Supplement 1: Model Details

Our goal was to tease apart symptom-general and symptom-specific changes over a psychosocial intervention. The bifactor model is a hierarchical model designed to separate out the general and specific variance in a measure.[1] We attempted to estimate a bifactor model in addition to latent growth curves within a single-level model, but faced convergence issues. We thus split the process into two steps:

1) We first estimated the general and specific psychopathology factors at the within-level of a multilevel confirmatory bifactor analysis. This summarized how symptoms covaried over the study period for each individual.

2) We then estimated factor scores (Bayesian plausible values) of the general and specific psychopathology factors for each individual at each time-point. Factor scores were analyzed using a multilevel growth model, which included both within-person growth curves and between-person differences in within-person growth curves (i.e. random effects).

We describe the multilevel confirmatory bifactor analysis followed by the multilevel growth model in more depth below.

1) Multilevel Factor Model

We used multilevel factor analysis[2,3] to estimate within-person general and specific psychopathology factors over the study period (See Figure S1). Multilevel factor analysis is typically used to estimate separate factor structures for the within-person and between-person portions a covariance matrix. However, we used multilevel factor analysis to reduce the computational demands of estimating bifactor dimensions over time, since ‘time’ is treated continuously rather than discretely. In other words, a single factor can be estimated across time-points rather than repeatedly at each time-point. Data were arranged with repeated observations in long-format (e.g., vertically) and multiple items in the wide format (e.g., horizontally):

Subject	Time	Item 1	Item 2	…	Item 20
1	1	y_{11}	y_{11}		y_{11}
1	2	y_{12}	y_{12}		y_{12}
1	3	y_{13}	y_{13}		y_{13}
1	4	y_{14}	y_{14}		y_{14}
2	1	y_{21}	y_{21}		y_{21}
2	2	y_{22}	y_{22}		y_{22}
2	3	y_{23}	y_{22}		y_{23}
2	4	y_{24}	y_{24}		y_{24}
⋮					
683	4	$y_{683\ 4}$	$y_{683\ 4}$		$y_{683\ 4}$

Each item was specified at the within-level (level 1). We did not allow for variances at the between-level (level 2), but corrected the standard errors for the nesting of observations within subjects using a subject ID cluster variable. The model can be expressed as follows:

$$Y_{ijt} = v_{W_{ijt}} + \Lambda_{\mathrm{W}}\eta_{\mathrm{W}_{ijt}} + \varepsilon_{\mathrm{W}_{ijt}}$$

where *Y* is a matrix reflecting the observed responses on each item, *j* = 1,…,*J*, at each time-point, *t* = 1,…,*T* across individuals, *i* = 1,…,*N*, $v_{W_{ijt}}$is a vector of within-level item thresholds; Λ_{W} is a within-level factor loading matrix, $\eta_{\mathrm{W}_{ijt}}$ is a vector of factors which vary randomly across time-points and items within subjects, and $e_{ijt}$is the within-person error. The $\Lambda_{\mathrm{W}}\eta_{\mathrm{W}_{ijt}}$term can be expressed more fully as:

$$\Lambda_{\mathrm{W}}\eta_{\mathrm{W}_{ijt}} = \lambda_{Wgeneral_j}\theta_{Wgeneral_{it}} + \lambda_{Wspecific1_j}\theta_{Wspecific1_{it}} + \lambda_{Wspecific2_j}\theta_{Wspecific2_{it}} + \lambda_{Wspecific3_j}\theta_{Wspecific3_{it}} + \lambda_{Wspecific4_j}\theta_{Wspecific4_{it}}$$

where $\lambda_{W_j}$are within-level factor loadings for each item and $\theta_{W_{it}}$ are within-level factor vectors which vary across subjects and time-points for the general factor, $general$, and specific factors, $specific1, \dots, specificK$, where *K* = 4 in the current model.

Our notation implies that this was a three-level model, with repeated observations at the lowest level (‘time’) nested in each item (‘item’), nested within individuals (‘subject’). However, when implementing the model in Mplus, we included each item as a different within-level variable (see the data structure table above), making it a multi-indicator two-level factor model. Nonetheless, the models are equivalent.

