Published in final edited form as:

Autism. 2021 October ; 25(7): 1872–1884. doi:10.1177/13623613211004054.

Psychometric properties of the Chinese Parent Version of the Autism Spectrum Rating Scale: Rasch analysis

Weili Yan¹, Richard J Siegert², Hao Zhou¹, Xiaobing Zou³, Lijie Wu⁴, Xuerong Luo⁵, Tingyu Li⁶, Yi Huang⁷, Hongyan Guan⁸, Xiang Chen⁹, Meng Mao¹⁰, Kun Xia¹¹, Lan Zhang¹⁰, Erzhen Li⁸, Chunpei Li¹, Xudong Zhang¹, Yuanfeng Zhou¹, Andy Shih¹², Eric Fombonne¹³, Yi Zheng¹⁴, Jisheng Han¹⁵, Zhongsheng Sun¹⁶, Yong-hui Jiang¹⁷, Yi Wang¹

¹Children's Hospital of Fudan University, China

²Auckland University of Technology, New Zealand

³Sun Yat-Sen University, China

⁴Harbin Medical University, China

⁵The Second Xiangya Hospital of Central South University, China

⁶Children's Hospital of Chongqing Medical University, China

⁷West China Hospital of Sichuan University, China

⁸Capital Institute of Paediatrics, China

⁹The 2nd Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, China

¹⁰Chengdu Women and Children's Hospital, China

¹¹Central South University, China

¹²Autism Speaks, USA

¹³Oregon Health & Science University, USA

¹⁴Capital Medical University, China

¹⁵Peking University, China

¹⁶Beijing Institutes of Life Science, Chinese Academy of Sciences, China

¹⁷Department of Genetics, Pediatrics, & Neuroscience, Yale University School of Medicine, New Haven, CT, USA

Abstract

Article reuse guidelines: sagepub.com/journals-permissions

Corresponding authors: Richard J Siegert, Department of Psychology, Faculty of Health and Environmental Sciences, Auckland University of Technology, Room AR321, Private Bag 92006, Auckland 0627, New Zealand. richard.siegert@aut.ac.nz, Yi Wang, Department of Neurology, Children's Hospital of Fudan University, 5th Building, No. 399, Wanyuan Road, Minhang District, Shanghai 201102, China. yiwang@shmu.edu.cn.

Supplemental material

Supplemental material for this article is available online.

The recent adaptation of a Chinese parent version of the Autism Spectrum Rating Scale showed the Modified Chinese Autism Spectrum Rating Scale to be reliable and valid for use in China. The aim of this study was to test the Modified Chinese Autism Spectrum Rating Scale for fit to the Rasch model. We analysed data from a previous study of the Modified Chinese Autism Spectrum Rating Scale which comprised 1593 non-cases and 420 children diagnosed with autism spectrum disorder. We used super items based on groups of locally dependent items and item deletion when necessary to achieve good fit to the model for each of the three subscales identified by Zhou et al. and for the full 59-item Modified Chinese Autism Spectrum Rating Scale. The resulting conversion tables enable the use of genuine unidimensional, interval level scores for the total score and three subscales. Reliability was high with Person Separation Index values ranging from 0.83 to 0.89 for the three subscales and 0.79 for the total scale. In addition, we were able to identify a full-scale version of the Autism Spectrum Rating Scale and its three subscales that are all free of differential item functioning in relation to the five person factors recorded namely age, sex, caseness, relative and city. In future studies, the Teacher version of the Modified Chinese Autism Spectrum Rating Scale needs examination with Rasch analysis.

Lay abstract

The Autism Spectrum Rating Scale is a behavioural rating scale completed by parents and teachers that is useful for identifying children with an autism spectrum disorder. The development of a modified Autism Spectrum Rating Scale suitable for use in China is important for the identification of children in China with an autism spectrum disorder. In this study, we examined the Modified Chinese Autism Spectrum Rating Scale using a statistical technique known as Rasch analysis. Rasch analysis tests whether the questionnaire meets the standards for modern scientific measurement. We used Rasch analysis to examine data from 2013 children in China including 420 diagnosed with an autism spectrum disorder who had been rated by a parent or grandparent. After removing a small number of items (questions), the Modified Chinese Autism Spectrum Rating Scale met the stringent criteria for Rasch measurement. The availability of a reliable and precise tool for assessing behaviours characteristic of an autism spectrum disorder in Chinese children will improve the identification and diagnosis of autism spectrum disorder in China, thus enabling better provision of support services.

Keywords

autism spectrum disorders; Autism Spectrum Rating Scale; China; parent version; psychometrics; Rasch analysis; school-age children

Introduction

The Centers for Disease Control and Prevention estimated the overall prevalence of autism spectrum disorder (ASD) in the United States as 16.8 per 1000 in 2014 with the condition four times more common in boys Baio et al. (2018). While research on ASD in China is relatively new, a recent study in three cities in China reported prevalence rates comparable with the West (Sun et al., 2019) and with a population of 1.4 billion this represents a significant number of people with ASDs requiring clinical, educational and social support. The development of standardized tools for the identification of ASDs which are appropriate

for use in China is an important step towards servicing this need. The recent adaptation and validation of a Chinese parent version of the Autism Spectrum Rating Scale (ASRS) showed the ASRS to be reliable and valid for use with children and parents in China (H. Zhou, Zhang, Zou, et al., 2017).

The ASRS

The ASRS is a 71-item rating scale for the assessment and quantification of symptoms of ASDs (Goldstein and Naglieri (2013); Goldstein & Naglieri, 2010). The ASRS was developed for use in the United States and has forms for rating children by either a parent or a teacher. This study reports Rasch analysis of the Chinese parent version of the ASRS, with a sample of 2013 children aged 6–13 years from China, including 420 with a diagnosis of ASD. The ASRS has robust psychometric properties although to our knowledge it has not been examined with modern item response theory methods particularly Rasch analysis. To date, the psychometric evaluation of the ASRS has involved Classical Test Theory (CTT) methods (DeVellis, 2006).

Goldstein and Naglieri (2012), who developed the ASRS, reported a three-factor structure best characterized the dimensionality of the ASRS in a large North American sample (clinical and non-clinical) aged 6–18. They used principal axis factoring with an oblique rotation to compare factorial congruence by sex, ethnicity, age and clinical status. They interpreted the three rotated factors as: Unusual Behaviours (26 items, Eigenvalue = 34.9), Self-Regulation (16 items, 4.2) and Social Communication (23 items, 3.8). All three factors had high internal consistency ($\alpha > 0.93$) and correlations among the three factors ranged from 0.53 to 0.65.

