Mindreading, Psychopathology and Social Adjustment in Middle Childhood and Adolescence

Dataset Description

Rory T. Devine, School of Psychology, University of Birmingham, UK.

Information about this study is available on the Open Science Framework:

https://osf.io/8x73r/

Participants

We recruited 1100 English-speaking children between the ages of 8 and 13 years from 37 classrooms in state-funded primary and secondary schools in the UK as part of a pre-registered study of children's ToM, mental health, and social adjustment. Sample size was based on a Monte Carlo simulation with 2000 replications carried out in *Mplus* Version 8 (Muthèn & Muthèn, 2017). The simulation included 5 latent factors: a bifactor model of mental health; a theory of mind factor (Devine & Hughes, 2016); and a social competence factor. Standardized direct effects were set to .25 (a medium effect) and indirect effects between P Factor and social adjustment via mindreading were set to .06. A sample size of 1000 participants provided sufficient power to detect significant unique direct and indirect effects of mental health on Theory of Mind (.76 - .99) and Social Adjustment (.90 - .95).

Of the 1100 children in participating classrooms, 31 children were excluded because their caregivers did not provide consent for their participation and/or the children were unable to participate in the study unaided by a classroom assistant. A further 49 children declined to participate in the study. Of 1020 children, 556 identified as girls and 453 identified as boys (11 children did not label themselves as boys or girls). Children were aged between 8.27 and 13.27 years (*M* age = 10.36, *SD* = 1.27). Participants were socio-economically diverse: 23.2% (of 770 children) were eligible for Free School Meals and 28.9% (of 772 children)

spoke languages in addition to English. The sample was ethnically diverse (based on data from 730 children): 51.5% White, 31.5% Asian or Asian British, 8.1% Black, Black British, Caribbean or African, 6% Mixed or Multiple Ethnicity, and 2.9% 'Other Ethnic Group'. One fifth (18.2% of 768 children) had a statement of special educational needs.

Procedure

Children participated in two whole-class sessions, approximately 1 week apart. Each session, led by two research assistants, lasted between 60 and 90 minutes and included a fixed-order battery of tasks paced by the researcher. Children faced a large screen and completed all tasks individually on a computer through the online data-collection platform PsyToolkit (Stoet, 2017). Teachers completed a questionnaire about each child. The study was approved by the University of Birmingham Research Ethics Committee. Information about the testing procedures can be found at: https://osf.io/8x73r/

Missing Data

Of the 1020 children (93% participation rate) included in the study, 890 (87.3%) children participated in both study visits and 130 (12.7%) children participated in one study visit. Teachers completed questionnaires for 786 (77.1%) of the children. Missing teacher questionnaires were attributed to the cessation of testing due to the Covid-19 pandemic restrictions. The majority of participants (N=720, 70.6%) participated in both testing sessions and their teachers completed the questionnaire. A further 170 children (16.7%) participated in two testing sessions but their teachers did not return a questionnaire. Of the remaining children, 66 (6.5%) participated in one testing session and their teacher returned a questionnaire and 64 (6.3%) participated one testing session and their teacher did not return a questionnaire.

