to date reveals a number of inconsistencies in the clinical autism phenotype across males and females. For example, some studies have indicated that autistic females display more severe social and communication difficulties compared to autistic males (Hartley and Sikora, 2009), while others have indicated that these difficulties are less severe in autistic women (McLennan et al., 1993), or report no sex differences (Wilson et al., 2016). Likewise, some studies suggest that autistic males show more stereotyped and repetitive behaviours than females (Hartley and Sikora, 2009; Hattier et al., 2011; Van Wijngaarden-Cremers et al., 2014) and

[1]University of New South Wales, Australia
[2]King's College London, UK
[3]Vrije Universiteit Amsterdam, The Netherlands
[4]VU University Medical Center, The Netherlands

Corresponding author:
Rachel Grove, School of Psychiatry, University of New South Wales, Level 1, 30 Botany Street, UNSW, Sydney, NSW 2052, Australia.
Email: r.grove@unsw.edu.au

that females on the spectrum display more socially acceptable special interests (Gould and Ashton-Smith, 2011). However, others argue that these interests and behaviours are similar across sex (Harrop et al., 2015). It has also been suggested that in order to receive a diagnosis, women are required to display more impairment in functioning than men (Dworzynski et al., 2012), yet other studies fail to confirm these findings when controlling for IQ (Holtmann et al., 2007; Pilowsky et al., 1998).

The conflicting results regarding sex differences in ASC may be due to methodological barriers related to recruitment and assessment. The recruitment of large samples of autistic females is difficult, and a significant proportion of studies lack the power to detect anything other than large effects (Mandy et al., 2012). This is particularly relevant for individuals without an intellectual disability and older autistic adults (Mandy et al., 2012). In addition, research may not capture the full range of women on the spectrum, given the delay in diagnosis experienced by autistic females (Giarelli et al., 2010; Rutherford et al., 2016), and few studies to date evaluating sex differences in adult samples.

There is therefore a need for consistent studies using standardised measures to evaluate sex differences in large samples of autistic adults. Currently available standardised assessment tools have historically been developed based on the 'male' phenotype of ASC. For example, the Autism Diagnostic Interview Revised (Lord et al., 1994) was developed based on a sample of 20 children, containing four females. Similarly, Rutter et al. (2003) argue that the core diagnostic symptoms of autism may be biased towards males, thus reducing the sensitivity of diagnostic and assessment measures for females. While there have been some attempts to develop specific measures to capture ASC in women (Kopp and Gillberg, 2011), others argue that current assessment tools are not sensitive enough to capture the 'female' autism phenotype (Halladay et al., 2015; Lai et al., 2015). Due to this potential bias in instruments used to evaluate ASC, it has been argued that exploring differences in mean or total scores is not useful and that sex differences need to be evaluated at a more detailed or item level (Kopp and Gillberg, 2011; Lai et al., 2015). Collecting broader information that captures the autism phenotype outside the core diagnostic criteria allows for specific qualitative differences across sex to be explored (Lai et al., 2015).

Multiple group factor analysis provides a systematic method for determining sex differences in autistic traits. This method can provide a fine-grained analysis of the way a measure evaluates traits in autistic males and females, as well as determine whether there is any bias across sex at the item level. A comprehensive understanding of whether these underlying constructs vary by sex has important implications for the definition of autism. It is therefore imperative to evaluate whether screening and assessment measures that are used commonly in research and clinical practice are biased across sex.

The Autism Spectrum Quotient (AQ; Baron-Cohen et al., 2001) is a well-validated assessment tool that measures quantitative traits of autism. Previous research has shown that there are sex differences in AQ scores within the general population, with males typically scoring higher than females (Baron-Cohen et al., 2001; Hoekstra et al., 2008; Ruzich et al., 2015). However, a recently developed short form of the AQ showed sex differences across both general population and clinical samples (AQ-Short; Hoekstra et al., 2011). In this study, men scored higher than women in the general population. However, autistic females scored higher on the AQ-Short than males (Hoekstra et al., 2011). This is consistent with previous research outlining that women with autism may self-report more difficulties than men (Lai et al., 2011).

Given the inconsistencies in previous research, it is important to study sex differences in the autism phenotype within an adult sample, containing a large number of females on the spectrum. This study will capitalise on an existing participant group available via the Netherlands Autism Register (NAR). The NAR is an online database containing a large sample of both autistic males and females. This will allow enough power for a detailed comparison of the factor structure of the AQ-Short across sex and to detect meaningful differences at the item level. This study will utilise the NAR data to evaluate the factor structure of the AQ-Short within a large adult sample of autistic men and women. It will also systematically evaluate whether specific items of the AQ-Short are more sensitive to assessing autism in males than females, in order to determine whether items on the AQ-Short are biased towards the male autism phenotype.

