We thank [AUTHORS] for their thoughtful comments on our article about changes in the psychopathology bifactor dimensions over a psychosocial intervention.[1] [AUTHORS] expressed concerns over our interpretation of the p factor and our selection of the bifactor model over alternative models. We will address each concern and highlight the issues they raise for the field more broadly.

How do we interpret the p factor?

[AUTHORS] discuss how the widespread interpretation of the p factor as a causal entity is problematic because factors rely on an unproven assumption of causality. We agree that authors (including ourselves) could be more mindful of these assumptions in our writing. It might be helpful to distinguish between two distinct but related entities: 'general psychopathology', the construct hypothesised to explain the positive co-occurrences among mental health problems, and the 'p factor', a statistical representation of these positive co-occurrences that is no more 'real' than any dispersion statistic for representing individual differences. Unless we validate the p factor against external criteria (as we attempted), or better still against prospective measures of causal mechanisms, we are at risk of making interpretative leaps beyond the data.

The p factor is first and foremost a statistical re-expression of the covariance among psychopathology variables. If this covariance accurately reflects co-occurrences in people's experiences of mental health problems, then the p factor will, statistically speaking, represent covariation in mental health problems and its underlying construct, i.e. general psychopathology. Like any measure, the p factor is influenced by the methods used to estimate it,[2] but the target construct remains constant. This is not to say that general psychopathology is unidimensional; there are various risk factors for psychopathology that interact in complex ways for each individual.[3] However, the p factor provides a tool for isolating these broad influences and investigating their treatment targets. Contrary to [AUTHORS] suggestion that an artifactual p factor could limit its utility in studying mechanisms for improving treatment, specific factors and their treatment targets can still be investigated free from the common method variance (which is conflated in other models).

Why did we choose the revised bifactor model?

We chose a 'revised' bifactor model with cross-loadings over a standard bifactor model without cross-loadings and a correlated factors model. As [AUTHORS] point out, the correlated factors model fit slightly better than the standard bifactor model, which is surprising given the bias towards bifactor models in model comparisons.[4] This was likely influenced by constraining the shared variance beyond general factor, and was hence resolved by freeing the cross-loadings. On a related note, we did not revise the correlated factors model because the cross-loadings were a result of shared variance beyond the p factor and hence not present in an exploratory correlated factors model, which conflates the general and specific variance.

Nonetheless, our modelling decisions raise two issues with the current practice of bifactor modelling. The first, discussed by [AUTHORS], is choosing a model using model fit indices alone. Researchers (including ourselves) can be criticized for prioritizing the bifactor model's superior fit–which can occur for non-substantive reasons[4]–without justifying its theoretical basis.[5] The notion of a severity dimension that is distinct from styles of

symptomatic expression appears to have emerged a posteriori with the resurgence of bifactor models, but it has a history in personality research and clinical practice. Still, we would argue that theoretical justification is not the only reason why a model might be preferred over another. Models might be selected because they are practically useful, even if they do not represent the true data-generating mechanism. Total and subscale scores in psychopathology measures tend to be underpinned by a single dimension, even if they sample a diverse range of problems.[2] The bifactor model allows us to capture variation in responding that we partially impose with the design of our measures; clinical outcomes look rather different when we take this into consideration.[1]

The second issue concerns shared variance beyond the *p* factor. Bifactor models with cross-loadings or specific factor correlations are becoming popular, but there is a danger in freeing these covariances to improve model fit without considering the consequences. Estimating shared variance threatens the interpretation of specific factors and implies model mis-specification due to unmodelled factors.[5] Yet not estimating the shared variance can also lead to model mis-specification and inflate general factor loadings.[6] Justifying the inclusion of shared variance beyond the *p* factor with past research (as we attempted) is important to avoid capitalizing on sample-specific error, but further work is needed to identify its methodological and theoretical impact.

References

1. Constantinou MP, Goodyer IM, Eisler I, et al. Changes in General and Specific Psychopathology Factors Over a Psychosocial Intervention. *J Am Acad Child Adolesc Psychiatry*. 2019;58(8):776-786. doi:10.1016/j.jaac.2018.11.011
2. Constantinou MP, Fonagy P. Evaluating Bifactor Models of Psychopathology with Model-Based Reliability Statistics. *PsyArXiv*. Pre-print. doi:10.31234/osf.io/6tf7j
3. Cicchetti D. Risk, Disorder, and Adaptation. In: Cicchetti D, ed. Developmental Psychopathology. 3rd ed. New York, NY: John Wiley & Sons; 2016:726-792.
4. Greene AL, Eaton NR, Li K, et al. Are fit indices used to test psychopathology structure biased? A simulation study. *J Abnorm Psychol.* In press. doi:10.1037/abn0000434.
5. Sellbom M, Tellegen A. Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*. In press. doi.org/10.1037/pas0000623
6. Murray AL, Johnson W. The limitations of model fit in comparing the bi-factor versus higher-order models of human cognitive ability structure. *Intelligence*. 2016;41(5):407-422. doi:10.1016/j.intell.2013.06.004