Can the Social Behavior Questionnaire help meet the need for dimensional, transdiagnostic measures of childhood and adolescent psychopathology?

Murray, A ; Eisner, Manuel ; Ribeaud, Denis

DOI: https://doi.org/10.1027/1015-5759/a000442

Can the Social Behavior Questionnaire help meet the need for dimensional, transdiagnostic measures of childhood and adolescent psychopathology?

Abstract

Background: The shift towards transdiagnostic and dimensional approaches to psychopathology research has created a growing need for psychometric assessments that reflect this conceptualisation.

Aims: We aimed to test whether an omnibus measure of psychopathology: the Social Behavior Questionnaire (SBQ), has suitable properties to serve as a dimensional, transdiagnostic assessment.

Method: We used an item response theory approach to evaluate the reliable ranges of measurement of the psychopathology dimensions measured by the SBQ.

Results: For the dimensions of ADHD, Prosociality, Internalising and Externalising, the SBQ can provide a reliable measure for below average to very high levels in a normative sample.

Keywords: transdiagnostic, dimensional, psychopathology, social behaviour questionnaire, item response theory

Traditional diagnostic classification systems view psychopathological disorders as categorical and distinct; however, accumulating evidence has provided ample reason to question these assumptions (e.g. Kruger & Eaton, 2015). Phenotypic and genetic analyses suggest that many common psychopathological disorders may simply represent the extreme ends of quantitative traits that show meaningful variation in both clinical and non-clinical ranges (e.g. Wray, Lee, Mehta, Vinkhuyzen, Dudbridge & Middeldorp, 2014; Walton, Ormel & Krueger, 2011). In addition, high levels of comorbidity have suggested that systems of classification that acknowledge the presence of broader, transdiagnostic factors such as Internalising and Externalising may provide a more useful characterisation of psychopathology than traditional classification systems such as DSM 5 and ICD-10 (e.g. Kruger & Eaton, 2014). The dimensional and transdiagnostic approach to psychopathology research implied by such observations depends on assessment tools that can reliably measure symptoms across a range of diagnostic domains a in both clinically diagnosed and 'sub-clinical' individuals. Failing to capture the full range of symptom levels expressed in the population can impede the detection of associations with other relevant variables or the detection of change over time (e.g. Reise & Haviland, 2005). It can also result in spurious or masked statistical interactions due to truncated score distributions (e.g. Kang & Waller, 2005).

Most previous evaluations of the reliability of psychopathological assessments have focussed on overall test reliability; not whether the assessment can measure an appropriate *range* of values with sufficient reliability. This latter property can be assessed using item response theory (IRT) models which recognise that measurement precision can vary across trait levels. Applying IRT models to sets of items can identify the range of trait values for which these items are sufficiently reliable (henceforth the 'reliable range' of a set of items). In this study, we evaluated whether the Social Behavior Questionnaire (SBQ; Tremblay et al., 1991): an omnibus psychopathology inventory has appropriate reliable ranges of measurement to be used as a transdiagnostic, dimensional measure of psychopathology in children and adolescents.

Method

Participants and Measures

Data were from the Zurich Project on the Social Development of Children and Youth (z-proso; see http://www.cru.ethz.ch/en/projects/z-proso.html.). Z-proso began in 2004 when the children entered school aged around 7. Informed consent was obtained from parents at the beginning of data collection and from the children from age 13 onwards. Children were included in the target sample if they attended one of 56 schools selected for participation. Schools were selected using a stratified random sampling method that took into account school size and location. Teacher ratings of the children were obtained at 8 time points covering the entire school career of the children (age 7 to 15). For this study, the sample for each wave comprised participants who had at least some data on the SBQ at that wave. Sample sizes ranged from N=977 to N=1346.

We focussed on 39 SBQ items that were administered at all waves because these can be directly compared across time. These measure Prosociality including helping and empathy; Internalising including anxiety and depression; attention-deficit hyperactivity disorder (ADHD); non-aggressive conduct problems including stealing, lying, vandalism and opposition/defiance; and aggression including physical, indirect, instrumental/dominance, and reactive aggression. The SBQ was administered in German. Paraphrases of item content in English can be found in Murray, Eisner & Ribeaud (2016). All responses were on a 5-point Likert scale from *Never* to *Very Often*. The SBQ was administered to three raters: the child, a parent and a teacher; however, only teacher ratings were obtained across all waves in a consistent format and we, therefore, focus only on these in the current study.

Statistical Procedure

We used IRT to evaluate the reliable ranges of the SBQ dimensions. IRT provides a framework for linking item responses to underlying latent attributes in a mathematically precise manner. Specifically, the probabilities of endorsing response categories of items measuring a common trait are modelled as functions of respondents' trait levels and item properties. Depending on the phenotype, we used either unidimensional or bi-factor models with logit link functions, allowing

discrimination parameters allowed to vary across items and thus giving rise to graded response models (e.g. see Gibbons et al., 2007). In bi-factor models, each item loads on two dimensions: one specific and one general. We used bi-factor models to avoid inflated test information due to violations of local independence due to the presence of subscales (e.g. Chen & Thissen, 1997). We fit separate models for ADHD, Prosociality, Internalising and Externalising. The latter two are commonly studied as transdiagnostic factors; however, preliminary analyses, suggested that ADHD and Prosociality items did not fit well within either. Separate models were, therefore, also fit for these phenotypes. Further details are provided in Supplementary Materials.

Given the developmental nature of the dataset, we fit models separately at each wave without imposing invariance constraints. We used robust maximum likelihood estimation (MLR) to take account of clustering of participants within classes and non-normality of indicators. Models are summarised in Figures 1-4. For scaling and identification, we fixed the latent factor means and variances to 0 and 1 respectively. Model fit was evaluated by examining standardised residuals for the univariate response distributions of items. These provide a measure of the deviation between observed and model-implied response distributions. Standardised residuals $>|1.96|$ suggest significant deviations from model-implied distributions. Global fit measures are also available for IRT models, including the Pearson chi-square and the likelihood ratio tests; however, their *p*-values are accurate only for small numbers of possible response patterns and are, therefore, unlikely to be accurate for the models tested in the current study. Other promising global fit statistics for IRT models are still undergoing development and testing (e.g. Maydeu-Olivares, 2013). To provide information on global fit we, therefore, report fit statistics obtained from confirmatory factor analyses (CFA) of the covariance structures corresponding to each IRT model. These were estimated using weighted least squares means and variances (WLSMV) estimation to take account of the ordered-categorical nature of items. All models were estimated in *Mplus 7.0* (Muthén & Muthén, 2010).

To assess the reliable ranges of the SBQ dimensions, we examined their test information curves. We focussed on the general dimensions because the specific dimensions were measured with too few items to support adequate reliable ranges of measurement. Test information for the general

dimensions was computed conditional on the specific dimensions. As .70 is often considered a minimum level of acceptable classical test theory reliability, we evaluated the range of trait values for which the test information was greater than the corresponding test information value of > 3.33. The choice of the .70 threshold reflects the fact that SBQ is not used for diagnostic purposes but to track developmental changes in normative samples. As such, it is more important to have an acceptable level of reliability across a wide range of phenotypic levels than high reliability around a diagnostic threshold. For tests used for diagnostic purposes, higher reliability threshold would be required; however, the range of phenotypic values for which this needs to be met could be much narrower.