Method

Participants

The sample consisted of 550 adults with a formal diagnosis of *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV) pervasive developmental disorder or DSM-5 autism spectrum disorder. All diagnoses were provided by a qualified clinician, independently from this study. Participants included 265 males and 285 females. The sample was recruited via the NAR a large online database that collects information from autistic individuals and their families. Participants who were over the age of 16 years and who reported an IQ above 70 were selected for the study. Participants were asked to choose a range that best reflected their IQ score based on either a previous IQ assessment (n = 338) or a self-reported estimate (n = 212). These scores (provided in Table 1) were used as a proxy measure for IQ.

Measures

The AQ-Short (Hoekstra et al., 2011) was administered to all participants in the sample. The AQ-Short is a 28-item measure that evaluates autistic traits, including social skills, attention switching, a preference for routines, imagination and a fascination with numbers and patterns. Items are scored on a 4-point Likert scale with response options 'definitely agree', 'slightly agree', 'slightly disagree' and 'definitely disagree'. A total of 13 items are included in the AQ-Short where a 'disagree' response indicates the presence of autistic traits. These items are reverse scored. Scores on the AQ-Short range from 28 to 112, with higher scores indicating greater endorsement of autistic traits. Previous research has indicated a two-factor hierarchical factor structure for the AQ-Short, with a higher order social behaviour factor (23 items) consisting of the social skills, routine, switching and imagination items and a numbers and patterns factor (5 items). The AQ-Short has previously been evaluated in both Dutch and English general population samples and in English individuals with a diagnosis of Asperger syndrome (Hoekstra et al., 2011). The AQ-Short shows reliability scores in line with the full version of the AQ (Murray et al., 2015). The AQ-Short has good sensitivity and specificity and has been shown to correlate highly with the original 50-item version of the measure (Hoekstra et al., 2011). For a review of the translation process, see Hoekstra et al. (2008).

Analytic strategy

Confirmatory factor analyses (CFAs) were conducted on the AQ-Short in order to evaluate differences in the factor structure across males and females. All models were estimated using the weighted least squares mean and variance adjusted estimator for categorical variables with theta parameterisation. Model fit indices including the comparative fit index (CFI; Bentler, 1987), Tucker-Lewis index (TLI; Tucker and Lewis, 1973) and the root mean square error of approximation (RMSEA; Steiger and Lind, 1980) were calculated in order to compare the relative fit of the CFA models. RMSEA scores $\leqslant 0.08$ are indicative of good model fit, with scores $\leqslant 0.05$ indicating excellent fit to the data (Browne and Cudeck, 1993). It is recommended that CFI and TLI scores are above 0.92, with scores >0.95 indicating excellent fit to the data (Hu and Bentler, 1999; Marsh et al., 2004). However, recent research has also shown that the CFI and TLI are impacted by the number of indicators in a model (Cheung and Rensvold, 2002). Therefore, within this study, a value of $\geqslant 0.90$ was accepted as indicative of good model fit.

Following the factor structure identified in Hoekstra et al. (2011), a two-factor hierarchical model was implemented in which the social skills, routine, switching and imagination factors were predicted to load on a higher order social behaviour factor, and the numbers and patterns items to load on a distinct factor. This model was estimated both within the total sample and across males and females separately (Models 1-3). Multiple group CFA models were then implemented in order to determine whether the factor structure of the AQ-Short is the same across sex, as well as to explore any potential subtle item differences between autistic males and females. To do this, a number of models with differing levels of measurement invariance were estimated (Models 4-8). Measurement invariance (Meredith, 1993) evaluates whether the same construct is being measured across groups, in this case across autistic men and women. First, a model was implemented in order to test for configural invariance (Model 4). Obtaining configural invariance in this multiple group model would indicate that the underlying or latent constructs (in this case social behaviour and numbers and patterns) are conceptualised in the same way in both autistic men and women. Next, metric invariance was evaluated by constraining the factor loadings to be equal across sex (Models 5-7). A metric invariant model indicates that the strength of the relationship between the individual AQ-Short items and the latent constructs (or autistic traits) is the same across groups. Finally, scalar invariance was evaluated by constraining the item thresholds to be equal across groups (Model 8). Scalar or threshold invariance is required in order for latent mean comparisons to be conducted (Meredith, 1993). A scalar invariant model implies that individuals who display the same level of autistic traits on the latent variables (i.e. scores on the social behaviour or numbers and patterns factor) will obtain the same score on the observed variable (or AQ-Short item) regardless of whether they are male or female. If an item is not scalar invariant, it is biased against either males or females, resulting in a total score that is not completely comparable across sex. For example, a measure of depression may contain an item evaluating frequency of crying. Women tend to cry more often than men, regardless of whether they have a diagnosis of depression. With the inclusion of this item, women would be more likely to score high on this depression scale than men, even if their severity of depression is the same. This type of item bias would indicate that this measure is not equally sensitive to picking up clinically significant traits of depression across sex. Multiple group models were used in this study over item response theory models as they provide more sensitive fit statistics for comparison when using large samples.

