

The Role of Nonlinear Factor-to-Indicator Relationships in Tests of Measurement Equivalence

Daniel J. Bauer
University of North Carolina

Measurement invariance is a necessary condition for the evaluation of factor mean differences over groups or time. This article considers the potential problems that can arise for tests of measurement invariance when the true factor-to-indicator relationship is nonlinear (quadratic) and invariant but the linear factor model is nevertheless applied. The factor loadings and indicator intercepts of the linear model will diverge across groups as the factor mean difference increases. Power analyses show that even apparently small quadratic effects can result in rejection of measurement invariance at moderate sample sizes when the factor mean difference is medium to large. Recommendations include the identification of nonlinear relationships using diagnostic plots and consideration of newly developed methods for fitting nonlinear factor models.

Keywords: measurement, invariance, nonlinear, factor analysis, structural equation modeling

Measurement invariance refers to the equivalent measurement of a construct or set of constructs in two or more groups or over time via a common instrument (i.e., a set of observed variables or indicators). When measurement invariance holds, differences observed between groups or over time reflect true differences on the constructs of interest. If measurement invariance does not hold, however, any observed differences might simply reflect inequities of measurement and not true differences on the relevant constructs (Horn & McArdle, 1992; Meredith, 1993; Ployhart & Oswald, 2004). Establishing measurement invariance has thus become an integral area of psychometric research. Within the factor analytic tradition, early attention was focused on this issue by Meredith (1964). More recently, this literature has expanded greatly by reformulating questions of measurement invariance within a confirmatory factor analytic or structural equation modeling approach (Jöreskog, 1971; Meredith, 1993; Sörbom, 1974).

Studies of measurement invariance usually involve the comparison of independent samples, typically two. A multiple-groups confirmatory factor analysis is then used to test the tenability of equality constraints on various components of the measurement model (e.g., factor loadings, indicator

intercepts, residual variances), often in a sequential fashion (Bollen, 1989, pp. 355–369; Widaman & Reise, 1997). Less common are tests of longitudinal invariance, wherein a single sample is administered the same measurement instrument on multiple occasions (Meredith & Horn, 2001; Vandenberg & Lance, 2000). Although procedures for testing measurement invariance are by now well established in practice, several methodological issues continue to receive attention. These include the relation of within- and between-group measurement properties (Borsboom, Mellenbergh, & Van Heerden, 2002; Lubke, Dolan, Kelderman, & Mellenbergh, 2003), the use and power of equivalence tests (Cheung & Rensvold, 2002; Meade & Lautenschlager, 2004), the permissibility of differences in indicator intercepts (Millsap, 1998), and the difficulties that may arise when only a subset of indicators function equivalently over groups or time (i.e., partial invariance; Byrne, Shavelson, & Muthén, 1989; Cheung & Rensvold, 1999; Millsap & Kwok, 2004; Steenkamp & Baumgartner, 1998).

The purpose of this article is to consider an issue that has heretofore received comparatively little attention in the measurement invariance literature. Specifically, the factor analysis model used to evaluate measurement invariance assumes that the relationships of the observed measures to the factors are linear. In the single-sample context, pronounced nonlinearity in these relationships may occur if, for instance, the manifest variables are dichotomous or ordinal and incorrectly treated as continuous. In such instances, difficulty factors may emerge in the linear factor model (McDonald, 1967; Wherry & Gaylord, 1944) and more preferable alternatives are to analyze polychoric correla-

Additional materials are on the Web at <http://dx.doi.org/10.1037/1082-989X.10.3.305.supp>

Correspondence concerning this article should be addressed to Daniel J. Bauer, Department of Psychology, Campus Box 3270, University of North Carolina at Chapel Hill, North Carolina 27599-3270. E-mail: dbauer@email.unc.edu

tions or to use nonlinear item response models (Jöreskog & Moustaki, 2001; Mislevy, 1986; Muthén, 1984; Muthén & Christofferson, 1981; Waller, Tellegen, McDonald, & Lykken, 1996). However, even with continuous observed variables, the relationships between the factors and indicators may not be strictly linear. A key implication of the linear model is that the strength of the relationship between a factor and measured variable is constant at least within the observed range of the data. This would not be the case if, for instance, there was a floor or ceiling effect for the measure or, more generally, if increasing levels of the latent factor produce either diminishing or increasing differences in the measured variable reflecting the factor. For example, in making ratings of aggression, observers might have greater difficulty judging differences in low levels of aggression than in high levels of aggression. Similarly, Mooijart and Bentler (1986) comment that, for attitude scale data, participants often provide more extreme opinion ratings (at the ends of the scale) than would be expected on the basis of a linear model.

Methods for estimating nonlinear factor analysis models have been available for many years and have recently seen rapid developments (see Wall & Amemiya, in press, for a review). These methods are rarely implemented in practice, however, despite the availability of specialized software (Etezadi-Amoli & McDonald, 1983; Wall & Amemiya, in press). Such neglect may reflect the view that minor departures from linearity are relatively innocuous, as a linear model will still provide a useful first-order approximation to the true function. However, when the goal of the investigation is to evaluate measurement invariance, a linear approximation can present certain difficulties. Namely, in this article, I will demonstrate that if the factor-to-indicator relationship is invariant but evinces some curvature and if the factor means differ over groups, then the loadings and intercepts obtained by fitting a linear factor model will differ, causing invariance tests to be rejected. Power curves show that at modest sample sizes, even seemingly minor curvature in factor-to-indicator relationships can lead to rejection of factorial invariance in a linear factor model if the factor mean difference is of modest to large size. Although this exposition is focused on between-groups tests of measurement invariance, similar problems would arise in longitudinal tests of measurement invariance.

Consequences of Unmodeled Nonlinearity for Factorial Invariance

Let us begin by considering a relatively simple five-indicator latent factor model for a single sample. Denoting the indicators as x_1 through x_5 and the latent factor as ξ , the model for person i may be written as

$$\begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \\ x_{4i} \\ x_{5i} \end{pmatrix} = \begin{pmatrix} \tau_1 \\ \tau_2 \\ \tau_3 \\ \tau_4 \\ \tau_5 \end{pmatrix} + \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \lambda_5 \end{pmatrix} \xi_i + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \\ \varepsilon_{5i} \end{pmatrix}. \quad (1)$$

This model is, in effect, a simultaneous linear regression of each indicator on the latent factor and hence the symbols τ , λ , and ε simply designate regression intercepts, slopes (loadings), and residuals. A key point for this article is that the regression of each indicator on the latent factor is assumed to be linear in form.

A more general expression for the factor model is

$$\mathbf{x}_i = \boldsymbol{\tau} + \boldsymbol{\Lambda} \boldsymbol{\xi}_i + \boldsymbol{\varepsilon}_i. \quad (2)$$

Here, \mathbf{x}_i is a $p \times 1$ vector of scores on p manifest variables (or factor indicators) for individual i , $\boldsymbol{\tau}$ is a $p \times 1$ vector of indicator intercepts and $\boldsymbol{\Lambda}$ is a $p \times q$ matrix of factor loadings describing the regression of the manifest variables on the q latent common factors in the $q \times 1$ vector $\boldsymbol{\xi}_i$. The $p \times 1$ vector $\boldsymbol{\varepsilon}_i$ holds the indicator residuals net the influence of the common factors. These residuals include both variability due to random measurement error and true-score variability that is unique to each indicator. In addition, the distribution of the latent factors $\boldsymbol{\xi}_i$ is captured via the $q \times 1$ vector of latent means $\boldsymbol{\kappa}$ and $q \times q$ covariance matrix $\boldsymbol{\Phi}$. The residuals $\boldsymbol{\varepsilon}_i$ are assumed to have zero expectation and a $p \times p$ covariance matrix $\boldsymbol{\Theta}$. The latent factors and residuals are assumed to be uncorrelated.