Results

Model fits and all input and output files are provided in Supplementary Materials. With the exception of the RMSEA values for the Prosociality and ADHD models, global fit statistics generally suggested that models fit well by conventional criteria i.e. TLI and CFI >0.95 and RMSEA<.08 (e.g. Hu & Bentler, 1999; Schermelleh-Engel et al., 2003). Standardised residuals for univariate response distributions were mostly <|1.96|; however, some specific areas of misfit were identified. In ADHD, for example, the highest response categories showed some evidence of misfit for 4 of the items at age 7. In addition, for Externalising, there was a concentration of misfit within the Age 11 wave and for Internalising, there was a concentration of misfit within the Age 9 wave. We did not make any model modifications on the basis of misfit identified so as to avoid capitalisation on chance. Full model results including specification, parameter values and item response functions are available on request from the first author.

Reliable ranges of measurement for each trait at each time point are provided in Table 1 in terms of the range of trait values for which the classical test theory reliability would be >.70. All dimensions showed a reliable range of measurement that spanned a reasonably wide range of phenotypic levels. Test information curves for the dimensions are provided in Supplementary Figures 1-4.

Discussion

We evaluated the reliable ranges of measurement of the Social Behavior Questionnaire. Given the shift towards a dimensional and transdiagnostic approach to conceptualising psychopathology, these kinds of investigations are important for ensuring that psychopathology measurement tools keep pace with evolving theoretical models. We found that the SBQ items generally provided reliable measures of ADHD, Prosociality, Internalising and Externalising for a sufficient range of trait levels to support their use in non-clinical populations. The results of the current study should allow researchers to gauge whether the studied subscales have adequate reliability for the trait levels anticipated in a given study and to help interpret the results of such studies. For example, null results may not be attributable to a lack of intervention effect when the subscale used is poorly calibrated to the levels of traits expressed in the sample.

Nonetheless, the skewness of the test information curves ADHD, Internalising and Externalising suggests that the SBQ could benefit from the addition of items measuring the 'positive' end of the psychopathology dimensions. Items capturing the tendency to experience positive mood states could complement the current Internalising items; items capturing affiliative behaviours, self-control, gratification delay or inhibition could complement the Externalising items; and items capturing sustained attention and situation-appropriate activity levels could complement the ADHD items. However, it may not always be meaningful to attempt to measure the 'low' or 'adaptive' end of psychopathological trait continua. Reise & Waller (2009) argued that many psychopathological traits are 'quasi-traits', defined as a unipolar trait where one end of the continuum represents severity while the other represents the absence of the trait. It will be important to address this possibility; however, doing so will require at least an attempt to write items measuring low/adaptive- end variation in psychopathological traits. If such items prove impossible to construct this may indicate support for the quasi-trait hypothesis.

Limitations

We evaluated the reliable range of the SBQ with regards to a certain set of latent factors; other researchers may seek to combine the SBQ items in different ways and would need to evaluate

the reliable range of the chosen combinations. The reliable range of measurement is also only one consideration in evaluating the appropriateness of a measurement tool for a given purpose: we did not consider other properties that may bear on construct validity, e.g., content or criterion validity. There are also a number of dimensions of psychopathology that the SBQ does not cover, such as eating disorders, psychotic disorders, autism, and most personality disorders.

Conclusions

The reliable range of measurement for the SBQ is adequate for use in general population samples for the traits of ADHD, Prosociality, Internalising and Externalising. The SBQ can help answer the call for dimensional, transdiagnostic psychopathology measures; however, the addition of items measuring certain trait levels would be beneficial.

References

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*,*22*, 265-289.

Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., ... & Stover, A. (2007). Full-information item bifactor analysis of graded response data. *Applied Psychological Measurement*, *31*, 4-19.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6,* 1-55.

Kang, S. M., & Waller, N. G. (2005). Moderated multiple regression, spurious interaction effects, and IRT. *Applied Psychological Measurement*, *29*, 87-105.

Krueger, R. F., & Eaton, N. R. (2015). Transdiagnostic factors of mental disorders. *World Psychiatry*, *14*, 27-29.

Maydeu-Olivares, A. (2013). Goodness-of-fit assessment of item response theory models. *Measurement: Interdisciplinary Research and Perspectives, 11,* 71-101.Murray, A. L., Booth, T., McKenzie, K., & Kuenssberg, R. (2015). What range of trait levels can the autism-spectrum quotient (AQ) measure reliably? An item response theory analysis. *Psychological Assessment.* Online First.

Murray, A. L., Eisner, M., & Ribeaud, D. (2016). The development of the general factor of psychopathology 'p-factor' through childhood and adolescence. *Journal of Abnormal Child Psychology*, 1-14.

Muthén, L.K., & Muthén, B.O. (2010). Mplus User's Guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén.

Reise, S. P., & Haviland, M. G. (2005). Item response theory and the measurement of clinical change. *Journal of Personality Assessment*, *84*, 228-238.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27-48.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online,* 8, 23-74.

Tremblay, R. E., Loeber, R., Gagnon, C., Charlebois, P., Larivee, S., & LeBlanc, M. (1991). Disruptive boys with stable and unstable high fighting behavior patterns during junior elementary school. *Journal of Abnormal Child Psychology*, *19*, 285-300.

Walton, K. E., Ormel, J., & Krueger, R. F. (2011). The dimensional nature of externalizing behaviors in adolescence: Evidence from a direct comparison of categorical, dimensional, and hybrid models. *Journal of Abnormal Child Psychology*, *39*, 553-561.

Wray, N. R., Lee, S. H., Mehta, D., Vinkhuyzen, A. A., Dudbridge, F., & Middeldorp, C. M. (2014). Research Review: Polygenic methods and their application to psychiatric traits. *Journal of Child Psychology and Psychiatry*,*55*, 1068-1087.

Table 1: Reliable ranges of measurement for SBQ dimensions from test information curves in standard deviation units

	ADHD		Prosociality		Internalising		Externalising	
Age	**Lower**	**Upper**	**Lower**	**Upper**	**Lower**	**Upper**	**Lower**	**Upper**
7	-2.11	4.14	-3.05	2.72	-1.13	4.03	-1.07	5.58
8	-1.33	3.25	-3.24	2.50	-0.95	4.69	-0.95	5.35
9	-1.23	2.89	-3.25	2.29	-1.24	5.18	-1.10	5.22
10	-1.24	2.86	-3.23	2.73	-1.21	4.47	-0.81	5.55
11	-1.34	2.98	-3.00	2.60	-1.20	3.87	-0.78	4.76
12	-1.21	2.98	-3.17	2.59	-1.17	3.90	-0.72	5.18
13	-1.47	3.15	-2.85	2.85	-1.25	4.64	-0.55	5.63
15	-1.40	3.16	-2.90	3.02	-1.17	4.34	-0.62	5.21

