Psychological Methods
2003, Vol. 8, No. 1, 16–37

1082-989X/03/$12.00 DOI: 10.1037/1082-989X.8.1.16

On Specifying the Null Model for Incremental Fit Indices in Structural Equation Modeling

Keith F. Widaman and Jane S. Thompson
University of California, Davis

In structural equation modeling, incremental fit indices are based on the comparison of the fit of a substantive model to that of a null model. The standard null model yields unconstrained estimates of the variance (and mean, if included) of each manifest variable. For many models, however, the standard null model is an improper comparison model. In these cases, incremental fit index values reported automatically by structural modeling software have no interpretation and should be disregarded. The authors explain how to formulate an acceptable, modified null model, predict changes in fit index values accompanying its use, provide examples illustrating effects on fit index values when using such a model, and discuss implications for theory and practice of structural equation modeling.

Over the last two decades, structural equation modeling (SEM) has steadily gained popularity as a method of data analysis in the social sciences. SEM practitioners usually rely on statistical programs designed specifically for such analyses, programs including LISREL (Jöreskog & Sörbom, 1993, 1996), EQS (Bentler & Wu, 1995), AMOS (Arbuckle & Wothke, 1999), and RAMONA (Browne & Mels, 1990), among others. SEM software allows the user to specify and to test even very complex structural models with relative speed and ease.

One notable feature of most SEM software programs is the ready availability of a large number of indices of the goodness of fit of a model to data. Members of one of the more useful classes of fit indices are termed *incremental fit indices* because such indices reflect the increment in fit of a given substantive model over that of a null model. Automatic calculation of fit indices is a welcome feature of SEM software packages. However, these SEM programs automatically use, implicitly, a standard, generic null model when computing incremental fit indices. Instances in which this generic null model is an inappropriate comparison model are not uncommon; in these cases, SEM software programs will fail to correct this mistaken assumption for the user.

In this article, we discuss first the general linear structural model on which SEM is based. Then, we discuss the traditional chi-square (χ^2) test of significance of model fit and the use of incremental fit indices. In doing so, we explain the concept of nested models and the role of the null model in the computation of these fit indices. Next, we highlight situations in which the traditional null model is an inappropriate comparison model and describe the likely effects of using a different, more appropriate null model. We then provide two empirical illustrations representing situations in which the traditional null model is an incorrect comparison model and close with discussion of implications of our results for the use of incremental fit indices.

A General Linear Structural Equation Model

In this section, we review briefly one general form of the mean and covariance structure model. The

Keith F. Widaman and Jane S. Thompson, Department of Psychology, University of California, Davis.

The Appendix is available in the online version of this article in PsycARTICLES.

The present work was supported in part by intramural grants from the Academic Senate of the University of California, by National Institute of Child Health and Human Development Grant HD-22953, and by United States Office of Education Grants G0085300208 and H023C80072 (Donald MacMillan, Principal Investigator). We thank Michael Browne and Robert MacCallum for discussing several key issues with us and Chris Fraley, Katie Gibbs, and Manfred Schmitt for comments on previous versions of this article.

Correspondence concerning this article should be addressed to Keith F. Widaman, Department of Psychology, University of California, One Shields Avenue, Davis, California 95616-8686. E-mail: kfwidaman@ucdavis.edu

purpose of this presentation is to provide a background for discussion of model fit to the data. Bollen (1989b) and Byrne (1998) provided introductions to this topic, and Browne and Arminger (1995) provided a more complete discussion. Readers new to SEM should pay special attention to the definition of the $\boldsymbol{\tau}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\Theta}$ matrices in Equations 1–4 below and to the role of sample size in Equation 6. An in-depth understanding of Equations 1–5 is less important to achieve on first reading.

If p manifest (i.e., measured or observed) variables for a random person are in raw score form (i.e., not in mean-deviation form) and are placed in a column vector $\mathbf{Y}$, the linear structural equation model can be written as

$$\mathbf{Y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}\,(\mathbf{I} - \mathbf{B})^{-1}\,(\boldsymbol{\alpha} + \boldsymbol{\zeta}) + \boldsymbol{\varepsilon}, \quad (1)$$

where $\boldsymbol{\tau}$ is a $p \times 1$ vector of intercepts for the p manifest variables, $\boldsymbol{\Lambda}$ is a $p \times r$ matrix of loadings of the p manifest variables on the r latent variables, $\mathbf{I}$ is an $r \times r$ identity matrix, $\mathbf{B}$ is an $r \times r$ matrix of regression weights for predicting latent variables from other latent variables, $\boldsymbol{\alpha}$ is an $r \times r$ vector of intercepts for the r latent variables, $\boldsymbol{\zeta}$ is an $r \times 1$ vector of residual scores (in mean-deviation form) on the latent variables for the person, and $\boldsymbol{\varepsilon}$ is a $p \times 1$ vector of unique factor scores for the individual.

The linear structural model in Equation 1 implies the following moment structure model:

$$\begin{aligned}\boldsymbol{\Sigma}^* = {} & [(\boldsymbol{\tau} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha})(\boldsymbol{\tau} + \boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\alpha})'] \\ & + [\boldsymbol{\Lambda}(\mathbf{I} - \mathbf{B})^{-1}\boldsymbol{\Psi}(\mathbf{I} - \mathbf{B}')^{-1}\boldsymbol{\Lambda}' + \boldsymbol{\Theta}], \end{aligned} \quad (2)$$

where $\boldsymbol{\Sigma}^*$ is a $p \times p$ population moment matrix, $\boldsymbol{\Psi}$ is an $r \times r$ matrix of covariances among latent variable residuals, $\boldsymbol{\Theta}$ is a $p \times p$ matrix (usually diagonal) of covariances among unique factors, and all other symbols are as defined above.

If the model in Equation 2 is fit to an observed sample moment matrix $\mathbf{S}^*$, one has

$$\begin{aligned}\mathbf{S}^* \approx {} & [(\hat{\boldsymbol{\tau}} + \hat{\boldsymbol{\Lambda}}(\mathbf{I} - \hat{\mathbf{B}})^{-1}\hat{\boldsymbol{\alpha}})(\hat{\boldsymbol{\tau}} + \hat{\boldsymbol{\Lambda}}(\mathbf{I} - \hat{\mathbf{B}})^{-1}\hat{\boldsymbol{\alpha}})'] \\ & + [\hat{\boldsymbol{\Lambda}}(\mathbf{I} - \hat{\mathbf{B}})^{-1}\hat{\boldsymbol{\Psi}}(\mathbf{I} - \hat{\mathbf{B}}')^{-1}\hat{\boldsymbol{\Lambda}}' + \hat{\boldsymbol{\Theta}}] \\ = {} & \hat{\boldsymbol{\Sigma}}^*, \end{aligned} \quad (3)$$

where a caret (ˆ) was added to matrices with best-fit parameter estimates, and all other symbols are as defined above. In Equation 3, the observed moment matrix $\mathbf{S}^*$ is approximated by the linear model with best-fit parameter estimates, and this model leads to an estimate of the population moments among variables, $\hat{\boldsymbol{\Sigma}}^*$, assuming the model is correct in the population.

If the manifest variables are measured in mean-deviation form, then the $\boldsymbol{\tau}$ and $\boldsymbol{\alpha}$ parameters fall out of Equations 1–3, and Equation 3 simplifies to the typical covariance structure model:

$$\mathbf{S} \approx \hat{\boldsymbol{\Lambda}}\,(\mathbf{I} - \hat{\mathbf{B}})^{-1}\,\hat{\boldsymbol{\Psi}}\,(\mathbf{I} - \hat{\mathbf{B}}')^{-1}\,\hat{\boldsymbol{\Lambda}}' + \hat{\boldsymbol{\Theta}} = \hat{\boldsymbol{\Sigma}}, \quad (4)$$

where $\mathbf{S}$ is a $p \times p$ sample covariance matrix, $\boldsymbol{\Sigma}$ is a population covariance matrix, and all other symbols are as defined above.