Under these assumptions, the model-implied mean vector and covariance matrix of \mathbf{x} are given as

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{\tau} + \boldsymbol{\Lambda} \boldsymbol{\kappa} \\ \boldsymbol{\Sigma} &= \boldsymbol{\Lambda} \boldsymbol{\Phi} \boldsymbol{\Lambda}' + \boldsymbol{\Theta}. \end{aligned} \quad (3)$$

It is important to note that not all parameters in $\boldsymbol{\tau}$, $\boldsymbol{\Lambda}$, $\boldsymbol{\kappa}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ can be estimated. One reason is that the latent factors, being unobserved, have no natural metric, and hence their means and variances are arbitrary. There are two common choices for establishing the scale of the latent factors. The first choice is to set the mean and variance of the factor to zero and one, respectively, thus standardizing the factor. The second possibility is to choose a particular indicator to anchor the scale of the latent factor by setting the intercept and loading of this indicator to zero and one, respectively (e.g., setting $\tau_1 = 0$ and $\lambda_1 = 1$ for the model in Equation 1). The latent factor then draws its (unstandardized) metric from the scaling indicator.

In addition to setting the scale of the latent variables, each parameter in the model must be uniquely identified by the means, variances, and covariances of the indicators. That is, there must be sufficient information in the data to locate a single optimal value for each estimate. Identification of these parameters can be ascertained through covariance

algebra or through rules that apply to specific model structures. For instance, for a single-factor model with uncorrelated residuals, the three-indicator rule establishes that the model is *just-identified*, meaning that there is just enough information to obtain a unique estimate for each parameter (Bollen, 1989, p. 244). With more indicators for the factor, the model is *overidentified*, meaning that there are enough restrictions on the means and covariances of the indicators to allow a test of the plausibility of these restrictions against the sample means and covariances with a chi-square test of overall model fit. Similarly, a two-indicator rule applies to multifactor models where the factor loadings form independent clusters (Bollen, 1989, p. 244; McDonald & Ho, 2002). In this case, two indicators per factor yield an overidentified model if the factors are correlated, the residuals of the indicators are uncorrelated, and there are no cross-loadings (all indicators load on only one factor).

Further details on the structure of the linear factor analysis model and issues of identification, estimation, and model testing are available from several excellent texts, including Bollen (1989), Kaplan (2000), or Kline (2005). This model is now elaborated to allow for multiple groups and tests of measurement invariance across groups.

Multiple Groups Factor Analysis

Suppose that data have been collected on the same manifest variables for two independent groups and that a linear factor model is specified for each group. Using a parenthetical superscript to differentiate between the two groups, the factor models for Groups 1 and 2 would be specified as

$$\begin{aligned}\boldsymbol{\mu}^{(1)} &= \boldsymbol{\tau}^{(1)} + \boldsymbol{\Lambda}^{(1)}\boldsymbol{\kappa}^{(1)} \\ \boldsymbol{\Sigma}^{(1)} &= \boldsymbol{\Lambda}^{(1)}\boldsymbol{\Phi}^{(1)}\boldsymbol{\Lambda}^{(1)'} + \boldsymbol{\Theta}^{(1)}\end{aligned}\quad (4)$$

and

$$\begin{aligned}\boldsymbol{\mu}^{(2)} &= \boldsymbol{\tau}^{(2)} + \boldsymbol{\Lambda}^{(2)}\boldsymbol{\kappa}^{(2)} \\ \boldsymbol{\Sigma}^{(2)} &= \boldsymbol{\Lambda}^{(2)}\boldsymbol{\Phi}^{(2)}\boldsymbol{\Lambda}^{(2)'} + \boldsymbol{\Theta}^{(2)}.\end{aligned}\quad (5)$$

These equations generalize straightforwardly to cases with more than two groups.

Typically, when one fits a multiple-groups factor model, the goal is to determine if the factor structure (or measurement model) is the same or different over groups. A hierarchy of levels of measurement invariance can be differentiated by reference to Equations 4 and 5 (Meredith, 1993; Widaman & Reise, 1997). Typically, the lowest level of equivalence to be considered is whether the general factor structure is invariant over groups, a condition known as *configural invariance* (Horn & McArdle, 1992). For configural invariance to hold, the model form must be identical over groups (in terms of zero and nonzero parameters), but the values of the nonzero parameters in the model can all

potentially differ. Configural invariance generally suggests that the factors represent the same theoretical constructs across groups, but these constructs cannot necessarily be compared directly across groups because of possible inequalities of measurement. To make such comparisons, not only must the form of the model be identical, the values of many of the estimates must also be equal.

The next level of invariance requires the factor-loading matrices to be equivalent over groups (i.e., $\boldsymbol{\Lambda}^{(1)} = \boldsymbol{\Lambda}^{(2)}$), a condition known as *weak factorial invariance*. If weak factorial invariance holds, then this permits unambiguous comparisons of the factor covariance matrices, $\boldsymbol{\Phi}^{(1)}$ and $\boldsymbol{\Phi}^{(2)}$. For instance, one might be interested in knowing whether two factors are more correlated in one group than another. If weak factorial invariance does not hold, the result of this comparison will depend on how the factors are scaled (e.g., which indicators are chosen to scale the factors by setting their loadings to one). Under weak factorial invariance, the comparison of $\boldsymbol{\Phi}^{(1)}$ and $\boldsymbol{\Phi}^{(2)}$ will instead yield substantively identical results no matter how the factors are scaled (Widaman & Reise, 1997). Similarly, differences in the factor means of the groups, $\boldsymbol{\kappa}^{(1)}$ and $\boldsymbol{\kappa}^{(2)}$, are often also of interest. For comparisons of the factor means to be valid, *strong factorial invariance* is required such that the intercepts of the models are also equal (i.e., $\boldsymbol{\tau}^{(1)} = \boldsymbol{\tau}^{(2)}$). Again, if this condition holds then the comparison of factor means is independent of the particular choice of scaling for the latent factors, and if it does not, then the comparison will vary depending on how the metric of the latent variables is set (Widaman & Reise, 1997). Finally, *strict factorial invariance* exists if, in addition to the above conditions, the residual variances of the indicators are equal across groups (i.e., $\boldsymbol{\Theta}^{(1)} = \boldsymbol{\Theta}^{(2)}$). This latter condition implies that all of the differences in the means and covariances of the indicators across the two groups arise from differences in the latent variables. The progressive imposition of these constraints produces a sequence of nested model comparisons that permit the evaluation of each level of invariance via a likelihood ratio chi-square test.

If some but not all of the factor loadings and/or intercepts significantly differ across groups, then this is referred to as *partial measurement invariance*. Some authors assert that comparisons of factor means should *not* be made under partial invariance (Bollen, 1989; Meredith, 1993; Ployhart & Oswald, 2004), whereas others have argued that small differences in a few factor loadings should not prevent the assessment of factor mean differences over groups (Byrne et al., 1989; Steenkamp & Baumgartner, 1998). Taking a more pragmatic approach to the issue, Millsap and Kwok (2004) suggested that the acceptability of partial invariance should depend on the measure's purpose. They considered the special case in which one selects individuals on the basis of unit-weighted composite scores, noting the potentially detrimental consequences of partial invariance for selection accuracy.

Nonlinear Factor-to-Indicator Relationships

One assumption that is key to the factor model reviewed above is that the regression of each manifest variable on the latent variables is linear. However, let us now consider the consequences of fitting a linear factor model when the true relationships between the manifest and latent variables are not linear. As an analytically convenient example, assume that the relationship of the latent factor to one indicator is well approximated by a quadratic function and that the factor is normally distributed. Note that a quadratic model will not approximate all nonlinear functions well (e.g., a logistic) and is used here only as a case-in-point. For this single manifest variable x and latent factor ξ , the measurement model might then be

$$x_i = \tau + \lambda \xi_i + \omega \xi_i^2 + \varepsilon_i, \tag{6}$$

where ω is used to designate the quadratic effect. To simplify the presentation, no equations are given for the other indicators in the model, some of which might also be nonlinearly related to the factor, so that the matrix notation used in the previous sections can be avoided.