The age of participants in our study ranged from 16 to 77 years. It was therefore important to explore the effect of age on the analyses. Age was centred for each sex separately by subtracting the mean age of males and females from the age of the participants within each group. This ensured that age was completely independent of sex. Centred age was included as a covariate in all subsequent models. Similarly, time since diagnosis (years passed

since the formal autism diagnosis was made) was included as a covariate for each latent variable in all models, as exploratory analyses of the data suggested a relationship between this variable and the AQ-Short. All analyses were estimated in Mplus version 7 (Muthén and Muthén, 2012).

Results

Demographic information is provided in Table 1. The male sample was significantly older than the female sample. Males had been diagnosed with autism for a significantly longer period of time than females. Mean scores on the AQ-Short for each group are provided in Table 1. Women scored significantly higher on the social behaviour subscale ($p < 0.05$) and significantly lower on the numbers and patterns scale ($p < 0.05$) than the male sample.

Fit indices and model comparisons for the CFA and multiple group models are provided in Table 2. Across all models, there was a significant effect of centred age on the social behaviour factor for males, with autistic trait scores increasing with age. Time since diagnosis had a significant effect on both the social behaviour and number and patterns factor in females, with fewer autistic traits reported across both factors in women who received their ASC diagnosis a long time ago. Initial analyses identified a number of items with correlated residuals. Based on the recommendations of Cole et al. (2007), two items containing similar wording were allowed to correlate in all models. Results from the CFA indicated that the two-factor hierarchical model displayed a good fit to the data in the total sample, males only and females only (Models 1-3). A multiple group CFA allowing the factor loadings and item thresholds to be freely estimated across groups (Model 4) also provided an adequate fit to the data. These findings indicate that configural variance was obtained.

To evaluate measurement invariance, chi-square difference testing between the constrained models (Models 5-8) and the freely estimated model (Model 4) was implemented. However, there is evidence that the chi-square statistic is impacted by sample size, therefore making this statistic overly sensitive to model misfit (Brannick, 1995; Cheung and Rensvold, 2002; Kelloway, 1995). Others have suggested that evaluating the difference between CFI and RMSEA indices from the freely estimated and constrained models provides a more robust indication of measurement invariance (Chen, 2007; Cheung and Rensvold, 2002). Difference scores were calculated between the CFI and RMSEA. $\Delta RMSEA \geqslant 0.01$ and $\Delta CFI \leqslant -0.005$ are indicative of a significant change in model fit between the nested models (Chen, 2007).

Metric invariance was evaluated across Models 5-7. Within Model 5, the factor loadings of the social behaviour factor were fixed across sex. Chi-square comparison tests indicated that the fit of this model did not decrease significantly from the freely estimated model (Model 4). The ΔCFI and $\Delta RMSEA$ values also indicated no significant deterioration of model fit, suggesting that the strength of the relationship between the AQ-Short items and the social behaviour factor (i.e. the factor loadings) was the same across groups. Model 6, in which the numbers and patterns factor loadings were constrained across sex, also indicated no substantial variation in fit. This was also confirmed in Model 7, in which all factor loadings were shown to be invariant across the sample. This indicates that the relationship between the items of the AQ-Short and the two-factor hierarchical structure is equivalent for autistic males and females.

Scalar invariance was evaluated in Model 8 by fixing the item factor loadings and item thresholds to be equivalent across sex. This model resulted in a significant decrease in fit, indicating that there are significant differences in the

Table 1. Demographic information and mean scores on the AQ-Short by sex.

	Males (n = 265)	Females (n = 285)	Total sample (n = 550)
	Mean (SD)	Mean (SD)	Mean (SD)
Age (years)	47.6 (13.0)**	39.9 (11.6)	43.6 (12.9)
Time since diagnosis (years)	8.0 (5.4)*	6.9 (5.3)	7.4 (5.4)
AQ total score	82.0 (12.4)	83.2 (11.0)	82.6 (11.7)
AQ social behaviour	68.0 (10.4)	70.0* (9.4)	69.0 (9.9)
AQ numbers/patterns	14.0* (3.8)	13.2 (3.6)	13.6 (3.7)
IQ proxy	%	%	%
>130	26.0	17.9	21.8
116–130	43.4	42.8	43.1
86–115	28.7	37.5	33.3
71–85	1.9	1.8	1.8

AQ: Autism Spectrum Quotient; IQ: intelligence quotient; SD: standard deviation.
*$p < 0.05$ and **$p < 0.01$ denote significant difference between males and females.

