


# Over reliance on model fit indices in confirmatory factor analyses may lead to incorrect inferences about bifactor models: A cautionary note

**Author:**Tyrone B. Pretorius<sup>1</sup> **Affiliation:**

<sup>1</sup>Department of Psychology,  
Faculty of Community and  
Health Sciences, University of  
the Western Cape, Cape  
Town, South Africa

**Corresponding author:**

Tyrone Pretorius,  
tpretorius@uwc.ac.za

**Dates:**

Received: 29 Sept. 2020

Accepted: 21 Jan. 2021

Published: 12 Mar. 2021

**How to cite this article:**

Pretorius, T.B. (2021). Over  
reliance on model fit indices  
in confirmatory factor  
analyses may lead to  
incorrect inferences about  
bifactor models: A cautionary  
note. *African Journal of  
Psychological Assessment*,  
3(0), a35. [https://doi.  
org/10.4102/ajopa.v3i0.35](https://doi.org/10.4102/ajopa.v3i0.35)

**Copyright:**

© 2021. The Authors.  
Licensee: AOSIS. This work  
is licensed under the  
Creative Commons  
Attribution License.

This brief article attempts to describe the importance of relying not just on model fit indices, but also on bifactor, confirmatory factor analysis to examine the factor structure of instruments presumed to be multidimensional. Three ancillary bifactor indices (explained common variance, omega hierarchical and percentage uncontaminated correlations) were calculated for three instruments that have been described as multidimensional in published research. One of these instruments, the Normative Beliefs about Aggression Scale (NOBAGS) demonstrated strong evidence of multidimensionality. The second instrument, Problem-Solving Inventory demonstrated some evidence of multidimensionality, but must be considered essentially unidimensional because of lack of sufficient evidence. The third instrument, Cyberchondria Severity Scale demonstrated essential unidimensionality with little evidence of multidimensionality. These findings support the argument that using only model fit statistics may lead researchers to draw incorrect conclusions about the dimensionality of an instrument.

**Keywords:** ancillary bifactor indices; bifactor models; model-fit indices; multidimensionality; unidimensionality.

Researchers may be interested in the scores produced by measuring instruments only to the extent that they can be used to test a proposed hypothesis. However, a critical first step in using a measuring instrument is to conduct an examination of its psychometric properties, such as reliability and validity. The investigation of the psychometric properties of an instrument should be undertaken even when the purpose of the research is not necessarily focused on the investigation of the psychometric properties. Researchers need to be certain that the scores obtained from any instrument are useful. This is critical as measurement properties are not inherent qualities of tests but rather of scores (Zangaro, 2019). As such, measuring instruments can have different properties in different applications and with different samples in different contexts. One such property of measuring instruments is validity, which refers to the extent that a measuring instrument actually measures the construct it claims to measure (Clark & Watson, 1995).