Suppose that the following misspecified linear factor model is fit to data generated from the above equation:

$$x_i = \tau^* + \lambda^* \xi_i + \varepsilon_i^* \tag{7}$$

The superscript * indicates a difference in value from Equation 6. The starred components to this equation include not just the intercept and slope, but also the residuals, which will differ because of the omission of the quadratic effect. If ξ has mean κ and variance ϕ , then the coefficients τ^* and λ^* can be expressed as

$$\begin{aligned} \lambda^* &= \lambda + 2\omega\kappa \\ \tau^* &= \tau + \omega(\phi - \kappa^2). \end{aligned} \tag{8}$$

A more detailed derivation of Equation 8 is given in the Appendix, which also provides a derivation of the variance of ε^* relative to ε . There are two important results of these derivations: (a) The factor loading is a function of the factor mean, and (b) the intercept is a function of the factor mean and variance. These results have serious implications for testing measurement invariance.

If x is measured in multiple samples and the factor mean κ differs across groups, then Equation 8 implies that the estimated loadings from a linear factor analysis model will also differ. Specifically, if Equation 6 holds in both groups and ξ is normally distributed within groups, then the difference in the linear factor loadings and intercepts will be

$$\begin{aligned} \lambda^{*(2)} - \lambda^{*(1)} &= 2\omega(\kappa^{(2)} - \kappa^{(1)}) \\ \tau^{*(2)} - \tau^{*(1)} &= \omega[(\kappa^{(1)})^2 - (\kappa^{(2)})^2 + \phi^{(2)} - \phi^{(1)}]. \end{aligned} \tag{9}$$

Clearly, as the factor mean difference between the groups increases, the slopes and intercepts of the linear factor model will also diverge. An example is shown in Figure 1, where even apparently modest curvature in the factor-to-indicator relationship produces a visible difference in the linear regression slopes and intercepts for the two groups. In the absence of a factor mean difference, the linear factor loadings will be invariant, but the intercepts will continue to differ if the factor variances are different.¹

Equation 9 suggests that tests of weak or strong measurement invariance in a linear factor model may fail even if the generating function is invariant over groups if this function is not strictly linear and the groups differ in their factor means. In some respects, the rejection of invariance in the linear factor model under these conditions would be accurate—if one applies a linear approximation to the regression function in each group, the intercepts and loadings will indeed differ. However, the conclusion drawn from this result may be that the manifest variable is not equally reflective of the latent factor in both groups, when in fact the true factor-to-indicator regression function is identical (but misspecified) for both groups. An open question is how much curvature is required for invariance tests to be rejected in the linear factor model. In the following section I report a small study aimed at answering this question for the quadratic case.

When Invariance Tests Fail

On the basis of the preceding analytical derivations, the power of the likelihood ratio test to reject measurement invariance in the linear factor model was evaluated under variations in three conditions: the extent of curvature in the factor-to-indicator relationship (ω), the factor mean difference ($\kappa^{(2)} - \kappa^{(1)}$), and the sample size. The following population model was used in all conditions:

$$\begin{pmatrix} x_{1i} \\ x_{2i} \\ x_{3i} \\ x_{4i} \\ x_{5i} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} (\xi_i) + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \omega \end{pmatrix} (\xi_i^2) + \begin{pmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \\ \varepsilon_{3i} \\ \varepsilon_{4i} \\ \varepsilon_{5i} \end{pmatrix}. \tag{10}$$

Note that only the regression of x_5 on ξ was specified as curvilinear. The distribution of ξ was specified as normal with variance of 1.0 and mean of $\kappa^{(g)}$ (where $g = 1, 2$

¹ The residual covariance matrices may also differ over groups if the factor variances differ over groups, having implications for tests of strict measurement invariance (see Appendix). This difference is not considered further in this article because the equivalence of the factor loadings and intercepts is generally thought to be of much greater importance than the equivalence of the residual variances.

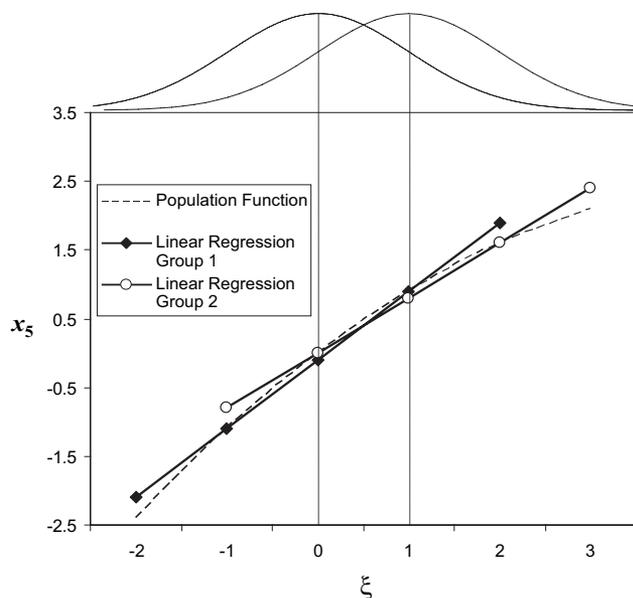


Figure 1. Linear regression of an item (x_s) on a factor (ξ) in each of two groups (solid lines) when the population generating function is truly quadratic (dashed line; $\omega = -.10$) and the standardized effect size of the factor mean difference is $d = 1.0$.

indexes group). The residuals of the indicators, ε , were specified to be normally distributed and independent with a variance of .25 each. This specification set the reliabilities of the linear indicators at .80.

Four levels of each condition were crossed in a $4 \times 4 \times 4$ factorial design. Specifically, ω was set to either $-.025$, $-.050$, $-.075$, or $-.100$. Although there are no clear norms for what the magnitude of quadratic effects might be, these levels were intentionally selected to assure monotonicity of the function over the observed range of the data (i.e., the quadratic function does not attain its maximum or change direction within the observed range of the data) and to represent apparently modest curvature in the factor-to-indicator relationship. To obtain a standardized measure of effect size, Cohen's f^2 was calculated within-groups for each level of ω , resulting in values of .005, .020, .045, and .080, respectively. According to Cohen (1988, pp. 410–414), an f^2 of .02 is a small effect, an f^2 of .15 is a medium effect, and an f^2 of .35 is a large effect. By this measure, the quadratic effect used in the simulation ranged from trivial to somewhere between small and medium. Also, the factor mean in the first group, $\kappa^{(1)}$, was set to 0 in all models to provide an anchor point for the factor mean difference and $\kappa^{(2)}$ was then set to either .25, .50, .75, or 1.0. Given the factor variance of one, these values are in the metric of Cohen's d , where .2, .5, and .8 represent small, medium, and large effects, respectively (Cohen, 1988, pp. 20–27). The larger effect size of 1.0 used in the present study was chosen in part because of the argument of Hancock (2001) that,

when comparing latent means, Cohen's standards should be adjusted upward because the mean difference is no longer attenuated by measurement error. The last condition varied in the simulation was the per group sample size, set at $N^{(1)} = N^{(2)} = 100, 200, 400, \text{ and } 800$.

