**Supplementary Materials 2: Cluster-corrected factor analysis**

In z-proso, multiple children were rated by the same teachers depending on the class in which the children were taught. Given this clustering, the raw item covariance matrices at each wave reflect a mixture of covariance due to the factor structure at the teacher (between-group) and child (within-group) level. We, therefore, corrected the item covariance matrix for covariance at the between-group level prior to analysis using the method proposed by Muthén (1994) and outlined below. To evaluate the extent to which the variation in item responses was due to influences at the within- versus between- group level, we computed Muthén’s (1994) intraclass correlation coefficient for the SBQ items. Magnitudes of >.05 suggest that there is sufficient variation of at the between-group level for multi-level analysis to be necessary (Dyer, Hanges & Hall, 2005). The average item ICCs for the 8 measurement waves are provided in Table 1 and in all cases are >.05.

To evaluate the within-group factor structure, we began by obtaining an estimate of the pooled within-group correlation matrix using *Mplus 6.*11(Dyer, Hanges & Hall, 2005; Muthén & Muthén, 2010; Muthén, 1994). This provides an estimate of the correlation between items corrected for rater effects**.** We then conducted an exploratory factor analysis using the corrected correlation matrix for one of the middle waves of data (Age 10) and checked that similar solutions were found at the two most distant waves (waves Age 7 and Age 15). The results from these exploratory analyses were used to guide our decision as to how many factors to extract in the main analysis.

We used the minimum average partial (MAP; Velicer, 1976) test, parallel analysis with principal components analysis (PA-PCA; Horn, 1965) and visual inspection of a scree plot to guide factor retention. These three methods have been recommended as among the most accurate under a range of conditions (Crawford et al., 2010; Velicer, Eaton & Fava, 2000). If the methods suggested diverging numbers of factors, we examined the corresponding factor solutions and used the theoretical interpretability of these as a further criterion for choosing a factor structure. We estimated factor solutions with varying numbers of factors using minimal residuals (minres) extraction with a bi-factor rotation (Jennrich & Bentler, 2011). All of this was conducted using the ‘psych’ package in R statistical software (R Core Team, 2014; Revelle, 2014).

Using the within-group pooled correlation matrix for wave Age 10, PA-PCA suggested 6 and the MAP test suggested 8 factors to retain. A scree plot (Figure S1) suggested a hierarchical structure with one strong general factor, 3-5 relatively specific factors of potential substantive magnitude and a number of additional specific factors, likely reflecting minor factors. We, therefore, examined several solutions ranging in their number of specific factors. Beyond 6 factors (1 general, 5 specific), minor factors began to appear. We, therefore, focussed on two solutions: a 5 factor (1 general, 4 specific), and 4 factor (1 general, 3 specific) model. When 1 general and 3 specific factors were extracted, these could be characterised as ‘Internalising, ‘ADHD’, and ‘Externalising’. The Internalising factor was defined by anxiety, depression and low pro-sociality; the ADHD factor was defined by attention deficits and hyperactive behaviours; and the Externalising factor was characterised by aggressive, defiant, oppositional behaviours and conduct problems. When 1 general and 4 specific factors were extracted, 3 of these specific factors were similar to those identified in the factor solutions with 3 specific factors and the fourth was defined by high pro-sociality and the pro-sociality items no longer loaded strongly on the Internalising factor. Given that the 1 general, 4 specific factor solution allowed a distinction between pro-sociality and internalising to be made while the 1 general, 3 specific factor solution blurred these constructs, we preferred the factor solution with 1 general and 4 specific factors on theoretical grounds. We were able to confirm that similar results held using the within-group pooled correlation matrices waves Age 7 and Age 15 when the children were at their youngest and oldest respectively. The optimal structure at all three points was judged to be one in which there was 1 general factor and 4 specific factors.

**References**

Crawford, A. V., Green, S. B., Levy, R., Lo, W. J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of parallel analysis methods for determining the number of factors. *Educational and Psychological Measurement, 70,* 885-901.

Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, *16*, 149- 167.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179-185.

Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika*, *76*, 537-549.

Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological Methods & Research*, *22*, 376-398.

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

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.R-project.org>.

Revelle, W. (2014). psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <http://CRAN.R>project.org/package=psych Version =1.4.8.

Velicer, W.F., Eaton, C.A., & Fava, J.L. (2000). Construct explication through Factor or Component Analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In Goffin, R. D., & Helmes, E. (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy.* Boston: Kluwer. (Pp. 41-71).

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321-327.

**Figures**

**Figure S1**

**F:\z-proso\p-factor 10 waves\p factor Figure 2.tiff**