Supplementary materials 2 - Appendix

In this appendix, we describe procedures to establish measurement invariance and measurement model fit, as well as report item parameter estimates and plots showing the reliability and the fit of the measurement model.
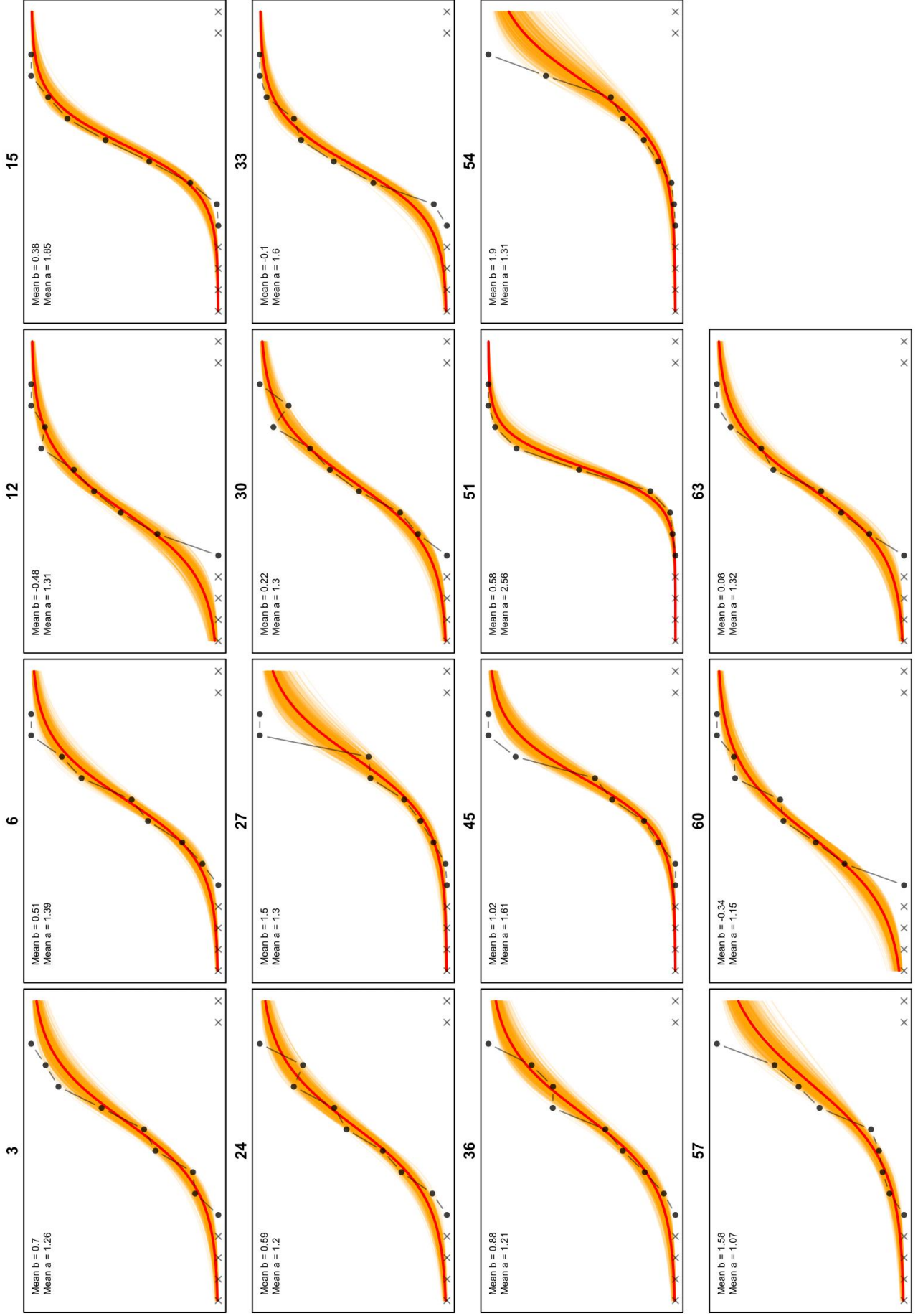
Establishing measurement invariance using leave-one-out crossvalidation

