



Published in final edited form as:

*J Child Psychol Psychiatry*. 2017 September ; 58(9): 1062–1064. doi:10.1111/jcpp.12801.

## Lingering questions about the Social Responsiveness Scale short form. A commentary on Sturm et al. (2017)

Aaron Kaat<sup>1</sup> and Cristan Farmer<sup>2</sup>

<sup>1</sup>Department of Medical Social Sciences, Northwestern University, Chicago, IL, USA

<sup>2</sup>National Institute of Mental Health Intramural Research Program, National Institutes of Health, Bethesda, MD, USA

The Social Responsiveness Scale (SRS) was originally developed to quantify autistic traits, represented by a single score indicating both likelihood of an autism spectrum disorder (ASD) diagnosis and severity of symptoms (Constantino, 2011; Constantino et al., 2003, 2004; Constantino & Todd, 2000, 2003; Constantino, Przybeck, Friesen, & Todd, 2000). However, the dual usage as a diagnostic screener and quantification of severity have resulted in an identity crisis. The SRS has been criticized for lacking specificity, at least among referred children, and for exhibiting inflated scores due to characteristics not specific to an ASD diagnosis, such as problem behaviors and low IQ (Aldridge, Gibbs, Schmidhofer, & Williams, 2012; Hus, Bishop, Gotham, Huerta, & Lord, 2013; Norris & Lecavalier, 2010). These findings challenge the construct validity of the SRS as a Level-2 diagnostic screener. Used to quantify severity of impairment, the SRS boasts few rigorous psychometric data to validate its structure, which was not empirically derived. Several studies have purported to assess the factorial validity of the SRS with principal components analysis (rather than the more appropriate factor analysis; Borsboom, 2006), and these have focused generally on confirming its unidimensionality and have failed to acknowledge evidence for secondary factors (Bölte, Poustka, & Constantino, 2008; Constantino et al., 2004; Wigham, McConachie, Tandos, & Le Couteur, 2012). An exception was a study by Frazier and colleagues (2014), who derived a multi-factor solution using item parcels. Recently, Sturm, Kuhfeld, Kasari, and McCracken (2017, this issue) sought to empirically derive a factor structure for the SRS (including both unidimensional and multidimensional models), shorten it, and improve its psychometric properties. They did so using item response theory (IRT) in what they described as an integrative data analysis (IDA; Curran & Hussong, 2009), which is a method of pooling item-level datasets from two or more sources.

This effort is an excellent contribution to the field, as it is the first empirically sound and clinically useful refinement of the popular scale. The proposed 16-item short form is a brief, unidimensional scale, which the authors note “may be used as a quantitative measure to track social communication over development, treatment, or yield even more precise association with genetic and biological markers” (Sturm et al., 2017). This is in line with the oft-repeated assertion that the SRS measures ASD traits that are *continuously distributed*

throughout the general population, with a categorical ASD diagnosis only representing the most severe end of the continuum (Constantino, 2011; Constantino et al., 2000, 2003, 2004, 2006; Constantino & Todd, 2003; Frazier et al., 2009). Although the *continuous* distribution of ASD traits is conceptually consistent with the dimensional approach to understanding psychopathology, in practice, most of our psychometric and statistical methods additionally specify that this distribution is *normal*. When a population sampled by a study can be reliably grouped into subpopulations (e.g., males and females, behavior problems or not), an additional concern arises: that measurement invariance holds, or to state it another way, that there is no differential item functioning (DIF) between subpopulations. These two points are the crux of this communication.

First, the IRT model used by Sturm et al. assumes that the latent trait (i.e., social impairment) is normally distributed in the population. If this is violated, other techniques must be used (e.g., latent mixture modeling, Muthen & Asparouhov, 2006; unipolar IRT, Lucke, 2013; or estimating the latent densities, Woods, 2007; Woods & Thissen, 2017). Further, we note that the IDA technique used by Sturm et al. involves the pooling of data samples from several sources. In IDA, all statistical assumptions, including normality, must be tested within each of the data sources (i.e., National Database for Autism Research, Autism Genetic Resource Exchange, Simons Simplex Collection, and the Interactive Autism Network), as well as within subsamples from each data source (e.g., male versus female) (Curran & Hussong, 2009; Curran et al., 2008). There is no indication in Sturm et al. (2017) that this assumption testing was completed. By lumping the samples and assuming only one latent (normal) population distribution for probands and controls, the item parameters may be substantially biased and are therefore uninterpretable.