2) Multilevel Growth Model

We estimated Bayesian plausible values (i.e. a distribution of factor scores) for the general and specific within-level factors described above. We thus had several estimates of each subject’s score on each factor at each time-point (e.g., $\hat{\theta}_{it}$), which were averaged over using multiple imputation. For simplicity, we refer to a single set of factor scores. Data were formatted with repeated observations for each factor in long format (e.g., vertically) and each factor in wide format (e.g., horizontally):

Subject	Time	θ_p	$\theta_{antisocial}$	$\theta_{anxiety}$	$\theta_{attention}$	θ_{mood}
1	0	y_{10}	y_{10}	y_{10}	y_{10}	y_{10}
1	1	y_{11}	y_{11}	y_{11}	y_{11}	y_{11}
1	2	y_{12}	y_{12}	y_{12}	y_{12}	y_{12}
1	3	y_{13}	y_{13}	y_{13}	y_{13}	y_{13}
2	0	y_{20}	y_{20}	y_{20}	y_{20}	y_{20}
2	1	y_{21}	y_{21}	y_{21}	y_{21}	y_{21}
2	2	y_{22}	y_{22}	y_{22}	y_{22}	y_{22}
2	3	y_{23}	y_{23}	y_{23}	y_{23}	y_{23}
⋮						
683	3	$y_{683\ 3}$	$y_{683\ 3}$	$y_{683\ 3}$	$y_{683\ 3}$	$y_{683\ 3}$

We estimated a two-level parallel process growth model using factor scores as outcome variables (See Figure S2). The simultaneous analysis of growth in each factor, *f* = 1,…,*F*, is denoted with a superscript (items in the multilevel factor model described above were also analyzed simultaneously, but denoted with a subscript). The within-level or level 1 portion of the model can be written as:

$$y_{it}^{(f)} = \beta_{0_i}^{(f)} + \beta_{1_i}^{(f)} Time_{it} + \beta_{2_i}^{(f)} Time^{2}{}_{it} + \varepsilon_{it}^{(f)}$$

where $y_{it}^{(f)}$ reflects factor scores for each individual, i = 1,…,N at each time-point, t = 0,…,T for a given factor, $\beta_{0_i}^{(f)}$ reflects the intercept or baseline factor scores for each individual when t = 0 (for each factor), $\beta_{1_i}^{(f)}$ and $\beta_{2_i}^{(f)}$reflect the linear and quadratic slopes of time on each factor, respectively, which vary randomly across individuals, $Time_{it}$and $Time^{2}{}_{it}$ reflect the observed values of time (0, 1, 2, 3) and time-squared (0, 1, 4, 9) for each individual at each time-point; and $\varepsilon_{it}^{(f)}$ reflects the individual- and time-specific residuals.

The between-level or level 2 part of the model can be expressed as

$$\begin{aligned}
\beta_{0_i}^{(f)} &= \gamma_{00}^{(f)} + \gamma_{01}^{(f)} c.Age_i + U_{0i}^{(f)} \\
\beta_{1_i}^{(f)} &= \gamma_{10}^{(f)} + \gamma_{11}^{(f)} c.Age_i + U_{1i}^{(f)} \\
\beta_{2_i}^{(f)} &= \gamma_{20}^{(f)} + \gamma_{21}^{(f)} c.Age_i + U_{2i}^{(f)}
\end{aligned}$$

where $\gamma_{00}^{(f)}$, $\gamma_{10}^{(f)}$, and $\gamma_{20}^{(f)}$ are the overall mean intercept, mean linear slope of time, and mean quadratic slope of time, respectively, across individuals for each factor; $\gamma_{01}^{(f)}$,$\gamma_{11}^{(f)}$, and $\gamma_{21}^{(f)}$ are the effect of between-person differences in baseline age (centred) on the intercept, linear time slope, and quadratic time slope for each factor, respectively; $c.Age_i$ reflects each person's baseline age centred using the sample mean age at baseline; and $U_{0i}^{(f)}$, $U_{1i}^{(f)}$, and $U_{2i}^{(f)}$ reflect person-specific deviations from the overall intercept, linear slope of time, and quadratic slope of time, respectively, for each factor.