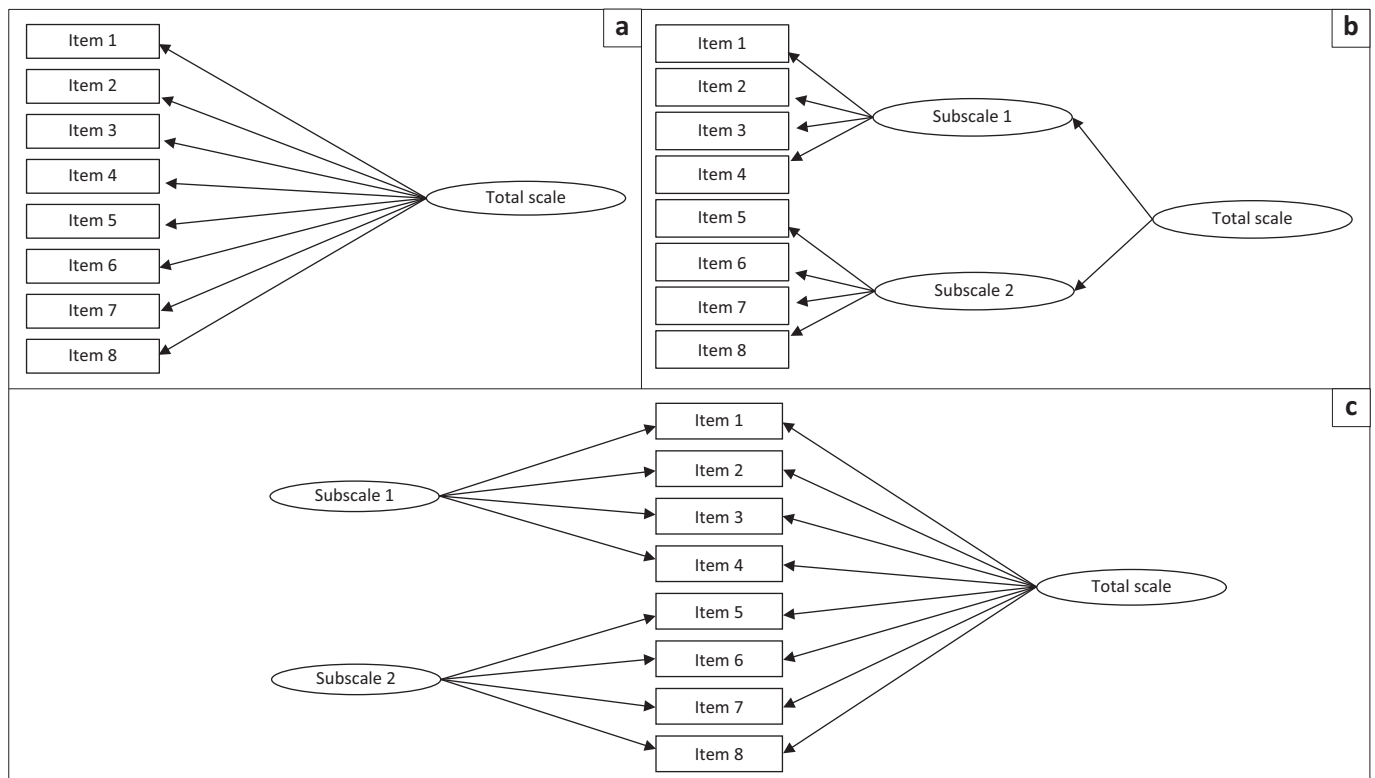
Construct validity is a type of validity that is often examined by use of confirmatory factor analysis (CFA) (Brown & Moore, 2012). If the hypothesised underlying structure of an instrument is replicated, the replication is considered indicative of construct validity. However, it should be noted that confirming that the structure of an instrument holds is only one part of validity. It is necessary to show that an instrument has both internal and external validity before using the instrument in practice.

In CFA, the scale items are regarded as observed measurements and the hypothesised factors are regarded as latent variables. If an instrument is hypothesised to have a total score and subscale scores, CFA will typically be used to examine several models of the structure of the instrument to determine which models best fits the data. Studies generally compare three conceptualisations of the factor structure of an instrument that is hypothesised to consist of a total scale and several subscales: a one-factor model, a second-order factor model and a bifactor model (see Figure 1). For example, Reynolds and Keith (2017) used a one-factor model, a second-order factor model and a bifactor model to examine the structure of the Wechsler Intelligence Scale for Children.

In the one-factor model, all items load on a total scale score. In the bifactor model, however, items load on both a total scale score (referred to as the 'general factor') and several subscale scores (referred to as 'specific factors': Reise et al., 2013). In the second-order model, items load first on

**Read online:**

Scan this QR  
code with your  
smart phone or  
mobile device  
to read online.



**FIGURE 1:** Three conceptualisations of the factor structure of a hypothetical eight-item scale. (a) One-factor model, rectangles are observed variables. Ellipse is latent variable; (b) Second-order factor model and (c) bifactor model.

several subscales, and the subscale scores in turn load on a total scale. Therefore, the relationship between the total scale and the observed items is mediated by the subscales in the second-order model (Brown & Moore, 2012).