Variables in Dataset

Variable Name	Type of Data	Description
ID	Numeric	Unique Child Identity Number
Part	Categorical	Participation Record: 1 = 1 visit only, 2 = 1 visit and teacher questionnaire, 3 = 2 visits and no teacher questionnaire, 4 = 2 visits and teacher questionnaire.
School	Categorical	School ID Number
CID	Categorical	Classroom ID Number
SCHFSM	Continuous	Proportion of children on free school meals taken from the Annual School Census.
SFSM	Categorical	Proportion of children on free school meals categorised as High (1) vs Average/Low (0)
EthB	Categorical	Child Ethnicity Binary (1 = Majority Ethnicity, 2 = Numerical Minority Ethnicity)
ETH	Categorical	Child Ethnicity (1 = White, 2 = Asian, 3 = Black, 4 = Mixed, 5 = Other)
Teach	Categorical	Teacher Questionnaire Received (0 = No, 1 = Yes)
Session	Categorical	Number of Sessions Completed (1 = 1 session, 2 = 2 sessions)
Gender	Categorical	Child Gender (0 = Girl, 1 = Boy)
AgeR	Continuous	Age in Years
FSM	Categorical	Free School Meals (0 = No, 1 = Receives Free School Meals)
EAL	Categorical	English As Additional Language (0 = No, 1 = Yes)
SEND	Categorical	Statement of Special Educational Needs and Disability (0 = No, 1 = Yes)
VA	Continuous	Verbal Ability (Age Standardized Score from the Mill Hill Vocabulary Scale)
SF1 – SF6	Categorical	Theory of Mind: Silent Film Task Items (0 = Fail, 1 = Partial Score, 2 = Correct)
SS1 – SS5	Categorical	Theory of Mind: Strange Stories Task Items (0 = Fail, 1 = Partial Score, 2 = Correct)
TR1MA – TR3MA	Continuous	Theory of Mind: Triangles Task Items (0 – 7 points for each item)
CBCLFSAD	Continuous	Child Behaviour Checklist Teacher Report Form: Anxious/Depressed Factor Score
CBCLFSWD	Continuous	Child Behaviour Checklist Teacher Report Form: Withdrawn/Depressed Factor Score
CBCLFSRB	Continuous	Child Behaviour Checklist Teacher Report Form: Rule Breaking Factor Score
CBCLFSAB	Continuous	Child Behaviour Checklist Teacher Report Form: Aggressive Behaviour Factor Score
CBCLFSIA	Continuous	Child Behaviour Checklist Teacher Report Form: Inattention Factor Score
CBCLFSIH	Continuous	Child Behaviour Checklist Teacher Report Form: Impulsivity/Hyperactivity Factor Score
CBCLFSSP	Continuous	Child Behaviour Checklist Teacher Report Form: Social Dependence Factor Score

CBCLFSTP	Continuous	Child Behaviour Checklist Teacher Report Form: Thought Problems Factor Score
CASTFSNI	Continuous	Childhood Autism Syndrome Test Teacher Report: Repetitive Behaviours and Interests
CASTFSCS	Continuous	Childhood Autism Syndrome Test Teacher Report: Communication Skills
CASTFSRS	Continuous	Childhood Autism Syndrome Test Teacher Report: Reciprocal Social Interaction
SMS_MEAN	Continuous	Social Maturity Scale Mean Score (Teacher Rated)
SP_Z_R	Continuous	Social Preference Score (Peer Nomination)
RCP_R	Continuous	Sociability and Leadership Mean Score (Peer Nomination)
ERFSCORE	Continuous	Emotion Recognition Task Latent Factor Score
EFFSCORE	Continuous	Executive Function Task Battery Latent Factor Score
SFTOT	Continuous	Silent Film Task (Summed Total Score)
SSTOT	Continuous	Strange Stories Task (Summed Total Score)
TRITOT	Continuous	Triangles Task (Summed Total Score)
EMRTOT	Continuous	Emotion Recognition Task (Summed Total Score)

Measures

Mental Health and Neurodevelopmental Diversity.

Teachers completed the *Child Behavior Checklist Teacher Report Form (CBCL TRF)* for children aged 6 to 18 years (Achenbach & Rescorla, 2013). There were 99 items across 8 subscales: anxious-depressed, withdrawn, aggressive behavior, rule-breaking behavior, inattention, hyperactivity-impulsivity, social dependence, and thought problems. We omitted four low frequency items that were not developmentally appropriate (i.e., talks of suicide, thinks of sex, uses tobacco, and uses drugs) (Achenbach & Rescorla, 2001) and we did not administer the somatic complaints subscale. Teachers completed the 20-item *Childhood Autism Syndrome Test* (CAST) (Ronald et al., 2006). The CAST captures behaviors associated with autism (e.g., differences in communication, social interaction, and repetitive behaviours and restricted interests) and has been used in large-scale studies with teachers.

Theory of Mind

In the *Silent Film Task* (Devine & Hughes, 2013) children watched five short film clips on a large screen depicting instances of deception, misunderstanding, and false belief.

Children responded to a single question about each clip, which required an explanation of a character's behavior. Children received 2 points for accurate mentalizing given the context, 1 point for partially correct responses, and 0 points for inaccurate or irrelevant responses (Devine et al., 2023). In the *Strange Stories Task* (Happé, 1994), the researcher read aloud five short vignettes, involving deception, misunderstanding and double bluff. Children answered an open-ended question about the characters' behavior. Accurate mentalizing received 2 points, partially correct responses received 1 point, and inaccurate responses received 0 points (White et al., 2009). In the *Frith-Happé Triangles Animation Task* (Castelli et al., 2000), children watched three short animations, each featuring interactions between two cartoon triangles involving instances of 'sneaking', 'pretending' and 'tricking'. Children were asked to describe what happened. Answers were rated for intentionality (a score from 0 to 5 indicating the degree to which the child attributed mental states to the triangles) and appropriateness (a score from 0 to 2 indicating the degree to which the child's description captured the key events of the clip). Intentionality and appropriateness scores were summed for each response giving a score of 0 to 7 points for each clip.