The covariance structure for the random effects across factors was unrestricted. That is, we freely estimated the covariances between the random intercepts, linear slopes, and quadratic slopes for each factor, as well as between factors, forming a 15 x 15 unrestricted covariance matrix:

$$V\begin{bmatrix} U_{0i}^{(1)} \\ U_{1i}^{(1)} \\ U_{2i}^{(1)} \\ U_{0i}^{(2)} \\ U_{1i}^{(2)} \\ U_{2i}^{(2)} \\ \vdots \\ U_{2i}^{(5)} \end{bmatrix} = \begin{bmatrix}
\tau_{00}^{(1)} & & & & & & & \\
\tau_{10}^{(1)} & \tau_{11}^{(1)} & & & & & & \\
\tau_{20}^{(1)} & \tau_{21}^{(1)} & \tau_{22}^{(1)} & & & & & \\
\tau_{00}^{(2,1)} & \tau_{01}^{(2,1)} & \tau_{02}^{(2,1)} & \tau_{00}^{(2)} & & & & \\
\tau_{10}^{(2,1)} & \tau_{11}^{(2,1)} & \tau_{12}^{(2,1)} & \tau_{10}^{(2)} & \tau_{11}^{(2)} & & & \\
\tau_{20}^{(2,1)} & \tau_{21}^{(2,1)} & \tau_{22}^{(2,1)} & \tau_{20}^{(2)} & \tau_{21}^{(2)} & \tau_{22}^{(2)} & & \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \\
\tau_{20}^{(5,1)} & \tau_{21}^{(5,1)} & \tau_{22}^{(5,1)} & \tau_{20}^{(5,2)} & \tau_{21}^{(5,2)} & \tau_{22}^{(5,2)} & \dots & \tau_{22}^{(5)}
\end{bmatrix}$$

References

1. Reise SP. The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*. 2012;47(5):667–696.
2. Muthén BO. Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*. 1991;28(4):338-354.
3. Muthén BO. Multilevel covariance structure analysis. *Sociological methods & research*. 1994;22(3):376-398.

Supplement 2: Longitudinal Measurement Invariance Testing

We used multilevel factor analysis with ‘time’ at the within-level and ‘subject’ at the between-level to estimate general and specific psychopathology factors at the within-level over time (see Supplement 1). A disadvantage of this modeling approach is that it was not possible to test for measurement invariance in the conventional sense, i.e. by holding factor loadings and item intercepts/thresholds constant at each time-point. This is because ‘time’ is an inherent feature of model parameters, e.g., a within-level factor loading reflects the way in which an item is predicted to covary with other items across time. In contrast, the conventional measurement invariance test relies on a single-level model, where factors are estimated at each time-point, and hence model parameters can be freely estimated or held constant at each time-point. In the multilevel approach, factor loadings and item intercepts/thresholds are assumed to be invariant. For example, an item intercept is the mean of that item over the within-level (e.g., time) when a given factor equals zero.

The reviewers and authors agreed that some type of invariance testing should be undertaken to support the assumption that change was mainly attributable to the factors and not changes in measurement properties. This is despite the fact that full or partial measurement invariance shown with the conventional approach would demonstrate properties of the parameters that are not immediately transferable to the multi-level approach. A factor loading in one model is not the same as a factor loading in the other. Moreover, full or partial invariance shown using the conventional approach cannot be carried over to the multilevel model, since there are simply no parameters to hold constant. That said, the results of both single-level and multi-level growth models should ultimately converge, and so invariance observed using one approach should roughly translate to the other.

We encountered convergence issues when estimating a single-level model with wide-formatted data. We believe this was mainly due to model complexity (e.g., simultaneously estimating four bifactor models in addition to growth factors is computationally taxing). We thus estimated the general and specific psychopathology factors for two adjacent time-points within the same single-level model, which converged successfully. However, when we attempted to assess metric invariance (e.g., equal factor loadings between the adjacent time-points), chi-square difference values between models were negative, which is possible but improper and non-meaningful.[1]

As an alternative, we tested the invariance of individual factor loadings between two adjacent time-points using Wald chi-square tests via the MODEL CONSTRAINT command in Mplus. We found that all factor loadings showed metric invariance except for those associated with the mood factor between time 2 (post-treatment) and time 3 (6-months follow-up), Wald $\chi^2(4) = 11.54$, $p = .021$ (the Wald test includes all mood items for brevity but each item was initially tested individually).