Figure 1: ADHD model

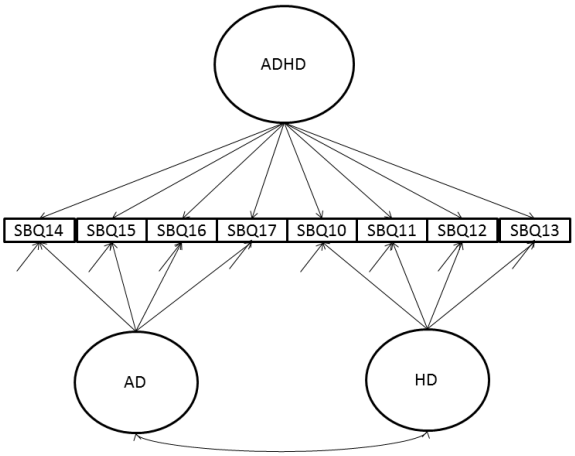


Figure 2: Prosociality model

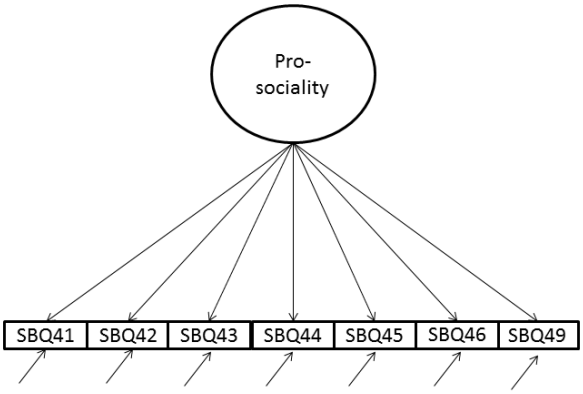


Figure 3: Internalising model

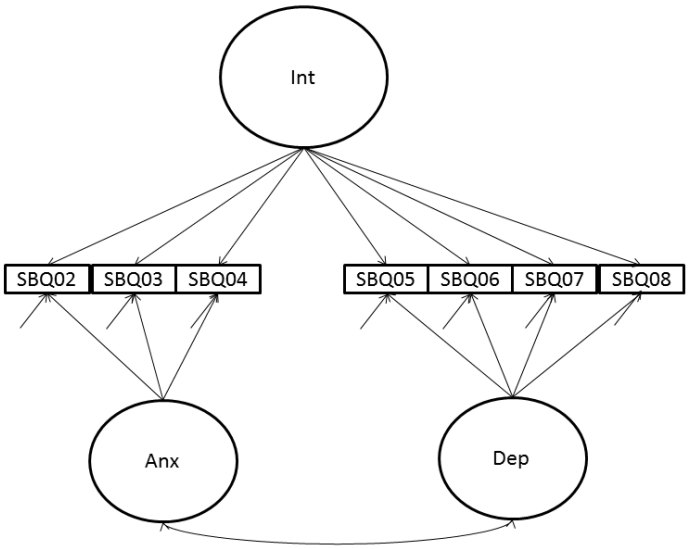
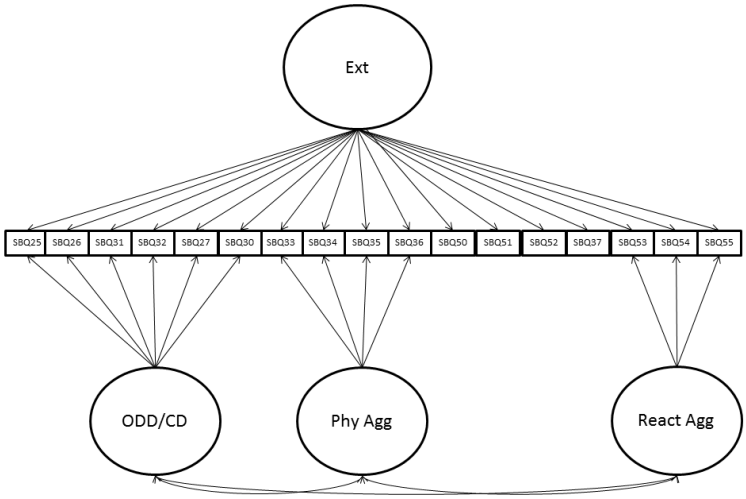


Figure 4: Externalising model



Supplementary Materials

Model Fitting

The models fit to the data are summarised in manuscript Figures 1-4. These were based on a series of preliminary exploratory factor analyses and will thus require independent verification in future research in other samples. The purpose of the analyses was also considered in selecting the best measurement model; namely, that the primary interest was in obtaining information about the general dimensions free from conflation with the specific dimensions reflecting the narrower subscales (e.g. Murray & Johnson, 2013). The fitted models correspond closely to the intended structure for the SBQ. In the ADHD models, a general ADHD factor was specified together with specific factors for the dimensions of Hyperactivity/Impulsivity and Attention Deficit. In the model for Internalising, a general Internalising factor was specified together with Anxiety and Depression specific factors. In the Externalising model, a general Externalising factor was specified together with specific Oppositional Defiant Disorder (ODD)/Conduct Disorder (CD), Reactive Aggression, and Physical Aggression factors. Initially, a Proactive Aggression specific factor was specified but this was removed to facilitate convergence. Due to the complexity of the bi-factor model, convergence problems are common and some model modifications may have to be made for pragmatic reasons (e.g. Maydeu-Olivares & Coffman, 2006). Similarly, although CD and ODD may be treated as distinct disorders in traditional diagnostics frameworks, their correlation was so close to unity in the current sample that they were combined into a single dimension to facilitate model estimation (e.g. APA, 2013). Finally, for Prosociality, a unidimensional model was judged sufficient to describe the covariation among items, therefore, a bi-factor structure was not necessary. In the bifactor models, the general factors were orthogonal to the specific factors but the specific factors were allowed to correlate.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Maydeu-Olivares, A., & Coffman, D. L. (2006). Random intercept item factor analysis. *Psychological Methods*, *11*, 344-362.

Murray, A. L., & Johnson, W. (2013). The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*, *41*, 407-422.

Table 1: Univariate standardised residuals for bi-factor graded response model for ADHD