We do not raise this point for purely pedantic purposes, but instead because we see evidence that individuals with and without ASD are not from the same population, which suggests that it was inappropriate to lump the data sources as done therein. The SRS scores are *not* normally distributed in the general population. As a demonstration of this, we show that in at least one data source used in the IDA—the Simons Simplex Collection, which comprises simplex ASD probands and their unaffected siblings—the distribution is clearly multimodal (see Figures 1a and 1b).

Interestingly, the shape of the overall distribution in Figure 1a (solid density line) is similar to the distribution described by the scale authors (e.g., Constantino, 2011). It would seem, then, that at least two non-normal distributions are masquerading as one skewed distribution. In the Simons Simplex Collection data, we find evidence for three types of individuals: siblings who have no social impairment (21% of siblings; Figure 1b, dotted density line); siblings who have little impairment with a highly skewed distribution (79% of siblings; Figure 1b, dashed line); and a group of ASD probands, for whom a relatively normal distribution is suggested (Figure 1b, dot-dash line). We are not the first to suggest a mixture distribution: at least two studies suggest the presence of subpopulations, documented as multiple latent classes (Constantino, Przybeck, Friesen, & Todd, 2000; Frazier et al., 2009). We note that the Simons Simplex Collection is not a general population sample, and ascertainment bias may have occurred. Nonetheless, given the consistency across

publications, we suspect that a multimodal distribution would have emerged had subpopulation normality been tested when developing this short form.

Given this evidence for multiple subpopulations rather than one single population, our second point is that we must then be concerned about the consistency of measurement properties across these groups. Assuming that DIF did not occur between datasets—which they should have tested in the context of an IDA (Curran & Hussong, 2009)—the authors were successfully able to cull offending DIF items for those with or without language impairment, males and females, those with high or low IQ, and behavioral issues. However, DIF by diagnostic category (probands versus controls) was not examined. This was curious, as the only other study to address the consistency of psychometric properties between those with and without ASD failed to demonstrate measurement invariance (a structural equation modeling approach similar to DIF) (Frazier et al., 2014). We attempted to quantify DIF of the short form, using proband and sibling data from the Simons Simplex Collection and the same DIF methods as Sturm et al., but we found that we could not complete the analyses without eliminating additional items.

The work done by Sturm et al. is at the cutting-edge of psychometric techniques, and the field will benefit from continued work like this. However, necessary IDA background analyses and important characteristics of the SRS were overlooked. What are the implications of ignoring DIF-by-diagnosis and non-normality in subpopulations? In the non-ASD population, ignoring these psychometric problems will result in inappropriate conclusions because the SRS development samples have included individuals with a range of impairment (or lack thereof), researchers may assume that the scores indicate whether a person has gotten better or worse, or that their symptoms are some magnitude better or worse than another's. However, because the distributional data suggest non-normal subpopulations that are not comparable due to DIF, all one can really say about an individual score is whether it is consistent with what we would expect from someone who has ASD. If we include individuals *without* ASD in our item calibrations and other psychometric exercises, as was done in the original SRS and in the more psychometrically advanced short form, we have no way to know what the scores of individuals *with* ASD mean (i.e., how the score relates to the level of the latent construct)—the parameters are biased.

In summary, although the distributional data suggest that the short form is not valid for use as a quantitative measure in individuals *without* ASD, the lack of DIF across important ASD subpopulations (sex, problem behavior, verbal ability, IQ) indicates that with further psychometric refinement, it may be a good choice for quantifying the severity of impairment in (and only in) individuals known *a priori* to have ASD.