We then tested for scalar invariance by comparing individual item thresholds between two adjacent time-points using Wald chi-square tests, while simultaneously testing for differences among all factor loadings (the latter was intended to mimic equality constraints on all factor loadings, which is a prerequisite when testing scalar invariance). Each of the 20 items had two thresholds (threshold A and B) which were compared at three adjacent time-points (time 1 vs. time 2, time 2 vs. time 3, time 3 vs. time 4), resulting in 120 tests. To minimize family wise error rates, we corrected the alpha level for the number of tests conducted on a single threshold between two adjacent time-points using the Bonferroni method (e.g., α/k, where α

is the type I error rate and k is the number of tests). Therefore, $\alpha = .003$ ($\alpha/k = .05/20$) when testing the equivalence of one of the two thresholds for each of the 20 items between two adjacent time-points.

Threshold A was invariant for 80% of items between time 1 and 2, while threshold B was invariant for 60% of items. Between time 2 and 3, threshold A was invariant for 90% of items, while threshold B was invariant for 95% of items. Finally, 100% of items showed invariance in threshold A and B between time 3 and 4. Non-invariance of item thresholds was thus mainly apparent between time 1 (baseline) and 2 (post-treatment), which may be because pre-treatment distributions can deviate from post-treatment distributions.[2,3] Three of the nine items (33%) that showed non-invariance in threshold A between time 1 and 2 also showed non-invariance in threshold B (e.g., SDQ items 5 and 12, and MFQ item 5). Therefore, the majority of non-invariance appeared sporadic rather than systematic.

In all, our conventional measurement invariance analysis demonstrates partial longitudinal measurement invariance, but caution is warranted when extending these findings to the multilevel model.

References

1. Satorra A, Bentler P. M. A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika.* 2001;66(4):507-514.
2. Hedeker D, Gibbons R. D. *Longitudinal Data Analysis*. 2006; John Wiley & Sons, Hoboken, NJ.
3. Vickers A. J. Parametric versus non-parametric statistics in the analysis of randomized trials with non-normally distributed data. *BMC medical research methodology*. 2005;5(1)5-35.

Supplement 3: Sensitivity Analysis – Growth Model Without Cross-loadings

We re-ran the multilevel growth model described in the paper using Bayesian plausible values from a bifactor model that did not include cross-loadings (see paper for model fit and Table S5 for factor loadings). Our goal was to determine the influence of cross-loadings on the direction and significance of the growth curves, particularly for the specific anxiety and antisocial factors. The decline in antisocial scores may have been driven by an increase in the negatively weighted anxiety item which cross-loaded onto the antisocial factor. Similarly, anxiety scores may have increased because of a decrease in the negatively weighted antisocial item or attention items which cross-loaded.

In the multilevel growth model without cross-loadings, the anxiety factor continued to show a significant linear increase over the study period (β = .34, *p* < .001, 95% CI [.18, .51]; see Figure S3b). The increase was stronger in magnitude than the model that included cross-loadings, most likely because of SDQ item 16's boost in loading strength from no longer cross-loading on the antisocial factor. Overall, it does not appear that the antisocial and attention items that cross-loaded on the anxiety factor underpinned its increase over time.

In contrast, the antisocial factor still declined over the study period (β = -.05, *p* = .614, 95% CI [-.22, .13]) but at a weaker magnitude which was no longer significant (see Figure S3a). Hence, it appears that the negatively weighted SDQ item 16 ('I am [not] nervous in new situations') contributed much to the decline in antisocial scores. However, to say that antisocial scores declined because of an increase in anxiety may not be entirely accurate, because SDQ item 16 loaded more strongly onto, and hence better represents, the antisocial factor than the anxiety factor. We would argue that in the context of the antisocial factor, SDQ item 16 reflects fearlessness more than separation anxiety (the original item meaning). Furthermore, forcing SDQ item 16 to load exclusively onto the anxiety factor despite its affinity to the antisocial factor may have supressed the latter's growth curve in the parallel process growth model.