Two methods were used to estimate power. First, using the approach of Satorra and Saris (1985), analytical power estimates were computed using the population mean vectors and covariance matrices implied by Equation 10 for the two groups. This approach provides power estimates for a likelihood ratio test between a model that perfectly reproduces the population means and covariances for the two groups and a second model that fails to do so. In this instance, the two-group linear factor model requiring only configural invariance provides perfect fit (through the parameter translation given in Equation 8 and the Appendix).² In contrast, a model imposing either weak or strong factorial invariance will fail to fit, as shown in Equation 9, and it is this discrepancy in fit that is used to calculate the power of the likelihood ratio test at various sample sizes (see Satorra & Saris, 1985; for extensions to multiple groups models, see Hancock, 2001, and Kaplan, 1989). One complication of this approach is that it assumes multivariate normality for the indicators, a condition that will not hold here given the quadratic relation of the fifth indicator to the factor. The nonnormality will be mild and hence can be expected to have little impact on the likelihood ratio test, but prudence dictates that an empirical approach to power estimation also be taken. As such, for each cell of the design, 500 data sets were simulated and fit by each model and the proportion of likelihood ratio tests rejected for each condition was tabulated. Data were simulated using the SAS data system and all models were fit in Mplus 3 (Muthén & Muthén, 2004). For all models, x_1 was chosen as the anchor indicator to set the scale of the latent factors.

The question of most interest was under what conditions would tests of measurement invariance be rejected if a linear factor model (fixing $\omega = 0$) was fit to the data? The first test considered was the likelihood ratio test between a model imposing weak factorial invariance relative to a model imposing only configural invariance (a 4- df test). Figure 2 shows the power curves obtained from both the analytical approach and the Monte Carlo study (using an α level of .05). The estimates from both approaches were quite consistent. When the factor mean difference was small ($\kappa^{(2)} - \kappa^{(1)} = .25$), the rejection rate was relatively low even at the highest level of curvature considered here (i.e., under 25%). However, as ω , $\kappa^{(2)} - \kappa^{(1)}$, and N increased, the probability of rejecting the equality constraints on the factor loadings also increased. The most extreme pairing of values is represented by the solid line in the lower right

² An anonymous reviewer helpfully pointed out that this would allow for the analytical approach to power estimation used here.

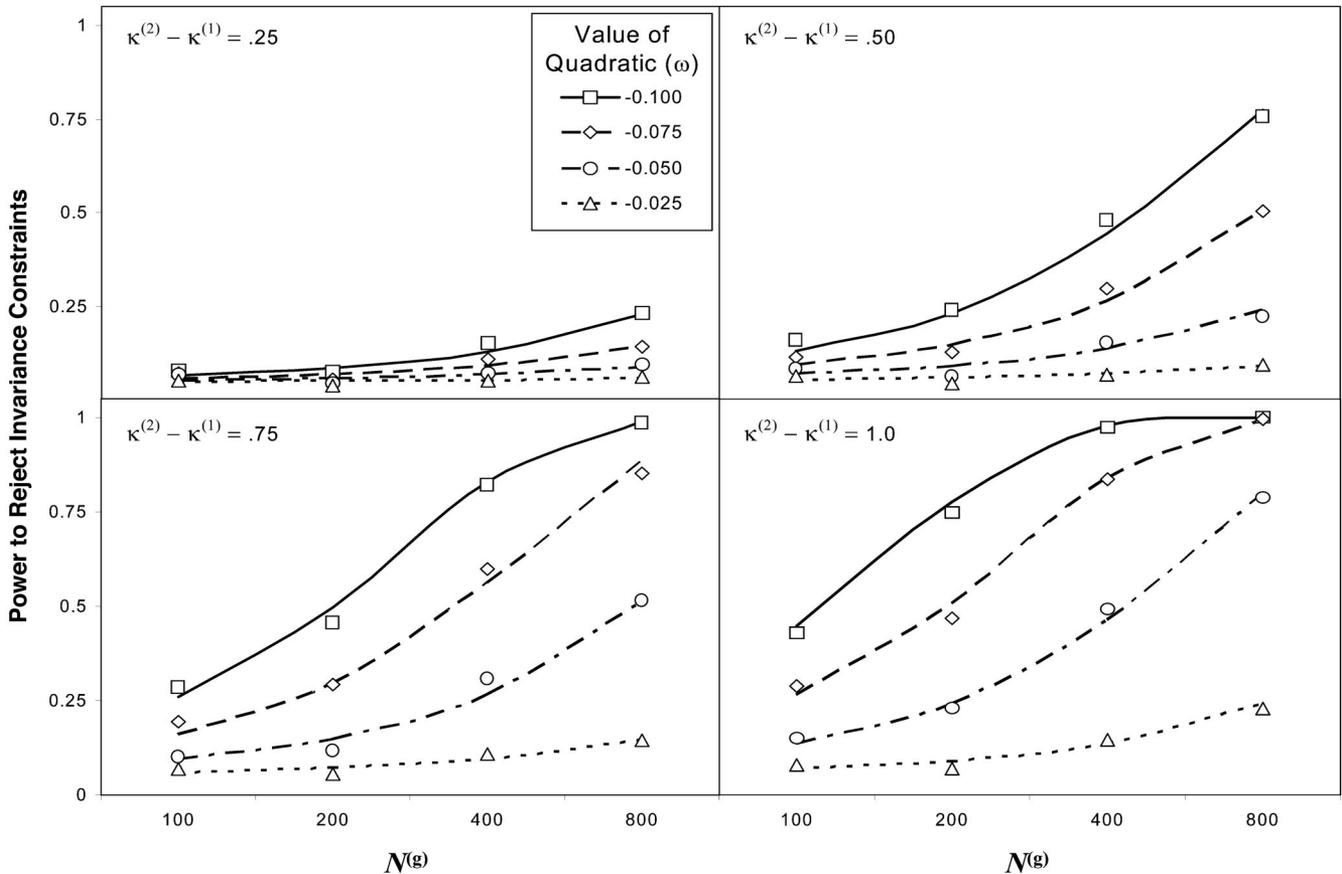


Figure 2. Test of weak invariance for the linear factor model (i.e., $\Lambda^{(1)} = \Lambda^{(2)}$). Lines reflect analytical power estimates; points reflect empirical power estimates obtained from the Monte Carlo study. Power is plotted as a function of within-group sample size.

panel of Figure 2, and these are also the values depicted in Figure 1. Although the curvature of the population function shown in Figure 1 appears quite modest, a linear model fit to this function that imposes weak measurement invariance would be rejected approximately 44% of the time in favor of one imposing only configural invariance even with a sample size of just 100 cases per group. At a sample size of 200 cases per group, the rejection rate rises to 78%. The rejection rate approaches 100% at sample sizes of 400 and 800 cases per group.

The second test considered was the likelihood ratio test between a model imposing strong factorial invariance relative to a model imposing only weak factorial invariance (again a 4-*df* test). These results are not shown because the power hovered at about .05 in all conditions, or quite close to the designated α level. This pattern of results was initially surprising because, in addition to requiring loading equality, strong factorial invariance requires intercept equality, a condition that Equation 9 suggests would not hold for x_5 . However, subsequent exploration of the results showed that the intercepts estimated for x_5 under weak factorial invariance tended to converge to similar values, despite the fact

that they were not forced to be equal (because of the imposition of equality on the factor loadings for this indicator).³ Adding formal equality constraints to these estimates then produced little further misfit, and hence they were not rejected. The chi-square test of overall fit for the strong invariance model continued to result in rejection rates that followed the same trends in Figure 1. Freely estimating the loadings and intercepts for x_5 in both groups (i.e., permitting partial invariance) resulted in good overall model fit in all conditions.