The Chinese version of the ASRS was developed by the team of the Research Special Fund for the Public Welfare Industry of Health of China using standard translation and back-translation methods with permission from Goldstein and Naglieri and approval from Multi-Health Systems (H. Zhou, Zhang, Luo, et al., 2017). The Chinese ASRS was initially piloted with a small sample of five Chinese parents of children in the 6–12 years age range who attended a local autism centre in Shanghai. All 71 items were translated using culturally equivalent expressions in the Chinese environment. H. Zhou, Zhang, Luo et al. (2017) recently reported an exploratory factor analysis of a Chinese version of the ASRS on a large sample of clinical and non-clinical children aged 5–11 who were rated by a parent, grandparent or other family member. After removing 12 unsatisfactory items with low or cross-loadings, they used weighted least mean squares factor extraction with Geomin (oblique) rotation, on the remaining 59 items. They found a similar three-factor solution although the order of the factors emerging was reversed – Social Communication (21 items, Eigenvalue = 17.1), Self-Regulation (14 items, 6.8) and Unusual Behaviours (24 items, 3.0). In this article, we used Rasch modelling (de Ayala, 2009; Rasch, 1960) to examine the performance of the Modified Chinese Autism Spectrum Rating Scale (MC-ASRS) – a 59-item Chinese version of the ASRS (B. Zhou et al., 2017; H. Zhou, Zhang, Luo, et al., 2017).

Rasch analysis

Rasch analysis offers several distinct advantages for the evaluation of any psychometric instrument (Pallant & Tennant, 2007; Tennant & Conaghan, 2007). Briefly, the Rasch model provides a template for the construction of interval scale measurement. When the data fit the Rasch model, they can be said to meet the requirements for a unidimensional scale where the items can be ordered hierarchically in terms of their ‘difficulty’. An ‘easy’ item is one that is endorsed by most participants and a ‘difficult’ item is endorsed by relatively few. Similarly, the process orders all the participants in terms of their ability along the same log-odds scale that it orders the items in terms of their difficulty. For example, on a scale of anxiety, all participants would be ordered from the least to the most anxious participant.

Good fit to the Rasch model is tested by a chi-square goodness-of-fit test. A number of attributes are tested in this process including: appropriate stochastic ordering of items, freedom from item bias (i.e. differential item functioning (DIF)), appropriate ordering of response options in polytomous items and testing local independence assumptions, including unidimensionality. This transformation (into an interval scale) is presented graphically in a person-item threshold distribution which gives an informative visual account of how well the items map on to the range of ability in the sample. For example, floor and ceiling effects are immediately apparent. Another advantage of Rasch analysis for clinical measures is the opportunity, within the same analytical framework, to identify item bias or DIF. This is where responding to an item systematically differs according to a personal or demographic characteristic *for people at the same level of the latent trait* being measured. For example, males or females with similar overall levels of anxiety responded differently (e.g. higher or lower) to an item on an anxiety questionnaire in a systematic way. In this study, we will examine each individual ASRS item for possible DIF or bias with respect to five person factors namely age, sex, caseness, rater’s relationship to child and geographical location (city). It is particularly important to identify DIF according to sex given the current interest in sex/gender differences in the diagnosis of ASD (e.g. Wood-Downie et al., 2021).

In practical terms, Rasch-derived scores offer clinicians and researchers at least three important advantages over raw scores (Hobart & Cano, 2009):

- The use of a genuine interval scale rather than an ordinal scale means that parametric statistics can be employed without violating their basic assumptions.
- The process of Rasch analysis uses the same logistic scale to order all persons in terms of the latent trait being measured and all items in terms of how easy or difficult they are.
- All items are examined for DIF or bias – which is when groups (e.g. males and females) respond differently in a systematic way to an item despite having the same overall level of the latent construct being measured.

The aim of this study was to test the Chinese version of the ASRS identified by H. Zhou, Zhang, Zou, et al. (2017) for fit to the Rasch model. When good fit is demonstrated, this permits the calculation of genuine interval level scores which improves the precision of a measure, justifies the calculation of individual change scores and meets the assumptions

required for parametric statistical analyses. In addition, we compared the raw ordinal scores and the Rasch interval scores for sensitivity and specificity.

Method

Sample

This analysis was based on a multi-centred pilot study of ASD completed between January and July of 2014 in four Provincial capital cities. To ensure the representativeness of the study sample, the protocol used a cluster sampling method, selecting local regular (i.e. not private) primary schools. All students of given age in the selected regular schools were eligible and invited for the pilot study to comprise the general sample. Those with an ASRS full-scale T-score over 60 were excluded from the general sample. During the same period, diagnosed ASD cases were recruited from each city at site PI's clinics to serve as the case group. An additional 220 cases of ASD identified from the subsequent main study were also included for the case group. All ASD cases met the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5) criteria; the methodology details of recruitment, questionnaire-based investigation and diagnosis have been described elsewhere (B. Zhou et al., 2017; H. Zhou, Zhang, Luo, et al., 2017; H. Zhou, Zhang, Zou, et al., 2017). The 2013 children rated comprised 1593 non-cases and 420 children diagnosed with ASD (see Table 1). The cases were mostly boys (83%) and the non-cases were more evenly balanced (51% male). Age ranged from 6 to 13 with a mean of 8.84 (SD = 1.83). Children were rated by their mother (1270, 63.1%), father 600 (29.8%), grandparent 56 (2.8%) or other relative 15 (0.7%). Data were collected in four cities namely Shanghai (469), Guangzhou (450), Changsha (385) and Harbin (474) with 235 missing data on this variable. Children with local residency and who had been living in the community for more than 6 months were eligible for this study. Descriptive statistics for ASD cases and non-cases are presented separately. in Table 1.

Measure

The ASRS asks parents and teachers to estimate the frequency of specific behaviours relevant to the diagnosis of ASDs during the past 4 weeks. It uses a 71-item rating scale with items scored from 0 (never) to 4 (very frequently). Twenty-two items are reverse scored. Separate forms exist for parent/teacher ratings and for children aged 2–5 and aged 6–18 years. This study is based on the parent's rating version of the ASRS.