As for the other factors, the p factor continued to decline over time (β = -.47, *p* < .001, 95% CI [-.60, -.34]), which, like the anxiety factor, was stronger in magnitude than the model featuring cross-loadings (see Figure S3a). Removing the cross-loadings appears to have strengthened changes in the general variance**,** perhaps because the general factor may absorb the variance associated with unmodelled cross-loadings.[1] Moreover, the quadratic slope for the p factor was now significant, albeit just (β = .04, *p* = .045, 95% CI [.01, .08]). The mood (β = -.04, *p* = .638, 95% CI [-.21, .13]) and attention (β = .02, *p* = .779, 95% CI [-.12, .16]) factors both decreased slightly in their baseline values compared to the model with cross-loadings, but continued to show little change over time (see Figure S3c).

References

1. Murray A, Johnson W. The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. Intelligence. 2013;41(5):407-422.

Table S1. Standardized Factor Loadings for the Mood and Feelings Questionnaire (Exploratory Within-level Factor Analysis)

	Factor	
Scale/Item	Self-Attitudes	Mood
1. I felt miserable or unhappy.	0.35	**0.36**
2. I didn't enjoy anything at all.	0.36	**0.34**
3. I felt so tired I just sat around and did nothing.	0.02	**0.61**
4. I was very restless.	-0.01	**0.68**
5. I felt I was no good anymore.	0.69	**0.34**
6. I cried a lot.	0.65	0.14
7. I found it hard to think properly or concentrate.	0.38	0.26
8. I hated myself.	0.85	0.02
9. I was a bad person.	0.72	0.00
10. I felt lonely.	0.78	0.03
11. I thought nobody really loved me.	0.84	-0.02
12. I thought I could never be as good as other kids.	0.83	-0.08
13. I did everything wrong.	0.81	-0.05

Note: Top five items loading ≥ .32 on the mood factor are in bold and were used in the primary model.

Table S2. Correlation Matrix of Bayesian Plausible Values for the General (*p*) and Specific (Anxiety, Mood, Antisocial, Attention) Psychopathology Factors

	p	Anxiety	Mood	Antisocial	Attention
p	—				
Anxiety	-0.042	—			
Mood	0.002	0.048	—		
Antisocial	0.06	-0.079	-0.004	—	
Attention	0.003	-0.016	-0.025	0.034	—

Note: The average number of observations over 100 imputations was 2,732 for 683 cases. Correlations between factors were set at zero in the original model.

Table S3. Within-level Polychoric Correlation Matrix. Items are Arranged by Specific Factor (eg, 1-5 = Anxiety, 6-10 = Mood, 11-15 = Antisocial, and 16-20 = Attention)

Item	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
1. SDQ 3	—																			
2. SDQ 8	0.43	—																		
3. SDQ 13	0.45	0.56	—																	
4. SDQ 16	0.29	0.44	0.39	—																
5. SDQ 24	0.40	0.51	0.42	0.40	—															
6. MFQ 1	0.39	0.43	0.55	0.29	0.33	—														
7. MFQ 2	0.26	0.25	0.40	0.21	0.22	0.54	—													
8. MFQ 3	0.30	0.26	0.29	0.21	0.19	0.38	0.45	—												
9. MFQ 4	0.30	0.25	0.32	0.18	0.18	0.42	0.39	0.50	—											
10. MFQ 5	0.36	0.44	0.57	0.29	0.32	0.63	0.61	0.45	0.46	—										
11. SDQ 5	0.29	0.24	0.34	0.23	0.17	0.35	0.25	0.24	0.30	0.37	—									
12. SDQ 7	0.01	0.01	0.18	0.00	-0.01	0.11	0.18	0.08	0.17	0.20	0.36	—								
13. SDQ 12	0.19	0.04	0.28	-0.01	0.05	0.15	0.17	0.11	0.22	0.20	0.43	0.29	—							
14. SDQ 18	0.23	0.19	0.29	0.16	0.13	0.24	0.23	0.16	0.27	0.28	0.35	0.25	0.35	—						
15. SDQ 22	0.08	0.03	0.24	0.03	0.11	0.12	0.14	0.03	0.16	0.21	0.21	0.24	0.41	0.39	—					
16. SDQ 2	0.25	0.20	0.23	0.24	0.14	0.20	0.13	0.19	0.40	0.22	0.41	0.15	0.27	0.28	0.14	—				
17. SDQ 10	0.27	0.22	0.25	0.30	0.17	0.18	0.15	0.17	0.36	0.24	0.39	0.18	0.28	0.31	0.17	0.68	—			
18. SDQ 15	0.24	0.26	0.24	0.38	0.17	0.17	0.14	0.19	0.29	0.24	0.46	0.24	0.24	0.31	0.15	0.55	0.57	—		
19. SDQ 21	0.07	0.03	0.14	0.07	0.02	0.11	0.14	0.10	0.18	0.22	0.38	0.43	0.26	0.27	0.21	0.29	0.31	0.37	—	
20. SDQ 25	0.02	0.04	0.11	0.18	0.00	0.11	0.11	0.06	0.14	0.17	0.20	0.37	0.09	0.13	0.10	0.22	0.23	0.34	0.41	—