Latent difference score modelling requires measurement invariance across informants (de Haan et al., 2018). Following the procedure outlined by Verhagen and Fox (2013), we fitted IRT models to the CBQ data with item parameters allowed to differ between parent and adolescent informants and models with item parameters constrained to be equal. We then compared the fit of these models for each item using approximate leave-one-out cross-validation (Vehtari et al., 2019). Seven items (CBQ item numbers 9, 18, 21, 39, 42, 48, and 66) were found to have substantially better fit with item parameters allowed to differ between parents and adolescents, and these items were removed in the analysis.
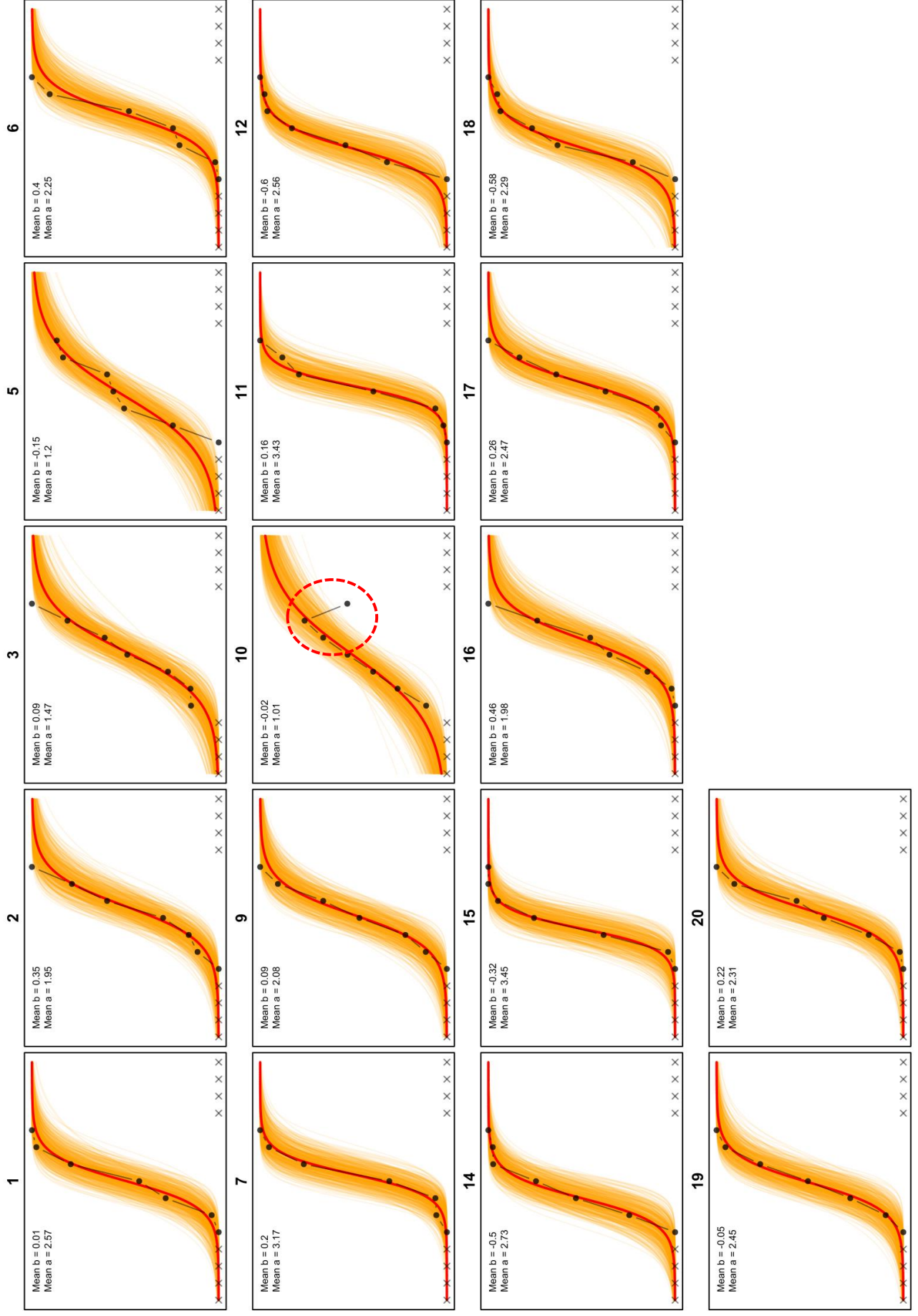
Evaluating IRT model fit by plotting