Analysis

Rasch analysis was completed using RUMM2030 software (Andrich et al., 2009). The Likelihood-ratio test was performed initially to determine the most suitable type of Rasch model (i.e. Rating Scale or Partial Credit). The Rating Scale Model assumes distances between item response thresholds are the same across all items whereas the Partial Credit Model allows more flexibility. If the Likelihood-ratio test is significant, then it rejects the Rating Scale model and the Partial Credit is appropriate. The Rasch Partial Credit Model discrimination parameters are fixed and are all equal across all item thresholds.

Rasch analyses were conducted initially for the total 71-item ASRS and then separately for each of the three subscales previously identified by H. Zhou, Zhang, Luo et al. (2017) who used exploratory factor analysis. Finally, we also tested a single 59-item scale comprising three superitems representing the three subscales, to determine if a total or full-scale Rasch score was also legitimate.

Rasch analysis is completed in an iterative fashion until satisfactory overall and individual item fit to the model and evidence of unidimensionality is obtained. Overall and individual item chi-square fit statistics that demonstrate interaction between the item and a trait should be non-significant ($p > 0.05$). However, the overall item–trait interaction chi-square and p-values generated by RUMM2030 software may be misleading in larger samples (Müller & Kreiner, 2015), and Hagell and Westergren (2016) suggested that estimation of type I errors is only accurate if $n < 500$. Consequently, we tested each of the best fitting models obtained using the full sample on a randomly selected subsample of $n = 250$ (e.g. see Medvedev et al., 2018). In addition, we calculated a second fit statistic for each analysis – the root mean square of approximation (RMSEA) – which is less influenced by sample size with RMSEA < 0.2 reflecting good fit to the Rasch model (Tennant & Pallant, 2012).

For an excellent fit to the Rasch model, the overall item and person fit residuals would be expected to have a mean around 0.00 and a standard deviation equal to approximately 1.00. Individual items should have fit residuals ranging between -2.50 and $+2.50$. Unidimensionality is tested using the method developed by Smith (2000). This method uses a principal component analysis of the residuals to create two subtests and then compares each participant on both subtests using the t-test. If more than 5% of participants are significantly different on the two subtests, then unidimensionality cannot be concluded.

Local dependency between items is examined by inspection of the correlations among item residuals and can be corrected by combining highly correlated items into a testlet or super-item (Nilsson & Tennant, 2011; Wainer & Kiely, 1987). When the basic criteria for fit to the Rasch model are met, the person-item threshold distribution of a final solution is investigated to establish how well the range of item difficulties covers the range of individual abilities on a latent trait. Finally, ordinal-to-interval transformation scores are calculated for users to transform ordinal data to an interval level scale.

In analysing the MC-ASRS and each of its three subscales, we followed these steps:

1. Examining the item responses to determine the most appropriate version of the Rasch model. The Rating Scale Model (Andrich, 1978) assumes the distances between item response category thresholds are all the same whereas the unrestricted Partial Credit Model (Masters, 1982) assumes these can vary within and across items. The appropriate model is identified by the Likelihood Ratio Test. If this test is significant ($p < 0.05$), then the Partial Credit Model must be used.
2. Checking overall model fit (item–trait interaction) and individual item fit.

3. Rescoring of items with disordered thresholds to see if this improves overall fit. A disordered threshold occurs when persons higher in the latent trait measured do not consistently endorse higher response options (i.e. 0, 1, 2, 3) for an item.
4. Resolving local dependency through creation of super-items. Local dependency occurs when responding on an item is influenced by another item independently of the latent variable concerned.
5. Removal of individual items where necessary to improve model fit.
6. Testing for unidimensionality (Smith, 2002). This is done by inspecting loadings on the first principal component of the residuals after the 'Rasch factor' has been removed and then creating two subtests based on the highest positive and negative loading items. All participants are then compared on the two subtests using the t-test and if more than 5% of tests are significant multidimensionality is inferred.
7. Testing for DIF by the person factors (age, sex, caseness, relative, city). DIF is present when persons at the same level of the latent variable being measured score systematically differently on an item, for example, when boys and girls with the same overall level of autistic spectrum behaviours score differently on an item.
8. After a satisfactory fit was achieved, unidimensionality confirmed and all DIF eliminated – raw score to interval level conversion tables were constructed.

Once the conversion tables were completed, we calculated interval scores for all participants and compared their sensitivity and specificity with that of their raw total scores using receiver operating characteristic (ROC) curve analysis.

Community involvement

The local community was actively involved in this study, including parents, teachers, non-state-owned special schools, charitable foundations and local government. They provided important support in the field, facilitated data collection and advised on local implementation of the survey tool (ASRS). However, they were only involved in the data collection and survey tool localization/implementation, but *not* in the study design or interpretation of results.

Results

Initial examination of full 71-item ASRS and 59-item Chinese version

Table 2 presents the overall fit statistics for each of the main analyses of the full ASRS and the three subscales. The suitability of the Partial Credit Model for Rasch analysis was confirmed by the significant Likelihood-ratio test (χ^2 (df 209) = 6016.10, $p < 0.001$). The initial analysis of the full 71 items showed poor overall fit to the model as shown by a large and significant chi-square value (see Table 1, Analysis 2). Of the 71 items, 22 showed poor individual item fit with fit residuals outside the accepted range (i.e. >-2.5 and <2.5). In addition, the full 71-item scale failed the test for unidimensionality. As the chi-square

statistic is highly sensitive to sample size and goodness of fit seeks to achieve $p > 0.05$, we repeated the analysis with a randomly selected subsample of 250 children. While this sample size adjustment lowered the chi-square value, it still indicated poor model fit with a large, significant chi-square (Table 2, Analysis 2). Consequently, we next examined the reduced set of 59 items identified by H. Zhou, Zhang, Luo et al. (2017) using traditional factor analysis methods. However, this analysis also resulted in a large, significant chi-square value and indications of multidimensionality (Table 2, Analysis 3). Consequently, the next series of iterative analyses focused on testing fit to the Rasch model for each of the three ASRS subscales identified by H. Zhou, Zhang, Luo et al. (2017) in the Chinese version.

Examination of the 21 items of the Social Communication Scale of the MC-ASRS

The initial Rasch analysis of the 21 social communication items identified as *Factor 1* by Zhou et al. resulted in a large, significant chi-square value (Table 2, Analysis 6). Table 3 presents the individual item fit statistics from this analysis with the 21 items ranked according to item difficulty (location) along with each item's category frequencies. A fit residual below -2.5 or above $+2.5$ is typically regarded as reflecting poor fit. The 'easiest' item is Item 32 *notice social cues* and the most 'difficult' is Item 31 *play with others*.