Two research assistants scored a reliability set comprised of data from 30 participants for each item of the Strange Stories, the Silent Film and Triangles tasks. Inter-rater reliability (Krippendorf's α) for coding each item ranged from .85 to .1.00 for the Strange Stories items and .87 to 1.00 for the Silent Film task items. Intra-class correlations for Triangles task Intentionality ratings ranged from .82 to .99 and from .74 to .76 for Appropriateness ratings. Having established inter-rater reliability, each response was scored by one research assistant only.

Social Adjustment

From a list of all children in the classroom, children rated each other using the Sociability and Leadership Scale of the *Revised Class Play* (Masten et al., 1985). Children

chose one person in their class (other than themselves) who they thought would be best at playing each of 15 ‘roles’ in a play (e.g., ‘someone who is a good leader’, ‘someone who plays fair’). The total number of nominations received by each child was summed together and standardized within each classroom ($\alpha = .79$).

To measure *Peer Social Preference*, children nominated up to three children in their class who they ‘most like’ and ‘least like’ to spend time with (Coie et al., 1982). The total number of nominations received by each participating child was standardized within the classroom to account for differences in classroom size. Social preference, the degree to which a child is liked by their peers, was calculated by subtracting classroom standardized “least like” nominations from standardized “most like” nominations (van den Berg et al., 2020).

Teachers completed the *Peer Social Maturity Scale* (Peterson et al., 2007), which measures peer-oriented social behaviors by asking teachers to rate children relative to their same-age peers in 8 domains (e.g., assertion, leadership, coping with peers, understanding others’ needs). High scores indicated better peer social interaction skills. Item scores were averaged to create a social competence score ($\alpha = .92$).

Executive Function

Children completed direct assessments of EF based on a protocol developed by Obradovic et al. (2018). In the *Digit Span Backwards Task* (Obradović et al., 2018), children viewed a sequence of between 2 and 5 numbers displayed one at a time and then had to type the sequence in reverse order. Accuracy across 8 trials was measured by summing the number of correctly recalled sequences. The *Hearts and Flowers Task* (Davidson et al., 2006) consisted of two blocks of 12 trials. In the Hearts (control) condition, children pressed a key on the same side as a heart appearing on their screen. In the Flowers (inhibition) condition, children pressed the key on the opposite side to a flower appearing on their screen. In the *Fish Flanker Task* (Rueda et al., 2004) children were required to ‘feed the fish’ in the middle

while ignoring fish on either side. In the 22 congruent trials all fish faced in the same direction. In the 12 incongruent trials children had to ignore the surrounding fish because the fish in the middle faced in the opposite direction to the other fish. In both tasks, we calculated the rate correct score (i.e., total correct trials/second) for each condition by summing the total number of correct trials in each condition and dividing this by the total time (i.e., the sum of all reaction times).

Verbal Ability

Children completed the multiple-choice section of the *Mill Hill Vocabulary Scale* (Rust, 2008) to measure verbal ability. Children selected a synonym for 20 target words from six possible response options each and received 1 point for each correctly identified word. Total scores were age-standardized.

Emotion Recognition

Children completed an *Emotion Recognition* task (Dadds et al., 2018) using 30 images from the *Developmental Emotional Faces Stimuli Set* (Meuwissen et al., 2017). Children viewed images of child, adolescent, and adult male (50%) and female (50%) faces and indicated whether the face was happy, sad, angry, fearful, or neutral. The number of correct responses for each emotion was calculated with scores for each emotion ranging from 0 to 6.

Data Reduction

Confirmatory Factor Analysis of Teacher Ratings of Children's Mental Health and Neurodevelopmental Diversity

Participants were a community-based sample recruited through schools. Using the standard scoring criteria (i.e., summed scores for each subscale), we compared the participants in our sample against established norms for the CBCL TRF Syndrome Scales

(Achenbach & Rescorla, 2001). Table S4 shows the proportion of boys and girls falling in the 'normal', 'borderline', and 'clinical' ranges for each syndrome scale.