Many researchers rely solely on model fit indices to compare and select the best-fitting model (Morgan et al., 2015). The most common fit indices are the chi-squared ( $\chi^2$ ), root mean square error of approximation (RMSEA), comparative fit index (CFI), standardised root mean square residual (SRMR), Tucker-Lewis Index (TLI), Goodness-of-Fit Index (GFI), Relative Fit Index (RFI), Normed Fit Index (NFI), Bollen's Fit Index (BL89) and Akaike's information criterion. If the fit indices indicate that the one-factor model is the best fit for the data, researchers often conclude that the scale is unidimensional, whereas fit indices that support either a second-order or bifactor model are taken as evidence of multidimensionality. However, Rodriguez et al. (2016b) have called these conclusions an 'overly simplistic conceptualization of the dimensionality of psychological data' (p. 231). There is also growing scepticism about relying on fit indices alone. For example, Morgan et al. (2015) describes these fit indices as useful but cautions that 'the exclusive use of approximate fit statistics is perilous' (p. 17). Judgements about the dimensionality of a measuring instrument based solely on model fit indices are problematic for two reasons. Firstly, it has been demonstrated that these indices generally favour bifactor models even in instances where the item loadings on general and specific factors may be relatively low (Bornovalova et al., 2020). Secondly, model fit indices fail to capture the relative strength of the general factor and specific factors (Reise et al., 2013).

If model fit indices support the bifactor model as the best fit, at least three possible conclusions can be drawn: (1) the instrument is essentially unidimensional, because the specific factors do not account for specific unique variance other than that explained by the general factor; (2) some limited evidence of multidimensionality exists, but is not sufficient to exclude a unidimensional interpretation; or (3) the specific factors account for sufficient reliable variance in addition to the variance accounted for by the general factor to support the interpretation of the instrument as multidimensional. To examine the dimensionality of an instrument, Rodriguez et al. (2016a) have urged researchers to calculate ancillary bifactor indices in addition to model fit indices. Ancillary bifactor indices include explained common variance (ECV), Omega hierarchical (OmegaH) and percentage of uncontaminated correlations (PUC). Indices such as these enable an evaluation of dimensionality. It is also possible to compute McDonald's omega: a model-based estimate of reliability. Explained common variance refers to the percentage of variance amongst all items that can be explained by each factor (ECV for general factor and ECV\_S for specific factors). Percentage of uncontaminated correlations measures the number of unique correlations amongst items that can be explained by the general factor alone. OmegaH measures the proportion of systematic variance in total scores that can be attributed to individual differences on the general factor (Rodriguez et al., 2016a).

The purpose of this commentary is to demonstrate the importance of calculating ancillary bifactor indices in addition to model fit indices to examine the dimensionality

of an instrument. To this end, bifactor indices were calculated for three published papers that concluded that a bifactor model was the best fitting model for the study data.

## Method

Three published studies were selected to demonstrate the three possible outcomes of examining the dimensionality of an instrument, as described in the introduction. The studies are described below:

Padmanabhanunni (2017) examined the factor structure of the Normative Beliefs about Aggression Scale (NOBAGS: Huesmann et al., 2011). A CFA confirmed that a bifactor model with a total scale (approval of aggression) and two subscales (retaliation beliefs and general beliefs) was the best fitting model ( $\chi^2 > 0.05$ , GFI, RFI, NFI  $> 0.95$  and RMSEA = 0.05).

Heppner et al. (2002) examined the generalisability of problem-solving appraisal amongst black South Africans and investigated the psychometric properties of the Problem Solving Inventory (PSI: Heppner, 1988). The results of CFA ( $\chi^2 < 0.05$ , CFI  $> 0.95$ , NFI  $> 0.90$ , BL89  $> 0.95$  and RMSEA = 0.08) supported the hypothesised bifactor structure of the PSI as a total scale of problem-solving appraisal and three subscales (problem-solving confidence, approach-avoidance style and personal control).

Norr et al. (2015) examined a bifactor model of the Cyberchondria Severity Scale (CSS: McElroy & Shevlin, 2014) which assesses anxiety and behaviours associated with seeking online health information. A CFA confirmed a bifactor structure ( $\chi^2 > 0.05$ , CFI  $> 0.95$ , RMSEA = 0.07) consisting of a total cyberchondria scale and four subscales (reassurance, excessiveness, distress and compulsion).

## Analysis

The standardised regression loadings reported in the three studies were used to calculate the bifactor indices necessary to assess the instruments' dimensionality. The Bifactor Indices Calculator (Dueber, 2017) was used for these calculations. The existing literature provides guidelines regarding the interpretation of these indices. Explained common variance provides an indication of the relative strength of factors, such that a higher ECV ( $> 0.80$ : Rodriguez et al., 2016b) is associated with a strong general factor and indicates that the instrument is essentially unidimensional. OmegaH indicates the proportion of systematic variance in total scores that is attributable to individual differences on the general factor. It has been suggested by Rodriguez et al. (2016b) that an OmegaH greater than 0.80 indicates that the instrument is essentially unidimensional. Finally, it has also been recommended that researchers consider ECV and OmegaH in conjunction with PUC, and Reise et al. (2013) suggest that PUC values lower than 0.80, together with general ECV values greater than 0.60 and OmegaH of the general factor greater than 0.70 would indicate that the presence of some multidimensionality that is not strong