If moment structure models are fit to data using Equation 3, estimates of the means of the manifest variables, $\hat{\boldsymbol{\mu}}$, are reproduced, using best-fit parameter estimates, by $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\tau}} + \hat{\boldsymbol{\Lambda}}\,(\mathbf{I} - \hat{\mathbf{B}})^{-1}\,\hat{\boldsymbol{\alpha}}$, where all symbols are as defined above. Estimates of population covariances among manifest variables are reproduced using Equation 4, which is a subset of Equation 3.

Indices of the Goodness of Fit of Models to Data

The Likelihood Ratio Chi-Square Statistic

For evaluating model fit to data, the chi-square statistic is the first fit index evaluated and reported by most investigators. The chi-square statistic is based on the discrepancy between the sample moment (or covariance) matrix and the model-implied moment (or covariance) matrix. When fitting moment structure models, the discrepancy function must incorporate misfit in modeling both means and covariances. An appropriate discrepancy function F is the normal theory maximum likelihood discrepancy function (cf. Brown & Arminger, 1995):

$$\begin{aligned}F = {} & (\overline{\mathbf{Y}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}\,(\overline{\mathbf{Y}} - \boldsymbol{\mu}) + \log|\boldsymbol{\Sigma}| - \log|\mathbf{S}| \\ & + \mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S}) - p, \end{aligned} \quad (5)$$

where $\overline{\mathbf{Y}}$ is a column vector of sample means on the p manifest variables, $\boldsymbol{\mu}$ is a column vector of model-implied means, log represents the natural log, $|\,|$ denotes the determinant of a matrix, tr is the trace operator that returns the sum of diagonal elements of a matrix, and other symbols are as defined above. The first term on the right side of Equation 5 represents the contribution to F of model misfit to the means on manifest variables. The remaining four terms on the right side of Equation 5 represent model misfit to the covariances among manifest variables. The discrepancy function F is bounded below by zero and will attain a value of zero only when $\boldsymbol{\Sigma} = \mathbf{S}$ and $\overline{\mathbf{Y}} = \boldsymbol{\mu}$.

The maximum-likelihood solution is obtained when F in Equation 5 is at its minimum for a given model. The likelihood ratio chi-square statistic is then computed as

$$\chi^2 = (N - 1)\,F, \tag{6}$$

where N is sample size, and other symbols are as defined above.[1] If t is the number of parameter estimates in a model, the chi-square statistic in Equation 6 has degrees of freedom (df) equal to $[p\,(p + 3)/2 - t]$ for a moment structure model or $df = [p\,(p + 1)/2 - t]$ for a covariance structure model. The chi-square statistic from Equation 6 is distributed approximately as a central chi-square if N is large, the model is correct, and all distributional assumptions (e.g., linearity and normality) are satisfied. Further, the null hypothesis tested by this chi-square statistic is that the model is exactly correct in the population.

Practical Fit Indices

Like any test of statistical significance, the chi-square statistic is a direct function of sample size and model misfit (cf. Equation 6). Hence, ill-fitting models might be judged to have adequate fit by this criterion if sample size were small. Conversely, the chi-square test of significance might suggest rejection of models associated with trivial discrepancies between the observed and model-implied covariance matrices and mean vectors if sample size were very large.

Consequently, alternative fit indices have been developed under the assumption that they might yield intuitively understandable indices of goodness of fit, yet be less sensitive to sample size (although sensitivity to sample size is reasonable in some situations; cf. Cudeck & Henly, 1991). These alternative fit indices are often termed *practical fit indices,* offering a contrast with the chi-square statistic, which is a statistical fit index. At least two general classes of practical fit indices may be distinguished: incremental and absolute fit indices. Each incremental fit index provides a measure of the proportional improvement in fit of a substantive model relative to a null model that is nested within the substantive model. In contrast, each absolute fit index yields a direct measure of model fit, with no consideration of the fit of the substantive model relative to a null model. Both incremental and absolute fit indices may, however, involve an implicit or explicit comparison with other models for the data, such as the saturated model.

During the past decade or so, several investigations of large numbers of SEM fit indices have been conducted (Bentler, 1990; Hu & Bender, 1998, 1999; Marsh, Balla, & McDonald, 1988). These studies have led to the following general conclusions: First, many absolute fit indices are relatively poor indicators of practical fit, as they are related too strongly to sample size. The root-mean-square error of approximation (RMSEA; Browne & Cudeck, 1993; Steiger & Lind, 1980) and the centrality index (CI; McDonald, 1989) represent two notable exceptions to this trend. Both the RMSEA and the CI are relatively independent of sample size and thus perform well as indices of practical fit. Second, in contrast, most of the commonly used incremental fit indices exhibit relative independence from sample size and thus are useful indices of practical fit. Given the generally good performance of incremental fit indices and our contention that values of these fit indices are calculated incorrectly in many cases, we restrict our attention in this article to the class of incremental fit indices. Furthermore, because absolute fit indices do not use the fit of a null model in their calculation, consideration of absolute fit indices is beyond the scope of this article.

Incremental Fit Indices

The most commonly used incremental fit indices are (a) the Tucker–Lewis index (TLI; Tucker & Lewis, 1973), also called the *nonnormed fit index* (NNFI; Bentler & Bonett, 1980); (b) an adjustment to the TLI (Bollen, 1986), called the *relative fit index* (RFI; Jöreskog & Sörbom, 1993); (c) the *normed fit index* (NFI; Bentler & Bonett, 1980); (d) a modification of the NFI (Bollen, 1989a), termed the *incremental fit index* (IFI; Jöreskog & Sörbom, 1993); (e) the *relative noncentrality index* (RNI; McDonald & Marsh, 1990); and (f) the *comparative fit index* (CFI; Bentler, 1990). These fit indices can be calculated as follows:

$$\mathrm{TLI} = \frac{(\chi^2_0/df_0) - (\chi^2_k/df_k)}{(\chi^2_0/df_0) - 1} = \frac{F_0/df_0 - F_k/df_k}{F_0/df_0 - 1/(N-1)}; \tag{7}$$

$$\mathrm{RFI} = \frac{(\chi^2_0/df_0) - (\chi^2_k/df_k)}{(\chi^2_0/df_0)} = \frac{F_0/df_0 - F_k/df_k}{F_0/df_0}; \tag{8}$$

[1] Technically, under normal theory maximum likelihood estimation, the multiplier for the chi-square statistic should be N, rather than $(N - 1)$ (see Browne & Arminger, 1995). However, with large sample size, the use of $(N - 1)$, rather than N, as multiplier has a trivial effect on the resulting chi-square value. Moreover, use of $(N - 1)$ as multiplier puts the resulting chi-square value on the same metric as the Wishart maximum-likelihood chi-square that is most commonly computed when models are fit only to the covariances among manifest variables.

$$\text{NFI} = \frac{\chi_0^2 - \chi_k^2}{\chi_0^2} = \frac{F_0 - F_k}{F_0}; \qquad (9)$$

$$\text{IFI} = \frac{\chi_0^2 - \chi_k^2}{\chi_0^2 - df_k} = \frac{F_0 - F_k}{F_0 - (df_k/(N-1))}; \qquad (10)$$

$$\text{RNI} = \frac{(\chi_0^2 - df_0) - (\chi_k^2 - df_k)}{(\chi_0^2 - df_0)} = 1 - \frac{(\chi_k^2 - df_k)}{(\chi_0^2 - df_0)} = 1 - \frac{F_k - [df_k/(N-1)]}{F_0 - [df_0/(N-1)]}; \qquad (11)$$

$$\text{CFI} = 1 - \frac{\max[(\chi_k^2 - df_k), 0]}{\max[(\chi_k^2 - df_k), (\chi_0^2 - df_0), 0]} = 1 - \frac{\max(\{F_k - [df_k/(N-1)]\}, 0)}{\max(\{F_k - [df_k/(N-1)]\}, \{F_0 - [df_0/(N-1)]\}, 0)} \qquad (12)$$

where χ_0^2, df_0, and F_0 signify, respectively, the chi-square, degrees of freedom, and minimum fit function value for the null model Model M_0, and χ_k^2, df_k, and F_k represent the chi-square, degrees of freedom, and minimum fit function value, respectively, for a substantive model of interest, Model M_k.