These results may be considered conservative because the values of ω used in the simulation appear to be small and only one in five indicators was nonlinearly related to ξ . Extrapolating from these results, it can be expected that even greater curvature in the regression of x_5 on ξ would lead to high rejection rates in smaller samples and with a smaller factor mean difference. Furthermore, the import of the single nonlinear indicator is to some extent obscured by

³ The expected intercept differences were observed when the model imposed only configural invariance.

the presence of four other linear indicators. A higher ratio of nonlinear to linear indicators would result in still higher rates of rejection for tests of factorial invariance and require further release of equality constraints on the factor loadings and intercepts to obtain good model fit. In addition to allowing partial invariance for these indicators, to attain good overall model fit, the investigator may also be forced to make post hoc model modifications to account for the unexplained covariance between the indicators due to the omitted nonlinear effect, such as allowing correlated residuals or introducing an additional common factor (see Appendix).

Limitations and Conclusions

The results presented here are necessarily limited in a number of respects. First, to aid in the analytical derivations, within-group normality and a quadratic factor-to-indicator relationship was assumed. Although the exact results given in Equations 8 and 9 would not hold for other population models, other nonlinear functions would similarly affect invariance testing with the linear factor model. That is, for each group, the linear factor model will provide a first-order approximation to the portion of the nonlinear function represented in that group's data. If the data for the two groups cover different regions of the nonlinear function, for instance by differing in their factor means, then the linear approximations obtained for the two groups will differ in their intercepts and slopes (factor loadings).

A second limitation of the present research is the scope of the power study, which was restricted to one of many possible population models. Despite this limitation, the results were nevertheless clear and consistent with the analytical derivations. Apparently modest curvature in the relationship of a single indicator to the factor can result in the rejection of tests of measurement invariance, particularly as the factor mean difference increases and the sample size grows. Given the strong dependence of these results on the factor means, explained analytically by Equations 8 and 9 for the quadratic case, this problem will most likely arise in those applications in which a large factor mean difference is expected. Such situations might include the comparison of age groups, pre- and posttest scores in an intervention study, or ethnic groups assessed by culturally sensitive measures, that is, in precisely those circumstances that we most wish to meet the requirement of measurement invariance so that we can compare the factor means unambiguously. Even when the factor mean difference is not of central concern, for instance when interest centers on whether the indicators are functioning similarly across groups or time, concluding that there is partial invariance may lead to substantively erroneous conclusions (e.g., that there are cultural differences in the interpretation of the measure) if the relationship is really invariant but nonlinear.

Recommendations

The implication of these results for empirical applications is that, when measurement invariance is rejected in a linear factor model, the reason may be that the true factor-to-indicator relationship is nonlinear, as opposed to inequality of measurement. Substantive theory may provide an indication of which of these two possibilities is most likely. In other cases, researchers will have to try to infer the more realistic model from the data. One preliminary strategy is to examine a scatterplot matrix of the observed variables. Nonlinear relationships between the observed variables indicate that the linear factor model is inappropriate and a nonlinear factor model should be considered (Yalcin & Amemiya, 2001). Interpretation of the scatterplot matrix may also be aided by superimposing loess (or lowess) regression lines against the data points (Cohen, Cohen, West, & Aiken, 2003, pp. 111–114). Loess regression locally weights the data points to provide a smooth nonparametric curve describing the relationship between each pair of variables (Cleveland, Devlin, & Grosse, 1988). Figure 3 displays such a plot with one of the simulated data sets (a single factor model with a quadratic effect on the fifth indicator, $\omega = -.1$, $\kappa^{(2)} = 1$, and $N^{(g)} = 200$). As can be seen, the scatterplots for x_1 , x_2 , x_3 , and x_4 all appear linear, whereas the plots for x_5 show slight but consistent curvature. From the scatterplot matrix alone, however, it is difficult to assess whether this curvature reflects a common nonlinear effect or differential linear relations in the two groups.

To obtain better resolution on this relationship, x_5 was regressed on x_1 , x_2 , x_3 , and x_4 to generate predicted values. The predicted values for x_5 represent an optimal *linear* combination of x_1 , x_2 , x_3 , and x_4 and capture variance in x_5 that is shared with these other indicators. Given that these indicators all load on a single factor, this shared variance is due to the common influence of that factor (assuming the residuals of the factor model are uncorrelated). Thus, the predicted values for x_5 are similar to factor score estimates. Similarly, the residuals from the linear regression of x_5 on the other indicators represent estimates of the residuals from the linear regression of x_5 on the latent factor, or ε_5^* . These residuals include a unique component as well as model error, as shown in Equation 11 of the Appendix.⁴ It is the model error due to the use of a linear equation that we seek to identify. The left panel of Figure 4 thus plots the observed values of x_5 against the predicted values. To highlight potentially nonlinear effects, a loess line is again fit to the data. As can be seen, although the data points are shifted to the right for the second group (consistent with the higher factor mean for this group), the nonlinear relationship be-

⁴ Alternative methods of estimating factor scores and observational residuals were also evaluated (as discussed by Bollen & Arminger, 1991), but these more complex methods did not perform as well as the method presented here.

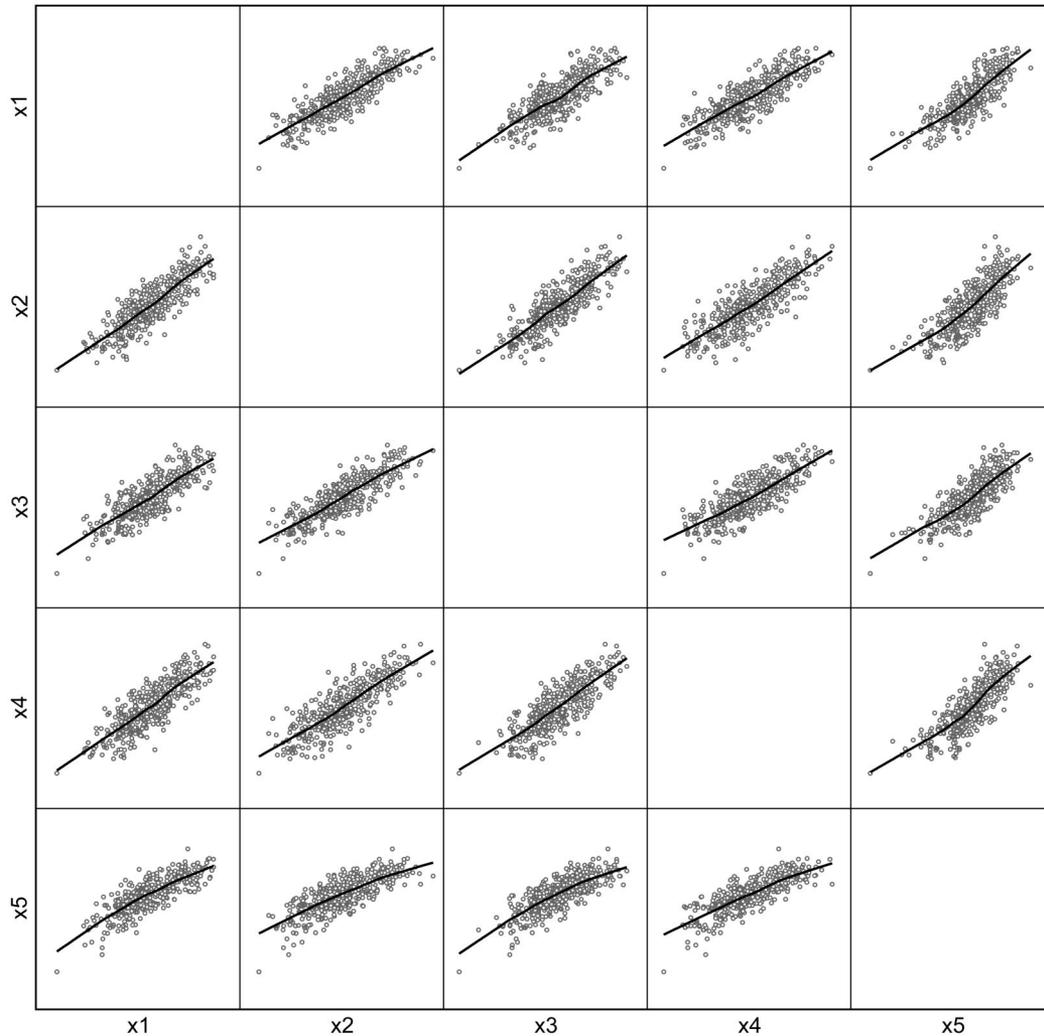


Figure 3. Scatterplot matrix and loess regression lines fit to the aggregated data from the two groups. Note the slight but consistent curvature in the relation of x_5 to the other factor indicators.

tween x_5 and the other indicators appears to be shared by both groups. This nonlinear trend is also evident in the right panel of Figure 4, which plots the residuals for x_5 versus the predicted values (i.e., removing the linear relationship). Overall, these procedures correctly identified x_5 as the sole nonlinear indicator and hence appear to be useful in diagnosing potential nonlinear effects in the factor model.