There were no items with disordered thresholds; hence, to improve fit, we created three super-items to reduce local dependency as a source of misfit. Three super-items were created (8, 31, 42; 28, 39; 33, 70) based upon high residual correlations among these items. Inspection of Table 3 shows that these items also had highly similar semantic content. This creation of three super-items resulted in an improved but nonetheless significant chi-square value ($\chi^2 = 566.729(153)$, $p < .001$). However, unidimensionality was demonstrated and there were no remaining high residual correlations present. The presence of six items with fit residuals $> +2.5$ suggested removing items to achieve satisfactory fit was necessary. After removal of items 5 and 61, Super item 3 (items 33 + 70) and then items 12, 23 and 32 in an iterative fashion, a significant but much reduced chi-square was observed ($\chi^2(99) = 261.63$) with acceptable unidimensionality (Table 2, Analysis 7). Repeating this analysis with a random subsample of 250 resulted in good fit with a small non-significant chi-square ($\chi^2(44) = 46.88$, $p = 0.35$) and unidimensionality (see Table 2, Analysis 8). However, testing for DIF across all five Person Factors (i.e. age, sex, caseness, relative, city) revealed significant DIF for item 15 by City. Consequently, this item was removed resulting in good overall model fit ($\chi^2(40) = 50.76$, $p = 0.11$), unidimensionality and no DIF observed for any of the four person factors.

Having demonstrated satisfactory fit to the Rasch model, a table for converting raw ordinal scores to interval scale scores for Social Communication was then constructed. Table 6 (Supplementary Material) presents the range of possible raw scores, their corresponding logit scores and the logit scores transformed to match the range of the 13-item Social Communication scale (Analysis 9, Table 2). Figure 2 presents the distribution of person locations (i.e. ability) on the Social Communication subscale plotted against the distribution of item threshold locations (i.e. difficulty) on the same logit scale. Cases and non-cases are plotted separately and show quite distinct distributions as expected. The social

communication subscale is generally well targeted to cover the range of ability with some evidence of small floor and ceiling effects.

Examination of the 14 items of the Self-Regulation Subscale of the MC-ASRS

The initial analysis of the 14 Self-Regulation items resulted in a significant chi-square value (Table 2, Analysis 10). Table 4 presents the individual item fit statistics from this analysis with the 21 items ranked according to item difficulty (location) along with each item's category frequencies. As four items had disordered thresholds, these were rescored but this had only minimal effect on overall model fit ($\chi^2(126) = 466.73$). Hence, the items' residual correlation matrix was inspected for evidence of local dependency – defined as a residual correlation 0.2 greater than the average of the correlations. Three super-items were created (see Table 2, Analysis 11). This resulted in an improved if significant fit ($\chi^2(99) = 345.57$, $p < .001$) with Person Separation Index = 0.87 and unidimensionality. However, repeating this analysis with a random sample of 250 did not result in satisfactory fit ($\chi^2(44) = 81.25$, $p < 0.001$), so the next step was to remove individual items demonstrating significant misfit. Item 6 was thus removed, and this resulted in a reduced chi-square value (Table 2, Analysis 12; $\chi^2(90) = 277.29$, $p < 0.01$) with PSI = 0.86 and unidimensionality. The analysis was then repeated on a random sample of 250 and good fit was evident ($\chi^2(40) = 53.74$, $p = 0.07$; Table 2, Analysis 13). However, as Item 1 showed DIF for both Caseness and Sex, it was necessary to remove this item to achieve good overall fit, unidimensionality, high PSI and complete absence of DIF (Table 2, Analysis 14). Figure 1 presents a graphical display of the evidence that Item 1 demonstrated DIF by sex. This figure shows that boys systematically score higher than girls at the same overall level of autistic behaviour across five class intervals.

A table for converting raw ordinal scores to interval scale scores for Self-Regulation was then constructed (see Table 7, Supplementary Material). Figure 3 presents the distribution of person locations (i.e. ability) plotted against the distribution of item threshold locations (i.e. difficulty) on the same logit scale with Cases and non-Cases plotted separately. Targeting looks excellent for the Self-Regulation subscale with very little evidence of floor or ceiling effects (see Table 6, supplementary material, for Ordinal to Interval Conversion Table for Social Communication).

Examination of the 24 items of the Unusual Behaviour Subscale of the MC-ASRS

The initial analysis of the 24 Unusual Behaviour items resulted in a significant chi-square value ($\chi^2(216) = 1562.23$). Table 5 presents the individual item fit statistics from this analysis with the 24 items ranked according to item difficulty (location) along with each item's category frequencies. Seven items had disordered thresholds and were rescored. However, this has no effect on goodness of fit and so item residual correlations were then examined for evidence of local dependency. Two super items were created comprising items 18, 19, 66 and 64 and items 20 and 21, respectively. This improved the chi-square value ($\chi^2(180) = 1263.34$) which was still significant, however. As there was no evidence of local dependency, it was necessary to delete items with high positive fit residuals. After removing item 51, the first Super item and then item 24, a much reduced if still significant chi-square value was observed ($\chi^2(153) = 422.81$, $p < 0.001$). Repeating this analysis with a random

sample of 250 participants resulted in a non-significant chi-square ($\chi^2_{68} = 77.98$, $p = 0.19$, a high PSI (0.89) and unidimensionality. However, there was evidence of DIF on one item (Super item 2) for the person factor Caseness. Removal of this item resulted in overall good fit, a high PSI with acceptable unidimensionality and no evidence of DIF on any person factors.

A table for converting raw ordinal scores to interval scale scores for Unusual Behaviours is presented in Table 8 (Supplementary Material). Figure 4 displays the distribution of person locations (i.e. ability) plotted against the distribution of item threshold locations (i.e. difficulty) for Cases and non-Cases. Once again targeting looks very good with some suggestion of a slight floor effect for both groups.

Examination of the full 59-item MC-ASRS total score using three super items

Having found a good fit to the Rasch model for each of the three subscales of the ASRS, we then sought to see if the full 59-item scale could also fit the model by creating three super items based on the three subscales (Social Communication, Self-regulation, Unusual Behaviour). This resulted in a small but significant chi-square value (Table 2, Analysis 4), which when repeated with a randomly selected sample of 250 children demonstrated acceptable fit, unidimensionality, a high PSI and no DIF on any of the five person factors (Table 2, Analysis 5). An ordinal to interval conversion table for the full scale 59-item Chinese Version ASR is provided in Table 9 (Supplementary Material).