Nested Models in SEM

The Concept of Nested Models

The statistical and practical fit indices described above provide useful measures for evaluating the relative fit of nested models. Under parameter nesting, a common definition of nested models is this: One model, Model M_k, is nested within another model, Model M_l, if one may arrive at the parameter vector for Model M_k by placing constraints on the parameter vector for Model M_l. The constraints invoked when moving from Model M_l to Model M_k may be quite varied, including constraining parameters to zero, invoking equality constraints, and so on. However, the important consideration is that the parameter vector for the nested Model M_k can be obtained by placing constraints on the parameter vector for the more inclusive Model M_l.

Given the preceding definition, two conditions will be satisfied if Model M_k is nested within Model M_l. First, Model M_k will have fewer parameter estimates and therefore a larger number of degrees of freedom than does Model M_l. Second, the parameter vector for Model M_k cannot include new parameter estimates that do not appear in the parameter vector for Model M_l. If either of these conditions is violated, then Model M_k cannot be nested within Model M_l.

A Continuum of Nested Models

Another common conception in the use of SEM is that of a continuum of nested models, which can be represented graphically as

$$M_0 \,.........\, M_k \,.........\, M_l \,..........\, M_s,$$

a continuum bounded by Models M_0 and M_s.

At one end of this continuum is Model M_s, the saturated model. Technically, to speak of the saturated model as if only one form of saturated model specification were possible is a misnomer, as a wide array of models of very different form satisfy the definition of the saturated model. To qualify as a saturated model, a model must (a) make as many parameter estimates as can be made from the data, a number of estimates that equals the number of unique elements of the covariance matrix and mean vector (if these are included in the analysis); and (b) reproduce exactly the covariances among and means on the manifest variables. If these conditions hold, a saturated model will have a chi-square value of zero with 0 *df.*

At the other end of the continuum is Model M_0, the null model. Once again, several options exist for specifying a null model. Any version of the null model must (a) estimate as few parameters as are reasonable for the data, and (b) reproduce a nonzero variance and mean (if included in the analysis) for each manifest variable.

Falling between the two ends of the continuum are interesting substantive models, two of which are identified above as Models M_l and M_k. If, as above, the parameter vector for Model M_k can be obtained by placing one or more constraints on the parameter vector for Model M_l, then Model M_k is nested within Model M_l. If all substantive models can be arrayed along a single continuum, as above, then the nesting relations among models are simple to characterize. At times, however, a linear continuum cannot be used to arrange models because a series of competing models has no simple nesting pattern. In such cases, a tree structure of models may be necessary; a tree structure with utility for certain classes of model is described in a later section. Regardless of whether a linear continuum or more complex tree structure is required to portray the nesting relations among models, any acceptable ordering of models should satisfy the following two conditions: (a) any given substantive model (e.g., Model M_k) should be more restricted than and therefore nested within the saturated model, Model M_s; and (b) the null model, Model M_0, should be more restricted than and therefore nested within any given substantive model, Model M_k, as well as, by substitution, within the saturated model.

The Form of the Null Model

We used the subscript zero for the null model in this and preceding sections to indicate that the null model considered here represents the hypothesis that the manifest variables are mutually independent and therefore reproduces zero, or null, covariances among all manifest variables. That is, in the traditional independence null model, the covariances among all manifest variables are constrained to zero, even though the variance and mean of each manifest variable are not constrained in any way and are therefore freely estimated. Tucker and Lewis (1973) used just such an independence model when developing the first incremental fit index, and Bentler and Bonett (1980) emphasized the independence null model when developing their incremental fit indices.

On the other hand, some researchers have argued that an independence null model is too restrictive a comparison model. For example, Sobel and Bohrnstedt (1985) argued that the null model should include estimates of certain parameters (e.g., factor loadings) that have been verified in prior research. Recently, Rigdon (1998a; see also Marsh, 1998; Rigdon, 1998b) argued that an equal-correlation model was a more appropriate null model than was an independence model.

Detailed consideration of claims regarding these competing null models is beyond the scope of this article, although these issues clearly deserve additional attention. For the present, we consider only independence null models that reproduce zero or null covariances among manifest variables. An independence null model is the most widely accepted null model and is the model used by existing SEM software when calculating incremental fit indices. However, as we show below, in many research applications, the standard independence null model is not nested within all substantive models under consideration, is therefore an inappropriate comparison model, and must be reformulated to be an acceptable null model.

The Use of Nested Models in Evaluating Fit

Likelihood Ratio Test Statistics

As is well known, if one model, M_k, is nested within another model, M_l, the difference in fit between the two models can be evaluated using a nested model chi-square difference test based on the likelihood ratio test statistics for the two models. Using this test, the difference in chi-square values, or $\Delta\chi^2$, is distributed as a chi-square variate with degrees of freedom equal to the difference in degrees of freedom, or Δdf, for the two models. If this $\Delta\chi^2$ is significant, then a statistically significant decrement in model fit accompanies the constraints on Model M_l that result in Model M_k.

Because the saturated model has 0 *df* and a chi-square value of zero, the chi-square statistic for any given substantive model, M_k, is directly a test of the statistical difference in fit between the substantive and saturated models. Although rarely considered, a chi-square difference test also could be computed between the substantive model and the null model nested within it. Such a test would provide a statistical indicator of improved fit of the substantive model over that of the null model. However, a statistical comparison between the substantive and null models has a valid interpretation only if the null model is nested within the substantive model and, by substitution, within all substantive models considered for a set of data. By extension, incremental fit indices also require the null model to be formally nested within all substantive models because these fit indices involve an explicit comparison with a null model.

Incremental Fit Index Values

The incremental fit indices in Equations 7–12 share a common characteristic: Each provides a measure of the proportional improvement in fit of a substantive model relative to a null model that is nested within the substantive model. This highlights the need for the null model to be nested within all substantive models considered for a given set of data, as only then can the proportional improvement in fit of competing substantive models be defined in a globally consistent fashion.

One way of conceiving of incremental fit indices is as providing indices of the placement of given substantive models on a normalized, 0–1 continuum, where the null model has a value of zero and an ideal model that fits perfectly in the population has a value of unity. The incremental fit indices in Equations

7–12 are distinguished by the ways in which they define the end points and the metric of the continuum. To define the end points of the continuum, the metric of the continuum, and the placement of a given substantive model on the continuum, one needs (a) the chi-square and degrees of freedom for the null model, for an ideal model, and for a substantive model or (b) the function minimum F for the null, ideal, and substantive models and sample size (cf. Equations 7–12). The statistics for an ideal model for the data are needed for the following reasons: An ideal model is correct in the population. Given this, if N is large and all assumptions (e.g., multivariate normality) are satisfied, then sampling variability is the only source of misfit of an ideal model in a sample. That is, sampling variability will lead to a nonzero value of the discrepancy function F and is therefore the only reason the chi-square statistic for an ideal model differs from zero. In such situations, the expected value of the sample estimator of the chi-square statistic is equal to its degrees of freedom, so the expected value of the sample estimator of the χ^2/df ratio for an ideal model is unity.