Item Category	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 15
SBQ10								
1	0.46	-0.49	-0.41	-0.97	0.62	-0.17	0.58	-0.52
2	0.65	-0.61	1.24	1.40	0.88	0.79	0.94	0.05
3	-0.27	0.99	-0.04	0.09	-0.83	0.11	-1.02	0.23
4	-0.49	0.86	-0.62	-0.39	-0.82	-0.75	-0.81	0.58
5	-1.03	-0.58	-0.83	-0.18	-0.61	-0.52	-0.52	0.01
SBQ11								
1	0.61	-0.42	-0.37	-0.99	0.98	-0.09	0.87	-0.89
2	0.13	-0.54	1.00	0.99	0.70	1.17	0.69	0.51
3	-0.17	0.84	0.02	0.12	-0.97	-0.48	-0.80	0.26
4	-0.02	0.84	-0.38	0.46	-1.09	-0.69	-1.11	0.66
5	-1.04	-0.48	-0.75	-0.58	-0.60	-0.56	-0.70	-0.07
SBQ12								
1	0.29	-1.23	-1.69	-1.92	1.70	-0.14	0.44	-1.50
2	0.05	1.37	3.49	2.47	0.57	1.06	1.07	1.18
3	0.30	0.19	-0.27	0.22	-2.09	0.09	-0.45	-0.23
4	0.08	0.71	-1.66	0.19	-0.72	-0.75	-1.57	1.24
5	-1.26	-1.00	-0.41	-0.99	-0.76	-1.15	-0.60	0.40
SBQ13								
1	0.09	-0.71	-1.31	-1.43	1.76	0.14	0.69	-1.16
2	0.55	1.10	3.12	2.09	0.53	1.04	0.99	0.99
3	0.04	-0.28	-0.45	-0.04	-2.40	-0.56	-1.02	-0.21
4	-0.28	0.28	-1.23	0.50	-0.54	-0.56	-1.32	1.38
5	-0.96	-0.37	-0.65	-1.41	-0.54	-1.10	-0.50	-0.20
SBQ14								
1	0.64	-1.26	-1.05	-1.53	0.54	-0.52	0.29	-0.58
2	1.32	1.94	1.69	2.44	1.22	1.26	0.64	0.05
3	0.68	0.50	-0.13	-0.24	-0.36	0.03	0.63	0.24
4	-1.83	-0.83	0.23	0.51	-1.13	-0.93	-0.82	1.18
5	-2.44	-0.90	-1.37	-1.65	-1.30	-0.23	-2.06	-0.96
SBQ15								
1	0.25	-1.77	-0.61	-1.95	0.30	-0.69	0.26	-0.21
2	1.44	2.21	0.49	2.40	0.77	1.21	-0.02	-0.64
3	1.27	0.71	0.85	-0.27	0.53	0.37	1.64	0.29
4	-1.11	0.03	0.06	1.37	-0.81	-0.73	-0.56	1.25
5	-2.88	-1.56	-1.31	-1.55	-1.43	-0.47	-2.45	-0.80
SBQ16								
1	0.23	-2.18	-0.83	-2.08	0.03	-0.72	0.20	-0.38
2	1.75	3.06	1.02	3.53	1.62	1.54	0.36	-0.74
3	1.13	0.39	-0.01	-1.25	-0.10	0.27	1.27	0.63
4	-2.08	-0.20	1.02	1.46	-0.96	-1.47	-0.54	1.45
5	-2.46	-1.68	-1.53	-1.91	-1.46	0.10	-2.71	-0.88
SBQ17								
1	0.33	-1.23	-0.67	-1.59	0.35	-0.46	0.26	-0.29
2	1.73	1.50	1.07	1.87	0.75	1.07	0.33	-0.87
3	0.47	0.87	0.29	0.19	0.05	0.07	0.93	0.90
4	-1.87	-0.80	-0.30	0.50	-0.71	-0.78	-0.92	0.95
5	-2.35	-1.02	-1.04	-1.37	-1.37	-0.30	-1.88	-0.79

Note. Fitted model is shown in Figure 1.

Table 2: Univariate standardised residuals for unidimensional graded response model for Prosociality

Item Category	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 15
SBQ41								
1	-0.30	-0.86	-1.20	-0.20	-0.04	-0.19	-0.32	-0.29
2	0.46	1.07	1.21	1.50	1.40	0.71	0.82	0.99
3	0.29	0.70	1.47	-0.28	-0.55	-0.10	-0.22	0.09
4	-0.40	-1.20	-1.66	-1.00	-0.76	-0.31	-0.37	-1.07
5	-0.25	-0.10	-0.21	0.02	0.18	-0.15	0.03	0.37
SBQ42								
1	-0.61	-0.66	-1.06	-0.69	-0.51	-0.28	-0.32	-0.57
2	1.06	1.16	1.00	1.65	1.69	0.72	0.51	1.07
3	0.01	0.20	1.37	-0.26	-0.48	-0.08	0.10	0.11
4	-0.61	-1.12	-1.65	-1.03	-0.88	-0.08	-0.30	-1.15
5	-0.08	0.28	-0.06	0.38	0.40	-0.45	-0.18	0.66
SBQ43								
1	-0.59	-1.28	-1.56	-0.93	-1.25	-0.47	-0.23	-0.92
2	-0.10	0.07	0.10	0.86	1.40	0.10	-0.25	0.13
3	0.09	2.03	2.29	1.15	0.65	0.41	0.45	1.00
4	0.73	-1.52	-0.99	-1.42	-1.04	0.01	-0.10	-0.94
5	-0.69	-0.15	-0.75	0.17	0.20	-0.37	-0.07	0.50
SBQ44								
1	-0.22	-0.58	-0.49	-0.30	-0.30	-0.09	-0.32	-0.52
2	-0.38	-0.35	-0.64	-0.06	0.16	-0.20	-0.14	0.12
3	0.61	1.31	1.42	0.88	0.59	0.35	0.62	0.48
4	-0.18	-0.95	-0.98	-0.93	-0.76	-0.20	-0.45	-0.68
5	-0.16	0.07	0.21	0.33	0.31	0.05	0.13	0.71
SBQ45								
1	-0.66	-1.15	-1.55	-0.87	-0.88	-0.39	-0.11	-0.49
2	1.19	1.52	1.12	1.23	0.80	0.61	0.17	0.45
3	-0.06	0.89	1.83	-0.03	0.21	-0.22	0.13	0.48
4	-0.28	-1.73	-1.83	-0.75	-0.69	0.06	-0.08	-0.70
5	-0.47	-0.07	-0.18	0.23	0.37	-0.20	-0.23	0.18
SBQ46								
1	-0.95	-1.57	-1.85	-1.36	-1.36	-0.43	-0.44	-0.87
2	0.64	0.80	0.48	0.80	0.70	0.20	0.13	0.77
3	0.01	1.65	2.31	0.90	0.57	0.11	0.17	0.57
4	0.48	-1.67	-1.35	-1.09	-0.66	0.23	0.06	-1.10
5	-0.81	0.03	-0.63	0.19	0.27	-0.45	-0.21	0.54
SBQ49								
1	-0.32	-0.95	-0.96	-0.55	-0.62	0.00	-0.18	-0.68
2	-0.25	-0.17	-0.58	-0.17	0.22	-0.42	-0.61	-0.44
3	0.66	1.43	1.65	0.98	0.85	0.40	0.33	0.64
4	-0.31	-1.18	-1.00	-0.92	-0.94	-0.03	0.01	-0.27
5	-0.16	0.10	-0.07	0.32	0.32	-0.16	0.26	0.30

Note. Fitted model is shown in Figure 2.

