



Best Practices in the Use of Bifactor Models: Conceptual Grounds, Fit Indices and Complementary Indicators

Buenas Prácticas en el uso de Modelos Bifactor: Teoría de Base, Índices de Ajuste e Indicadores Complementarios

Pablo Ezequiel Flores-Kanter *^{1,2}, Sergio Dominguez-Lara³,
Mario Alberto Trógolo^{1,2}, Leonardo Adrián Medrano^{1,2}

1 - Universidad Siglo 21, Córdoba, Argentina.

2 - Facultad de Psicología, Universidad Nacional de Córdoba, Argentina.

3 - Universidad de San Martín de Porres, Perú.

Introduction
Bifactor Model:
Basic Issues
Example: Factor
structure of
PANAS
Conclusion
References

Recibido: 18/05/2018 **Revisado:** 24/07/2018 **Aceptado:** 06/09/2018

Abstract

Bifactor models have gained increasing popularity in the literature concerned with personality, psychopathology and assessment. Empirical studies using bifactor analysis generally judge the estimated model using SEM model fit indices, which may lead to erroneous interpretations and conclusions. To address this problem, several researchers have proposed multiple criteria to assess bifactor models, such as a) conceptual grounds, b) overall model fit indices, and c) specific bifactor model indicators. In this article, we provide a brief summary of these criteria. An example using data gathered from a recently published research article is also provided to show how taking into account all criteria, rather than solely SEM model fit indices, may prevent researchers from drawing wrong conclusions.

Keywords: *Confirmatory factor analyses, bifactor models, PANAS, complementary statistical fit indices*

Resumen

En los últimos años, el uso de modelos bifactor ha tenido un creciente protagonismo en las literaturas de personalidad, de psicopatología y de evaluación psicológica. Los estudios empíricos que aplican este procedimiento generalmente hacen determinaciones sobre la estructura final utilizando únicamente índices de ajuste obtenidos a partir de modelos de ecuaciones estructurales, lo que puede derivar en conclusiones erróneas. Por este motivo, recientemente se han propuesto diferentes criterios a la hora de evaluar el ajuste de modelos bifactor que incluyen a) consideraciones teóricas, b) índices de ajuste global del modelo y c) indicadores específicos del modelo bifactor. En este artículo se repasan brevemente estos criterios. Asimismo, se ofrece un ejemplo utilizando datos provenientes de una publicación reciente con el fin de evidenciar cómo las decisiones basadas en los índices globales de ajuste, sin tener en cuenta las bases teóricas y los indicadores específicos del bifactor, pueden llevar a los investigadores a obtener conclusiones sesgadas.

Palabras clave: *Análisis factorial confirmatorio, modelos bifactor, PANAS, índices estadísticos complementarios*

Introduction

Despite the increasing use of bifactor models in psychology and related sciences, several researchers have noticed common errors in the bifactor analyses performed, particularly in the interpretation of model fit based solely in SEM (structural equation model) model fit indices (Gignac, 2016; Rodriguez, Reise, & Haviland, 2016). Based on literature review, the aim of the present article is to recommend multiple criteria when assessing bifactor models, including a) conceptual grounds, b) overall model fit indices, and c) additional, specific bifactor model indicators. Using data from a recent publication on factorial structure of the Positive and Negative Affect Schedule (Seib-Pfeifer, Pugnaghi, Beauducel, & Leue, 2017), we illustrate errors and misinterpretation that can result from exclusively relying on SEM model fit indices without taking into consideration certain theoretical basis and/or specific bifactor model fit indices.

Bifactor Model: Basic Issues

Within the framework of confirmatory factor analysis, albeit not exclusive to this approach, bifactor models (BM) represent one of the options available to the applied researcher for simultaneously testing the extent to which a particular set of items are explained by a general underlying factor and group-level factors. By doing so, BM constitutes a means to determine whether the construct being measured by a scale can be viewed primarily as either unidimensional or multidimensional. Bifactor models should be distinguished from hierarchical models with higher-order factors. Particularly, because in the latter the group-level factors represent dimensions of a general factor, whereas in BM, group-level factors are hypothe-