TLI. When formulating the TLI, Tucker and Lewis (1973) noted that the minimum fit function value F and its associated chi-square are both in a metric of sums of squared residuals. They then divided each model chi-square by its degrees of freedom to yield a measure of misfit that is in a mean square metric, and this also served as a strong correction for model complexity. The expected value of the sample estimator of the χ^2/df ratio is unity for an ideal model, so the denominator of the TLI (see Equation 7) represents the length of the continuum running from the null model to an ideal model, a continuum that has a mean square, or χ^2/df, metric. By subtracting the χ^2/df ratio for a substantive model from the χ^2/df for the null model and then dividing by the length of the null-model-to-ideal-model continuum, the TLI provides an index of the relative placement of a substantive model along the continuum. A substantive model with a TLI of .95 falls 95% of the way, in mean-square metric units, along the continuum from the null model (TLI $=$ 0) to an ideal model for the data (which has $\chi^2 = df$ and therefore $\chi^2/df = 1$ and TLI $= 1.0$).

RFI. The mean square metric pioneered by the TLI is retained in the RFI, but the upper end point of the null-model-to-ideal-model continuum is defined in a different fashion (see Equation 8). To force values of the RFI to fall between 0 and 1, the ideal model must have a χ^2/df ratio of zero. The only model having a chi-square statistic with a sample expected value of zero is a saturated model, the χ^2/df ratio for which is undefined (i.e., 0/0) because it has 0 *df.* Despite this problem, an ideal model is implicitly defined as one with a χ^2/df ratio of zero. Thus, a substantive model with an RFI of .95 falls 95% of the way, in mean-square metric units, on the continuum from the null model (RFI $=$ 0) to this redefined ideal model for the data (having $\chi^2/df = 0$ and RFI $=$ 1).

NFI and IFI. In the NFI, the null-model-to-ideal-model continuum is redefined in a sums of squares metric, embodied in the chi-square (see Equation 9). If one also redefines the ideal model as a model that fits the data perfectly, a substantive model with an NFI of .95 falls 95% of the way, in sum-of-square metric units, along the continuum from the null model (NFI $=$ 0) to an ideal model for the data (with $\chi^2 = 0$ and NFI $=$ 1). The IFI (Equation 10) has a form that is very similar to the NFI, the only change being the redefinition of the ideal model as a model with chi-square equal to the degrees of freedom for the given substantive model. The length of the null-model-to-ideal-model continuum changes if competing substantive models have different degrees of freedom values, so computing the difference in IFI values for competing nested models is technically invalid, though often not too misleading. We note that the IFI is the only incremental fit index with this problematic feature. Nonetheless, a substantive model with an IFI of .95 falls 95% of the way, in sum-of-square metric units, along the continuum from the null model (IFI $=$ 0) to this redefined ideal model for the data (having $\chi^2 = df$ for the substantive model, hence IFI $=$ 1).

RNI and CFI. In the RNI and CFI[2] (Equations 11 and 12, respectively), the null-model-to-ideal-model continuum is defined in a sums of squares metric similar to that of the NFI (cf. Equation 9), albeit couched in terms of noncentrality. The *noncentrality parameter* (NCP) is a parameter of a theoretical distribution, in this case the chi-square distribution. In SEM, if a model is correct in the population, N is

[2] The RNI (McDonald & Marsh, 1990) and CFI (Bentler, 1990) are essentially identical in theoretical justification and similar in computation. When RNI values fall between 0 and 1, the RNI and CFI yield identical values. Bentler (1990) added side conditions to keep CFI values within the 0–1 continuum, whereas RNI values are not constrained to fall within this range.

large, and all assumptions are satisfied, then sample values of the likelihood ratio test statistic are distributed as a central chi-square with appropriate degrees of freedom (cf. Equation 6 and its discussion). A central chi-square has an NCP equal to zero. However, if a model exhibits misfit in the population, as do virtually all empirical models, or if assumptions are not satisfied, then sample values of the likelihood ratio test statistic are distributed as a noncentral chi-square, with an NCP greater than zero if N is large. The NCP is related to the degree of model misfit in the population, with greater misfit reflected in a larger NCP. In a sample, the NCP for a model is estimated by the model $\chi^2 - df$. Hence, the NCP represents excess model misfit beyond that expected by sampling variability alone. The expected value of the sample estimator of the NCP is zero for an ideal model because an ideal model fits perfectly in the population. Because of this, a model with an RNI or CFI of .95 falls 95% of the way, in sum-of-square metric units, along the continuum from the estimated noncentrality of the null model (RNI = CFI = 0) to the noncentrality associated with an ideal model (having $\chi^2 = df$, NCP = 0, and RNI = CFI = 1.0).

The Nature of the Independence Null Model

The Standard Null Model

One way to specify the standard independence null model Model M_0 is to delete all parameter estimates associated with latent variables. For example, if the covariance structure model in Equation 4 is used, fixing the three parameter matrices $\mathbf{\Lambda}$, $\mathbf{B}$, and $\mathbf{\Psi}$ to be null matrices leaves the diagonal $\mathbf{\Theta}$ matrix as the only matrix containing parameter estimates. Estimating only the p diagonal elements of $\mathbf{\Theta}$ leads to a null model that reproduces exactly the variance of each manifest variable and null covariances among all manifest variables. If the model in Equation 4 is extended to a multiple-group setting, then p estimates of the variances of the manifest variables are made in each of the G groups, for a total of pG parameter estimates across the groups.

If the moment structure model in Equation 3 is fit to the means on and covariances among manifest variables, defining the four parameter matrices $\mathbf{\Lambda}$, $\mathbf{B}$, $\boldsymbol{\alpha}$, and $\mathbf{\Psi}$ to be null matrices leaves the $\boldsymbol{\tau}$ vector and diagonal $\mathbf{\Theta}$ matrix as the only matrices with parameter estimates. A model of this form uses $2p$ parameter estimates to reproduce both means and variances of manifest variables but represents covariances among manifest variables as null, or zero. If this model is extended to a multiple-group setting, $2pG$ estimates of mean and variance parameters are made across the groups.

Specifying an Acceptable Null Model

The standard null model described above is appropriate for many single-sample analyses. However, many single-group models and most multiple-group models require a modified null model, labeled here Model M_0', to ensure that this modified null model is nested within the most restrictive substantive model for the data. Two conditions must hold for a model to be an acceptable independence null model. First, an acceptable null model must represent covariances among manifest variables as null, or zero. Second, and the key distinction here, if any within-group and/or between-group constraints on estimates of manifest variable means or residual variances are invoked in any substantive models under consideration, these constraints must be included in an acceptable null model. These constraints on means and residual variances will typically be operationalized as constraints on the $\boldsymbol{\tau}$ and $\mathbf{\Theta}$ matrices that are the only matrices with parameter estimates in the standard null model. Thus, a modified, acceptable null model is more highly constrained than the standard null model. One further complication is the presence of more than one acceptable null model. We describe below considerations that govern selection of one of these multiple null models when computing incremental fit indices for a given set of data.