A more general two-step strategy to diagnosing nonlinear effects in a factor model would be as follows. First, produce a scatterplot matrix from all of the indicators of a single factor, as in Figure 3. On the basis of this matrix, partition the factor indicators into two subsets, one subset within which all indicators appear to be linearly related to one another, and another subset for which the assumption of linearity is more dubious. Second, regress each indicator in the second subset on all of the indicators in the

first subset within a multiple linear regression model. Save the predicted values and residuals from the regression model to produce diagnostic plots such as Figure 4 for each questionable indicator. Use these plots to aid in judging whether there is a nonlinear relationship between the indicator and factor or there are linear relationships in each group with different slopes. In the latter case, a linear factor model with partial invariance may be called for. In the former case, however, one must address the nonlinear nature of the relationship. Two options for doing so can be considered: applying a nonlinear transformation to the indicator to make the relationship more linear, so that the original model structure can be retained, or revising the model structure to include the nonlinear effect. If one selects the latter option and proceeds to estimate the nonlinear factor model, it would seem most reasonable to select an indicator from the first

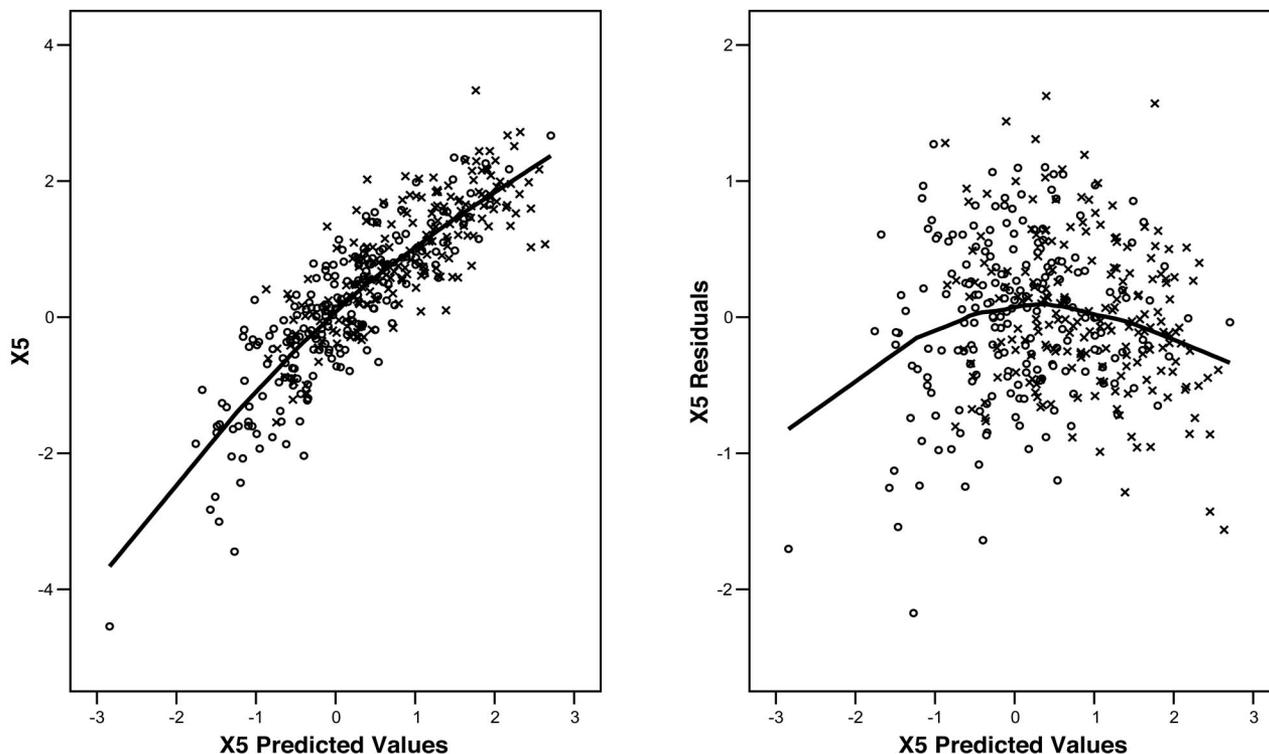


Figure 4. Plot of raw scores (left) and residuals (right) for indicator x_5 against predicted values for x_5 obtained from the optimal linear combination of x_1 , x_2 , x_3 , and x_4 . Data points for the two groups are distinguished by different symbols (o's for Group 1 and x's for Group 2). The line superimposed on each plot is obtained from a loess fit to the aggregated data.

subset to serve as the anchor for scaling the latent factor. If there are multiple factors in the model, these procedures could be repeated for the indicators of each factor in turn.

Unfortunately, given compelling evidence of nonlinearity in factor-to-indicator relationships, it has, until recently, been relatively difficult to fit a nonlinear factor analysis model. Following the early contributions of McDonald (1967), Etezadi-Amoli and McDonald (1983), and Mooijaart and Bentler (1986), however, new methods of estimation for nonlinear structural equation models have been developed, including the maximum-likelihood-based methods of Klein and Moosbrugger (2000), Klein and Muthén (2003), Lee and Zhu (2002), and Yalcin and Amemiya (2001); the method of moments approach of Wall and Amemiya (2000); and the Bayesian approaches of Arminger and Muthén (1998) and Zhu and Lee (1999). Of these methods, the approximate maximum-likelihood approach of Klein and Moosbrugger (2000) has been incorporated into the commercial software program Mplus and can be used with multisample models (Muthén & Muthén, 2004). To my knowledge, this is the only software presently available that permits the simultaneous estimation of a nonlinear factor model in two or more groups. Examples of Mplus input files are

provided at <http://dx.doi.org/10.1037/1082-989X.10.3.305.supp> showing the specification of a quadratic effect in a factor model fit to the simulated data set shown in Figures 3 and 4. It should be noted, however, that the Klein and Moosbrugger approach in Mplus is limited to the estimation of quadratic functions (or, in other contexts, product interaction models), whereas some other approaches to nonlinear factor analysis permit the estimation of more complex functions (but have not yet been extended to the analysis of multiple samples). If a more complex function is required, it may be necessary to conduct a nonlinear factor analysis in each group independently and then compare the obtained estimates in a more-or-less ad hoc manner.

Two other issues are also worth considering when estimating nonlinear factor models. First, some methods for fitting nonlinear factor models are more reliant on the assumption of normality for the latent factors than others. For these methods, failure to meet the assumption of normality may increase the risk of identifying spurious nonlinear effects (see Marsh, Wen, & Hau, 2004). Second, as in multiple regression analysis (Aiken & West, 1991, chap. 5), grand mean centering may prove useful for interpretational and computational reasons when modeling nonlinear effects.