**TABLE 1:** Bifactor indices for three studies.

Bifactor indices	Padmanabhanunni (2017)	Heppner et al. (2002)	Norr et al. (2015)
ECV_GF	0.54	0.63	0.80
ECV_s1	0.20	0.14	0.06
ECV_s2	0.26	0.06	0.07
ECV_s3	-	0.16	0.03
ECV_s4	-	-	0.05
OmegaH	0.60	0.75	0.93
PUC	0.60	0.72	0.77

Note: Padmanabhanunni = 2 specific factors (retaliation beliefs and general beliefs), Heppner et al. = 3 specific factors (problem-solving confidence, approach-avoidance style and personal control), Norr et al. = 4 specific factors (reassurance, excessiveness, distress and compulsion).

ECV\_GF, explained common variance for general factor; ECV\_s, explained common variance specific factor; OmegaH, Omega Hierarchical; PUC, percent of uncontaminated correlations.

enough to rule out the interpretation of the instrument as essentially unidimensional.

## Results

The results of the ancillary bifactor analyses are reported in Table 1.

In the Padmanabhanunni (2017) study, the general factor of the NOBAGS accounted for 54% of the common variance, and the two specific factors accounted for 46% of the common variance (20% and 26%, respectively). OmegaH was 0.60, well below the cut-off of 0.80 suggested by Rodriguez et al. (2016b). When considered with PUC, the ECV of the general factor was below 0.60 and OmegaH was below 0.70. These bifactor indices clearly support the interpretation of the NOBAGS as multidimensional.

In the Heppner et al. (2002) study, the general factor of the PSI accounted for 63% of the variance, and the three specific factors accounted for 14%, 6% and 16% of the variance, respectively. OmegaH was below 0.80, which suggests that the instrument may have some multidimensionality. However, when considered with PUC, ECV was greater than 0.60, OmegaH was greater than 0.70 and PUC was lower than 0.80. These findings indicate that there is some evidence of multidimensionality, but the evidence is not strong enough to overrule the interpretation of the PSI as unidimensional.

Finally, in the Norr et al. (2015) study, the general factor of the CSS accounted for 80% of the variance, and just 20% of the variance was explained by the four specific factors. The variance explained by each of the four specific factors ranged from 3% to 7%. OmegaH was above 0.80, which indicates that this instrument is essentially unidimensional. Its unidimensionality was further confirmed when PUC, ECV and OmegaH were considered together (PUC  $< 0.80$ , ECV  $> 0.60$  and OmegaH  $> 0.70$ ).

## Conclusion

The aim of this commentary was to demonstrate that model fit indices alone provide insufficient evidence to draw conclusions

about the dimensionality of a measuring instrument. Three published papers that drew such conclusions based on fit indices of a bifactor model were subjected to ancillary bifactor analyses, in which ECV, OmegaH and PUC were used to determine the relative strength of the general factor and the specific factors. These analyses indicated that one instrument (NOBAGS) demonstrated sufficient evidence of multidimensionality. One instrument (PSI) demonstrated some evidence of multidimensionality, but the evidence was not strong enough to rule out the possibility of the instrument being unidimensional. The third instrument (CSS) did not demonstrate evidence of multidimensionality and was determined to be essentially unidimensional. These findings highlight the insufficiency of solely relying on CFA model fit indices to draw conclusions about the hypothesised structure of a measuring instrument. The model fit indices for these three studies, reported in the methods section above, showed acceptable fit indices for all three studies. However, the bifactor indices demonstrated that the assumption of multidimensionality is not tenable. Researchers, investigating bifactor models are urged to go beyond model fit indices and investigate the pattern of item loadings as well as calculating ancillary bifactor indices.

## Acknowledgements

### Competing interests

The author declares that he has no financial or personal relationships that may have inappropriately influenced him in writing this research article.

### Author's contribution

T.B.P. is the sole author of this research article.

### Ethical considerations

This article followed all ethical standards for research without direct contact with human or animal subjects.