Constraints on mean and/or residual variance parameter estimates might seem illogical in a null model in which manifest variables are represented as statistically uncorrelated. However, constraints on mean and/or residual variance estimates reflect substantive hypotheses about the scale (i.e., the mean and/or variance) of each manifest variable. Such hypotheses are separable from those regarding the latent structure underlying covariances among manifest variables. Further, the resulting null model can be a meaningful model in its own right. For example, low birth weight infants are at risk for many negative outcomes. Suppose a researcher assessed a sample of low birth weight infants on several measures of infant mental functioning, and scaled scores on each measure were in an IQ metric (i.e., $\mu = 100$ and $\sigma = 15$ in the population). Placing constraints on estimates of the mean parameters would allow the researcher to test whether low birth weight infants were affected equally in mean level across the measures in the analysis; similarly, constraints on residual variance parameter estimates would enable tests of equality of variance across the measures. These are meaningful hypotheses to test, regardless of whether the covariances among the measures were nonzero and consistent with a linear structural model of some form (e.g., factor analytic).

Below, we describe briefly some single-group and multiple-group situations to illustrate how within- and between-group constraints can be implemented in an acceptable null model. First, we describe likely effects on incremental fit index values if a modified null model is used.

Effects on Incremental Fit Indices of a More Restricted Null Model

In the preceding section, we argued that the standard null model is an inappropriate comparison model in certain research situations. In each of these cases, an acceptable null model has additional constraints on parameter estimates and therefore a larger number of degrees of freedom than the standard null model. With additional restrictions, the chi-square value for an acceptable null model must be equal to or greater than the chi-square for the standard null model. Here, we explore likely effects of change in chi-square and degrees of freedom for an acceptable null model on computed fit index values.

Change in the χ^2/df ratio for the null model. Relative to the χ^2/df ratio for the standard null model, the χ^2/df ratio for an acceptable null model may be smaller, the same, or larger. These outcomes would occur if the increase in chi-square value for the acceptable null model over that for the standard null model were proportionally less than, equal to, or greater than the increase in degrees of freedom, respectively. Any change in the χ^2/df ratio for the null model would have its largest direct effects on the TLI (Equation 7) and RFI (Equation 8), the two incremental fit indices that incorporate the strong χ^2/df correction for model complexity in their calculation. If the χ^2/df ratio for the acceptable null model were smaller than that of the standard null model, TLI and RFI index values for substantive models would be lower in absolute terms, but differences in fit between nested substantive models would be larger. If the χ^2/df ratio for the acceptable null model were larger than that of the standard null model, TLI and RFI index values for substantive models would be larger in absolute value, and differences in fit between nested substantive models would be smaller. Of course, if the χ^2/df ratio for the acceptable null model were identical to that of the standard null model, then TLI and RFI index values would be unchanged.

Change in chi-square for the null model. Given that an acceptable null model imposes additional restrictions beyond those in the standard null model, the overall chi-square for an acceptable null model must be equal to or greater than the chi-square for the standard null model. Values of the NFI and IFI (Equations 9 and 10, respectively) would be unchanged if the chi-square for an acceptable null model were equal to the chi-square for the standard null model. If the chi-square for an acceptable null model were greater than the chi-square for the standard null model, then all NFI and IFI fit index values would be larger in absolute value but show smaller differences between alternative nested models.

Change in the estimated NCP for the null model. Because the chi-square for an acceptable null model must be equal to or greater than the chi-square for the standard null model, and degrees of freedom for the former model must be greater than that for the latter, the estimated NCP for an acceptable null model may be smaller than, equal to, or larger than that for the standard null model. Change in the estimated NCP for the null model is most directly relevant to the RNI and CFI fit indices (Equations 11 & 12, respectively), the indices that rely directly on the estimated NCP for the null and substantive models in their calculation. If the estimated NCP for an acceptable null model were smaller than that for the standard null model, RNI and CFI fit index values would be lower in absolute magnitude, and differences between competing models would be larger or more pronounced. If the estimated NCP for an acceptable null model were larger than that for the standard null model, RNI and CFI fit index values would be higher in absolute magnitude, and differences between competing models would be smaller. Of course, if the estimated NCP for an acceptable null model were identical to that for the standard null model, RNI and CFI fit index values would be unchanged.

General comment. Note also that regardless of the difference in the χ^2/df ratio, chi-square, or estimated NCP between the standard and acceptable null models, values of all incremental fit indices will be affected less by choice of null model the nearer the fit index values were to unity when using the standard null model.

Research Applications Requiring a More Restricted Null Model

In this section, we describe single- and multiple-sample models for which the standard null model is an inappropriate comparison model and demonstrate how an acceptable null model is nested within substantive models.

Single-Sample Models

Psychometric models. As discussed by Jöreskog (1974), psychometric models can be evaluated using SEM software. If one considers the means of tests as well as their covariances, psychometric models that impose stringent requirements on test scores can be tested. To represent these models, constrain **B** and $\boldsymbol{\alpha}$ in Equation 2 to be null matrices, leaving

$$\boldsymbol{\Sigma}^* = \boldsymbol{\tau}\,\boldsymbol{\tau}' + \boldsymbol{\Lambda}\,\boldsymbol{\Psi}\,\boldsymbol{\Lambda}' + \boldsymbol{\Theta}, \qquad (13)$$

where all symbols are as defined above.

The most common psychometric models considered are models for parallel, tau-equivalent, and congeneric tests. Under usual definitions, parallel tests have equal means, equal true score variances, and equal error variances. Tau-equivalent tests have equal means and equal true score variances, but error variances that may vary across tests. Congeneric tests are assumed to measure the same construct but have no constraints on means, true score variances, or error variances. We also distinguish essentially parallel and essentially tau-equivalent tests, which are identical to parallel and tau-equivalent tests, respectively, except that the stipulations about equality of means are dropped, so these models may be evaluated using only covariances among tests or by relaxing equality constraints on mean estimates.

The preceding models may be tested using Equation 13 in the following manner: First, fix $\boldsymbol{\Psi}$ to be a unit scalar. Then, (a) parallel tests have equal means (or intercepts) in $\boldsymbol{\tau}$, equal saturation with the latent construct (i.e., equal factor loadings) in $\boldsymbol{\Lambda}$, and equal unique factor variances in $\boldsymbol{\Theta}$; (b) essentially parallel tests are specified as are parallel tests, but equality constraints on means in $\boldsymbol{\tau}$ are relaxed; (c) tau-equivalent tests have equal intercepts in $\boldsymbol{\tau}$ and equal factor loadings in $\boldsymbol{\Lambda}$, but unique factor variances in $\boldsymbol{\Theta}$ that vary across tests; (d) essentially tau-equivalent tests are specified as are tau-equivalent tests, except that equality constraints on means in $\boldsymbol{\tau}$ are relaxed; and (e) congeneric tests have no constraints on intercepts, factor loadings, or unique variances. If a researcher has access only to covariances among tests, only comparisons among essentially parallel, essentially tau-equivalent, and congeneric tests can be performed (cf. Jöreskog, 1974).

The standard null model is an inappropriate comparison model for calculating either statistical difference tests or incremental fit indices for the psychometric models outlined above because the standard null model cannot be nested within the most constrained of the competing models. To see why, consider the tree structure for models shown in Figure 1. The tree structure in Figure 1 was based on the following considerations: Factor loadings in $\boldsymbol{\Lambda}$ can be freely estimated, estimated with constraints, or fixed at zero (as in an independence null model), but intercepts in $\boldsymbol{\tau}$ and residual variances in $\boldsymbol{\Theta}$ can only be freely estimated or estimated with constraints because fixing parameters to zero in these matrices is rarely, if ever, tenable. Given this, we entertained three specifications for elements in the $\boldsymbol{\Lambda}$ matrix: Structure 0, in which all factor loadings are fixed at zero; Structure 1, in which all factor loadings are estimated but constrained to be invariant (or equal); and Structure 2, in which all factor loadings are freely and separately estimated. Four sets of specifications were considered for the $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ matrices: Structure A, in which invariance (or equality) constraints are placed on all elements in $\boldsymbol{\tau}$ and on all elements in $\boldsymbol{\Theta}$; Structure B, with invariance constraints on all elements in $\boldsymbol{\tau}$ but no constraints on elements in $\boldsymbol{\Theta}$; Structure B′, with invariance constraints on all elements in $\boldsymbol{\Theta}$ but no constraints on elements in $\boldsymbol{\tau}$; and Structure C, with all elements in both $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ freely estimated. The numbers and letters assigned to structures provide a key to nesting relations, with structures having earlier ordinal status being more restricted than, and therefore nested within, structures having later ordinal status. That is, Structure 0 is nested within Structure 1, and Structure 1 is nested within Structure 2. Similarly, Structure A is nested within Structures B and B′, and Structures B and B′ are nested within Structure C. Because Structures B and B′ have the same ordinal status, neither is nested within the other, and no statistical comparisons can be made between these structures.