Following previous studies (e.g., Michelini et al., 2019), we recoded the items of the CBCL Teacher Report Form and CAST into binary scores (i.e., not true vs somewhat or very true). Next, we inspected the tetrachoric correlations between items and aggregated those items that exhibited high correlations (>.75) and recoded these as binary items (Achenbach & Rescorla, 2001). We used CFA with the WLSMV estimator in *Mplus* Version 8 (Muthèn & Muthèn, 2017) to examine the latent factor structure of each scale in the CBCL Teacher Report Form. We scaled each latent factor using the lead indicator. We evaluated model fit using the same three standard criteria reported above (Brown, 2015). Table S5 shows the model fit indices and key features of each of these models and the omega reliability statistic for each latent factor. We saved the factor scores from this model and used these in subsequent analyses. Following others (e.g., Harden et al., 2020), we adjusted these scores by regressing the scores onto age and gender and used the residualized scores in our analyses.

The latent factor structure of the CAST was also examined. We compared the fit of three models based on the standard scoring guidelines (Ronald et al., 2008). Specifically, we examined the fit of a one factor model where all items loaded onto one latent factor, a two-factor model where restricted, repetitive behaviours and interests items loaded onto one factor and items about social-communicative skills loaded onto a second correlated factor, and a three-factor model where items for each trait loaded onto separate but correlated latent factors representing restricted, repetitive behaviours and interests, social relationships, and communication skills. None of these models provided a good fit to the data (Table S6).

We carried out a categorical data exploratory factor analysis (EFA) using the WLSMV estimator with oblique Geomin rotation to estimate a first-order factor solution incorporating between 1 and 4 latent factors. A four-factor solution provided the best fit to the data, χ^2 (87)

= 131.48, CFI = 0.982, TLI = 0.968, RMSEA = 0.026. We used the parameter estimates from this solution to specify and estimate a CFA. The model provided an adequate fit to the data (Table S6) but one latent factor was not correlated with the others. This latent factor was comprised of items about the importance of others (e.g., 'cares how s/he is perceived by group', 'important to this pupil to fit in with peer group', 'people are important to him/her'). We therefore removed this latent factor from the model. The remaining latent factors captured individual differences in 'restricted, repetitive behaviors' ($\omega = .75$), 'reciprocated social behavior' ($\omega = .60$) and 'communication skills' ($\omega = .71$) (Table S7). We saved the factor scores from this model and used these in subsequent analyses. We used the age and gender residualized scores in our analyses.

Latent Factor Structure of Mental Health and Neurodiversity

We compared a set of competing models based on previous research using the CBCL. We used CFA with a robust maximum likelihood estimator in *Mplus* Version 8 (Muthèn & Muthèn, 2017) to estimate latent factor scores. We compared non-nested models using standard model fit indices (Brown, 2015) alongside the AIC and sample-size adjusted BIC, selecting the model with the lowest AIC and SABIC values (Caspi et al., 2024). We inspected the modification indices and respecified models to improve fit, where justifiable. To handle Heywood cases, we set the variance of items with negative residuals to close to 0 (Chen et al., 2001). To aid estimation of the higher-order and bifactor models, we set the scale of the latent factors by freely estimating the first factor loading and setting the latent factor variances to 1 (Geiser, 2013). Model fit statistics are shown in Table S8.

In the one factor model, all indicators loaded onto a single latent factor. In the three-factor model, mental health indicators loaded onto an internalizing (i.e., anxious, withdrawn) and externalizing latent factor (i.e., rule breaking, aggression) and indicators of neurodiversity (e.g., thought problems, social dependence, inattention, hyperactivity,

restricted behaviours and interests, communication, reciprocal social behaviour) loaded onto a separate latent factor. In the five-factor model, all indicators loaded onto separate correlated factors representing internalizing (i.e., anxious, withdrawn), externalizing (i.e., rule breaking, aggression), thought problems (i.e., thought problems, social dependence), attention deficits/hyperactivity (i.e., inattention, hyperactivity), and autism traits (i.e., restricted behaviours and interests, communication, reciprocal social behaviour). In the higher-order model, each of the five latent factors loaded onto a higher-order P Factor. In the bifactor model, all indicators loaded onto a general P Factor and onto five specific latent factors. Each factor was specified as orthogonal (i.e., the correlations between all factors were set to 0). Finally, we estimated an orthogonal S-1 bifactor model (Heinrich et al., 2023), where all indicators (except thought problems and social dependence) loaded onto both the P Factor and four condition-specific latent factors (i.e., internalizing, externalizing, attention deficit/hyperactivity, autism traits). The orthogonal S-1 bifactor model with 4 specific factors was selected as the best fitting model.