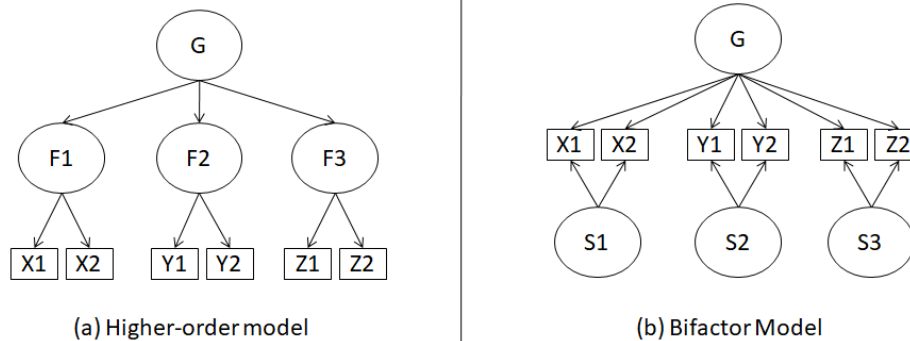
sized to be independent (i.e. orthogonal) and not lower-order factors of the general factor (see Figure 1). Accordingly, BM is especially useful for assessing the validity of an instrument intended to measure both the overall construct and its specific dimensions (Rodriguez et al., 2016).

Given that BM tends to outperform conventional confirmatory factorial analysis (CFA) models simply because of the way in which these are specified (Gignac, 2016; Morgan, Hodge, Wells, & Watkins, 2015), when a particular BM yields adequate model fit indices it is highly recommended to use complementary statistical indices for a more accurate interpretation. These additional, BM-specific fit indices are the Explained Common Variance (ECV), the Percentage of Uncontaminated Correlations (PUC), the hierarchical Omega coefficients (omega ω , omega subscale ω_s , omega hierarchical ω_h , and omega hierarchical subscale ω_{hs}) and the H coefficient (Rodriguez et al., 2016).

The ECV represents the proportion of the common variance attributable to the general factor. On the other hand, the ω_h reflects the proportion of the total variance explained by the general factor, while the ω_{hs} reflects the proportion of the total variance accounted for the specific factors after controlling for the influence of the general factor. High values on ECV ($> .60$) and ω_h ($> .70$) indicate that the variance of the indicators is substantially accounted for by the general factor and, therefore, the tenability of specific factors would be forced. Nonetheless, some authors (Bonifay, Lane, & Reise, 2017) have also pointed that if ω_{hs} explains a non-redundant amount of variance ($> .30$), then the specific factors could be retained along with the general factor, albeit considering the theoretical underpinnings of such decision. Also, there is an application of ECV at item level (ECV-I, Stucky, Thissen, & Edelen, 2013) with a similar interpretation to ECV.

Figure 1

Graphical representations of higher-order factor model and bifactor model.



Note. X1, X2, Y1, Y2, y Z1, Z3 = items; F1-F3 = factors; S1-S3 = group-level factors in bifactor model; G = general factor.

The PUC indicates the percentage of correlations not contaminated by multidimensionality (i.e., group-level factors) and moderates or supports the interpretation of ECV. According to [Rodriguez et al. \(2016\)](#), when ECV is $> .70$ and $PUC > .70$, the common variance can be regarded as essentially unidimensional, thus supporting the general factor. Finally, the H coefficient is a measurement of construct replicability and is defined as the extent to which a set of items represents a latent variable ([Hancock & Mueller, 2001](#)). This coefficient is calculated as the ratio of the percentage of variance explained by the latent variable to the percentage of variance unexplained by the latent variables. High H values ($> .80$) suggest that the latent variable is well-defined and adequately represented by the observed indicators, thus more likely to be replicated across studies.

Along with complementary model fit indices, theoretically-based reasoning should be applied when deciding in favor of a factor model other than overall model fit ([Morgan et al., 2015](#)), so that the interpretation of the resulting factors is conceptually well founded.

Example: Factor structure of PANAS

In their study, [Seib-Pfeifer et al. \(2017\)](#) concluded that the BM comprising positive affect (PA), negative affect (NA), and a general Affective Polarity factor had a better model fit compared to alternative two- and three-factor models. They made this assertion even when relevant BM statistical indices (e.g., ECV, ω_h , ω_{hs}) were not provided. Indeed, when BM statistical indices are calculated, conclusions differ markedly from [Seib-Pfeifer et al. \(2017\)](#). In particular, results show that the common variance accounted for by the general Affective Polarity factor is rather weak (ECV = 10.8 %), as it is also for the variance at item level (ECV- I_{mean} : .07 for PA and .16 for NA). The mean factor loading is moderate-to-high for both PA ($\lambda_{\text{mean}} = .52$) and NA ($\lambda_{\text{mean}} = .57$), and a significant amount of variance is also explained by the two specific factors (PA $\omega_{hs} = .78$; NA $\omega_{hs} = .72$). In contrast, both factor loadings and explained variance of the general Affective Polarity factor are close to zero ($\lambda_{\text{mean}} = .091$; $\omega_h = .049$). In addition, the PUC value is .53 and the H coeffi-

cient is acceptable for group-level factors ($H_{PA} = .80$; $H_{NA} = .77$) and low for the general factor ($H_G = .41$). All in all, these findings suggest that the conclusion drawn by Seib-Pfeifer et al. (2017) in their study, i.e that BM provides the most appropriate representation of the PANAS, is erroneous.