To evaluate the fit of the IRT model we plotted the observed probabilities of endorsing an item at different levels of the latent conflict variable together with item characteristic curves from the fitted model. We used the posterior means of the latent conflict variable and divided those into groups by 0.5 increments of the latent variable. For each item we then computed the proportion of endorsements in each group, giving the observed probability of endorsement at that approximate level of conflict. We plotted these probabilities along with 500 item characteristic curves sampled from the posterior distributions of the item parameters. This allowed us to assess the fit of the IRT model to the data, by comparing the observed probabilities of endorsement at different levels of latent trait with the posterior distribution of endorsement probabilities predicted by the model (Orlando & Thissen, 2000). These plots (see the following three pages) indicated acceptable or good fit of the IRT models for most items, except item 10 of the BHS. This item was also excluded from the analysis, bringing the final number of included items to 15 (CBQ) and 16 (BHS).

Items from the CBQ Perception of the Dyad Subscale (15)

3 Mean b = 0.7 Mean a = 1.26	**6** Mean b = 0.51 Mean a = 1.39	**12** Mean b = -0.48 Mean a = 1.31	**15** Mean b = 0.38 Mean a = 1.85
24 Mean b = 0.59 Mean a = 1.2	**27** Mean b = 1.5 Mean a = 1.3	**30** Mean b = 0.22 Mean a = 1.3	**33** Mean b = -0.1 Mean a = 1.6
36 Mean b = 0.88 Mean a = 1.21	**45** Mean b = 1.02 Mean a = 1.61	**51** Mean b = 0.58 Mean a = 2.56	**54** Mean b = 1.9 Mean a = 1.31
57 Mean b = 1.58 Mean a = 1.07	**60** Mean b = -0.34 Mean a = 1.15	**63** Mean b = 0.08 Mean a = 1.32	

Items from the Beck Hopelessness Inventory (17)