Table 2. Fit indices and model comparisons of confirmatory factor analysis of the AQ-Short.

Model	Description	Fit indices						
		RMSEA	CFI	TLI	χ^2	$\Delta\chi^2$ (*df*)	ΔRMSEA	ΔCFI
Two-factor hierarchical model								
1	Total sample (n = 550)	0.056	0.913	0.905	1086.495**			
2	Males (n = 265)	0.055	0.925	0.918	716.197**			
3	Females (n = 285)	0.055	0.903	0.893	736.769**			
Measurement invariance analyses								
4	Total sample (n = 550) free	0.055	0.915	0.907	1453.283**			
5	Total sample (n = 550) social behaviour factor loadings invariant	0.052	0.923	0.918	1413.971**	30.47 (23)	−0.003	0.008
6	Total sample (n = 550) numbers and patterns factor loadings invariant	0.054	0.917	0.910	1443.309**	3.03 (5)	−0.001	0.002
7	Total sample (n = 550) all factor loadings invariant	0.051	0.925	0.920	1407.365**	34.2 (28)	−0.004	0.010
8	Total sample (n = 550) factor loadings and item intercepts invariant	0.050	0.922	0.925	1507.150**	144.59 (107)**	−0.005	0.007

RMSEA: root mean square error of approximation; CFI: comparative fit index; TLI: Tucker–Lewis index; χ^2: chi-square statistic; $\Delta\chi^2$ (df): chi-square difference test; df: degrees of freedom; ΔRMSEA: difference score compared with unconstrained model; ΔCFI: difference score compared with unconstrained model.
**$p < 0.01$.

item thresholds of the AQ-Short for men and women on the spectrum. However, ΔRMSEA and ΔCFI scores were marginal, and within the recommended cut-off scores, ΔRMSEA $\geqslant 0.01$ and ΔCFI $\leqslant -0.005$ when compared with the freely estimated model. This indicates that while the chi-square difference testing indicated some misfit, the differences in item thresholds were not substantial.

Model 7, in which the factor loadings were equivalent across autistic males and females, while allowing the item thresholds to vary across groups, provided the best fit to the data. This metric invariance model is represented graphically in Figure 1.

In order to investigate the subtle differences in item thresholds between males and females, confidence intervals around the thresholds and modification indices for each group were examined. There were two items that were different between males and females (i.e. that contained thresholds where the confidence intervals did not overlap). Item 10 'I would rather go to a library than to a party' and item 26 'New situations make me anxious' were shown to display varying item thresholds across groups. Women were more likely to respond 'definitely agree' on these items, while men were relatively more likely to indicate they definitely disagreed (see Figure 2), resulting in higher average item scores (in line with autism) in women than men.

Discussion

This study evaluated the factor structure of the AQ-Short in a large sample of autistic adults. Results indicated that a two-factor hierarchical structure incorporating a social behaviour and numbers and patterns factor provided a good fit across sex. This highlights that the latent or underlying structure of the autism phenotype, as measured by the AQ-Short, is the same for autistic men and women. Measurement invariance investigations showed that the factor loadings were equivalent across groups, indicating that the relationship between the specific items of the AQ-Short and latent autistic traits did not differ by sex. This highlights that sex differences in item scores on the AQ-Short reflect meaningful variation in autistic traits across autistic men and women. However, analysis of the item thresholds identified two items that contained a subtle bias towards women on the spectrum.

Item 10 'I would rather go to a library than to a party' and item 26 'New situations make me anxious' were shown to display different item thresholds across groups. Female scores were more in line with those expected in autism on these two social behaviour items compared with men. This indicates subtle item bias on the AQ-Short that could erroneously make women appear more impaired on the social behaviour factor of this scale. This finding somewhat goes against the notion that assessment tools are biased towards a male expression of autism (Kopp and Gillberg, 2011; Kreiser and White, 2014; Rutter et al., 2003; Tierney et al., 2016), given that these two social behaviour items showed an *increased* sensitivity towards ASC in women. It could be that women with autism may be somewhat more aware of their social communicative difficulties than men (Lai et al., 2011), or identify more with their autism diagnosis, and