Table S4. Within-Level Standardized Factor Loadings for the Common Factor Model

	Factor
Scale/Item	General
SDQ	
3. I get a lot of headaches	0.52***
8. I worry a lot	0.55***
13. I am often unhappy	0.65***
16. I am nervous in new situations	0.46***
24. I have many fears	0.45***
5. I get very angry	0.61***
7. I [do not] usually do as I am told	0.34***
12. I fight a lot	0.40***
18. I often get accused of lying or cheating	0.48***
22. I take things that are not mine	0.31***
2. I am restless	0.61***
10. I am constantly fidgeting	0.64***
15. I am easily distracted	0.62***
21. I [do not] think before I do things	0.41***
25. I [do not] finish the work I am doing	0.31***
MFQ	
1. I felt miserable/unhappy	0.63***
2. I didn’t enjoy anything	0.57***
3. I felt so tired I just sat around and did nothing	0.49***
4. I was very restless	0.58***
5. I felt I was no good anymore	0.73***
M	0.52
SD	0.12

Note: *M* = mean; MFQ = Mood and Feelings Questionnaire; SDQ = Strengths and Difficulties Questionnaire.

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table S5. Within-Level Standardized Factor Loadings for the Correlated Factors Model and Factor Correlations

	Factor			
Scale/Item	Anxiety	Antisocial	Attention	Mood
SDQ				
3. I get a lot of headaches	0.63***			
8. I worry a lot	0.70***			
13. I am often unhappy	0.80***			
16. I am nervous in new situations	0.57***			
24. I have many fears	0.57***			
5. I get very angry		0.78***		
7. I [do not] usually do as I am told		0.46***		
12. I fight a lot		0.54***		
18. I often get accused of lying or cheating		0.60***		
22. I take things that are not mine		0.42***		
2. I am restless			0.74***	
10. I am constantly fidgeting			0.78***	
15. I am easily distracted			0.76***	
21. I [do not] think before I do things			0.54***	
25. I [do not] finish the work I am doing			0.42***	
MFQ				
1. I felt miserable/unhappy				0.74***
2. I didn’t enjoy anything				0.67***
3. I felt so tired I just sat around and did nothing				0.58***
4. I was very restless				0.64***
5. I felt I was no good anymore				0.86***
M	0.65	0.70	0.56	0.65
SD	0.10	0.15	0.14	0.16

	1.	2.	3.	4.

1. Anxiety	—			
2. Antisocial	0.43***	—		
3. Attention	0.43***	0.72***	—	
4. Mood	0.69***	0.52***	0.39***	—

Note: *M* = mean; MFQ = Mood and Feelings Questionnaire; *SD* = standard deviation; SDQ = Strengths and Difficulties Questionnaire.
$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table S6. Within-Level Standardized Factor Loadings for a Confirmatory Bifactor Model Without Cross-loadings

	Factor				
Scale/Item	General	Anxiety	Antisocial	Attention	Mood
SDQ					
3. I get a lot of headaches	0.49***	0.34***			
8. I worry a lot	0.46***	0.63***			
13. I am often unhappy	0.62***	0.40***			
16. I am nervous in new situations	0.42***	0.38***			
24. I have many fears	0.34***	0.59***			
5. I get very angry	0.67***		0.22***		
7. I [do not] usually do as I am told	0.35***		0.29***		
12. I fight a lot	0.37***		0.57***		
18. I often get accused of lying or cheating	0.48***		0.35***		
22. I take things that are not mine	0.27***		0.55***		
2. I am restless	0.47***			0.64***	
10. I am constantly fidgeting	0.51***			0.63***	
15. I am easily distracted	0.55***			0.48***	
21. I [do not] think before I do things	0.39***			0.27***	
25. I [do not] finish the work I am doing	0.28***			0.28***	
MFQ					
1. I felt miserable/unhappy	0.54***				0.47***
2. I didn't enjoy anything	0.45***				0.60***
3. I felt so tired I just sat around and did nothing	0.41***				0.47***
4. I was very restless	0.52***				0.35***
5. I felt I was no good anymore	0.66***				0.50***
M	0.46	0.47	0.40	0.46	0.48
SD	0.11	0.13	0.16	0.18	0.09
ω/ω_s	0.91	0.80	0.73	0.78	0.83