The model-based estimate of reliability, ω_H = 0.79, indicated that 79% of the variance in total scores across the original scales was accounted for by the P Factor. Model-based estimates of reliability indicated that once variance in P was taken into account, the percentage variance in subscale scores accounted for by specific factors was 56% for internalizing, ω_{HS} = 0.56, 29% for externalizing, ω_{HS} = 0.29, 30% for ADHD, ω_{HS} = 0.30, and 35% for autism, ω_{HS} = 0.35 (Rodriguez et al., 2016).

Confirmatory Factor Analysis of Social Adjustment

We used CFA with a robust maximum likelihood estimator to estimate social adjustment latent factor scores. Specifically, we tested a one factor model in which the scores from the Peer-assessed Revised Class Play Sociability and Leadership scale, Peer-rated Social Preference, and Teacher-rated Social Maturity loaded onto one latent factor. The model

was just-identified as the number of freely estimated parameters equalled the number of pieces of known information in the input variance-covariance matrix (Brown, 2015). Each indicator loaded significantly on the latent factor: Revised Class Play *Std. Est.* = .80, p <.0001, Social Preference *Std. Est.* = .66, p <.0001, Social Maturity *Std. Est.* = .52, p <.0001.

Confirmatory Factor Analysis of Executive Function Tasks.

Children performed better on congruent trials of the Fish Flanker task, M = 1.41, SD = 0.46, than on incongruent trials, M = 1.36, SD = 0.46, t (923) = 4.40, p < .001, d = 0.15, 95%CI [0.08, 0.21]. Likewise, performance on the no conflict trials, M = 2.09, SD = 0.68, was better than their performance on the conflict trials, M = 1.61, SD = 0.68, of the Hearts and Flowers Task, t (891) = 21.16, p < .001, d = 0.68, 95%CI [0.64, 0.78]. On the Backward Digit Span Task, performance declined across the task as the length of digit sequences increased, F (3, 2575) = 940.34, p < .001. There were significant differences between each trial type with 2-digit trials, M = 1.88, SD =0.39, being the easiest and 5-digit trials, M = 0.56, SD = 0.76, being the most challenging. All contrasts were significant with Cohen's d ranging from 0.41, 95%CI [0.34, 0.48] to 1.63, 95%CI [1.53, 1.73].

We used CFA with a robust maximum likelihood estimator to estimate latent factor scores for executive function. Specifically, we tested a one factor model in which the incongruent trials score of the Fish Flanker Task, conflict trials score of the Hearts and Flowers Task, and total score for the Backward Digit Span Task loaded onto one latent factor. The model was just-identified as the number of freely estimated parameters equalled the number of pieces of known information in the input variance-covariance matrix (Brown, 2015). Each indicator loaded significantly on the latent factor: Hearts and Flowers Task *Std. Est.* = .65, p <.0001, Fish Flanker Task *Std. Est.* = .60, p <.0001, Backward Digit Span Task *Std. Est.* = .42, p <.0001. We saved the factor scores from this model and used these in subsequent analyses.

Confirmatory Factor Analysis of the Emotion Recognition Task.

Children's performance on the emotion recognition task varied by the type of emotion depicted, F (4, 3540) = 319.584, $p < .001$. There were no significant differences in performance on sad, M = 4.46, SD = 1.35, and neutral faces, M = 4.49, SD = 1.55, t (885) = -0.44, p = .66, Cohen's d = -.02, 95%CI [-0.08, 0.05]. All other contrasts were significant ($p <$ .001) with children performing best on happy faces, M = 5.10, SD = 1.13, then neutral and sad faces, fearful faces, M = 3.82, SD = 1.56, and angry faces, M = 3.21, SD = 1.40, Cohen's d ranged from 0.34, 95%CI [0.27, 0.41] to 1.22, 95%CI [1.13, 1.31].

We used CFA with a robust maximum likelihood estimator to generate latent factor scores for emotion recognition. Specifically, we tested a one factor model in which accuracy scores for each emotion loaded onto one latent factor. A one-factor model fit the data well, χ^2 (4) = 14.02, CFI = 0.98, TLI = 0.94, RMSEA = 0.05. Standardized loadings ranged from .39 to .65 and all loadings were statistically significant, $p < .0001$. We saved the factor scores from this model and used these in subsequent analyses.