The Cartesian product of the three types of $\boldsymbol{\Lambda}$ structure with the four types of structure for $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ generates 12 potential models, arrayed as a tree structure in Figure 1 having three columns and four rows. The first, second, and third columns contain, respectively, models with Structures 0, 1, and 2 for $\boldsymbol{\Lambda}$. Thus, within

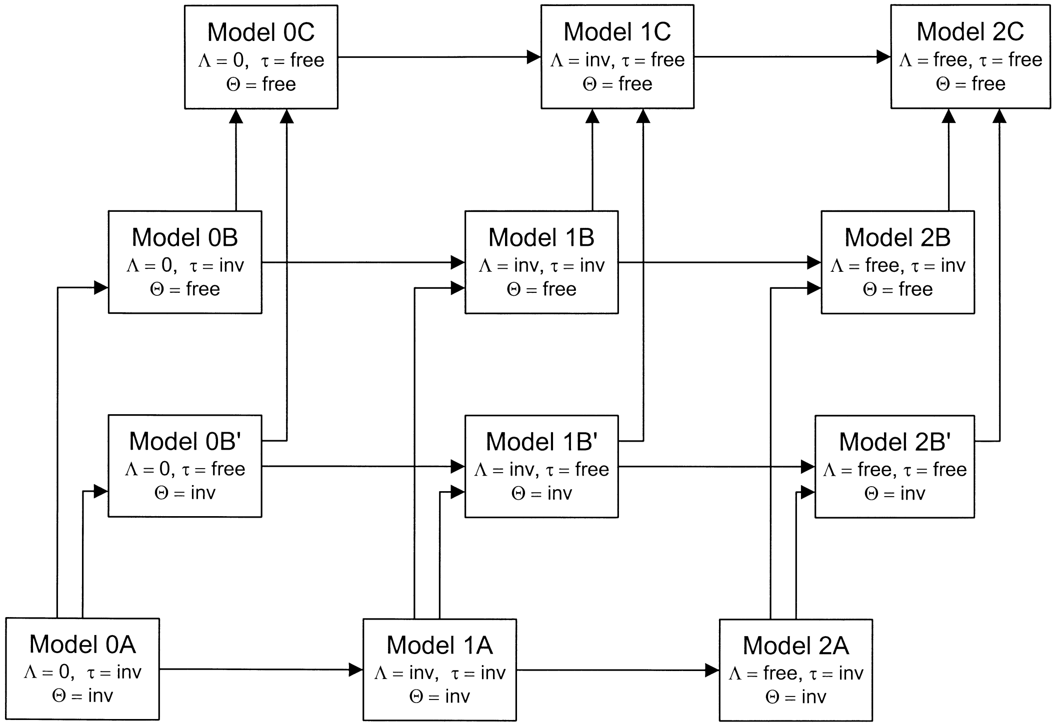

Figure 1. A tree structure of confirmatory factor models based on patterns of constraint placed on the $\boldsymbol{\Lambda}$, $\boldsymbol{\tau}$, and $\boldsymbol{\Theta}$ matrices (where 0 means that parameter estimates are fixed at zero, *inv* that parameter estimates are constrained to invariance, and *free* that parameters are separately and freely estimated). The three columns of the tree structure, from left to right, are defined by constraints on the $\boldsymbol{\Lambda}$ matrix: in Column 0, $\boldsymbol{\Lambda}$ elements are fixed at zero; in Column 1, $\boldsymbol{\Lambda}$ elements are estimated under invariance constraints; and in Column 2, $\boldsymbol{\Lambda}$ elements are freely estimated. The four rows of the tree structure, from bottom to top, are defined by constraints on the $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ matrices: in Row A, elements in both $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ are estimated under invariance constraints; in Row B′, elements in $\boldsymbol{\tau}$ are freely estimated, but $\boldsymbol{\Theta}$ elements are estimated under invariance constraints; in Row B, elements in $\boldsymbol{\tau}$ are estimated under invariance constraints, but $\boldsymbol{\Theta}$ elements are freely estimated; and in Row C, elements in both $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ are freely estimated.

each row, a model is nested within any model to its right. From bottom to top, the four rows of the tree structure contain, respectively, models with Structures A, B′, B, and C for $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$. Hence, within each column, a model is nested within models in higher rows, with the exception that no nesting relations hold between models in the second and third rows. Single-headed arrows represent simple nesting relations between models, as each arrow is drawn from a nested (i.e., more restricted) model to a less restricted model in which constraints on parameter estimates in a single matrix in the nested model are relaxed; arrows drawn between models conform to the nesting relations discussed above. Thus, Model 1A is nested within Model 1B because invariance constraints on elements in $\boldsymbol{\Theta}$ in the former model are relaxed in the latter. The least restricted model shown in Figure 1 is Model 2C at the upper right corner of the tree structure; the saturated model (not shown in Figure 1) would fall to the right and/or above Model 2C. Conversely, the most restricted model shown in Figure 1 is Model 0A, in the lower left corner of the tree structure; as the most restricted model, Model 0A is nested within all other models shown in Figure 1. In general, any model in Figure 1 that falls to the left of and/or below a second model is nested within the second model, except that models in the second and third rows have no nesting relations.

Note that all four models in the first column—Models 0A, 0B′, 0B, and 0C—are null models because all elements in $\boldsymbol{\Lambda}$ are constrained to zero and thus covariances among manifest variables are represented as null in all four models. Which null model should be used in a particular application depends on the nature of the substantive models considered for the data. If the most highly constrained model considered for a set of data were Model 1C or 2C, then Model 0C would be an acceptable null model as Model 0C is nested within Models 1C and 2C. Because Models 0B, 0B′, and 0A are nested within Model 0C, these additional null models would also be acceptable null models. Model 0C is perhaps the most logical choice as null model because it is the least restricted null model that is nested within the most restricted substantive model for the data. However, if a researcher ever considered either Model 1B or 2B for a set of data, Model 0C would no longer be an acceptable null model because Model 0C is not nested within Models 1B and 2B. Instead, given its nesting relation, Model 0B would likely be the null model of choice. Further, if the researcher ever considered either Model 1A or 2A for the data, then Model 0A would be the only acceptable null model because Model 0A is the only null model nested within Models 1A and 2A.

Model 0C is the standard independence null model used by SEM programs when computing incremental fit indices. Given the foregoing discussion, Model 0C will be an appropriate and acceptable null model for many applications but will not be acceptable for others. To determine the optimal null model, a researcher should consider all substantive models to be fit to a given set of data. The optimal acceptable null model must be nested within the most restricted substantive model specified for the data, must constrain covariances among manifest variables to zero, and should contain no additional constraints beyond those invoked in the most restricted substantive model. Once it is selected, this optimal acceptable null model should be used when computing incremental fit index values for all substantive models fit to the given set of data.