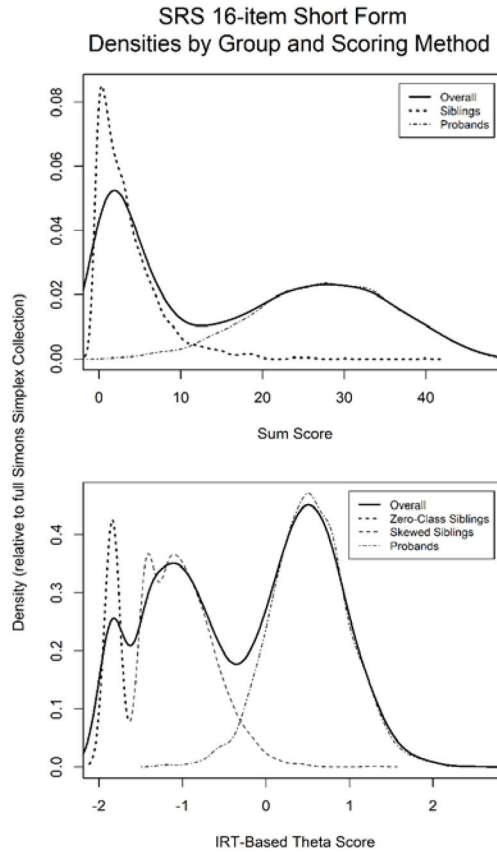
## Acknowledgments

This work was supported in part by the Intramural Research Program of the National Institute of Mental Health, National Institutes of Health (CAF). This commentary was invited by the Editors of JCPP and has been subject to internal review. The authors have declared that they have no competing or potential conflicts of interest in relation to this article.

## References

- Aldridge FJ, Gibbs VM, Schmidhofer K, Williams M. Investigating the Clinical Usefulness of the Social Responsiveness Scale (SRS) in a Tertiary Level, Autism Spectrum Disorder Specific Assessment Clinic. *Journal of Autism and Developmental Disorders*. 2012; 42(2):294–300. DOI: 10.1007/s10803-011-1242-9 [PubMed: 21516433]
- Bölte S, Poustka F, Constantino JN. Assessing autistic traits: cross-cultural validation of the social responsiveness scale (SRS). *Autism Research*. 2008; 1(6):354–363. DOI: 10.1002/aur.49 [PubMed: 19360690]
- Borsboom D. The attack of the psychometricians. *Psychometrika*. 2006; 71(3):425. [PubMed: 19946599]
- Constantino JN. The Quantitative Nature of Autistic Social Impairment. *Pediatr Res*. 2011; 69(5 Part 2 of 2):55R–62R.
- Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM, Brophy SL, ... Reich W. Validation of a Brief Quantitative Measure of Autistic Traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *Journal of Autism and Developmental Disorders*. 2003; 33(4):427–433. DOI: 10.1023/a:1025014929212 [PubMed: 12959421]
- Constantino JN, Gruber CP, Davis S, Hayes S, Passanante N, Przybeck T. The factor structure of autistic traits. *Journal of Child Psychology and Psychiatry*. 2004; 45(4):719–726. DOI: 10.1111/j.1469-7610.2004.00266.x [PubMed: 15056304]
- Constantino JN, Lajonchere C, Lutz M, Gray T, Abbacchi A, McKenna K, ... Todd RD. Autistic Social Impairment in the Siblings of Children With Pervasive Developmental Disorders. *American Journal of Psychiatry*. 2006; 163(2):294–296. DOI: 10.1176/appi.ajp.163.2.294 [PubMed: 16449484]
- Constantino JN, Todd RD. Genetic Structure of Reciprocal Social Behavior. *American Journal of Psychiatry*. 2000; 157(12):2043–2045. DOI: 10.1176/appi.ajp.157.12.2043 [PubMed: 11097975]
- Constantino JN, Todd RD. Autistic traits in the general population: A twin study. *Archives of General Psychiatry*. 2003; 60(5):524–530. DOI: 10.1001/archpsyc.60.5.524 [PubMed: 12742874]
- Constantino JN, Przybeck TPD, Friesen D, Todd RD. Reciprocal Social Behavior in Children With and Without Pervasive Developmental Disorders. *Developmental and Behavioral Pediatrics*. 2000; 21(1):2–11.
- Curran PJ, Hussong AM. Integrative Data Analysis: The Simultaneous Analysis of Multiple Data Sets. *Psychological methods*. 2009; 14(2):81–100. DOI: 10.1037/a0015914 [PubMed: 19485623]
- Curran PJ, Hussong AM, Cai L, Huang W, Chassin L, Sher KJ, Zucker RA. Pooling Data from Multiple Longitudinal Studies: The Role of Item Response Theory in Integrative Data Analysis. *Developmental psychology*. 2008; 44(2):365–380. DOI: 10.1037/0012-1649.44.2.365 [PubMed: 18331129]
- Frazier TW, Ratliff KR, Gruber C, Zhang Y, Law PA, Constantino JN. Confirmatory factor analytic structure and measurement invariance of quantitative autistic traits measured by the Social Responsiveness Scale-2. *Autism*. 2014; 18(1):31–44. DOI: 10.1177/1362361313500382 [PubMed: 24019124]
- Frazier TW, Youngstrom EA, Sinclair L, Kubu CS, Law P, Rezaei A, ... Eng C. Autism Spectrum Disorders as a Qualitatively Distinct Category From Typical Behavior in a Large, Clinically Ascertained Sample. *Assessment*. 2009; 17(3):308–320. DOI: 10.1177/1073191109356534 [PubMed: 20040725]
- Hus V, Bishop S, Gotham K, Huerta M, Lord C. Factors influencing scores on the social responsiveness scale. *Journal of Child Psychology and Psychiatry*. 2013; 54(2):216–224. DOI: 10.1111/j.1469-7610.2012.02589.x [PubMed: 22823182]
- Lucke, JF. Positive Trait Item Response Models. In: Millsap, RE, van der Ark, LA, Bolt, DM., Woods, CM., editors. *New Developments in Quantitative Psychology: Presentations from the 77th Annual Psychometric Society Meeting*; New York, NY: Springer New York; 2013. p. 199–213.
- Muthen B, Asparouhov T. Item response mixture modeling: Application to tobacco dependence criteria. *Addictive Behaviors*. 2006; 31(6):1050–1066. doi:<http://dx.doi.org/10.1016/j.addbeh.2006.03.026>. [PubMed: 16675147]

