PERGAMON

Personality and Individual Differences 28 (2000) 181–185

PERSONALITY AND INDIVIDUAL DIFFERENCES

www.elsevier.com/locate/paid

Reassessing the fit of the confirmatory factor analysis of the multidimensional students life satisfaction scale: comments on 'confirmatory factor analysis of the multidimensional Students' Life Satisfaction Scale'

Mark Shevlin[a,*], Jeremy N.V. Miles[b], Christopher Alan Lewis[a]

[a]*School of Behavioural and Communication Sciences, University of Ulster at Magee College, Derry, Northern Ireland, UK*

[b]*Institute of Behavioural Sciences, Derby University, Mickleover, Derby, DE3 5GX, UK*

Received 21 September 1998; accepted 17 February 1999

Abstract

Greenspoon and Saklofske [(1998). Confirmatory factor analysis of the multidimensional Students' Life Satisfaction Scale. *Personality and Individual Differences*, 25, 965–971] present the results of a confirmatory factor analysis of the multidimensional Students' Life Satisfaction Scale. In this paper, we argue that the results of this analysis do not demonstrate, as the authors claim, that the proposed model is able to satisfactorily account for the data. Recommendations are made regarding more appropriate methods of assessing model fit in confirmatory factor analysis. © 1999 Elsevier Science Ltd. All rights reserved.

The statistical benefits of confirmatory factor analysis (CFA) in comparison to exploratory factor analytic techniques are well documented in the psychometric literature (Bollen, 1989; Hayduk, 1987; Pedhazur & Schmelkin, 1991). In particular, confirmatory factor analysis allows a statistical test of how well an a priori specified factor model explains the observed pattern of sample correlations or covariances, commonly referred to as 'model fit'. Model fit can be assessed using indices of overall fit (e.g. χ^2, GFI, RMSR), incremental fit (e.g. NFI, CFI), the root mean square error of approximation (RMSEA) and Hoelter's critical N (CN). It has been suggested that model fit should be evaluated using information from these different families of

* Corresponding author.

PII: S0191-8869(99)00069-0

indices (Bollen & Long, 1993; Tanaka, 1993; Hoyle & Panter, 1995). In addition, alternative models can be compared using the AIC (Akaike, 1987) or nested models using the χ^2 difference test. Comprehensive review literature on issues relating to model fit is available in Bollen (1989) or Marsh, Balla and Hau (1996), and Hoyle and Panter (1995) present guidelines for reporting information regarding model fit.

Given the abundance of fit indices available to the researcher, and the extensive literature on evaluating model fit, the authors of this paper are not convinced by conclusions reached by Greenspoon and Saklofske (1998) that their "...model offers an adequate fit to the data" for the following reasons. We find ourselves unconvinced because the indices that Greenspoon and Saklofske choose to report do not provide sufficient evidence that the model does fit, and they fail to report other fit indices that may provide stronger evidence of model fit.

Greenspoon and Saklofske rely on the evidence of four fit indices to argue that their model fits; the GFI (Goodness of Fit Index) and AGFI (Adjusted Goodness of Fit Index) both developed by Jöreskog and Sörbom (1984), the RMR (Root Mean Residual) and χ^2, which both have a long history of use in factor analysis. We argue that the values of these indices reported by Greenspoon and Saklofske do not demonstrate an acceptable fit of the model to the data, and without additional information it is difficult to make any judgment regarding fit.

Greenspoon and Saklofske state that the conventional cut-off for the GFI and AGFI indices is 0.8. We would state that the term 'conventional cut-off', is poorly defined, and arguing for any global cut-off level leads to an exchange of opinions without any real advancement in knowledge (Barrett, personal communication). Instead researchers should look to empirical evidence derived from Monte Carlo simulations and evidence derived from an examination of the mathematical foundations of the indicators that they are employing. The results from a Monte Carlo simulation of the performance of the GFI by Shevlin and Miles (1998) suggests that using cut-offs greater than 0.9 may still lead to models being accepted which do not adequately account for the data. The analysis carried out by Shevlin and Miles showed that they found that when population factor loadings were 0.7 (approximately the same values as were found by Greenspoon and Saklofske), and sample sizes were 200 or 400, the median GFI value for a severely mis-specified model was 0.83. This is a larger value than that obtained by Greenspoon and Saklofske, although it should be noted that the model used in this study was considerably smaller (in terms of number of parameters estimated, number of variables and degrees of freedom) than that analysed by Greenspoon and Saklofske. Similarly, Hu and Bentler (1995) report the results of a study in which they found that the use of a cut-off of 0.9 for the GFI and AGFI leads to both type I and type II errors in model selection. A further problem in developing a suitable cut-off value for the GFI and AGFI is the fact that the mean of the sampling distribution for both of these indices is affected by the sample size (Anderson & Gerbing, 1984), with larger sample sizes tending towards larger mean values, although it should be noted that the value of the GFI and AGFI is not affected directly by sample size in any one model (Bollen, 1990).

The second alternative to the use of opinion in the selection of cut-off values is the examination of the mathematical foundations of a fit index. MacCallum and Hong (1997) examined the GFI and the AGFI from the perspective of power analysis, and found that because of the influence of degrees of freedom and sample size it is difficult to determine appropriate values for GFI and AGFI that will indicate model fit, or a lack of model fit for any particular model.

