Editorial

Some Guidelines Concerning the Modeling of Traits and Abilities in Test Construction

Over the last few years the majority of authors who submitted manuscripts for publication in European Journal of Psychological Assessment selected confirmatory factor analysis as their method for test construction. This is a very agreeable development. Confirmatory factor analysis is based on a well-developed model of measurement that is closely linked to the corresponding model of the covariance matrix. As a consequence, parameter estimation occurs in close agreement with the model of measurement as well as with the model of the covariance matrix. One very useful property of this method is that the model must provide a complete account for the variances and covariances of the items. This way, structural deviations from the basic assumptions of a model become apparent if the items of a prospective measure show other properties than the expected ones.

Unfortunately, many submissions reporting the results of confirmatory factor analysis are deficient in one way or the other, so that some guidelines in modeling traits and abilities in test construction may prove helpful for future submissions.

Some manuscripts report models that show an insufficient degree of fit or include features that cannot really be accepted as appropriate according to the present state of the art. Such manuscripts cause considerable uneasiness (Barrett, 2007). Other manuscripts list well-known limits for fit statistics at length. Since journal space is limited and there are always quests for more journal space, it does not seem reasonable to me that every second manuscript should provide the more or less same list of fit statistics and corresponding limits. In order to improve the efficient use of space, I would like to provide some general guidelines so that only deviations from these guidelines would make the presentation of additional information necessary.

First, the demonstration of a good model fit is an essential part of any manuscript reporting confirmatory factor analysis. Although a manuscript may also report the results achieved for other models showing an insufficient degree of fit, the model that is expected to justify the presentation

Karl Schweizer, Editor-in-Chief Goethe University Frankfurt, Germany

of a new measure to the scientific public must show a good fit. This appears to us to be a very reasonable provision since models showing mediocre or even bad model fit usually do not survive the next attempt at replication. Furthermore, the appropriateness of parameter estimates achieved for ill-fitting models is questionable. Extending an ill-fitting confirmatory factor model to a full structural equation model frequently leads to fancy path coefficients that can provide a distorted view of the empirical reality. Of course, there is also the danger of a strict and inflexible reliance on "cut-off" values for model fit (Goffin, 2007). Therefore, when securing good model fit some flexibility in editorial politics would also seem to be necessary.

Second, some basic standards exist concerning the model fit with which a manuscript must comply in order to warrant publication in a scientific journal. Although the complexity of models may justify a large number of fit statistics, there are core properties of models that are usually reflected by a few fit statistics quite well. The minimal set of fit statistics proposed by Kline (2005) seems to be best suited for this purpose. This minimal set includes (1) the model χ^2 , (2) the root mean square error of approximation (RMSEA) (Steiger & Lind, 1980), (3) the Bentler Comparative Fit Index (CFI) (Bentler, 1990), and (4) the standardized root mean square residual (SRMR). Other fit statistics may also be presented and discussed. Because of the special importance of the minimal set, it seems essential to provide some guidelines for model fit with respect to these statistics.

The model χ^2 is a fit statistic that has the advantageous property of being closely linked to a probability distribution. However, despite the exactness of this statistic, it was found to be "not valid in most cases" (Jöreskog & Sörbom, 1982, p. 408) and is thus not recommended because of a number of dependencies (Bentler, 2007). We prefer a derived fit statistic, the normed χ^2 , which is achieved by computing the ratio of the model χ^2 and the degrees of freedom (Wheaton, Muthén, Alwin, & Summers, 1977). Although the normed χ^2 also shows some dependency on sample size, it can provide valuable guidelines for the majority of manuscripts submitted to this journal. A normed χ^2 below 2 usually suggests good model fit and below 3 acceptable model fit (Bollen, 1989). Very large samples may necessitate the selection of larger limits. Although the second fit statistic of the minimal set (RMSEA) is not related to an established probability distribution, it has the advantage of being usually associated with a confidence interval. RMSEA values less than 0.05 were found to indicate a good model fit and less than 0.08 an acceptable model fit (Brown & Cudeck, 1993). The comparative fit index (CFI) indicates a good model fit for values in the range between 0.95 and 1.00, whereas values in the range of 0.90 to 0.95 signify acceptable fit (Bentler, 1990; Hu & Bentler, 1999). Finally, values of the standardized root mean square residual (SRMR) are expected to stay below 0.10 (Kline, 2005).

Third, there is the issue of correlated errors, which signify that covariation is present which cannot be explained by the model. Although there are rare examples of models integrating correlated errors in a systematic way, normally a good model does not include correlated errors. Therefore, convincing arguments should be provided for deviations from the principle of avoiding correlated errors. Recent developments have shown that there are characteristic sources of model misfit, and that consideration thereof can improve the model fit considerably. For example, there are the impairments of model fit due to item wording (DiStefano & Motl, 2006; Rauch, Schweizer, & Moosbrugger, 2007; Vautier, Raufaste, & Cariou, 2003) and position effect (Hartig, Hölzel, & Moosbrugger, 2007; Schweizer, Schreiner, & Gold, 2009). Future research may reveal further sources, the consideration of which may make correlated errors dispensable.

References

Barrett, P. (2007). Structural equation modeling: Adjusting model fit. Personality and Individual Differences, 42, 815–824.

Bentler, P.M. (1990). Comparative fit indices in structural models. *Psychological Bulletin*, 107, 238–246.

- Bentler, P.M. (2007). On test and indices for evaluating structural models. *Personality and Individual Differences*, 42, 825–829.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Brown, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.
- DiStefano, C., & Motl, R.W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13, 440–464.
- Goffin, R.D. (2007). Assessing the adequacy of structural equation modeling: Golden rules and editorial policies. *Personality* and Individual Differences, 42, 831–839.
- Hartig, J., Hölzel, B., & Moosbrugger, H. (2007). A confirmatory analysis of item reliability trends (CAIRT): Differentiating true score and error variance in the analysis of item context effects. *Multivariate Behavioral Research*, 42, 157–183.
- Hu, L.-T., & Bentler, P. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Jöreskog, K.G., & Sörbom, D. (1982). Recent developments in structural equation modeling. *Journal of Marketing Research*, 19, 404–416.
- Kline, R.B. (2005). Principles and practices of structural equation modeling. New York: Guilford.
- Rauch, W., Schweizer, K., & Moosbrugger, H. (2007). Method effects due to social desirability as a parsimonious explanation of the deviation from unidimensionality in LOT-R scores. *Personality and Individual Differences*, 42, 1597–1607.
- Schweizer, K., Schreiner, M., & Gold, A. (2009). The confirmatory investigation of APM items with loadings as a function of the position and easiness of items: A two-dimensional model of APM. *Psychology Science Quarterly*, 51, 47–64.
- Steiger, J.H., & Lind, J.M. (1980, June). Statistically based tests for the number of common factors. Paper presented at the meeting of the Psychometric Society, Iowa City, IA.
- Vautier, S., & Raufaste, E., & Cariou, M. (2003). Dimensionality of the Revised Life Orientation Test and the status of filler items. *International Journal of Psychology*, 38, 390–400.
- Wheaton, B., Muthén, B., Alwin, D., & Summers, G. (1977). Assessing reliability and stability in panel models. In D.R. Heise (Ed.), *Sociological methodology* (pp. 84–136). San Francisco: Jossey-Bass.

European Journal of Psychological Assessment 2010; 26(1):1-2