Conclusion

In a strict sense, a factor is a mathematical abstraction derived from the empirical covariance between a set of variables, which may (or may not) be interpreted as a common, substantive cause underlying a set of observable behaviors. The bifactor model is an alternative specification of the second order factor. In contrast to second-order models, the bifactor a) allows for the quantification of the direct effect of the general factor on observable variables without the need for such a relationship to be fully mediated by group factors, and b) facilitates independent evaluation of the merits of general and group factors. Underlying this model is the hypothesis that there is a general factor with causal influence on all items (Arias, Jenaro, & Ponce, 2018). Sometimes, the variance shared between different indicators may be artificial, the valence of the items (Bäckström & Björklund, 2016), or general features of those examined -not directly related to the construct- such as self-esteem (Davies, Connelly, Ones, & Birkland, 2015). These may increase the fit of the bifactor model, leading to the erroneous conclusion that there is a general, psychological, and psychometrically meaningful factor.

Bifactor models constitute a useful analytic method for the assessment of construct validity in psychological measuring. However, researchers should be wary when performing bifactor analysis and avoid using SEM model fit indices as the main or only criteria to judge the feasibility of a

BM, since conclusions based on such criteria can be misleading considering that complementary statistical fit indices provide researchers with relevant information concerning the salience of general and group-level factors. Such information, together with theoretical background, will allow for the selection of the most appropriate factor model and ensure the validity of the inferences drawn from the scale scores.

References

- Arias, V. B., Jenaro, C., & Ponce, F. P. (2018). Testing the generality of the general factor of personality: An exploratory bifactor approach. *Personality and Individual Differences, 129*, 17-23. doi: [10.1016/j.paid.2018.02.042](https://doi.org/10.1016/j.paid.2018.02.042)
- Bäckström, M., & Björklund, F. (2016). Is the general factor of personality based on evaluative responding? Experimental manipulation of item-popularity in personality inventories. *Personality and Individual Differences, 96*, 31-35. doi: [10.1016/j.paid.2016.02.058](https://doi.org/10.1016/j.paid.2016.02.058)
- Bonifay, W., Lane, S. P., & Reise, S. P. (2017). Three concerns with applying a bifactor model as a structure of psychopathology. *Clinical Psychological Science, 5*(1), 184-186. doi: [10.1177/2167702616657069](https://doi.org/10.1177/2167702616657069)
- Davies, S. E., Connelly, B. S., Ones, D. S., & Birkland, A. S. (2015). The general factor of personality: The “Big One,” a self-evaluative trait, or a methodological gnat that won’t go away? *Personality and Individual Differences, 81*, 13-22. doi: [10.1016/j.paid.2015.01.006](https://doi.org/10.1016/j.paid.2015.01.006)
- Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55*, 57-68. doi: [10.1016/j.intell.2016.01.006](https://doi.org/10.1016/j.intell.2016.01.006)
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit & D. Sorbom (Eds.), *Structural equation modeling: Present and future—A Fest-*

schrift in honor of Karl Joreskog (pp. 195-216). Lincolnwood, IL: Scientific Software International.

- 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. doi: [10.3390/jintelligence3010002](https://doi.org/10.3390/jintelligence3010002)
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, 98(3), 223-237. doi: [10.1080/00223891.2015.1089249](https://doi.org/10.1080/00223891.2015.1089249)
- Seib-Pfeifer, L.-E., Pugnaghi, G., Beauducel, A., & Leue, A. (2017). On the replication of factor structures of the Positive and Negative Affect Schedule (PANAS). *Personality and Individual Differences*, 107, 201-207. doi: [10.1016/j.paid.2016.11.053](https://doi.org/10.1016/j.paid.2016.11.053)
- Stucky, B. D., Thissen, D., & Edelen, M. O. (2013). Using logistic approximations of marginal trace lines to develop short assessments. *Applied Psychological Measurement*, 37(1), 41-57. doi: [10.1177/0146621612462759](https://doi.org/10.1177/0146621612462759)
-