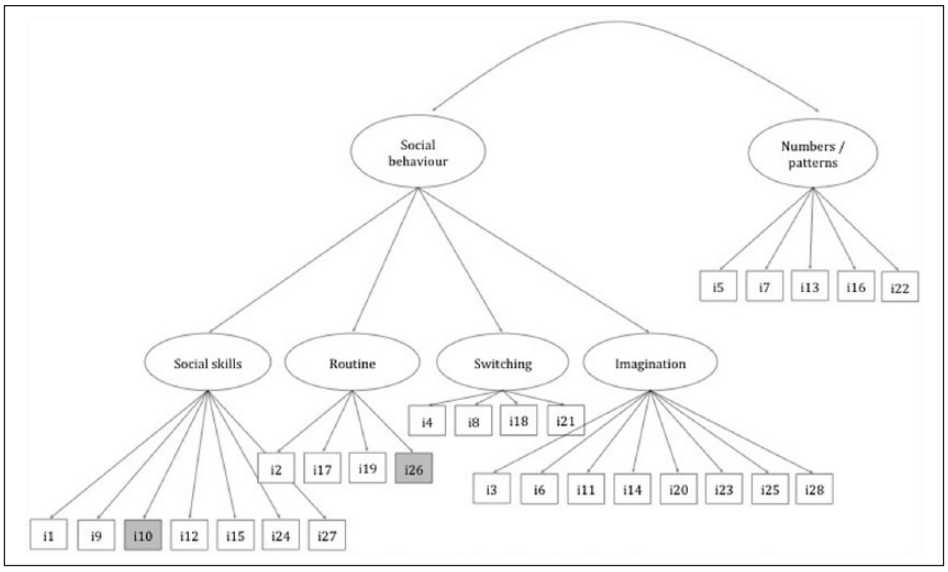

Figure 1. Two-factor hierarchical structure of the AQ-Short.

therefore are more able to report their associated social and communication difficulties. Perhaps given that knowledge of women on the spectrum without an associated intellectual disability is still emerging, autistic females may feel more need to justify their diagnosis than men. However, given that these differences in item thresholds were subtle and only found for two of the items, the items on the AQ-Short appear to measure the autism phenotype consistently across men and women.

Results indicated that age has a subtle yet significant positive association with AQ-Short social behaviour scores within the male sample, suggesting that self-reported social and communication difficulties might become slightly more pronounced later in life. There was also a significant relationship between time since diagnosis and scores on the AQ-Short in women, with a more recent diagnosis resulting in higher scores on the social behaviour and numbers and patterns items of the scale. This finding could reflect that females recently diagnosed with ASC are more likely to report difficulties due to living up to the expectations of the diagnosis. Because our study relied on self-report only, it is not possible to discern whether these modest age and time since diagnosis effects are due to self-reporting or would also be observed by others.

Initial examination of the data without the inclusion of covariates also highlighted a non-invariant item threshold on an additional item of the AQ-Short, evaluating a fascination with numbers. Upon further investigation, this item was found to be non-invariant due to age rather than sex. This subtle, yet significant, difference highlights the importance of accounting for age within analyses evaluating sex differences in ASC. It also shows that the AQ-Short is not necessarily biased towards men because it asks about a fascination for numbers and patterns rather than more social systems.

Mean score comparisons on the AQ-Short across sex were consistent with previous research evaluating the AQ-Short (Hoekstra et al., 2011; Lai et al., 2011), indicating that women reported more difficulties in social behaviour than men. However, the subtle item bias for two items included in the social behaviour scale suggests that these mean differences need to be interpreted with some caution. In addition, men scored higher than females on the numbers and patterns factor (a scale that was found to be measurement invariant across the sexes). This is consistent with previous research indicating that autistic males display more repetitive behaviours than females (Hattier et al., 2011). However, it should be noted that the AQ-Short does not capture all the diagnostic symptoms of autism outlined in DSM-5, particularly those relating to repetitive behaviours and sensory interests which may still include additional sex differences that are not captured in this study.

Limitations

Participants in this study had a previously confirmed diagnosis of ASC. While it is important to evaluate sex differences in the autism phenotype among autistic adults, individuals who may be missed in the diagnostic process were not included in the sample. Previous research has highlighted the importance of range restriction, and that evaluating individuals with a diagnosis potentially underestimates the relationship between the core features of autism (Murray et al., 2014). Within this study, it remains difficult to determine whether higher scores on the social behaviour items of the AQ-Short are indicative of greater