ω_H/ω_{Hs}	0.73	0.40	0.34	0.41	0.39
ECV/ECV$_s$	0.51	0.13	0.10	0.13	0.13

Note: ECV = Explained Common Variance; ECV$_s$ = Explained Common Variance subscale; *M* = mean; MFQ = Mood and Feelings Questionnaire; SD = standard deviation; SDQ = Strengths and Difficulties Questionnaire; ω = omega; ω_s = omega subscale; ω_H = omega hierarchical; ω_{Hs} = omega hierarchical subscale.

$^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table S7. Within-Level Standardized Factor Loadings for an Exploratory Bi-factor Model (Bi-Geomin Orthogonal Rotation)

	Factor				
Scale/Item	General	Anxiety	Antisocial	Attention	Mood
SDQ					
3. I get a lot of headaches	0.55	0.26	-0.07	0.01	0.09
8. I worry a lot	0.65	0.44	-0.30	-0.14	-0.03
13. I am often unhappy	0.74	0.17	0.00	-0.23	0.08
16. I am nervous in new situations	0.54	0.33	**-0.37**	0.06	-0.08
24. I have many fears	0.54	0.39	-0.24	-0.13	-0.07
5. I get very angry	0.58	-0.16	0.25	0.16	0.04
7. I [do not] usually do as I am told	0.30	**-0.50**	0.32	-0.04	0.02
12. I fight a lot	0.38	0.02	0.60	0.11	-0.03
18. I often get accused of lying or cheating	0.46	-0.03	0.33	0.09	0.01
22. I take things that are not mine	0.32	-0.01	0.48	-0.04	-0.07
2. I am restless	0.48	-0.04	0.05	0.66	0.05
10. I am constantly fidgeting	0.52	-0.06	0.04	0.62	-0.02
15. I am easily distracted	0.56	-0.25	-0.04	0.44	-0.10
21. I [do not] think before I do things	0.35	**-0.52**	0.23	0.14	-0.01
25. I [do not] finish the work I am doing	0.28	**-0.58**	-0.03	0.09	-0.03
MFQ					
1. I felt miserable/unhappy	0.60	0.08	-0.05	-0.20	0.39
2. I didn’t enjoy anything	0.46	-0.04	0.03	-0.18	0.54
3. I felt so tired I just sat around and did nothing	0.37	0.05	-0.07	0.06	0.54
4. I was very restless	0.45	0.02	0.07	0.23	0.51
5. I felt I was no good anymore	0.66	-0.04	0.00	-0.21	0.46

Note: Items in bold reflect cross-loadings meeting the threshold of .32. Model fit: CFI = .95, TLI = .91, RMSEA = .06, SRMR = .04. MFQ = Mood and Feelings Questionnaire; SDQ = Strengths and Difficulties Questionnaire.

Table S8. Correlations Between Random Intercepts, Random Linear Slopes, and Random Quadratic Slopes for the General (p) and Specific Psychopathology Factors