With regard to psychometric test models, the various test theory models identified above can be located within the tree structure shown in Figure 1. Here, invariance constraints are within-group constraints on estimates in parameter matrices, so the parallel test model is consistent with Model 1A, the tau-equivalent test model is consistent with Model 1B, the essentially tau-equivalent test model is consistent with Model 1C, and the congeneric test model is consistent with Model 2C. Assuming that analyses are based on the means on and covariances among tests, the standard null model, Model 0C, has $2p$ parameter estimates, which represent the mean and variance of each of the p tests. The parallel test model, Model 1A, has only three parameter estimates, as the means, factor loadings, and unique factor variances for tests are constrained to equality across tests. Because p must be greater than or equal to 2, the standard null model has more parameter estimates than the parallel test model, so cannot be nested within it. An acceptable null model, Model 0A, retains the equality constraints on the unique factor variances in $\boldsymbol{\Theta}$ and the mean vector in $\boldsymbol{\tau}$ that are present in Model 1A but constrains the factor loadings in $\boldsymbol{\Lambda}$ to zero. Thus, the optimal null model for this research application has only two parameter estimates and is formally nested within the parallel test model. The verbal statement for this null model is as follows: Each test has positive variance, and all of these variances are equal; each test has an unknown mean, and all of these means are equal; and the tests are mutually uncorrelated.

Example 1: Data from Votaw (*1948*). To illustrate these ideas, we examined data reported by Votaw in an evaluation of the reliability of scoring essays. A total of 126 students were tested; the instrument was an English composition test with three separate essays. The four forms used in our analyses were (a) Typed 1, a typed copy of the part 1 essay; (b) Written 1, a handwritten version of the part 1 essay; (c) Carbon 1, a carbon copy of the handwritten version of the part 1 essay; and (d) Typed 2, a typed copy of the part 2 essay. The four forms were given to a set of graders to be scored. Covariances among scores on the four forms were provided by Jöreskog and Sörbom (1996); the means of the ratings were reported by Votaw, except for the mean of Typed 2, which was therefore imputed. With both means on and covariances among the four forms (see Table 1), parallel and tau-equivalent test restrictions can be evaluated.

Using Equation 13, a one-factor model, with all four tests loading on the single common factor, was the basic model for the data. The $\boldsymbol{\Psi}$ matrix was a 1×1 matrix, with $\Psi_{11} = 1.0$ in all models. To illustrate effects on fit indices of fitting restricted substantive models, we evaluated six models from the tree structure in Figure 1:

Model 2C: $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4)$;
$\boldsymbol{\Lambda} = (\lambda_{11}, \lambda_{21}, \lambda_{31}, \lambda_{41})$;
$\boldsymbol{\Theta} = \text{diag}(\theta_{11}, \theta_{22}, \theta_{33}, \theta_{44})$;

Model 1C: $\boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4)$;
$\boldsymbol{\Lambda} = (\lambda_{11} = \lambda_{21} = \lambda_{31} = \lambda_{41})$;
$\boldsymbol{\Theta} = \text{diag}(\theta_{11}, \theta_{22}, \theta_{33}, \theta_{44})$;

Model 1B: $\boldsymbol{\tau} = (\tau_1 = \tau_2 = \tau_3 = \tau_4)$;
$\boldsymbol{\Lambda} = (\lambda_{11} = \lambda_{21} = \lambda_{31} = \lambda_{41})$;
$\boldsymbol{\Theta} = \text{diag}(\theta_{11}, \theta_{22}, \theta_{33}, \theta_{44})$;

Table 1
Example 1: Covariances Among Grades Assigned to Four Essays, With Means on the Essays, From Votaw (1948)

	Essay			
Essay	Typed 1	Written 1	Carbon 1	Typed 2
Typed 1	25.0704			
Written 1	12.4363	28.2021		
Carbon 1	11.7257	9.2281	22.7390	
Typed 2	20.7510	11.9732	12.0692	21.8707
M	14.9048	15.4841	14.4444	15.1234

Note. $N = 126$. Covariances are from *LISREL 8: User's Reference Guide* (p. 127), by K. G. Jöreskog and D. Sörbom, 1996, Chicago: Scientific Software International. Copyright 1996 by Scientific Software International. Reprinted with permission. Means for the first three variables are from "Testing Compound Symmetry in a Normal Multivariate Distribution," by D. F. Votaw Jr., 1948, *Annals of Mathematical Statistics, 19,* p. 471. Copyright 1948 by Institute of Mathematical Statistics. Permission to reprint was granted by the Institute of Mathematical Statistics. The mean for Typed 2 was not reported by either Jöreskog and Sörbom or Votaw; the value of 15.1234 was therefore imputed. Typed 1 = typed version of Essay 1; Written 1 = handwritten original of Essay 1; Carbon 1 = carbon copy of the handwritten original of Essay 1; Typed 2 = typed version of Essay 2.

$$\text{Model 1A: } \boldsymbol{\tau} = (\tau_1 = \tau_2 = \tau_3 = \tau_4);\quad \boldsymbol{\Lambda} = (\lambda_{11} = \lambda_{21} = \lambda_{31} = \lambda_{41});\quad \boldsymbol{\Theta} = \text{diag}(\theta_{11} = \theta_{22} = \theta_{33} = \theta_{44});$$

$$\text{Model 0C: } \boldsymbol{\tau} = (\tau_1, \tau_2, \tau_3, \tau_4);\ \boldsymbol{\Lambda} = 0;\quad \boldsymbol{\Theta} = \text{diag}(\theta_{11}, \theta_{22}, \theta_{33}, \theta_{44});$$

$$\text{Model 0A: } \boldsymbol{\tau} = (\tau_1 = \tau_2 = \tau_3 = \tau_4);\ \boldsymbol{\Lambda} = 0;\quad \boldsymbol{\Theta} = \text{diag}(\theta_{11} = \theta_{22} = \theta_{33} = \theta_{44}).$$

The substantive models we considered, from least to most restricted, were Models 2C, 1C, 1B, and 1A; these models represented, respectively, the congeneric, essentially tau-equivalent, tau-equivalent, and parallel test models because their patterns of constraints met the stipulations discussed above for such models. Given their patterns of constraint, Models 2C, 1C, 1B, and 1A had 12, 9, 6, and 3 parameter estimates, respectively. Model 0C was the standard null model, with $\boldsymbol{\Lambda}$ fixed to zero and a total of 8 parameter estimates: 4 in the $\boldsymbol{\tau}$ vector and 4 in the diagonal $\boldsymbol{\Theta}$ matrix. Model 0A, the optimal acceptable null model, had only 2 parameter estimates, invoking equality constraints on all elements in $\boldsymbol{\tau}$ and on all elements on the diagonal of $\boldsymbol{\Theta}$ because these constraints were invoked in the most restricted substantive model, Model 1A.