Items from the Beck Hopelessness Inventory (16)
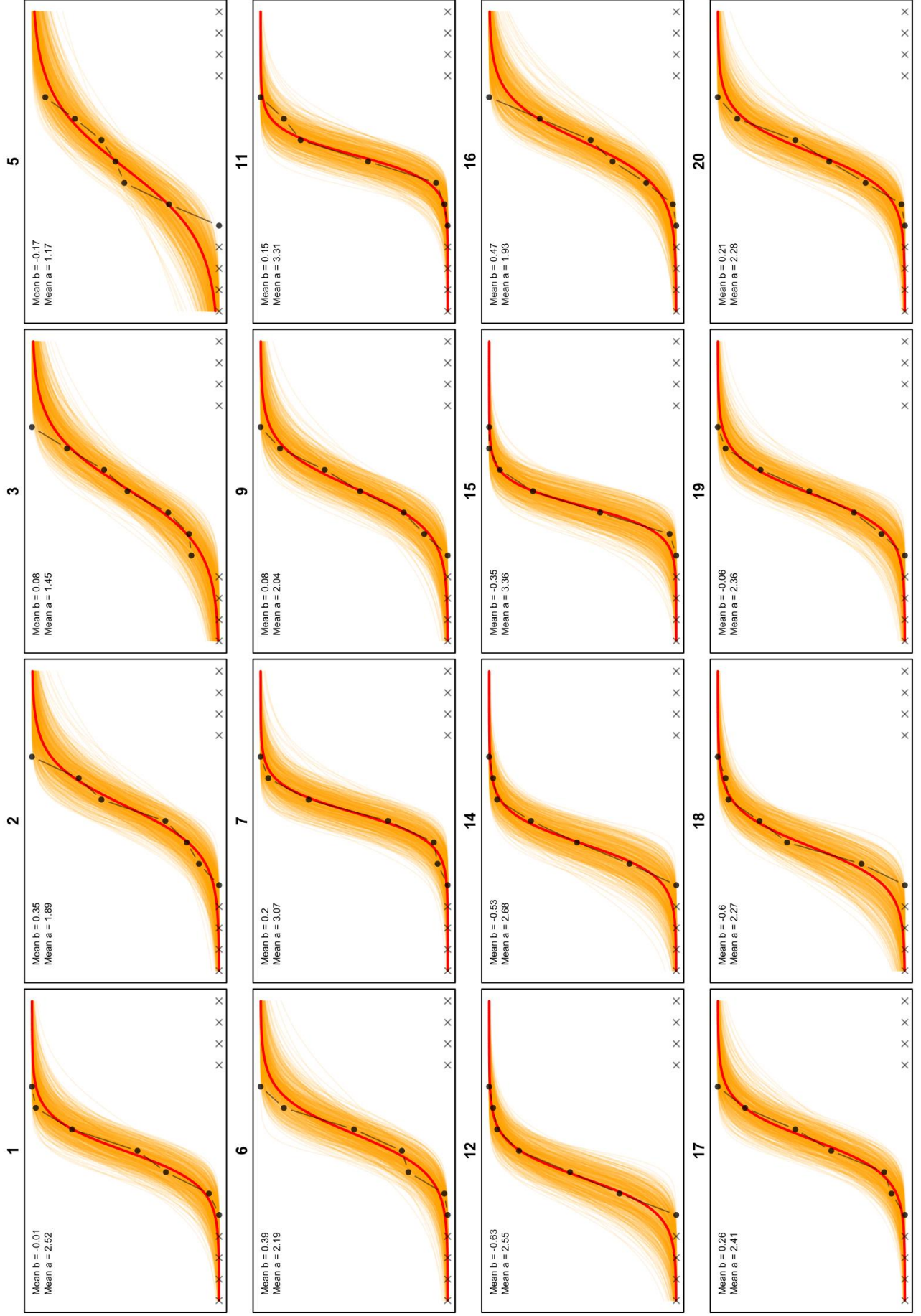
The posterior estimates for the IRT model item parameters are summarised in the following tables. The estimates are from the final measurement models, without the excluded items:

Conflict Behavior Questionnaire

Parameter	Mean	SD	66% CI	90% CI	ESS Bulk	ESS Tail
Item discrimination parameters						
Item 3	1.28	0.13	1.15 ; 1.40	1.07 ; 1.51	3541	3101
Item 6	1.40	0.15	1.26 ; 1.54	1.17 ; 1.64	2859	2831
Item 12	1.38	0.15	1.24 ; 1.52	1.15 ; 1.63	3083	2775
Item 15	1.93	0.19	1.75 ; 2.11	1.63 ; 2.25	2867	2987
Item 24	1.25	0.13	1.13 ; 1.37	1.04 ; 1.47	3640	2973
Item 27	1.46	0.16	1.30 ; 1.61	1.20 ; 1.74	3169	2912
Item 30	1.36	0.14	1.22 ; 1.49	1.14 ; 1.59	3061	3215
Item 33	1.56	0.16	1.41 ; 1.71	1.31 ; 1.84	3113	3147
Item 36	1.29	0.14	1.16 ; 1.42	1.07 ; 1.53	3661	2716
Item 45	1.63	0.17	1.47 ; 1.78	1.36 ; 1.91	3300	3693
Item 51	2.77	0.29	2.50 ; 3.04	2.33 ; 3.28	2256	2973
Item 54	1.54	0.18	1.36 ; 1.71	1.24 ; 1.84	3801	2680
Item 57	1.19	0.14	1.06 ; 1.33	0.97 ; 1.43	3675	2982
Item 60	1.25	0.13	1.12 ; 1.37	1.04 ; 1.47	3353	3298
Item 63	1.32	0.14	1.19 ; 1.45	1.10 ; 1.55	3477	3177
Item threshold parameters						
Item 3	0.62	0.09	0.53 ; 0.71	0.46 ; 0.78	2540	2532
Item 6	0.45	0.09	0.37 ; 0.53	0.31 ; 0.59	2904	3086
Item 12	-0.50	0.09	-0.58 ; -0.41	-0.64 ; -0.36	2214	2995
Item 15	0.29	0.07	0.23 ; 0.36	0.18 ; 0.41	2638	3101
Item 24	0.51	0.09	0.43 ; 0.60	0.37 ; 0.67	2593	3170
Item 27	1.30	0.13	1.18 ; 1.42	1.10 ; 1.52	2666	3266
Item 30	0.15	0.08	0.08 ; 0.23	0.02 ; 0.29	2556	3252
Item 33	-0.18	0.08	-0.25 ; -0.10	-0.31 ; -0.05	2638	2883
Item 36	0.77	0.10	0.68 ; 0.87	0.61 ; 0.95	2844	3342
Item 45	0.95	0.10	0.85 ; 1.04	0.79 ; 1.12	2427	3318
Item 51	0.47	0.07	0.41 ; 0.54	0.36 ; 0.59	2324	3163
Item 54	1.60	0.15	1.46 ; 1.74	1.36 ; 1.86	2711	2771
Item 57	1.37	0.15	1.23 ; 1.51	1.14 ; 1.64	3057	3339
Item 60	-0.38	0.09	-0.47 ; -0.30	-0.53 ; -0.24	2482	3123
Item 63	0.04	0.08	-0.03 ; 0.12	-0.09 ; 0.17	2795	2737

Beck Hopelessness Inventory