A comparison of sensitivity and specificity for raw scores and interval scores

The ROC analysis of the 59-item Chinese ASRS raw scores involving 352 cases and 1407 non-cases yielded an area under the curve (AUC) of 0.835 (95% confidence intervals (CIs): 0.808–0.861). Using the method that maximizes Youden's index (sensitivity + specificity – 1), we established that the optimal cut-off for raw scores was 90.5, which resulted in 71.9% sensitivity and a specificity of 81.3%. The analysis of Rasch transformed interval scores resulted in an AUC of 0.834 (95% CIs: 0.808–0.861). The optimal cut-off was 116.5 with identical sensitivity and specificity. However, this diagnostic comparison was completed solely to compare the two sets of scores in terms of predictive validity – since Rasch analysis is primarily about the internal structure or validity of a scale. It is important to note that the raw total score is never recommended to be used directly for screening ASD children. The T score of the ASRS should always be used for diagnosis of individual children.

Discussion

The aim of the present research was to examine the psychometric properties of the Chinese ASRS using Rasch analysis. We used super items based on groups of locally dependent items and then item deletion when necessary to achieve good fit to the model for each of the three subscales previously identified by H. Zhou, Zhang, Luo et al. (2017) and for the full 59-item MC-ASRS scale. The resulting conversion tables enable the use of genuine unidimensional, interval level scores for the total score and all three subscales. Reliability was high with PSI values, a statistic similar to Cronbach's alpha, ranging from 0.83 to 0.89 for the three subscales and 0.79 for the total scale.

The person-item threshold distribution maps the range of person ability or trait being measured (i.e. ASD) against the location or difficulty level of each individual item threshold – on the same logistic scale. These distributions (Figures 2 to 5) show that the ASRS and its three subscales are generally well targeted for assessing the range of autistic behaviour that might be encountered in children in China. At the same time, there was some evidence of a small ceiling effect on the Social Communication subscale with some children scoring above the location of the most difficult item threshold and some degree of floor effect on all three subscales. However, the floor effect is modest and arguably not a major limitation since it is only an issue at the low or asymptomatic end of the scale – in other words where there are few if any behavioural manifestations of ASD.

One important feature of the present work is that we are able to identify a full-scale version of the ASRS and its three subscales that are all free of DIF in relation to the five person factors recorded namely age, sex, caseness, relative rating the child, and city. DIF represents bias in an item and is a serious problem where it is not identified and addressed. The four Rasch scales identified here were all free of DIF for the five person factors (age, sex, caseness, relative, city) recorded. This means that meaningful comparisons can be made on the four interval scales according to the person factors without risk of bias, for example, when comparing mean scores for boys and girls. The three Rasch-derived subscales also involved a reduced number of items which has the advantage of providing a reliable, unidimensional score with a shorter scale and less burden on participants.

Another important contribution of this study is the detailed information it provides on the relative ‘difficulty’ of each item when used with children in China. A difficult item refers to one which relatively few children score high on. This is reflected in Tables 3 to 5 by items where relatively low numbers score in the response categories 3, 4 and 5. Conversely, an ‘easy’ item is one where a large number of children score above 2 on the Likert-type scale. In the Rasch output, it is the item location which reflects item difficulty with a high positive location meaning a more difficult item. For example, in Table 4, Self-Regulation, the items *Appear disorganized*, *Smell, taste or eat inedible objects*, and *Argue and fight with other children* are the three most difficult items in this subscale. The most difficult items are also likely to be the items most strongly indicating ASD.

While these results are encouraging, it remains to be demonstrated if and how Rasch scores can actually improve the precision or psychometric performance of the ASRS. For example, we found no meaningful difference between the scores in their diagnostic accuracy. At the same time, it has been argued that the advantages of Rasch scores are most evident with longitudinal data (i.e. repeated measures) and the present study is cross-sectional (Las Hayas et al., 2011). In particular, comparing individual change scores across time means subtracting one score from a score at an earlier time and this is not possible with rank order or ordinal data. Nonetheless, further work is required to establish the putative merits of Rasch scores over raw scores.

Limitations of the present study

There were several limitations to this study that need to be acknowledged and addressed in future research. The sample studied only included children from 6 to 13 years of age,

whereas the ASRS is designed for ages 6–18. It is important that future studies with the ASRS in China extend its use to include children from 14 to 18 and that appropriate validation and normative data is published for this age group. There is also a different version of the ASRS for younger children aged 2–5 and it could be valuable to translate and adapt that version for psychometric evaluation to enable its use in China.

This study would also have been enhanced if more information about both groups of children, cases and non-cases, had been collected – such as their language level, IQ and adaptive functioning. Unfortunately, this was not possible due to the scale of the study and the resources that demanded. Hence, future research on the Chinese version of the ASRS could profitably focus on exploring its relationships with these variables in the Chinese context.

Another limitation of this study is that it only includes data from the parent observation version of the MC-ASRS and it would be useful to also conduct Rasch analysis of the Teacher version. H. Zhou, Zhang, Zou et al. (2017) have already reported Chinese norms for the ASRS for children aged 6–12 for both Parent and Teacher ratings. Rasch analysis of both the Parent and Teacher ASRS rating scales will enable us to examine each of the 71 items for potential DIF across parent and teacher ratings, thereby eliminating potential sources of bias and hopefully improving agreement between home and classroom ratings.

It could also be argued that the use of super-items for good fit is a limitation since this was not achievable using just the individual items. The use of super-items or ‘testlets’ has become increasingly common in Rasch analysis as a method for addressing the issue of local dependency among items (Nilsson & Tennant, 2011; Wainer & Kiely, 1987). Local dependency means highly correlated items that can inflate reliability. In Rasch analysis, this process involves adding together responses from small groups of items which have high residual correlations among themselves. However, this can result in some loss of information since different combinations of scores on the individual items comprising a super-item yield the same super-item score (Wang & Wilson, 2005). It is a similar process to the creation of item parcels in confirmatory factor analysis which has its critics (Marsh et al., 2013).