Table S4. *Percentage of Children in 'Borderline' and 'Clinical' range based on CBCL TRF Syndrome Cut-off Ratings*

	Whole Sample (N = 782)			Girls (N = 428)			Boys (N = 354)		
	Normal	Borderline	Clinical	Normal	Borderline	Clinical	Normal	Borderline	Clinical
Anxious/Depressed	87.7	6.9	5.4	86.7	8.6	4.7	89.0	4.8	6.2
Withdrawn	94.8	4.0	1.2	96.0	3.3	0.7	93.2	5.1	1.7
Social Dependence	92.4	3.5	4.1	93.0	3.5	3.5	91.8	3.4	4.8
Thought Problems	92.4	2.6	5.0	93.9	2.1	4.0	90.7	3.1	6.2
Inattention	93.6	3.5	2.9	95.8	2.8	1.4	91.0	4.2	4.8
Impulsivity	93.2	3.3	3.5	95.1	2.1	2.8	91.0	4.8	4.2
Rule Breaking	92.2	4.6	3.2	93.7	3.3	3.0	90.4	6.2	3.4
Aggression	91.9	4.0	4.1	93.7	3.5	2.8	89.8	4.5	5.6

Table S5. Model Fit Statistics and Summary of Parameter Estimates for CBCL Teacher Report Form Scales.

Model	χ^2	df	RMSEA	CFI	TLI	Min. Loading	Max. Loading	ω
Internalizing Latent Factors	583.943	169	0.056	0.945	0.938			
Anxious/Depressed						.54	.89	.86
Withdrawn/Depressed						.49	.91	.73
Externalizing Latent Factors	606.894	298	0.036	0.990	0.989			
Rule Breaking						.40	.97	.80
Aggressive Behavior						.65	.98	.94
Other Problems	380.130	151	0.044	0.956	0.950			
Thought Problems						.72	.92	.80
Social Dependence						.48	.92	.79
Attention Problems	1271.556	298	0.065	0.972	0.969			
Inattentive						.73	.97	.94
Impulsive/Hyperactive						.72	.97	.93

Table S6. Model Fit Statistics and Summary of Parameter Estimates for CAST.

Model	χ^2	df	RMSEA	CFI	TLI
One Factor Model	1186.587	135	0.100	0.572	0.515
Two Factor Model	1162.064	134	0.099	0.581	0.522
Three Factor Model*	1140.950	132	0.099	0.589	0.524
CFA based on EFA	359.010	129	0.048	0.906	0.889
Three Factor Model	198.396	75	0.046	0.946	0.935

Note. *This model included a negative residual variance

Table S7. *Standardized WLSMV Parameter Estimates for Measurement Model of CAST Questionnaire.*

	RRBI		**Reciprocal Social Behavior**		**Communication**	
	Est.	**S.E.**	**Est.**	**S.E.**	**Est.**	**S.E.**
Memory for details	.68	.06				
Interest in topics	.81	.05				
Insistence on sameness	.77	.05				
Repetitive movements	.89	.06				
Takes things literally	.70	.04				
Style of communication	.74	.04				
Loses the listener	.81	.04				
Eye gaze, facial expression, voice	.86	.03				
Same interests as peers			.32	.08		
Easy to interact with others			.94	.02		
At least one good friend			.94	.02		
Social appropriateness					.85	.04
Social behavior on own terms					.90	.03
Turns conversation to favorite subject					.81	.04

Note. All loadings were statistically significant ($p < .01$).

Table S8. Fit Indices for Mental Health and Neurodiversity Models

	Model	χ^2	df	CFI	TLI	RMSEA	AIC	SABIC
1	One Factor Model	4524.081	44	0.425	0.281	0.361	8583.324	8632.374
2	Separate Conditions Model (3 Factors)	1237.743	39	0.846	0.783	0.198	4531.827	4588.309
3	Separate Conditions Model (5 Factors)*	475.921	35	0.943	0.911	0.127	3462.147	3524.573
4	Higher-Order Model (5 Factors)*	1003.080	43	0.877	0.842	0.169	4169.895	4220.431
5	Bifactor Model (Orthogonal)*	917.333	38	0.887	0.837	0.172	4059.805	4117.773
6	Bifactor Model (Orthogonal, Five Specific Factors, Cross-Loadings)*	208.868	33	0.977	0.962	0.083	3111.839	3177.239
7	Bifactor Model (Orthogonal, Four Specific Factors, Cross-Loadings)*	175.256	33	0.982	0.970	0.074	3068.884	3134.284

Note. *Set residual variance for some items close to 0 due to presence of Heywood Cases.