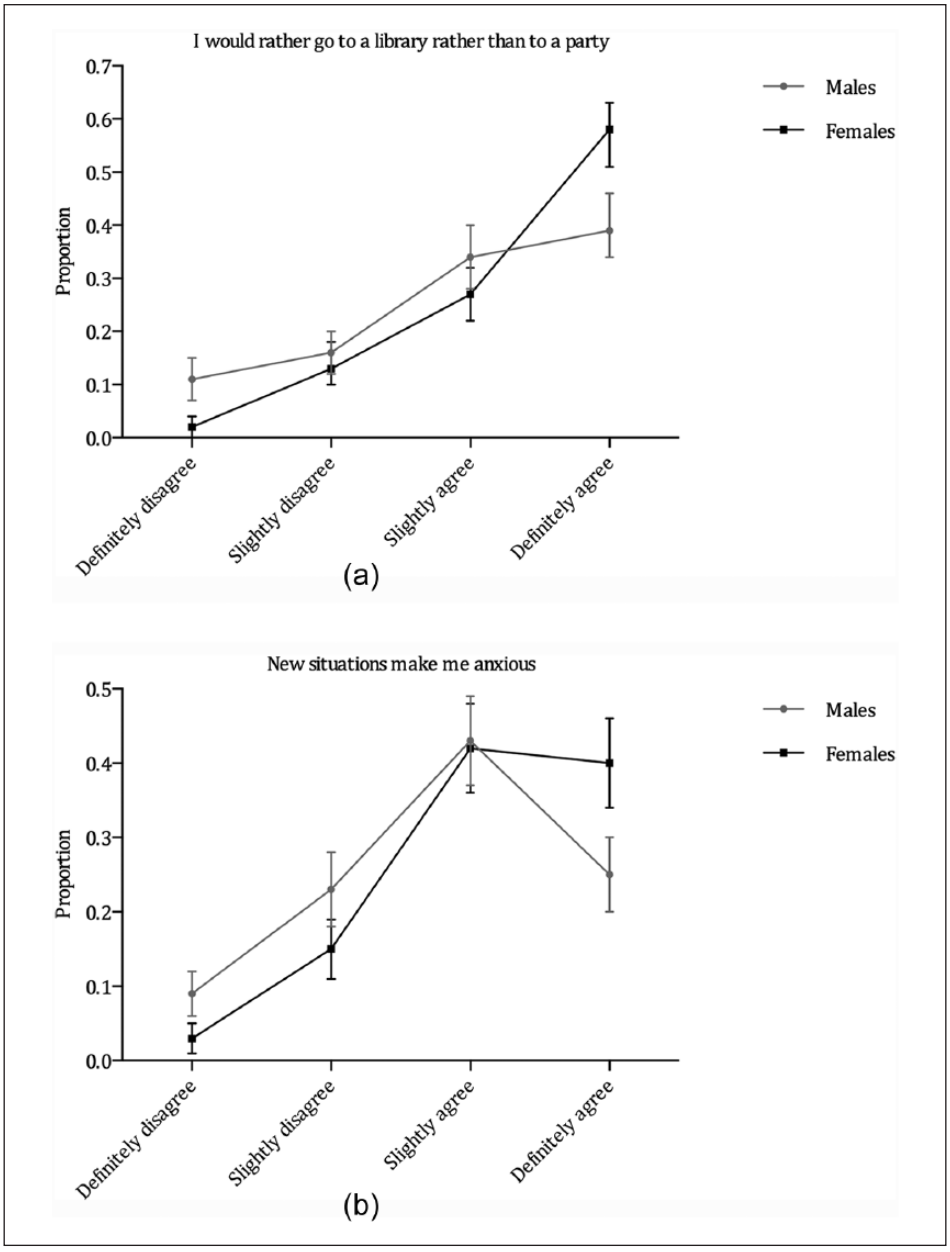

Figure 2. (a) Responses by sex AQ-Short item 10 and (b) responses by sex AQ-Short item 26.

severity of symptoms in identified females, or of under-identification of less affected females. Females with ASC thus remain a crucial group for future research. It was also not possible to confirm participants' diagnosis of ASC. However, previous research has shown that online research databases have the ability to recruit a representative sample that, on further testing, meet the diagnostic criteria for autism (Lee et al., 2010; Warnell et al., 2015). Moreover, the online nature of this study may also have decreased ascertainment bias, as it allowed the inclusion of individuals who may be unable or unwilling to take part in more time-consuming research protocols that require travelling to a lab or inviting research assistants into their home. Accessibility has been shown to be a factor that is important to autistic adults participating in research (Haas et al., 2016). The study would have benefited from the inclusion of more precise measures of IQ and language, as well as the inclusion of both observational and self-report data, rather than self-report information only. Future research evaluating sex differences in the adult autism phenotype via clinical observation and self-report is warranted. In addition, the inclusion of both Dutch and English samples to evaluate cross-cultural differences in the items of the AQ-Short would strengthen future research.

Conclusion

This study evaluated sex differences in the autism phenotype, as measured by the AQ-Short, in a large sample of autistic adults. Results revealed a two-factor structure incorporating a social behaviour and numbers and patterns factor. There was no evidence obtained to suggest that the AQ-Short is biased towards men because it asks about a fascination for numbers and patterns rather than more social systems. However, a subtle female bias was detected in two social behaviour items of the scale, showing an increased sensitivity towards ASC in women. This may be representative of an increased self-awareness in autistic females. Contrary to expectations, the underlying structure of the AQ-Short was equivalent for both autistic males and females, suggesting that the autism phenotype, as measured by the AQ-Short, is consistent across sex. Furthermore, the relationship between the individual items and autistic traits did not differ for males and females on the spectrum. This has implications for future research evaluating sex differences in the autism phenotype.