Greenspoon and Saklofske also cite the root mean square residual (RMSR) value of 0.8 as evidence of the fit of their model. The RMSR also suffers from interpretational difficulties, as it is difficult to specify a priori acceptable values of RMSR, which would indicate adequate fit in all circumstances. Additionally, RMSR suffers from a problem that it is affected by sample size. In a correctly specified model a larger sample size leads to a smaller value for RMSR (Anderson & Gerbing, 1984). Instead, Bollen (1989) states that "it would be extremely helpful to have a simultaneous significance test of the hypothesis that all population residuals are zero. [The χ^2 test] is such a significance test" (p. 262).

The χ^2 test reported by Greenspoon and Saklofske is highly statistically significant, implying that the model does not satisfactorily account for the data. This is dismissed by the authors on the grounds of excessive power due to the large sample size. Again, we risk venturing into the area of 'learned opinion' to determine whether or not a particular sample is sufficiently large to warrant an appeal to the effect of excessive power. We only note that when the effect of sample size may have improved the apparent fit of the model (such as with the GFI, AGFI and RMSR), the effect of a large sample is not mentioned. Instead of attempting to invoke the sample size effect, we feel that the authors should make use of a number of indices that have been used to try to move away from a dichotomous decision of reject/accept the model, and towards attempts to provide descriptions of model fit (Tanaka, 1993).

More acceptable ways of assessing model fit include using the Root Mean Square Error of Approximation (RMSEA; Steiger & Lind, 1980; Steiger, 1990) and a class of indices known collectively as the incremental fit indices (the first of which, the Non-Normed Fit Index [NNFI] and the Normed Fit Index [NFI] developed by Bentler & Bonett, 1980). The RMSEA provides "a measure of the *discrepancy per degree of freedom* for the model" (Browne & Cudeck, 1993, p. 144, italics in original). The calculation of the RMSEA uses the χ^2 value of the model, in conjunction with the sample size and a correction for the complexity of the model (degrees of freedom) to ensure that these factors do not affect the decision to reject or accept the model. An additional advantage of the RMSEA is that it has a known sampling distribution, and therefore confidence limits can be calculated. The significance value of the χ^2 tests the null hypothesis of exact fit, which will always be false, and therefore if the sample size is sufficiently large the power of the test will ensure that the model is rejected. To overcome this problem Browne and Cudeck propose a test of close fit, which tests the null hypothesis that RMSEA is greater than 0.05.

Incremental fit indices (see Marsh et al., 1996 for a review) are generally based on a comparison between the χ^2 value for the fitted model, and the χ^2 value for the null model. By comparing the hypothesised model with the null model, the incremental fit indices provide a measure of the degree of improvement that has occurred by fitting the specified model. The incremental fit indices are usually scaled between 0 (poor fit) and 1 (perfect fit). A major advantage of the incremental fit indices is that sample size effects, and other factors that adversely affect χ^2 will occur for both the null model and the model under test, and will not therefore directly affect χ^2.

In summary, it is difficult to justify the conclusion reached by Greenspoon and Saklofske based on the limited information provided in their paper. The information that is presented suggests that the proposed model is not an acceptable description of the sample data. The subsequent interpretation of the model by Greenspoon and Saklofske is analogous to reporting non-significant factorial ANOVA results from an experiment and proceeding to discuss main effects or interactions. Although the use of confirmatory factor analysis and structural equation modelling is relatively recent in the personality literature the critical evaluation of such techniques should be as rigorous as that for traditional statistical methods. As Cliff (1983) noted, structural equation modelling "...lends an air of unchallengeable sanctity to conclusions that would otherwise be subjected to the most intense scrutiny" (p. 116). This should not be the case.

Acknowledgements

This paper has benefited greatly from the comments of Paul Barrett and we would like to thank him for this assistance.

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317–332.

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness of fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika*, *49*, 155–173.

Bentler, P. M., & Bonett (1980). Significance tests and the analysis of goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K. A. (1990). Overall fit in covariance structure models: two types of sample size effects. *Psychological Bulletin*, *107*, 256–259.

Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Newbury Park, CA: Sage.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long, *Testing structural equation models*. Newbury Park, CA: Sage.

Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioural Research*, *18*, 115–126.

Greenspoon, P. J., & Saklofske, D. H. (1998). Confirmatory factor analysis of the multidimensional Student's Life Satisfaction Scale. *Personality and Individual Differences*, *25*, 965–971.

Hayduk, L. A. (1987). *Structural equation modeling using LISREL: essentials and advances*. Baltimore: John Hopkins University Press.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle, *Structural equation modeling: concepts, issues, and applications*. Newbury Park, CA: Sage.

Hu, L., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle, *Structural equation modeling: concepts, issues, and applications*. Newbury Park, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI users guide* (3rd ed.). Moorsville, IN: Scientific Software.

MacCallum, R. C., & Hong, S. (1997). Power analysis in covariance structure modeling using GFI and AGFI. *Multivariate Behavioural Research*, *32*(2), 193–210.

Marsh, H. W., Balla, J. R., & Hau, K. (1996). An evaluation of incremental fit indices: a clarification of mathematical and empirical properties. In G. A. Marcoulides, & R. E. Schumacker, *Advanced structural equation modeling: issues and techniques*. Mahwah, NJ: Erlbaum.

Pedhazur, E., & Schmelkin, L. (1991). *Measurement, design and analysis: an integrated approach*. New York: Holt, Rinehart & Winston.

Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, *25*(1), 85–90.

Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioural Research*, *25*, 173–180.

Steiger J. H., & Lind J. M. (1980). *Statistically based tests for the number of factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen, & J. S. Long, *Testing structural equation models*. Newbury Park, CA: Sage.