References

- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.
- Arminger, G., & Muthén, B. (1998). A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, *63*, 271–300.
- Benton, T., Hand, D., & Crowder, M. (2004). Two zs are better than one. *Journal of Applied Statistics*, *31*, 239–247.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Bollen, K. A., & Arminger, G. (1991). Observational residuals in factor analysis and structural equation models. *Sociological Methodology*, *21*, 235–262.
- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2002). Different kinds of DIF: A distinction between absolute and relative forms of measurement invariance and bias. *Applied Psychological Measurement*, *26*, 433–450.
- Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: a reconceptualization and proposed new method. *Journal of Management*, *25*, 1–27.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.
- Cleveland, W. S., Devlin, S. J., & Grosse, E. (1988). Regression by local fitting. *Journal of Econometrics*, *37*, 87–114.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Etezadi-Amoli, J., & McDonald, R. P. (1983). A second generation nonlinear factor analysis. *Psychometrika*, *48*, 315–342.
- Goodman, L. (1960). On the exact variance of products. *Journal of the American Statistical Association*, *64*, 708–712.
- Hancock, G. R. (2001). Effect size, power, and sample size determination for structured means modeling and MIMIC approaches to between-groups hypothesis testing of means on a single latent construct. *Psychometrika*, *66*, 373–388.
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117–144.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, *36*, 409–426.
- Jöreskog, K. G., & Moustaki, I. (2001). Factor analysis of ordinal variables: A comparison of three approaches. *Multivariate Behavioral Research*, *36*, 347–387.
- Kaplan, D. (1989). Power of the likelihood ratio test in multiple group confirmatory factor analysis under partial measurement invariance. *Educational and Psychological Measurement*, *49*, 579–586.
- Kaplan, D. (2000). *Structural equation modeling: Foundations and extensions*. Thousand Oaks, CA: Sage.
- Klein, A., & Moosbrugger, H. (2000). Maximum likelihood estimation of latent interaction effects with the LMS method. *Psychometrika*, *65*, 457–474.
- Klein, A., & Muthén, B. O. (2003). *Quasi maximum likelihood estimation of structural equation models with multiple interaction and quadratic effects*. Manuscript submitted for publication.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.
- Lee, S. Y., & Zhu, H. T. (2002). Maximum likelihood estimation of nonlinear structural equation models. *Psychometrika*, *67*, 189–210.
- Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model. *Intelligence*, *31*, 543–566.
- Marsh, H. W., Wen, Z., & Hau, K.-T. (2004). Structural equation models of latent interactions: evaluation of alternative estimation strategies and indicator construction. *Psychological Methods*, *9*, 275–300.
- McDonald, R. P. (1967). *Nonlinear factor analysis*. *Psychometric Monograph Number 15*. Richmond, VA: William Byrd Press.
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analysis. *Psychological Methods*, *7*, 64–82.
- Meade, A. W., & Lautenschlager, G. J. (2004). A Monte-Carlo study of confirmatory factor analytic tests of measurement equivalence/invariance. *Structural Equation Modeling*, *11*, 60–72.
- Meredith, W. (1964). Notes on factorial invariance. *Psychometrika*, *29*, 177–185.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, *58*, 525–543.
- Meredith, W., & Horn, J. (2001). The role of factorial invariance in modeling growth and change. In A. Sayer & L. Collins (Eds.), *New methods for the analysis of change* (pp. 203–240). Washington, DC: American Psychological Association.
- Millsap, R. E. (1998). Group differences in regression intercepts: implications for factorial invariance. *Multivariate Behavioral Research*, *33*, 403–424.
- Millsap, R. E., & Kwok, O.-M. (2004). Evaluating the impact of partial factorial invariance on selection in two populations. *Psychological Methods*, *9*, 93–115.
- Mislevy, R. J. (1986). Recent developments in the factor analysis of categorical variables. *Journal of Educational Statistics*, *11*, 3–31.
- Mooijaart, A., & Bentler, P. M. (1986). Random polynomial factor analysis. In E. Diday, Y. Escoufier, L. Lebart, J. Pages, Y. Schektman, & R. Tomassone (Eds.), *Data analysis and informatics IV* (pp. 241–250). Amsterdam: Elsevier Science.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, *49*, 115–132.

- Muthén, B., & Christofferson, A. (1981). Simultaneous factor analysis of dichotomous variables in several populations. *Psychometrika*, *46*, 407–419.
- Muthén, L. K., & Muthén, B. O. (2004). *Mplus user's guide* (3rd ed.). Los Angeles: Muthén & Muthén.
- Ployhart, R. E., & Oswald, F. L. (2004). Applications of mean and covariance structure analysis: Integrating correlational and experimental approaches. *Organizational Research Methods*, *7*, 27–65.
- Satorra, A., & Saris, W. E. (1985). The power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83–90.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, *27*, 229–239.
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, *25*, 78–90.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–69.
- Wall, M. M., & Amemiya, Y. (2000). Estimation for polynomial structural equations. *Journal of the American Statistical Association*, *95*, 929–940.
- Wall, M. M., & Amemiya, Y. (in press). A review of nonlinear factor analysis statistical methods. In R. Cudeck & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions*. Mahway, NJ: Erlbaum.
- Waller, N. G., Tellegen, A., McDonald, R. P., & Lykken, D. T. (1996). Exploring nonlinear models in personality assessment: Development and preliminary validation of a negative emotionality scale. *Journal of Personality*, *64*, 545–576.
- Weisstein, E. W. (1999). Normal distribution. *MathWorld—A Wolfram Web Resource*. Retrieved January 1, 2005, from <http://mathworld.wolfram.com/NormalDistribution.html>
- Wherry, R. J., & Gaylord, R. H. (1944). Factor pattern of test items and tests as a function of the correlation coefficient: Content, difficulty, and constant error factors. *Psychometrika*, *9*, 237–244.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: applications in the substance use domain. In K. J. Bryant, M. Windle, & S. G. West (Eds.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.
- Yalcin, I., & Amemiya, Y. (2001). Nonlinear factor analysis as a statistical method. *Statistical Science*, *16*, 275–294.
- Zhu, H.-T., & Lee, S.-Y. (1999). Statistical analysis of nonlinear factor analysis models. *British Journal of Mathematical and Statistical Psychology*, *52*, 225–242.

Appendix

Derivation of Equation 8

This appendix provides a proof of Equation 8 and some additional results using standard rules of covariance algebra (see Bollen, 1989, pp. 21–23), the moments of the normal distribution (Weisstein, 1999), and expectations of products of normal distributions (Benton, Hand, & Crowder, 2004; Goodman, 1960). To begin, let us assume that the true factor-to-indicator relationship is

$$x = \tau + \lambda\xi + \omega\xi^2 + \varepsilon. \quad (\text{A1})$$

In addition, let us make the common assumptions that the latent variable ξ is normally distributed with mean κ and variance ϕ , that ε is also normally distributed with mean 0 and variance θ , and that ε is uncorrelated with ξ .

Now suppose that the form of the fitted model posits that x is linearly related to ξ such that

$$x = \tau^* + \lambda^*\xi + \varepsilon^*. \quad (\text{A2})$$

Because of the omission of the ξ^2 predictor, the coefficients τ^* and

λ^* will differ from τ and λ from Equation A1. The residuals from the regression, ε^* , will also differ from ε and hence the variance of ε^* , denoted θ^* , will differ from θ .