- Norris M, Lecavalier L. Screening Accuracy of Level 2 Autism Spectrum Disorder Rating Scales. *Autism*. 2010; 14(4):263–284. DOI: 10.1177/1362361309348071 [PubMed: 20591956]
- Sturm A, Kuhfeld M, Kasari C, McCracken JT. Development and validation of an item response theory-based Social Responsiveness Scale short form. *Journal of Child Psychology and Psychiatry*. 2017; doi: 10.1111/jcpp.12731
- Wigham S, McConachie H, Tandos J, Le Couteur AS. The reliability and validity of the Social Responsiveness Scale in a UK general child population. *Research in Developmental Disabilities*. 2012; 33(3):944–950. doi:<http://dx.doi.org/10.1016/j.ridd.2011.12.017>. [PubMed: 22277583]
- Woods CM. Empirical Histograms in Item Response Theory With Ordinal Data. *Educational and Psychological Measurement*. 2007; 67(1):73–87. DOI: 10.1177/0013164406288163
- Woods CM, Thissen D. Item Response Theory with Estimation of the Latent Population Distribution Using Spline-Based Densities. *Psychometrika*. 2017; 71(2):281. doi: 10.1007/s11336-004-1175-8



**Figure 1. SRS Short Form Score Densities in the Simons Simplex Collection**

Score densities (i.e., proportion of the total sample with a given score) on the 16-item SRS short form are provided for the Simons Simplex Collection (probands  $n=2630$ ; siblings  $n=2273$ ). Panel 1a expresses the observed sum score distribution. Panel 1b illustrates the distribution of the IRT-based latent-trait *expected a posteriori* (EAP) (which is akin to a factor score in factor analysis). Both panels are clearly multi-modal, violating the assumption of an underlying latent normal distribution for the population, which results in the inadequacy of means and standard deviations for describing scores that are only ordinal (not interval or ratio) in nature. In both panels, the solid line is the overall sample density. In Panel 1a, the dotted line is for siblings and the dot-dash line represents probands. In Panel 1b, the dotted line shows the zero-class for siblings, the dashed line is the skewed distribution of siblings, and the dot-dash line represents the relatively normal distribution for probands.