Future directions

The adaptation and psychometric evaluation of the Chinese version of the ASRS represents an important step towards improving diagnostic and assessment services for people with ASD and their families who live in China. However, there is more work that needs to be done before the ASRS can be fully utilized in China. As noted above, the Teacher version of the MC-ASRS needs examination with Rasch analysis. Generalizability Theory analyses could also be fruitfully employed to identify how much measurement error is attributable to different measurement ‘facets’ (e.g. Person, rater, item and occasion) and to their interactions (e.g. rater $\times$ occasion) (Medvedev et al., 2017). In future studies, we intend to compare the factor structure of the ASRS from the data collected in China with other countries (e.g. New Zealand) to test for factorial invariance cross-culturally.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We would like to thank all the parents, children and families who participated in this project.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship and/or publication of this article: This research was supported by funding from the New Zealand Ministry of Education's New Zealand-China Tripartite Partnership Fund. This study was supported by Nation Health and Family Planning Commission of the People's Republic of China (Grant number: 201302002; [ClinicalTrials.gov](https://clinicaltrials.gov) number [NCT 02200679](https://clinicaltrials.gov)). Data collection by Dr Luo's team in Changsha City was supported the Key R & D Program of Hunan Province (2017SK2032).

References

- Andrich D, Sheridan B, & Luo G (2009). RUMM 2030. RUMM Laboratory.
- Baio J, Wiggins L, Christensen DL, Maenner MJ, Daniels J, Warren Z, Kurzius-Spencer M, Zahorodny W, Rosenberg CR, White T, Durkin MS, Imm P, Nikolaou L, Yeargin-Allsopp M, Lee L-C, Harrington R, Lopez M, Fitzgerald RT, Hewitt A, ... Dowling NF (2018). Prevalence of autism spectrum disorder among children aged 8 years – Autism and Developmental Disabilities Monitoring Network, 11 sites, United States, 2014. *MMWR Surveillance Summaries*, 67(No. SS-6), 1–23. 10.15585/mmwr.ss6706a1
- de Ayala RJ (2009). *The theory and practice of Item Response Theory*. The Guilford Press.
- DeVellis RF (2006). Classical test theory. *Medical Care*, 44(11, Suppl. 3), S50–S59. [PubMed: 17060836]
- Goldstein S, & Naglieri JA (2010). *Autism spectrum rating scales*. Springer.
- Goldstein S, Naglieri JA, Rzepa S, & Williams KW (2012). A national study of autistic symptoms in the general population of school-age children and those diagnosed with autistic spectrum disorders. *Psychology in the Schools*, 49(10), 1001–1015.
- Goldstein S, & Naglieri JA (2013). *Autism Spectrum Rating Scales™ (ASRS®) Technical Manual*. Multi-Health Systems Inc.
- Hagell P, & Westergren A (2016). Sample size and statistical conclusions from tests of fit to the Rasch model according to the Rasch Unidimensional Measurement Model (RUMM) program in health outcome measurement. *Journal of Applied Measurement*, 17, 416–431. [PubMed: 28009589]
- Hobart J, & Cano S (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technology Assessment*, 13(12), iii, ix–x, 1–177. 10.3310/hta13120
- Las Hayas C, Bilbao A, Quintana JM, Garcia S, Lafuente I, & IRTSS Cataract Group. (2011). A comparison of standard scoring versus Rasch scoring of the Visual Function Index-14 in patients with cataracts. *Investigative Ophthalmology and Visual Science*, 52(7), 4800–4870. [PubMed: 21527383]
- Marsh HW, Lüdtke O, Nagengast B, Morin AJS, & Von Davier M (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right – Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, 18, 257–284. [PubMed: 23834417]
- Masters G (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.
- Medvedev ON, Krägeloh CU, Narayanan A, & Siegert RJ (2017). Measuring mindfulness: Applying generalizability theory to distinguish between state and trait. *Mindfulness*, 8, 1036–1046. 10.1007/s12671-017-0679-0
- Medvedev ON, Turner-Stokes L, Ashford S, & Siegert RJ (2018). Rasch analysis of the UK Functional Assessment Measure on patients with complex disability after stroke. *Journal of Rehabilitation Medicine*, 50, 420–428.

- Müller M, & Kreiner S (2015). Item fit statistics in common software for Rasch analysis. Department of Biostatistics, University of Copenhagen. <http://www.pubhealth.ku.dk/bs/publikationer>
- Nilsson ÅL, & Tennant A (2011). Past and present issues in Rasch analysis: The Functional Independence Measure (FIMTM) revisited. *Journal of Rehabilitation Medicine*, 43, 884–891. [PubMed: 21947180]
- Pallant JF, & Tennant A (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46, 1–18.
- Rasch G (1960). Probabilistic models for some intelligence and attainment tests. University of Chicago Press.
- Smith EV Jr., (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 205–231. [PubMed: 12011501]
- Sun X, Allsion C, Wei L, Matthews FE, Auyeung B, Wu YY, Griffiths S, Zhang J, Baron-Cohen S, & Brayne C (2019). Autism prevalence in China is comparable to Western prevalence. *Molecular Autism*, 10, Article 7. 10.1186/s13229-018-0246-0
- Tennant A, & Conaghan PG (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheumatology*, 57, 1358–1362.
- Tennant A, & Pallant JF (2012). The Root Mean Square Error of Approximation (RMSEA) as a supplementary statistic to determine fit to the Rasch model with large sample sizes. *Rasch Measurement Transactions*, 25, 1348–1349.
- Wainer H, & Kiely GL (1987). Item clusters and computer adaptive testing: A case for testlets. *Journal of Educational Measurement*, 24, 185–202.
- Wang W-H, & Wilson M (2005). Assessment of differential item functioning in testlet-based items using the Rasch testlet model. *Educational and Psychological Measurement*, 65, 549–576.
- Wood-Downie H, Wong B, Kovshoff H, Mandy W, Hull L, & Hadwin JA (2021). Sex/gender differences in camouflaging in children and adolescents with autism. *Journal of Autism and Developmental Disorders*, 51, 1353–1364. [PubMed: 32691191]
- Zhou B, Zhou H, Wu L, Zou X, Luo X, Fombonne E, Wang Y, Yan W, & Xu X (2017). Assessing the accuracy of the Modified Chinese Autism spectrum Rating Scale and Social Responsiveness Scale for Screening Autism Spectrum Disorder in Chinese Children. *Neuroscience Bulletin*, 33(2), 168–174. [PubMed: 28258510]
- Zhou H, Zhang L, Luo X, Wu L, Zou X, Xia K, Wang Y, Xu X, Ge X, Jiang Y-Y, Fombonne E, Yan W, & Wang Y (2017). Modifying the Autism Spectrum Rating Scale (6–18 years) to a Chinese context: An exploratory factor analysis. *Neuroscience Bulletin*, 33(2), 175–182. [PubMed: 28238114]
- Zhou H, Zhang L, Zou X, Luo X, Xia K, Wu L, Wang Y, Xu X, Ge X, Jiang Y-H, Fombonne E, Yan W, & Wang Y (2017). Chinese norms for the Autism Spectrum Rating Scale. *Neuroscience Bulletin*, 33(2), 161–167. [PubMed: 28233147]

Figure 1.
Examination of item 1 for differential item function by sex.