### Funding information

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

### Data availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

## Disclaimer

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any affiliated agency of the author.

## References

- Bornoalova, M.A., Choate, A.M., Fatimah, H., Petersen, K.J., & Wiernik, B.M. (2020). Appropriate use of bifactor analysis in psychopathology research: Appreciating benefits and limitations. *Biological Psychiatry*, 88(1), 18–27. <https://doi.org/10.1016/j.biopsych.2020.01.013>
- Brown, T.A., & Moore, M.T. (2012). Confirmatory factor analysis. In *Handbook of structural equation modeling* (pp. 361–379). Retrieved from [https://www.researchgate.net/profile/Michael\\_Moore8/publication/251573889\\_Hoyle\\_CFA\\_Chapter\\_-\\_Final/links/0deec51f14d2070566000000/Hoyle-CFA-Chapter-Final.pdf](https://www.researchgate.net/profile/Michael_Moore8/publication/251573889_Hoyle_CFA_Chapter_-_Final/links/0deec51f14d2070566000000/Hoyle-CFA-Chapter-Final.pdf)
- Clark, L.A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309 – 319. Retrieved from [http://www.bwgriffin.com/gsu/courses/edur9131/content/Clark\\_validity\\_scaledevelopment.pdf](http://www.bwgriffin.com/gsu/courses/edur9131/content/Clark_validity_scaledevelopment.pdf)
- Dueber, D.M. (2017). *Bifactor indices calculator: A Microsoft excel-based tool to calculate various indices relevant to bifactor CFA models*. <https://doi.org/10.13023/edp.tool.01>
- Heppner, P.P. (1988). *The problem-solving inventory*. Palo Alto, CA: Consulting Psychologist Press.
- Heppner, P.P., Pretorius, T.B., Wei, M., Lee, D.G., & Wang, Y.W. (2002). Examining the generalizability of problem-solving appraisal in Black South Africans. *Journal of Counseling Psychology*, 49(4), 484. <https://doi.org/10.1037/0022-0167.49.4.484>
- Huesmann, L.R., Guerra, N.G., Miller, L., & Zelli, A. (2011). *The Normative Beliefs about Aggression Scale [NOBAGS]*. Retrieved from <https://rcgd.isr.umich.edu/aggr/Measures/NormativeBeliefsAboutAggScale.2011.pdf>
- McElroy, E., & Shevlin, M. (2014). The development and initial validation of the cyberchondria severity scale (CSS). *Journal of Anxiety Disorders*, 28(2), 259–265. <https://doi.org/10.1016/j.janxdis.2013.12.007>
- Morgan, G.B., Hodge, K.J., Wells, K.E., & Watkins, M.W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research?: A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, 3(1), 2–20. <https://doi.org/10.3390/jintelligence3010002>
- Norr, A.M., Allan, N.P., Boffa, J.W., Raines, A.M., & Schmidt, N.B. (2015). Validation of the Cyberchondria Severity Scale (CSS): Replication and extension with bifactor modeling. *Journal of Anxiety Disorders*, 31, 58–64. <https://doi.org/10.1016/j.janxdis.2015.02.001>
- Padmanabhanunni, A. (2017). The factor structure of the Normative Beliefs about Aggression Scale as used with a sample of adolescents in low socio-economic areas of South Africa. *South African Journal of Psychology*, 49(1), 27–38. <https://doi.org/10.1177%2F0081246317743185>
- Reise, S.P., Scheines, R., Widaman, K.F., & Haviland, M.G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling a bifactor perspective. *Educational and Psychological Measurement*, 73(1), 5–26. <https://doi.org/10.1177%2F0013164412449831>
- Reynolds, M.R., & Keith, T.Z. (2017). Multi-group and hierarchical confirmatory factor analysis of the Wechsler Intelligence Scale for Children – Fifth Edition: What does it measure? *Intelligence*, 62, 31–47. <https://doi.org/10.1016/j.intell.2017.02.005>
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016a). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, 21(2), 137. <https://doi.org/10.1037/met0000045>
- Rodriguez, A., Reise, S.P., & Haviland, M.G. (2016b). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223–237. <https://doi.org/10.1080/00223891.2015.1089249>
- Zangaro, G.A. (2019). Importance of reporting psychometric properties of instruments used in nursing research. *Western Journal of Nursing Research*, 41(11), 1548–1550. <https://doi.org/10.1177/0193945919866827>