Parameter	Mean	SD	66% CI	90% CI	ESS Bulk	ESS Tail
Item discrimination parameters						
Item 1	2.47	0.57	1.92 ; 3.00	1.63 ; 3.48	669	1192
Item 2	1.89	0.44	1.46 ; 2.30	1.23 ; 2.67	714	1363
Item 3	1.41	0.34	1.09 ; 1.73	0.89 ; 2.02	689	1414
Item 5	1.15	0.29	0.88 ; 1.42	0.72 ; 1.66	805	1496
Item 6	2.17	0.51	1.69 ; 2.64	1.43 ; 3.09	634	1174
Item 7	3.10	0.71	2.43 ; 3.77	2.04 ; 4.34	591	1519
Item 9	2.00	0.47	1.56 ; 2.43	1.30 ; 2.83	671	1511
Item 11	3.30	0.77	2.58 ; 4.01	2.16 ; 4.67	644	1437
Item 12	2.46	0.57	1.92 ; 3.00	1.63 ; 3.50	694	1342
Item 14	2.54	0.60	1.97 ; 3.11	1.68 ; 3.62	666	1354
Item 15	3.38	0.79	2.62 ; 4.12	2.21 ; 4.79	637	1285
Item 16	1.89	0.45	1.47 ; 2.31	1.25 ; 2.67	659	1358
Item 17	2.47	0.58	1.91 ; 3.01	1.62 ; 3.49	679	1361
Item 18	2.27	0.54	1.76 ; 2.76	1.50 ; 3.21	727	1289
Item 19	2.33	0.54	1.82 ; 2.84	1.54 ; 3.30	700	1568
Item 20	2.32	0.54	1.80 ; 2.82	1.52 ; 3.27	685	1528
Item threshold parameters						
Item 1	0.03	0.24	-0.20 ; 0.26	-0.36 ; 0.43	789	1580
Item 2	0.39	0.26	0.14 ; 0.64	-0.03 ; 0.82	670	1642
Item 3	0.13	0.25	-0.10 ; 0.38	-0.27 ; 0.56	802	1627
Item 5	-0.15	0.25	-0.39 ; 0.10	-0.56 ; 0.28	874	1671
Item 6	0.44	0.26	0.18 ; 0.70	0.03 ; 0.87	665	1578
Item 7	0.23	0.24	0.00 ; 0.47	-0.16 ; 0.63	721	1556
Item 9	0.13	0.24	-0.10 ; 0.36	-0.26 ; 0.53	763	1595
Item 11	0.20	0.24	-0.04 ; 0.43	-0.20 ; 0.59	732	1613
Item 12	-0.62	0.27	-0.87 ; -0.36	-1.09 ; -0.18	801	1371
Item 14	-0.51	0.26	-0.76 ; -0.26	-0.95 ; -0.09	796	1246
Item 15	-0.32	0.25	-0.55 ; -0.08	-0.72 ; 0.10	803	1358
Item 16	0.51	0.27	0.24 ; 0.76	0.08 ; 0.96	657	1697
Item 17	0.28	0.25	0.05 ; 0.53	-0.12 ; 0.70	690	1558
Item 18	-0.57	0.27	-0.82 ; -0.31	-1.02 ; -0.13	791	1314
Item 19	-0.02	0.24	-0.25 ; 0.21	-0.41 ; 0.38	770	1476
Item 20	0.23	0.24	0.00 ; 0.47	-0.17 ; 0.65	754	1465