Figure 2.
Person-item threshold distribution for 13-item Social Communication subscale (n = 2013).

Figure 3.
Person-item threshold distribution for 12-item Self-Regulation subscale (n = 2013).

Figure 4.
Person-item threshold distribution for 13-item Unusual Behaviour subscale (n = 2013).

Figure 5.
Person-item threshold distribution for 59-item ASRS (n = 1999).

Table 1.

Demographic characteristics of sample for ASD cases and non-cases.

	Participants N = 2013	ASD cases (n = 420)	Non-cases (n = 1593)	
Age (mean, median, range, SD)	8.79, 9.00, 6, 1.99	8.86, 9.00, 7, 1.78	8.86, 9.00, 7, 1.78	t = 0.66, df (604.5), p = 0.512
Sex (M:F)	347:72 (83%:17%)	813:780 (51%:49%)	813:780 (51%:49%)	$\chi^2 = 135.94(1)$, p < 0.001
Relative				
Mother	260 (62%)	1010 (63%)		
Father	93 (22%)	507 (32%)		
Grandparent	56 (13%)	31 (2%)		
Other	15 (3%)	10 (0.06%)		
Missing	0	35 (2%)		
City				
Shanghai	54	415		
Guangzhou	39	411		
Changsha	42	343		
Harbin	50	424		
Missing	235	0		

ASD: autistic spectrum disorder.

Table 2.

Summary of Rasch analyses of full 71-item ASRS and three subscales.

Analysis	Items	χ^2_{df} p	RMSEA	PSI	Uni	DIF
1. n = 2013	71 items	7544.59/639, p < 0.001	0.07	0.95	No	Yes
2. n = 250	71 items	1053.62/284, p < 0.001	0.04	0.94	No	Yes
3. n = 2013	59-Item Chinese version ASRS	5356.90/531, p < 0.001	0.10	0.95	No	Yes
4. n = 1999	59 items 3 Super items	106.89/27, p < 0.001	0.04	0.94	Yes	Yes
5. n = 250	59 items 3 Super items	17.89/12, p = 0.11	0.02	0.79	Yes	No
6. n = 1998	Social Communication 21 items	832.94/189, p < 0.001	0.04	0.93	No	Yes
7. n = 2013	Social Communication 14 items: 9 + 2 super items: 9, 10, 15, 43, 45, 47, 55, 56, 69, (8,31,42), (28, 39)	261.63/99/p < 0.001	0.03	0.89	Yes	Yes
8. n = 250	Social Communication 14 items: 9 + 2 super items	46.88/44/p = 0.35	0.02	0.89	Yes	Yes
9. n = 250	Social Communication 13 items: 8 + 2 super items: 9, 10, 43, 45, 47, 55, 56, 69, (8,31,42), (28, 39)	50.76/44, p = 0.11	0.02	0.89	Yes	No
10. n = 2005	Self-Regulation 14 items	466.96/126, p < 0.001	0.04	0.87	No	Yes
11. n = 2013	Self-Regulation 14 items: 8 + 3 super items: 1, 6, 16, 17, 27, 36, 60, 71 (30, 35), (37, 58), (44, 57)	345.57/99/p < 0.001	0.35	0.87	Yes	Yes
12. n = 2005	Self-Regulation 14 items: 7 + 3 super items: 1, 16, 17, 27, 36, 60, 71 (30, 35), (37, 58), (44, 57)	277.29, 90, p < 0.01	0.03	0.86	No	Yes
13. n = 250	Self-Regulation 13 items: 7 + 3 super items: 1, 16, 17, 27, 36, 60, 71 (30, 35), (37, 58), (44, 57)	53.74/40, p = 0.07	0.04	0.84	Yes	Yes
14. n = 250	Self-Regulation 12 items: 6 + 3 super items: 16, 17, 27, 36, 60, 71 (30, 35), (37, 58), (44, 57)	45.15/36, p = 0.14	0.03	0.83	Yes	No
15. n = 2013	Unusual Behaviour 24 items: 13, 18, 19, 20, 21, 22, 24, 25, 29, 38, 40, 41, 48, 49, 50, 51, 53, 54, 62, 63, 64, 65, 66, 67	1562.23/216, p < 0.001	0.05	0.88	No	Yes
16. n = 2013	Unusual Behaviour 24 Items: 18 + 2 super Items (18, 19, 66, 64) and (20, 21).	1263.34/180, p < 0.001	0.05	0.88	Yes	Yes
17. n = 1999	Unusual Behaviour: 18 items; 16 + 1 super item: 13, 22, 25, 29, 38, 40, 41, 48, 49, 50, 53, 54, 62, 63, 65, 67 and (20, 21)	422.81/153, p < 0.001	0.03	0.86	Yes	Yes
18. n = 250	Unusual Behaviour: 17 items; 16 + 1 super item;	77.98/68, p < 0.19	0.02	0.87	Yes	Yes
19. n = 250	Unusual Behaviour: 16 items: 13, 22, 25, 29, 38, 40, 41, 48, 49, 50, 53, 54, 62, 63, 65, 67	63.68/64, p = 0.48	0.00	0.86	Yes	No

RMSEA: root mean square of approximation; DIF: differential item functioning; ASRS: Autism Spectrum Rating Scale; PSI: person separation index.

Table 3.

Item location, item fit residual and category frequencies for the 21 social communication items of the ASRS ranked by location (item difficulty).