The coefficient λ^* can be solved for by noting that

$$\lambda^* = \frac{COV(\xi, x)}{VAR(\xi)} = \frac{E(\xi x) - E(\xi)E(x)}{\phi}. \quad (\text{A3})$$

To solve for λ^* , let us first substitute for x using Equation A1. The first term in the numerator of Equation A3 then becomes

$$\begin{aligned} E(\xi x) &= E[\xi(\tau + \lambda\xi + \omega\xi^2 + \varepsilon)] \\ &= E[\tau\xi + \lambda\xi^2 + \omega\xi^3 + \xi\varepsilon] \\ &= \tau E(\xi) + \lambda E(\xi^2) + \omega E(\xi^3) + E(\xi\varepsilon) \\ &= \tau\kappa + \lambda(\kappa^2 + \phi) + \omega(\kappa^3 + 3\kappa\phi) + 0 \\ &= \tau\kappa + \lambda\kappa^2 + \lambda\phi + \omega\kappa^3 + 3\phi\omega\kappa. \end{aligned} \quad (\text{A4})$$

A similar substitution in the second term of the numerator yields

$$\begin{aligned}
E(\xi)E(x) &= E(\xi)E(\tau + \lambda\xi + \omega\xi^2 + \varepsilon) \\
&= E(\xi)[\tau + \lambda E(\xi) + \omega E(\xi^2) + E(\varepsilon)] \\
&= \kappa[\tau + \lambda\kappa + \omega(\kappa^2 + \phi) + 0] \\
&= \tau\kappa + \lambda\kappa^2 + \omega\kappa^3 + \phi\omega\kappa. \tag{A5}
\end{aligned}$$

Subtracting Equation A5 from Equation A4 and inserting the result into Equation A3 gives

$$\lambda^* = \frac{\lambda\phi + 2\phi\omega\kappa}{\phi} = \lambda + 2\omega\kappa, \tag{A6}$$

as indicated in Equation 8. Here it is important to note that if the factor is scaled by setting the mean to zero, then λ^* will equal λ . As an interesting aside, this is also the result one would obtain by taking the first derivative of Equation A1 with respect to ξ and evaluating at the mean of ξ , or κ . That is, the slope given by Equation A6 is also the slope of the tangent line to the true quadratic function at the mean (see also Aiken & West, 1991, p. 65).

In a similar way, the coefficient τ^* can be determined by drawing on the familiar equation for the intercept of a regression line:

$$\begin{aligned}
\tau^* &= E(x) - \lambda^*E(\xi) \\
&= E(\tau + \lambda\xi + \omega\xi^2 + \varepsilon) - \lambda^*E(\xi) \\
&= \tau + \lambda E(\xi) + \omega E(\xi^2) + E(\varepsilon) - \lambda^*E(\xi) \\
&= \tau + \lambda\kappa + \omega(\kappa^2 + \phi) + 0 - \lambda^*\kappa. \tag{A7}
\end{aligned}$$

Substituting for λ^* via Equation A6, Equation A7 can be rewritten as

$$\begin{aligned}
\tau^* &= \tau + \lambda\kappa + \omega(\kappa^2 + \phi) - (\lambda + 2\omega\kappa)\kappa \\
&= \tau + \lambda\kappa + \omega\kappa^2 + \omega\phi - \lambda\kappa - 2\omega\kappa^2 \\
&= \tau + \omega\phi - \omega\kappa^2 \\
&= \tau + \omega(\phi - \kappa^2), \tag{A8}
\end{aligned}$$

as also indicated in Equation 8.

Interest may also center on the change in the residual variance as a function of the model misspecification. To determine this, the residuals ε^* must be expressed in terms of the original model parameters. Equations A1 and A2 show that

$$\begin{aligned}
\varepsilon^* &= x - (\tau^* + \lambda^*\xi) \\
&= (\tau + \lambda\xi + \omega\xi^2 + \varepsilon) - (\tau^* + \lambda^*\xi) \\
&= \varepsilon + [(\tau + \lambda\xi + \omega\xi^2) - (\tau^* + \lambda^*\xi)]. \tag{A9}
\end{aligned}$$

The grouping of terms given on the third line of Equation A9 has the intuitively pleasing interpretation that the residuals of the misspecified model are equal to the original residuals plus the discrepancy between the true and fitted regression function. That is, the residuals include both random error and model error. Sub-

stituting the results given in Equations A6 and A8 for λ^* and τ^* and subsequent simplification then yields

$$\begin{aligned}
\varepsilon^* &= \varepsilon + \tau + \lambda\xi + \omega\xi^2 - [\tau + \omega(\phi - \kappa^2) + (\lambda + 2\omega\kappa)\xi] \\
&= \varepsilon + \omega(\xi^2 - 2\kappa\xi - \phi + \kappa^2). \tag{A10}
\end{aligned}$$

The reader can verify that $E(\varepsilon^*)$ is zero.

Using the prior results, and assuming normality of ε , the variance of ε^* , denoted θ^* , can be expressed as

$$\begin{aligned}
\theta^* &= E(\varepsilon^{*2}) \\
&= E\{[\varepsilon + \omega(\xi^2 - 2\kappa\xi - \phi + \kappa^2)]^2\} \\
&= E(\varepsilon^2) - 4\kappa\omega E(\xi\varepsilon) - 2\phi\omega E(\varepsilon) + 2\kappa^2\omega E(\varepsilon) + 2\omega E(\xi^2\varepsilon) \\
&\quad + 4\kappa\phi\omega^2 E(\xi) + \kappa^4\omega^2 + \phi^2\omega^2 + \omega^2 E(\xi^4) - 2\kappa^2\phi\omega^2 \\
&\quad - 4\kappa\omega^2 E(\xi^3) - 4\kappa^3\omega^2 E(\xi) - 2\phi\omega^2 E(\xi^2) + 6\kappa^2\omega^2 E(\xi^2) \\
&= \theta + 4\kappa\phi\omega^2\kappa + \kappa^4\omega^2 + \phi^2\omega^2 + \omega^2(\kappa^4 + 6\kappa^2\phi + 3\phi^2) \\
&\quad - 2\kappa^2\phi\omega^2 - 4\kappa\omega^2(\kappa^3 + 3\kappa\phi) - 4\kappa^3\omega^2\kappa \\
&\quad - 2\phi\omega^2(\kappa^2 + \phi) + 6\kappa^2\omega^2(\kappa^2 + \phi) \\
&= \theta + 2\phi^2\omega^2. \tag{A11}
\end{aligned}$$

As one would expect, Equation A11 implies that $\theta^* > \theta$ for all values of ϕ and ω , with the inequality increasing with the degree of curvature in the true relationship and the degree of dispersion of the latent factor. That θ^* does not depend on the factor mean is a consequence of the symmetry of the quadratic function: A linear approximation will perform equally well at all points on the quadratic function. For other nonlinear functions, a linear approximation may vary in quality depending on what portion of the nonlinear function is being approximated (i.e., how nonlinear that portion of the function is). Hence, for other functions, θ^* may also depend on the factor mean.

Using derivations similar to Equation A11, it is also possible to show that if two indicators, say x_1 and x_2 , are both quadratically related to the latent factor ξ , then the covariance of the residuals obtained from fitting a partially invariant linear factor model will equal

$$COV(\varepsilon_1^*, \varepsilon_2^*) = COV(\varepsilon_1, \varepsilon_2) + 2\phi^2\omega_1\omega_2, \tag{A12}$$

where the subscripts differentiate the residuals and quadratic effects for the two indicators. If local independence holds in the population and $COV(\varepsilon_1, \varepsilon_2) = 0$, then this reduces to $COV(\varepsilon_1^*, \varepsilon_2^*) = 2\phi^2\omega_1\omega_2$. It is this apparent common variance unexplained by the linear factor model that can lead to the estimation of so-called difficulty factors to obtain good overall model fit. Alternatively, post hoc model modifications to allow for correlated residuals would also produce good model fit, in effect obscuring the misspecification of the nonlinear factor-to-indicator relationships. Note that in the multiple group context, if the factor variances differ between groups (or the quadratic effects), then the residual covariance will also differ.

Received September 8, 2004
Revision received May 24, 2005
Accepted May 27, 2005 ■