As we use latent variables from IRT models rather than sum-scores in the analysis, conventional estimates of reliability (such as coefficient alpha) are not directly applicable. According to item response theory the reliability of trait estimates is not uniform across the range of the latent variable, but can be described by the test information function as long as the item response model fits the data (de Ayala, 2009). We computed and plotted the test information function for the final item sets of the CBQ and the BHS, which indicated acceptable reliability across the observed range of the latent variables.
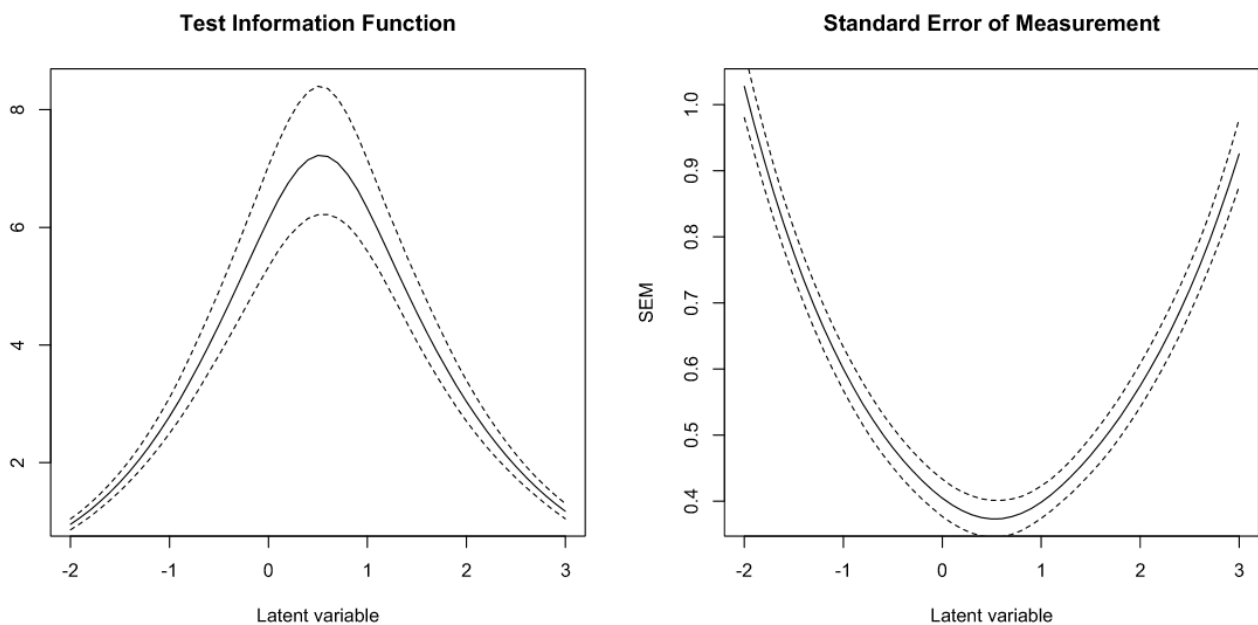
Test information curve and plot of the standard error of measurement across the latent variable for the 15 items from the CBQ Perception of the Dyad subscale
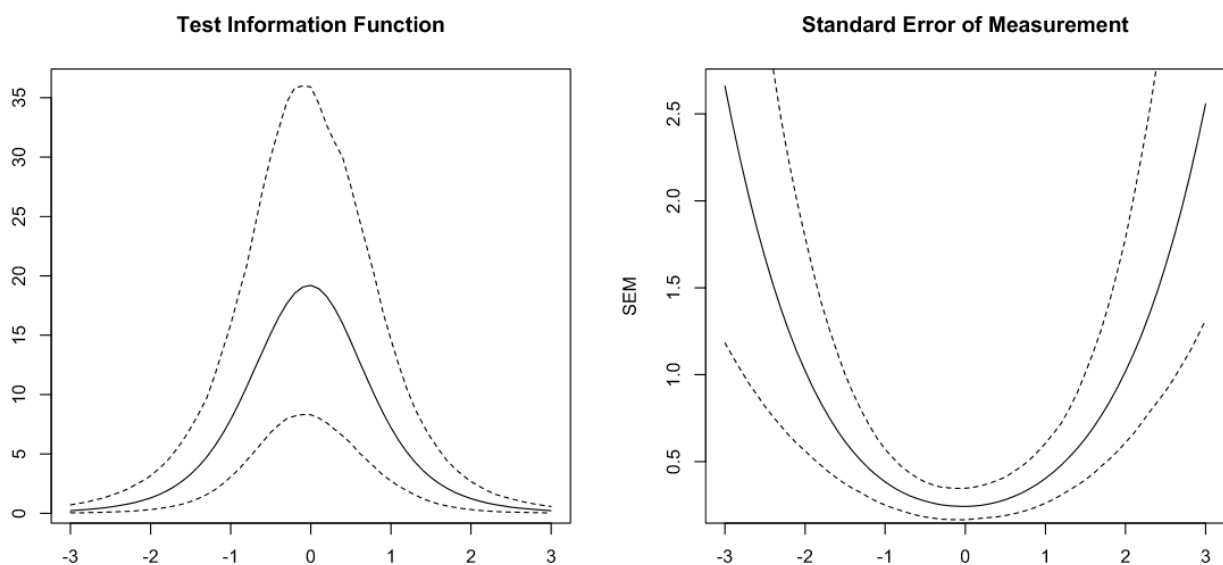
Test information curve and plot of the standard error of measurement across the latent variable for the 16 items from the BHS

References

de Ayala, R. J. (2009). *The theory and practice of item response theory*. Guilford Press.

de Haan, A., Prinzie, P., Sentse, M., & Jongerling, J. (2018). Latent difference score modeling: A flexible approach for studying informant discrepancies. *Psychological Assessment*, *30*(3), 358-369. https://doi.org/10.1037/pas0000480

Gabry, J., & Mahr, T. (2021). *bayesplot: Plotting for Bayesian Models*. In (Version R package version 1.8.1) https://mc-stan.org/bayesplot/

Orlando, M., & Thissen, D. (2000). Likelihood-Based Item-Fit Indices for Dichotomous Item Response Theory Models. *Applied Psychological Measurement*, *24*(1), 50-64. https://doi.org/10.1177/01466216000241003

Vehtari, A., Gabry, J., Yao, Y., & Gelman, A. (2019). *loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models.* In (Version 2.1.0) https://CRAN.R-project.org/package=loo

Verhagen, A. J., & Fox, J. P. (2013). Bayesian tests of measurement invariance. *British Journal of Mathemathical and Statistical Psychology*, *66*(3), 383-401. https://doi.org/10.1111/j.2044-8317.2012.02059.x