Table 3: Univariate standardised residuals for bi-factor graded response model for Internalising

Item Category	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 15
SBQ2								
1	-0.08	-0.38	-0.11	-0.27	-0.31	-0.03	-0.10	0.29
2	0.11	-0.32	0.10	-0.18	0.09	0.12	0.15	-0.30
3	0.38	0.90	0.69	0.78	0.16	0.05	0.47	0.57
4	-0.28	0.17	-0.20	-0.15	0.07	-0.21	-0.74	-0.46
5	-0.49	-0.59	-1.57	-0.76	0.04	-0.04	-0.39	-0.91
SBQ3								
1	0.00	-0.66	0.07	-0.32	-0.28	0.15	0.30	0.19
2	0.45	0.91	0.97	0.39	0.63	0.62	0.59	0.44
3	0.06	0.39	0.27	0.52	-0.19	-0.34	-0.22	0.00
4	-0.48	-0.85	-1.35	-0.40	-0.51	-1.02	-1.22	-1.11
5	-0.68	-0.60	-2.07	-1.02	0.13	-0.19	-0.65	-0.75
SBQ4								
1	0.00	-0.86	-0.07	-0.20	-0.56	-0.11	-0.15	0.25
2	-0.16	0.81	0.76	-0.08	0.69	0.62	0.67	0.20
3	0.61	0.62	0.16	0.68	-0.30	-0.37	0.41	0.10
4	-0.21	-0.57	0.06	0.13	0.10	-0.16	-1.43	-0.16
5	-0.94	-0.72	-2.77	-1.59	0.23	-0.42	-0.57	-1.65
SBQ5								
1	-0.32	-1.30	0.00	-1.21	-0.66	-0.22	0.32	-0.45
2	1.93	2.77	2.31	2.00	1.39	0.89	0.90	2.46
3	-1.21	-0.82	-0.63	-0.29	-0.97	-0.47	-0.56	-1.40
4	-1.29	-0.82	-2.42	-0.46	0.17	-0.54	-1.41	-1.17
5	0.28	-1.18	-2.14	-0.96	0.28	-0.14	-0.40	-0.76
SBQ6								
1	0.08	-0.69	0.38	-0.98	-0.63	-0.11	0.41	0.01
2	1.43	1.90	1.80	1.45	1.10	0.92	0.32	2.01
3	-0.83	0.16	-0.47	0.20	-0.81	-0.49	-0.04	-0.83
4	-1.81	-1.75	-2.04	-0.68	0.35	-0.78	-1.07	-1.83
5	1.09	-1.17	-2.40	-0.62	0.47	0.07	-0.32	-0.40
SBQ7								
1	0.10	0.06	1.09	-0.41	0.09	0.17	0.63	0.60
2	0.96	1.20	0.72	1.08	0.07	0.38	0.44	1.02
3	-0.98	-0.05	-0.88	-0.18	-0.59	-0.56	-0.84	-1.31
4	-0.87	-2.26	-2.24	-0.86	0.45	-0.18	-1.01	-1.12
5	0.28	-1.00	-1.64	-0.53	0.56	-0.22	0.06	-0.20
SBQ8								
1	0.83	0.10	1.01	-0.33	0.03	0.46	0.71	0.86
2	0.56	1.75	1.35	1.15	0.32	0.25	0.69	1.03
3	-1.68	-1.63	-1.60	-0.64	-0.65	-0.80	-1.25	-1.68
4	-0.87	-1.31	-2.13	-0.43	0.21	-0.35	-1.25	-0.98
5	0.58	-0.95	-1.82	-0.79	0.28	-0.15	-0.21	-0.72

Note. Fitted model is shown in manuscript Figure 3.

Table 4: Univariate standardised residuals for bi-factor graded response model for Externalising

Item Category	Age 7	Age 8	Age 9	Age 10	Age 11	Age 12	Age 13	Age 15
SBQ25								
1	0.17	0.32	0.67	0.67	1.93	0.13	0.33	0.34
2	-0.05	-0.02	-0.24	-0.29	-1.15	0.04	-0.32	0.03
3	-0.20	-0.45	-0.71	-0.66	-1.53	-0.19	-0.19	-0.44
4	-0.10	-0.42	-0.31	-0.28	-0.63	-0.30	0.12	-0.42
5	-	-	-0.15	-0.11	-	-	0.12	-0.05
SBQ26								
1	0.06	-0.40	0.10	-0.01	1.89	-0.05	-0.03	-0.20
2	0.24	0.69	0.19	0.29	-0.03	0.22	0.07	0.54
3	-0.10	0.11	-0.39	-0.12	-1.27	-0.03	-0.06	0.09
4	-0.47	-0.40	-0.08	-0.55	-1.14	-0.13	-0.03	-0.47
5	-0.22	-0.81	-0.22	0.18	-2.20	-0.27	0.05	-0.75
SBQ27								
1	-0.26	-0.64	-0.01	-0.03	1.53	0.01	0.16	-0.89
2	-0.12	0.32	0.15	0.21	0.04	-0.26	-0.10	0.48
3	0.59	0.73	0.05	0.19	-0.71	0.31	0.19	1.05
4	0.16	0.11	-0.22	-0.55	-1.79	0.22	-0.51	0.17
5	-0.51	-0.57	-0.36	-0.41	-2.06	-0.22	-0.12	-0.98
SBQ30								
1	-0.47	-0.64	-0.04	-0.12	1.47	0.04	0.00	-0.53
2	0.14	0.29	0.27	0.22	-0.16	-0.12	0.32	0.52
3	0.57	0.86	-0.14	0.22	-0.65	0.07	-0.14	0.50
4	0.13	-0.11	-0.11	-0.57	-1.58	0.23	-0.57	-0.64
5	-0.24	-0.69	-0.32	-0.19	-1.68	-0.31	-0.01	0.32
SBQ31								
1	0.06	-0.35	0.23	0.21	2.14	-0.15	0.25	-0.41
2	0.15	0.94	0.07	0.16	-0.57	0.36	-0.13	0.80

3	-0.10	-0.19	-0.43	-0.60	-1.54	-0.08	-0.44	0.03
4	-0.35	-1.09	-0.10	-0.37	-1.27	-0.05	0.04	-0.62
5	-0.19	-	-0.32	0.19	-2.11	-0.62	0.28	-0.67
SBQ32								
1	-0.17	-0.37	-0.36	0.06	1.58	0.05	0.15	-0.46
2	0.24	0.26	0.60	0.06	-0.04	-0.15	0.11	0.38
3	0.00	0.23	0.06	0.02	-1.07	0.11	-0.14	0.39
4	0.04	0.07	-0.45	-0.22	-1.39	0.10	-0.43	-0.10
5	-0.22	-0.13	-0.20	-0.30	-1.33	-0.16	0.03	-0.29
SBQ33								
1	0.09	-0.52	0.16	0.27	2.17	0.34	0.68	-0.26
2	0.16	0.90	0.23	-0.02	0.35	-0.14	-0.33	0.30
3	-0.56	0.19	-0.37	-0.17	-1.69	-0.47	-0.96	0.24
4	0.16	-0.47	-0.17	-0.48	-2.18	0.02	0.33	-0.62
5	0.41	-0.83	-0.27	0.34	-2.20	0.53	1.75	0.89
SBQ34								
1	0.27	-0.66	0.02	0.58	2.42	0.25	0.74	-0.36
2	0.07	1.05	0.41	-0.32	0.11	0.03	-0.58	0.31
3	-0.71	0.38	-0.59	-0.39	-1.90	-0.69	-0.75	0.56
4	0.17	-0.83	0.06	-0.36	-2.26	0.13	0.45	-0.88
5	0.36	-0.86	-0.20	0.38	-2.08	0.49	2.20	1.06
SBQ35								
1	0.34	-0.20	0.36	0.49	2.80	0.28	0.90	-0.18
2	-0.24	1.01	0.05	-0.31	-0.49	-0.03	-0.87	0.36
3	-0.24	-0.55	-0.66	-0.33	-2.09	-0.55	-0.66	0.03
4	0.00	-0.59	0.02	-0.29	-2.08	-0.04	0.58	-0.76
5	0.06	-0.58	-0.15	0.38	-1.85	0.31	2.26	0.97
SBQ36								
1	0.19	-0.33	-0.07	-0.05	2.27	0.28	0.44	-0.18
2	0.24	0.43	0.53	0.17	-0.09	0.23	0.02	0.33
3	-0.52	-0.06	-0.58	0.01	-1.65	-0.80	-0.73	0.16