	1.	2.	3.	4.	5.	6	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. $U_{0i}^{(p)}$	0.38***														
2. $U_{1i}^{(p)}$	-0.11	0.26													
3. $U_{2i}^{(p)}$	0.02	-0.08	0.03												
4. $U_{0i}^{(anxiety)}$	-0.02	0.02	0.00	0.22***											
5. $U_{1i}^{(anxiety)}$	0.07	-0.07	0.02	-0.14	0.27										
6. $U_{2i}^{(anxiety)}$	-0.01	0.02	-0.01	0.03	-0.08	0.03									
7. $U_{0i}^{(mood)}$	-0.01	0.07	-0.02	0.06	-0.05	0.01	0.16*								
8. $U_{1i}^{(mood)}$	0.09	-0.14	0.04	-0.04	0.09	-0.02	-0.14	0.27							
9. $U_{2i}^{(mood)}$	-0.02	0.04	-0.01	0.01	-0.02	0.01	0.03	-0.08	0.03						
10. $U_{0i}^{(anti)}$	-0.01	0.05	-0.01	-0.03	-0.01	0.00	-0.01	0.00	0.00	0.16*					
11. $U_{1i}^{(anti)}$	0.04	-0.07	0.02	-0.01	0.02	-0.01	0.03	-0.01	0.00	-0.14	0.31				
12. $U_{2i}^{(anti)}$	-0.01	0.02	-0.01	0.01	-0.01	0.00	-0.01	0.00	0.00	0.03	-0.09	0.03			
13. $U_{0i}^{(atten)}$	0.02	0.06	-0.02	-0.04	0.02	0.00	-0.03	0.02	-0.01	0.02	-0.03	0.01	0.24***		
14. $U_{1i}^{(atten)}$	0.00	-0.11	0.04	-0.01	0.05	-0.02	0.00	-0.01	0.00	0.00	0.05	-0.02	-0.17	0.33	
15. $U_{2i}^{(atten)}$	0.00	0.03	-0.01	0.00	-0.02	0.01	0.00	0.00	0.00	0.00	-0.01	0.01	0.04	-0.09	0.03

Note: Variances are on the diagonal. anti = specific antisocial factor; atten = specific attention factor; p = general psychopathology; U_{0i} = random intercept; U_{1i} = random linear slope; U_{2i} = random quadratic slope.

$^{***}p$ < .001; $^{**}p$ < .01; $^{*}p$ < .05.

Table S9. Regression Coefficients of the Random Effects for Each Factor on Baseline Age

Parameter	*B*	*p*	95% LL	95% UP
Random Intercept				
p	-0.03	0.24	-0.09	0.02
Anxiety	0.02	0.57	-0.05	0.08
Mood	-0.02	0.58	-0.08	0.04
Antisocial	-0.02	0.62	-0.09	0.05
Attention	-0.02	0.46	-0.09	0.04
Random Slope (Linear)				
p	0.06	0.16	-0.02	0.15
Anxiety	0.00	0.99	-0.11	0.11
Mood	0.03	0.62	-0.08	0.13
Antisocial	-0.02	0.73	-0.14	0.10
Attention	-0.02	0.74	-0.10	0.07
Random Slope (Quadratic)				
p	-0.02	0.11	-0.05	0.00
Anxiety	0.00	0.99	-0.03	0.03
Mood	0.00	0.82	-0.04	0.03
Antisocial	0.01	0.71	-0.03	0.04
Attention	0.01	0.64	-0.02	0.03

Note: *B* = partially standardized beta; LL = lower limit; UP = upper limit

Figure S1. Schematic of the Item-Level Multilevel Confirmatory Bi-factor Analysis With Cross-loadings

Note: Each box reflects an observed item from the Strengths and Difficulties Questionnaire (SDQ) or Mood and Feelings Questionnaire (MFQ). Each circle reflects a latent variable which was estimated at the within-level only. p = general psychopathology; Anx = anxiety; Anti = antisocial; Atten = attention.

Figure S2. Schematic of the Multilevel Growth Curve Model Using Bayesian Plausible Values for the Within-level Bifactor Dimensions

Note: General (p) and specific psychopathology factor scores were regressed onto linear and quadratic time variables. Random effects are illustrated by the black circles at the end of the path (random intercepts) and at the middle of the path labelled with an S (random slopes). At the between level, the random intercept (i), random linear slope (s), and random quadratic slope (s^2) for each factor were correlated, and also regressed on a centered age variable. p = general psychopathology; Anx = anxiety; Anti = antisocial; Atten = attention; c.Age = age centred.

Figure S3. Predicted and Observed Within-level Growth Curves for the p Factor and Specific Anxiety, Mood, Antisocial, and Attention Factor BPVs Estimated From a Model Without Cross-loadings.

Note: Average predicted trajectories (curves) and observed means (data points with error bars) for (A)the general psychopathology and specific antisocial factors, (B) the specific anxiety factor, and (C) the specific mood and attention factors. The zero-point reflects the factor mean. Error bars indicate 95% confidence intervals.