Acknowledgements

The authors would like to thank all the participants for being involved in the study and would like to thank Dr Mike Aitken for his helpful advice regarding the modelling strategy used in this paper.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The Netherlands Autism Register was financially supported by the Netherlands Autism Society (NVA), Het Autisme Fonds, Stichting Dijkverzwaring and CZ Fonds.

Note

1. A recent study found that the term 'autistic person/people' was commonly preferred by autistic adults (see Kenny et al., 2016). Identity-first language has therefore been used throughout this manuscript.

References

American Psychiatric Association (APA) (2013) *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*. Washington, DC: American Psychiatric Publishing, Inc.

Baird G, Simonoff E, Pickles A, et al. (2006) Prevalence of disorders of the autism spectrum in a population cohort of children in South Thames: the special needs and autism project (SNAP). *The Lancet* 368: 210–215.

Baron-Cohen S, Wheelwright S, Skinner R, et al. (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: 5–17.

Bentler PM (1987) Comparative fit indexes in structural models. *Psychological Bulletin* 107: 238–246.

Brannick MT (1995) Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior* 16: 201–213.

Browne MW and Cudeck R (1993) Alternate ways of assessing model fit. In: Bollen KA and Long JS (eds) *Testing Structural Equation Models*. Newbury Park, CA: SAGE, pp.136–162.

Centers for Disease Control and Prevention (2014) Prevalence of autism spectrum disorder among children aged 8 years: autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR Surveillance Summaries* 63: 1–21.

Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 14: 464–504.

Cheung GW and Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 9: 233–255.

Cole DA, Ciesla JA and Steiger JH (2007) The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods* 12: 381–398.

Dworzynski K, Ronald A, Bolton P, et al. (2012) How different are girls and boys above and below the diagnostic threshold for autism spectrum disorders? *Journal of the American Academy of Child and Adolescent Psychiatry* 51: 788–797.

Giarelli E, Wiggins LD, Rice CE, et al. (2010) Sex differences in the evaluation and diagnosis of autism spectrum disorders among children. *Disability and Health Journal* 3: 107–116.

Gould J and Ashton-Smith J (2011) Missed diagnosis or misdiagnosis? Girls and women on the autism spectrum. *Good Autism Practice* 12: 34–41.

Haas K, Costley D, Falkmer M, et al. (2016) Factors influencing the research participation of adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 46: 1793–1805.

Halladay AK, Bishop S, Constantino JN, et al. (2015) Sex and gender differences in autism spectrum disorder: Summarizing evidence gaps and identifying emerging areas of priority. *Molecular Autism* 6: 36.

Harrop C, Gulsrud A and Kasari C (2015) Does gender moderate core deficits in ASD? An investigation into restricted and repetitive behaviors in girls and boys with ASD. *Journal of Autism and Developmental Disorders* 45: 3644–3655.

Hartley SL and Sikora DM (2009) Sex differences in autism spectrum disorder: an examination of developmental functioning, autistic symptoms, and coexisting behavior problems in toddlers. *Journal of Autism and Developmental Disorders* 39: 1715–1722.

Hattier MA, Matson JL, Tureck K, et al. (2011) The effects of gender and age on repetitive and/or restricted behaviors and interests in adults with autism spectrum disorders and intellectual disability. *Research in Developmental Disabilities* 32: 2346–2351.

Hoekstra RA, Bartels M, Cath DC, et al. (2008) Factor structure, reliability and criterion validity of the autism-spectrum quotient (AQ): A study in Dutch population and patient groups. *Journal of Autism and Developmental Disorders* 38: 1555–1566.

Hoekstra RA, Vinkhuyzen AAE, Wheelwright 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: 589–596.

Holtmann M, Bölte S and Poustka F (2007) Autism spectrum disorders: sex differences in autistic behaviour domains and coexisting psychopathology. *Developmental Medicine & Child Neurology* 49: 361–366.

Hu L and Bentler PM (1999) Cutoff criterion for fit indices in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling* 6: 1–55.

Kelloway EK (1995) Structural equation modelling in perspective. *Journal of Organizational Behavior* 16: 215–224.

Kenny L, Hattersley C, Molins B, et al. (2016) Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism* 20: 442–462.

Kim YS, Leventhal BL, Koh YJ, et al. (2011) Prevalence of autism spectrum disorders in a total population sample. *American Journal of Psychiatry* 168: 904–912.

Kopp S and Gillberg C (2011) The autism spectrum screening questionnaire (ASSQ)-revised extended version (ASSQ-REV): An instrument for better capturing the autism phenotype in girls? A preliminary study involving 191 clinical cases and community controls. *Research in Developmental Disabilities* 32: 2875–2888.