Item	Statement	Location	FitResid ^a	ChiSq	Prob	0	1	2	3	4
Q32	Notice social cues	-1.246	5.413	79.906	0.000	88	314	781	575	226
Q43	Show interest in the ideas of others	-0.563	1.477	28.806	0.001	181	642	767	287	110
Q39	<i>Care about what other people think</i> ²	-0.526	-1.464	19.034	0.025	247	604	674	341	126
Q10	Engage in tasks require sustained	-0.302	3.753	14.786	0.097	271	642	683	316	81
Q23	Keep a conversation going	-0.291	4.009	8.644	0.471	395	715	488	257	130
Q45	Understand age-appropriate humour	-0.206	2.597	8.917	0.445	620	569	381	238	182
Q28	Understand how someone else feels ²	-0.188	-3.229	34.164	0.000	348	652	634	266	82
Q61	Look at others when interacting	-0.166	10.603	56.772	0.000	376	795	481	250	84
Q69	Show good peer interactions	0.026	-4.666	46.636	0.000	505	763	408	224	89
Q15	Understand others point of view	0.03	-1.891	21.597	0.010	465	737	508	207	75
Q56	Start conversations with other	0.058	0.706	25.492	0.002	477	686	532	202	74
Q47	Listen when spoken to	0.095	-2.505	35.225	0.000	393	822	529	194	54
Q5	Follow instructions	0.197	9.916	176.327	0.000	546	756	498	144	45
Q8	Share fun activities with other ¹	0.246	-3.968	50.742	0.000	583	873	312	150	75
Q12	Play with toys appropriately	0.25	3.828	37.04	0.000	647	727	383	165	60
Q42	Share enjoyment with others ¹	0.316	-5.358	55.289	0.000	662	798	332	119	75
Q55	Smile appropriately	0.336	1.314	17.665	0.039	608	798	386	147	47
Q70	Respond when spoken to by other ³	0.461	-0.176	20.492	0.015	781	693	307	157	51
Q33	Respond when spoken to by adults ³	0.462	2.248	45.796	0.000	678	799	368	109	36
Q9	Look at others when talking with	0.499	-0.042	3.407	0.946	655	723	426	152	36
Q31	Play with others ¹	0.511	-3.675	46.207	0.000	744	798	257	135	55

^aItem fit residuals > 2.5 or < -2.5 are considered to represent a mis-fitting item.

Superscripts ¹, ² and ³ represent items comprising Super-items 1, 2 and 3, respectively.

Table 4.

Item location, item fit residual and category frequencies for the 14 self-regulation items of the ASRS ranked by location (item difficulty).

Item	Statement	Location	FitResid ^a	ChiSq	Prob	0	1	2	3	4
Q36	Careless mistakes in school work	-1.177	-0.337	27.648	0.001	49	415	783	552	199
Q30	Become distracted	-0.877	-5.906	70.771	0.000	144	440	699	524	194
Q35	Problems paying attention to	-0.208	-6.269	59.456	0.000	447	617	542	249	145
Q37	Talk too much about things adults	-0.083	3.685	33.307	0.000	304	764	631	232	64
Q58	Ask questions that were off topic	-0.038	2.297	18.335	0.031	296	777	646	218	55
Q57	Fail to complete tasks	-0.001	-0.845	25.356	0.003	349	902	508	170	69
Q60	Interrupt or intrude on others	0.056	-1.287	10.536	0.309	330	768	660	187	49
Q16	Learn simple tasks but quickly	0.073	1.997	17.536	0.041	448	707	582	196	64
Q17	Use language immature for his or her age	0.096	0.124	9.487	0.394	573	688	469	183	88
Q71	Fidgety when asked to sit still	0.169	-1.681	12.754	0.174	628	638	495	154	82
Q44	Leave homework or chores unfinished	0.176	4.887	19.489	0.021	774	674	329	117	106
Q1	Appear disorganized	0.372	-2.984	28.244	0.001	761	701	346	122	68
Q27	Smell, taste or eat inedible objects	0.687	3.666	16.594	0.055	1138	498	208	117	39
Q6	Argue and fight with other children	0.753	6.154	117.452	0.000	678	961	314	40	11

^aItem fit residuals > 2.5 or < -2.5 are considered to represent a mis-fitting item.

Table 5.

Item location, item fit residual and category frequencies for the 24 unusual behaviour items of the ASRS ranked by location (item difficulty).

Item	Statement	Location	FitResid	ChiSq	Prob	0	1	2	3	4
Q51	Insist on certain routines	-0.9	16.624	586.101	0.000	113	312	555	743	280
Q24	Insist on doing things same way	-0.569	5.782	56.348	0.000	196	443	716	502	144
Q13	Strong reaction to any change	-0.422	3.172	17.876	0.037	205	607	807	286	96
Q62	Overreact to loud noises	-0.324	0.715	22.592	0.007	297	660	671	260	108
Q63	Become upset if routines changed	-0.236	-2.259	60.591	0.000	253	677	766	235	66
Q48	Focus on one subject too much	-0.184	-0.433	27.726	0.001	299	721	722	190	70
Q41	Not understand why others don't	-0.151	1.028	15.395	0.081	414	742	565	179	98
Q29	Overreact to common smells	-0.108	5.26	110.4	0.000	561	638	461	228	115
Q49	Need things to happen as expected	-0.085	1.237	7.951	0.539	319	659	671	289	53
Q40	Focus too much on details	-0.018	1.465	3.779	0.925	344	813	635	158	50
Q64	Choose to play alone	0.002	-2.724	29.978	0.000	588	712	420	186	94
Q54	Line up objects in a row	0.064	2.828	40.384	0.000	586	700	436	204	70
Q25	Overreact to touch	0.118	1.288	21.976	0.009	691	683	412	136	75
Q19	Social problems with children	0.126	-0.464	34.67	0.000	1022	516	212	130	123
Q67	Twirl, spin or bang objects	0.158	3.32	31.261	0.000	773	659	345	146	76
Q50	Talk too much about things other	0.172	-1.696	30.382	0.000	494	963	354	140	47
Q65	Insist on keeping certain objects	0.185	2.798	30.274	0.000	828	613	321	165	74
Q53	Fascinated with parts of objects	0.186	-3.529	57.315	0.000	555	727	514	158	48
Q22	Become obsessed with details	0.206	-3.915	58.448	0.000	682	691	430	134	62
Q38	Resist being touched or held	0.26	4.008	29.9	0.000	811	676	359	102	54
Q66	Have social problems with adults	0.314	-2.35	54.316	0.000	986	566	255	124	70
Q21	Repeat words/phrases out of	0.325	-6.525	114.12	0.000	954	620	247	116	70
Q18	Get into trouble with adults	0.426	-1.598	17.27	0.045	991	601	269	96	48
Q20	Use an odd way of speaking	0.453	-6.342	103.176	0.000	1078	567	211	91	58