4	-0.32	0.00	0.05	-0.16	-1.96	-0.35	-0.54	-0.60
5	0.11	-0.12	-0.37	-0.22	-1.63	0.53	1.55	-0.24
SBQ37								
1	0.20	-0.10	-0.06	0.16	2.45	0.34	0.71	0.06
2	0.19	0.44	0.40	0.37	-0.62	-0.01	-0.24	0.31
3	-0.42	-0.42	-0.45	-0.29	-1.80	-0.59	-0.95	-0.10
4	-0.44	-0.17	0.09	-0.53	-1.72	-0.40	-0.51	-1.10
5	0.09	-0.09	-0.60	-0.79	-1.51	0.84	1.88	0.43
SBQ50								
1	0.08	-0.24	-0.19	0.03	1.67	0.09	0.09	-0.42
2	0.03	0.24	0.42	-0.08	0.03	0.22	0.05	0.20
3	-0.07	0.09	-0.06	0.18	-1.22	-0.37	-0.20	0.73
4	-0.23	-0.06	-0.22	0.17	-1.75	-0.41	-0.48	-0.31
5	0.10	-0.21	-0.47	-0.59	-1.54	0.65	1.27	-0.74
SBQ51								
1	-0.77	-0.50	-0.48	-0.31	1.65	-0.03	0.05	-0.45
2	0.39	0.12	0.28	-0.14	0.22	0.20	0.10	0.23
3	0.43	0.51	0.38	0.47	-0.85	-0.08	-0.02	0.43
4	0.28	0.17	0.06	0.60	-1.76	-0.40	-0.66	0.32
5	0.20	-0.03	-0.16	-0.38	-1.93	0.47	0.59	-0.49
SBQ52								
1	0.39	-0.40	-0.17	-0.13	2.49	0.19	0.49	0.05
2	-0.04	0.47	0.57	0.17	-0.44	0.39	0.04	0.12
3	-0.36	0.02	-0.38	0.11	-1.50	-0.61	-0.89	0.22
4	-0.49	-0.01	-0.13	0.44	-1.91	-0.62	-0.91	-1.09
5	-0.05	-0.11	-0.42	-0.91	-1.93	0.96	1.38	0.56
SBQ53								
1	-0.58	-0.65	-0.38	0.10	0.62	0.13	-0.34	-0.34
2	0.45	0.32	-0.19	-0.16	1.22	-0.22	0.38	0.35
3	0.62	0.48	1.19	0.43	0.39	0.88	0.54	0.46
4	-0.74	-0.03	-0.26	-0.30	-1.40	-0.55	-0.56	-0.60

5	0.31	-0.09	-0.89	-0.33	-3.01	-0.86	-0.37	-0.02
SBQ55								
1	-0.17	-0.73	-0.27	-0.12	1.15	-0.01	-0.42	-0.48
2	0.82	1.15	0.66	0.65	1.14	0.49	0.79	0.98
3	-0.52	-0.33	0.52	-0.25	-0.75	-0.01	-0.02	-0.04
4	-0.46	-0.04	-1.19	-0.37	-1.80	-0.91	-0.60	-0.84
5	0.05	-0.17	-0.79	-0.45	-2.76	-0.03	-0.17	-0.05
SBQ54								
1	-0.30	-0.47	-0.03	0.11	0.66	0.23	-0.37	-0.34
2	0.31	0.26	-0.21	-0.12	1.00	-0.43	0.42	0.22
3	0.45	0.49	1.03	0.40	0.44	0.79	0.52	0.53
4	-0.73	-0.26	-0.62	-0.37	-1.52	-0.56	-0.65	-0.59
5	0.08	-0.12	-0.93	-0.31	-3.07	-0.56	-0.44	0.04

Note. Fitted model is shown in manuscript Figure 4.

Table 5: Global fits for ADHD bi-factor CFA

Wave	CFI	TLI	RMSEA	WRMR
Age 7	0.999	0.997	0.059	0.359
Age 8	0.999	0.997	0.096	0.838
Age 9	0.999	0.996	0.098	0.896
Age 10	0.998	0.994	0.114	1.019
Age 11	0.998	0.996	0.116	0.930
Age 12	0.999	0.997	0.083	0.634
Age 13	0.998	0.996	0.105	0.779
Age 15	0.998	0.996	0.094	0.821

Table 6: Global fits for Prosociality unidimensional CFA

Wave	CFI	TLI	RMSEA	WRMR
Age 7	0.953	0.930	0.149	1.863
Age 8	0.959	0.939	0.133	1.816
Age 9	0.965	0.947	0.119	1.592
Age 10	0.962	0.943	0.147	2.276
Age 11	0.960	0.940	0.147	1.608
Age 12	0.976	0.963	0.156	1.951
Age 13	0.975	0.963	0.133	1.517
Age 15	0.961	0.942	0.156	2.046

Table 7: Global fits for Internalising bi-factor CFA

Wave	CFI	TLI	RMSEA	WRMR
Age 7	0.999	0.997	0.036	0.269
Age 8	1.000	0.999	0.021	0.227
Age 9	0.999	0.998	0.034	0.247
Age 10	1.000	1.000	0.000	0.158
Age 11	1.000	1.000	0.010	0.186
Age 12	1.000	0.999	0.023	0.185
Age 13	0.999	0.998	0.039	0.268
Age 15	1.000	0.999	0.024	0.194

Table 8: Global fits for Externalising bi-factor CFA

Wave	CFI	TLI	RMSEA	WRMR
Age 7	0.991	0.989	0.054	1.153
Age 8	0.982	0.976	0.061	1.217
Age 9	0.986	0.981	0.061	1.333
Age 10	0.990	0.987	0.055	1.238
Age 11	0.986	0.981	0.059	1.183
Age 12	0.989	0.986	0.058	1.083
Age 13	0.989	0.986	0.049	0.988
Age 15	0.981	0.975	0.049	1.016

Figure 1:

Test information curves for bi-factor graded response model for ADHD

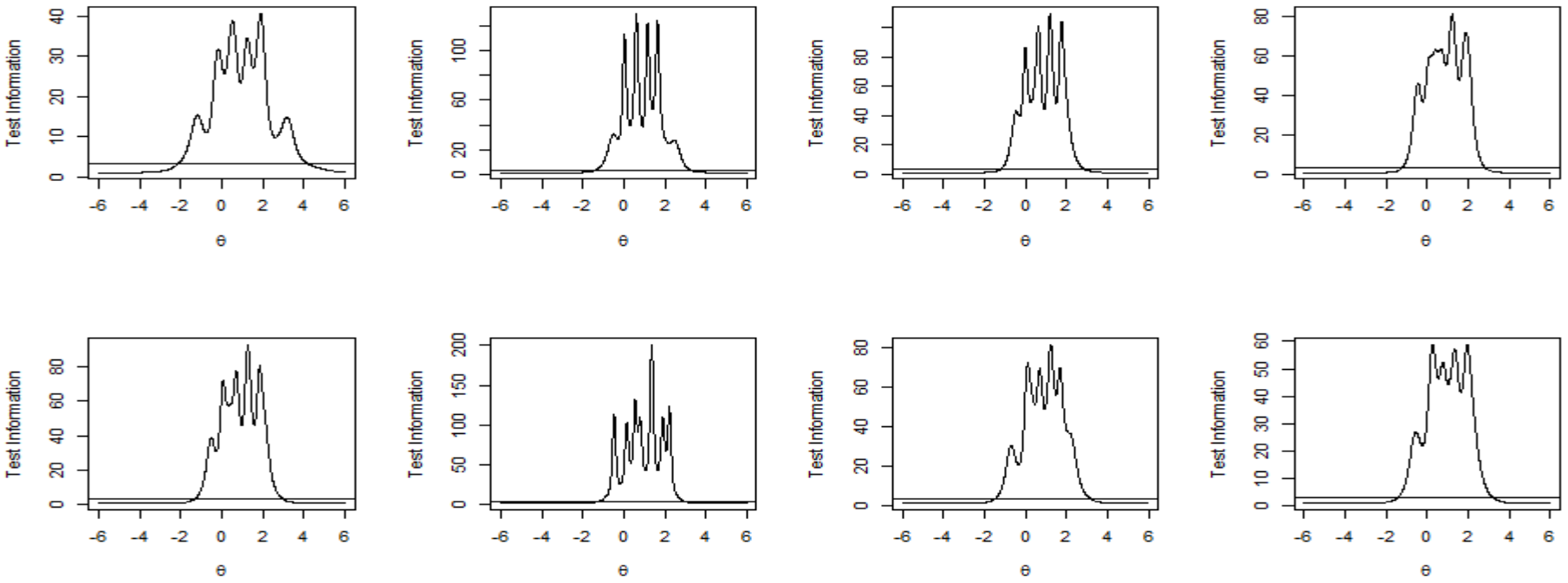


Figure 2: Test information curves for unidimensional graded response model for Prosociality

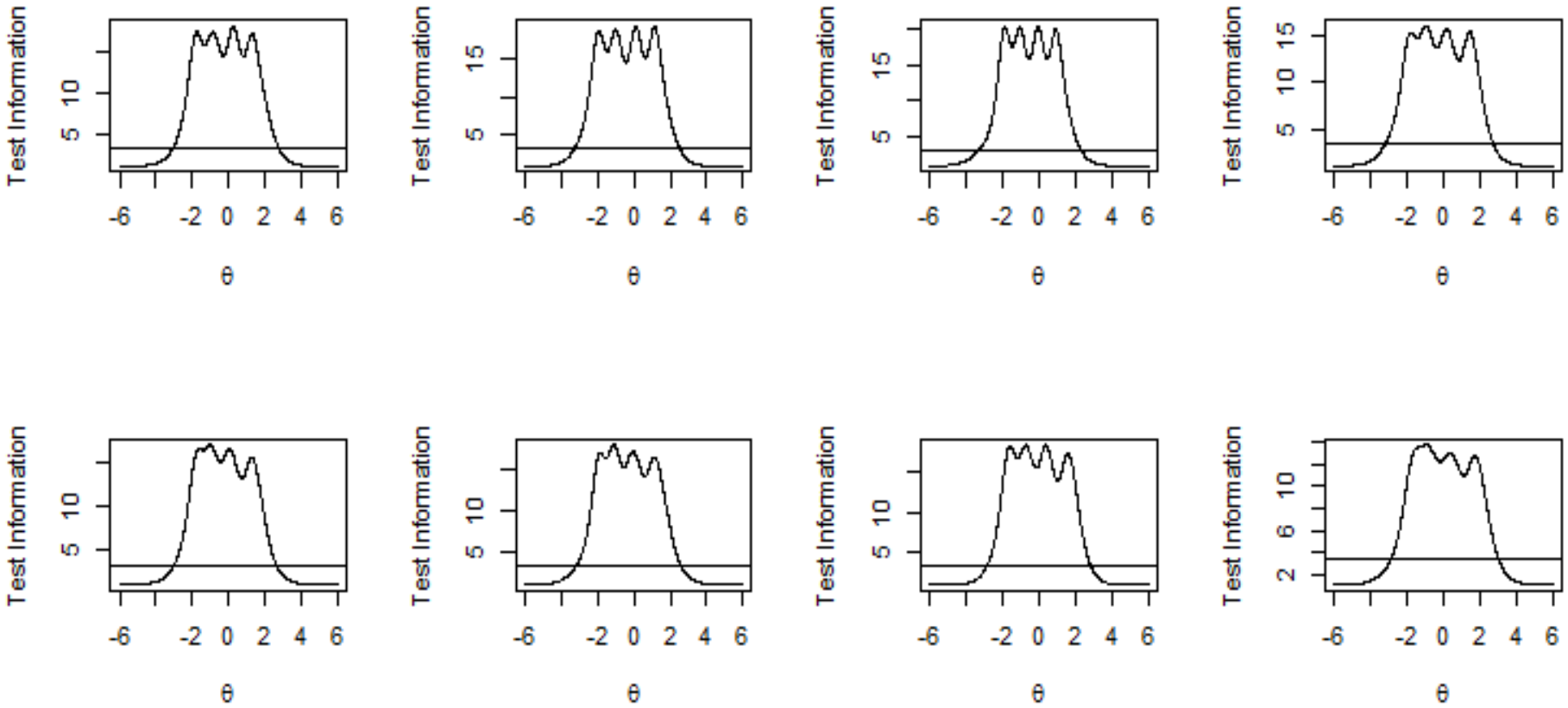


Figure 3:

Test information curves for bi-factor graded response model for Internalising

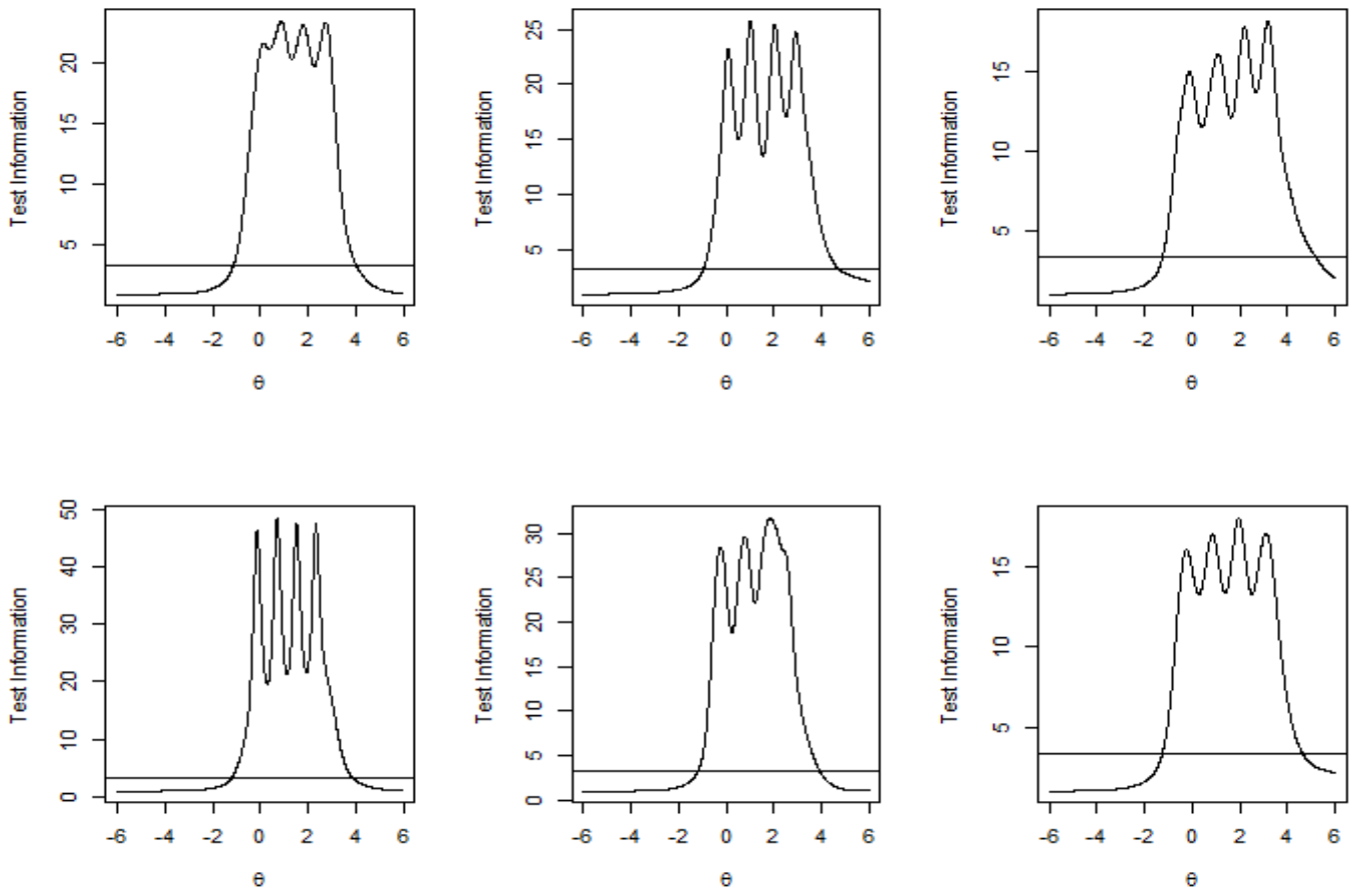


Figure 4:

Test information curves for bi-factor graded response model for Externalising.

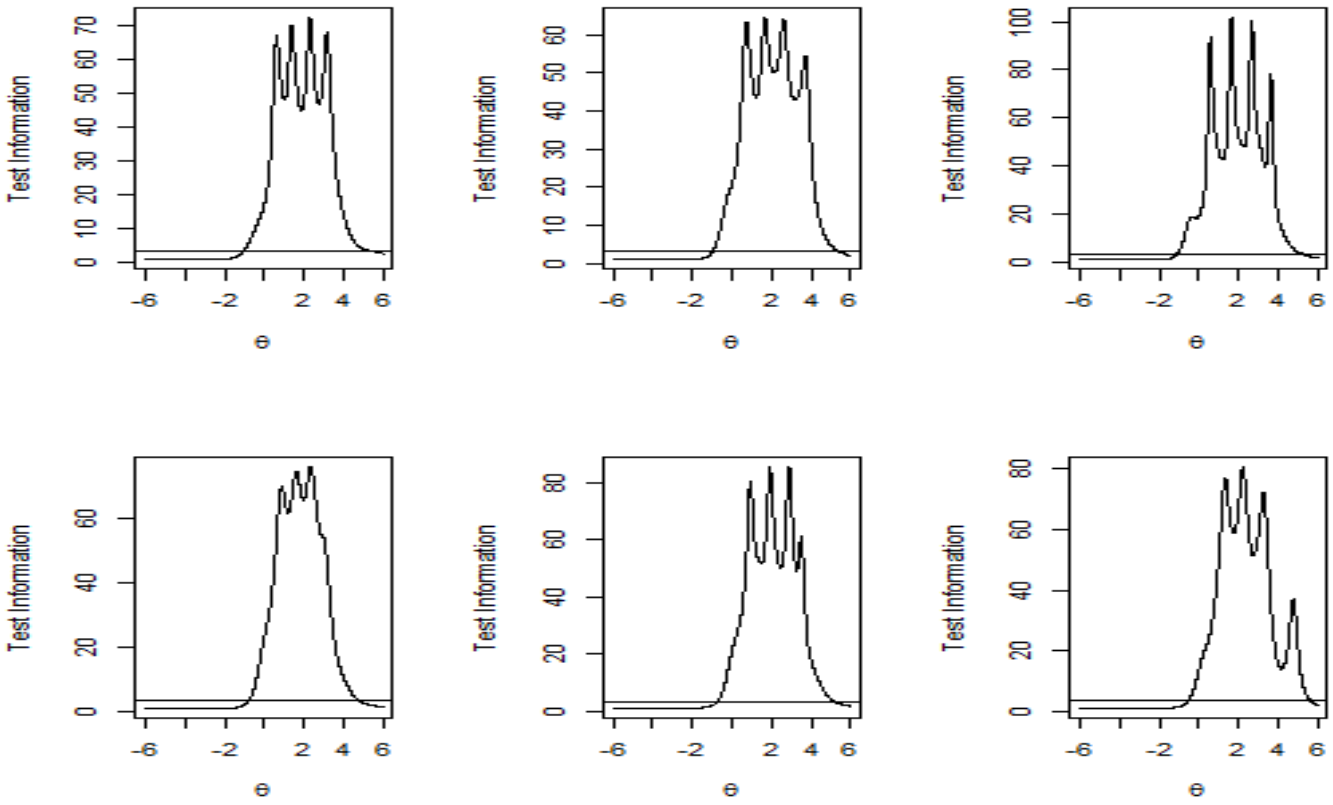


Figure Notes

Figure 5

Test information curves for ADHD at (top row from left to right) Age 7, Age 8, Age 9, Age 10, (bottom row from left to right) Age 11, Age 12, Age 13, Age 15

Figure 6

Test information curves for Internalising at (top row from left to right) Age 7, Age 8, Age 9, Age 10, (bottom row from left to right) Age 11, Age 12, Age 13, Age 15

Figure 7

Test information curves for Internalising at (top row from left to right) Age 7, Age 8, Age 9, Age 10, (bottom row from left to right) Age 11, Age 12, Age 13, Age 15

Figure 8

Test information curves for Externalising at (top row from left to right) Age 7, Age 8, Age 9, Age 10, (bottom row from left to right) Age 11, Age 12, Age 13, Age 15