Kreiser NL and White SW (2014) ASD in females: are we overstating the gender difference in diagnosis? *Clinical Child and Family Psychology Review* 17: 67–84.

Lai M-C, Lombardo MV, Auyeung B, et al. (2015) Sex/gender differences and autism: Setting the scene for future research. *Journal of the American Academy of Child and Adolescent Psychiatry* 54: 11–24.

Lai M-C, Lombardo MV, Pasco G, et al. (2011) A behavioral comparison of male and female adults with high functioning autism spectrum conditions. *PLoS ONE* 6: e20835.

Lee H, Marvin AR, Watson T, et al. (2010) Accuracy of phenotyping of autistic children based on internet implemented parent report. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* 153B: 1119–1126.

Lord C, Rutter M and Le Couteur A (1994) Autism diagnostic interview-revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders* 24: 659–685.

McLennan JD, Lord C and Schopler E (1993) Sex differences in higher functioning people with autism. *Journal of Autism and Developmental Disorders* 23: 217–227.

Mandy W, Charman T, Gilmour J, et al. (2011) Toward specifying pervasive developmental disorder-not otherwise specified. *Autism Research* 4: 121–131.

Mandy W, Chilvers R, Chowdhury U, et al. (2012) Sex differences in autism spectrum disorder: Evidence from a large sample of children and adolescents. *Journal of Autism and Developmental Disorders* 42: 1304–1313.

Marsh HW, Hau KT and Wen Z (2004) In search of golden rules: comment on hypothesis-testing approaches to setting cutoff values for fit indices and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling* 11: 320–341.

Mattila ML, Kielinen M, Linna SL, et al. (2011) Autism spectrum disorders according to DSM-IV-TR and comparison with DSM-5 draft criteria: an epidemiological study. *Journal of the American Academy of Child and Adolescent Psychiatry* 50: 583–592.

Meredith W (1993) Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58: 525–543.

Murray AL, McKenzie K, Kuenssberg R, et al. (2014) Are we under-estimating the association between autism symptoms? The importance of considering simultaneous selection when using samples of individuals who meet diagnostic criteria for an autism spectrum disorder. *Journal of Autism and Developmental Disorders* 44: 2921–2930.

Murray AL, McKenzie K, Kuenssberg R, et al. (2015) 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*. Epub ahead of print 16 October. DOI: 10.1177107319111561230.

Muthén LK and Muthén BO (2012) *Mplus*. 7th ed. Los Angeles, CA: Muthén & Muthén.

Pilowsky T, Yirmiya N, Shulman C, et al. (1998) The autism diagnostic interview-revised and the childhood autism rating scale: differences between diagnostic systems and comparison between genders. *Journal of Autism and Developmental Disorders* 28: 143–151.

Rutherford M, McKenzie K, Johnson T, et al. (2016) Gender ratio in a clinical population sample, age of diagnosis and duration of assessment in children and adults with autism spectrum disorder. *Autism* 20: 628–634.

Rutter M, Caspi A and Moffitt TE (2003) Using sex differences in psychopathology to study causal mechanisms: unifying issues and research strategies. *Journal of Child Psychology and Psychiatry* 44: 1092–1115.

Ruzich E, Allison C, Smith P, et al. (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.

Scott FJ, Baron-Cohen S, Bolton P, et al. (2002) Brief report: prevalence of autism spectrum conditions in children aged 5–11 years in Cambridgeshire, UK. *Autism* 6: 231–237.

Steiger JH and Lind JM (1980) Statistically based tests for the number of common factors. Paper presented at Meeting of the psychometric society. Iowa City, IA, June.

Tierney S, Burns J and Kilbey E (2016) Looking behind the mask: social coping strategies of girls on the autistic spectrum. *Research in Autism Spectrum Disorders* 23: 73–83.

Tucker LR and Lewis C (1973) A reliability coefficient for maximum likelihood factor analysis. *Psychometrika* 15: 149–162.

Van Wijngaarden-Cremers PJM, Van Eeten E, Groen WB, et al. (2014) Gender and age differences in the core triad of impairments in autism spectrum disorders: A systematic review and meta-analysis. *Journal of Autism and Developmental Disorders* 44: 627–635.

Warnell F, George B, McConachie H, et al. (2015) Designing and recruiting to UK autism spectrum disorder research databases: Do they include representative children with valid ASD diagnoses? *BMJ Open* 5: e008625.

Wilson CE, Murphy CM, McAlonan G, et al. (2016) Does sex influence the diagnostic evaluation of autism spectrum disorder in adults? *Autism*. Epub ahead of print 22 January. DOI: 10.1177/1362361315611381.