The chi-square values and fit statistics for each of the models described above are shown in Table 2. In the top half of Table 2, all values were calculated using the standard null model, Model 0C, and are identical to values reported by common SEM software. However, the standard null model, with 8 parameter estimates, cannot be nested within the most restricted substantive model, Model 1A, which contains equality constraints on elements in the $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ matrices that are not present in the standard null model and therefore has only 3 parameter estimates. Because the standard null model is an improper comparison model, incremental fit index values shown in

Table 2
Example 1: Fit Indices for Structural Models Fit to the Votaw (1948) Data

Model	No. est.	*df*	χ^2	χ^2/df	NCP	TLI	RFI	NFI	IFI	RNI and CFI[a]
Standard null, Model 0C										
0C: Standard null	8	6	272.49	45.42	266.49	0	0	0	0	0
2C: Congeneric	12	2	2.28	1.14	0.28	.997	.975	.992	.999	.999
1C: Ess. tau equiv.	9	5	40.42	8.08	35.42	.841	.822	.852	.868	.867
1B: Tau equivalent	6	8	44.97	5.62	36.97	.896	.876	.835	.860	.861
1A: Parallel	3	11	115.27	10.48	104.27	.787	.769	.577	.601	.609
Acceptable null, Model 0A										
0A: Acceptable null	2	12	277.83	23.15	265.83	0	0	0	0	0
2C: Congeneric	12	2	2.28	1.14	0.28	.994	.951	.992	.999	.999
1C: Ess. tau equiv.	9	5	40.42	8.08	35.42	.680	.651	.855	.870	.867
1B: Tau equivalent	6	8	44.97	5.62	36.97	.791	.757	.838	.863	.861
1A: Parallel	3	11	115.27	10.48	104.27	.572	.547	.585	.609	.608

Note. No. est. = the number of parameter estimates in the model; NCP = estimated noncentrality parameter; TLI = Tucker–Lewis Index; RFI = Bollen (1986) relative fit index; NFI = normed fit index; IFI = Bollen (1989a) incremental fit index; RNI = relative noncentrality index; CFI = comparative fit index; Ess. tau equiv. = essentially tau-equivalent test model.
[a] When RNI values fall between 0 and 1, RNI = CFI. This occurred for all models in this table; to conserve space, only a single value for the RNI and CFI is reported.

Table 3
Example 1: Indices of Difference in Fit for Alternative Structural Models Fit to Votaw (1948) Data

Model comparison	Δdf	$\Delta\chi^2$	ΔTLI	ΔRFI	ΔNFI	ΔIFI	ΔRNI and ΔCFI[a]
Standard null, Model 0C							
Model 2C vs. 1C: $\boldsymbol{\Lambda}$ invariance	3	38.14	−.156	−.153	−.140	−.131	−.132
Model 1C vs. 1B: $\boldsymbol{\tau}$ invariance	3	4.55	.055	.054	−.017	−.008	−.006
Model 1B vs. 1A: $\boldsymbol{\Theta}$ invariance	3	70.30	−.109	−.107	−.258	−.259	−.252
Acceptable null, Model 0A							
Model 2C vs. 1C: $\boldsymbol{\Lambda}$ invariance	3	38.14	−.314	−.300	−.137	−.129	−.132
Model 1C vs. 1B: $\boldsymbol{\tau}$ invariance	3	4.55	.111	.106	−.017	−.007	−.006
Model 1B vs. 1A: $\boldsymbol{\Theta}$ invariance	3	70.30	−.219	−.210	−.253	−.254	−.253

Note. The $\Delta\chi^2$, distributed as a chi-square variate with Δdf, enables a statistical test of the difference in fit of the models compared. All differences in practical fit index values were calculated to indicate the worsening of fit levels for the more restricted model. Thus, a negative value indicates that the more restricted model had worse fit relative to the less restricted model, and a positive value indicates that the more restricted model had better fit than did the less restricted model. TLI = Tucker–Lewis Index; RFI = Bollen (1986) relative fit index; NFI = normed fit index; IFI = Bollen (1989a) incremental fit index; RNI = relative noncentrality index; CFI = comparative fit index.
[a] When RNI values fall between 0 and 1, RNI = CFI, which occurred for all models. In such cases, ΔRNI = ΔCFI. To conserve space, only a single value for the ΔRNI and ΔCFI is reported.

the top half of Table 2 have no straightforward interpretation, are invalid, and should be disregarded. However, these values are presented to allow comparison with values computed using an acceptable null model.

In the bottom half of Table 2, all fit index values are based on the acceptable null model Model 0A, which is nested within all four substantive models, as shown above. The chi-square for the acceptable null model (277.83) differed little from that of the standard null model (272.49), but the acceptable null model had twice as many degrees of freedom as did the standard null model. As a result, the χ^2/df ratio for the acceptable null model (23.15) was approximately half the size of the χ^2/df ratio, for the standard null model (45.42). As predicted, this led to rather large differences in TLI and RFI values when comparing parallel fit values listed in the top half and the bottom half of Table 2. Because the chi-square values for the standard and acceptable null models were very similar, the NFI and IFI were affected only slightly. Further, because the estimated NCP for the standard null model (266.49) was virtually identical to that for the acceptable null model (265.83), RNI and CFI fit values were essentially identical regardless of choice of null model.

Differences in fit index values between competing models are shown in Table 3.[3] Three model comparisons were of most interest: (a) between the congeneric model and the essentially tau-equivalent model, (b) between the essentially tau-equivalent model and the tau-equivalent model, and (c) between the tau-equivalent model and the parallel test model. The top half of Table 3 provides values calculated using the incorrect, standard null model, whereas values calculated using the acceptable null model are shown in the bottom half of Table 3.

Values in Table 3 exhibited two clear trends. First, fit indices containing a strong correction for model complexity—the TLI and RFI—were, variably, equally sensitive (2C vs. 1C), more sensitive (1C vs. 1B), or less sensitive (1B vs. 1A) to differences in model fit than were fit indices containing no such correction for model complexity—the NFI, IFI, RNI, and CFI—but only when the improper, standard null model was used. When the acceptable null model was used, the TLI and RFI were either clearly more sensitive (2C vs. 1C and 1C vs. 1B) or approximately equally sensitive (1B vs. 1A) to differences in model fit than were the NFI, IFI, RNI, and CFI.

Second, the two indices with a strong correction for model complexity—the TLI and RFI—were greatly affected by choice of null model, having difference values under an acceptable null model approximately twice as large as those under the standard null model. In contrast, fit indices with no such correction for model complexity—the NFI, IFI, RNI, and CFI—were

[3] No formal cutoffs have been developed to denote practically important differences in incremental fit index values for competing models. Differences in fit index values are presented here to enable simple descriptive comparisons of relative model fit.

relatively unaffected by choice between the acceptable and standard null models. The TLI and RFI were greatly affected by choice of null model because the two null models, Models 0C and 0A, had similar chi-square values but rather different degrees of freedom, leading to a large difference in the χ^2/df ratios for the two models, which had predicted, sizeable effects on fit index values. Conversely, because Models 0C and 0A had similar chi-square and estimated NCP values, the NFI, IFI, RNI, and CFI were impacted very little by choice of null model, as hypothesized.

Other single-sample research situations requiring nonstandard null models. Many other single-sample research situations requiring nonstandard null models could be identified. To conserve space, we discuss here a single example, specifically a growth curve model fit to longitudinal data.[4] Assume the presence of five manifest variables, representing the scores on a single measure administered at five points in time at yearly intervals, leading to $5(5 + 3)/2 = 20$ means on and unique covariances among manifest variables. The standard null model would have $2p = 10$ parameter estimates, one each for the mean and for the variance of each manifest variable.

Substantive growth curve models usually have two latent variables: An Intercept factor reflecting general level of scores (e.g., performance at the first time of measurement), and a Growth factor representing change across time. To identify an acceptable null model more easily, one can delete the Growth factor from the model, leaving only a single latent variable, the Intercept factor. This "Intercept-only" model represents differences among individuals in their mean level on the manifest variables but allows no growth over time. Although an Intercept-only model is not often considered in empirical work, the model is a potentially reasonable model in some domains of inquiry. For example, in the study of personality during adulthood, a researcher might posit that persons exhibit substantial individual differences on a given personality attribute but that these individual differences are very stable across time. One way of specifying such a model is shown in Figure 2: (a) Following standard conventions, the unit constant is represented as a triangle, latent variables are represented as circles, manifest variables are represented as squares, directed paths (e.g., regression weights) are represented as straight, single-headed arrows, and nondirected paths (e.g., variances or covariances) are represented as curved, doubled-headed arrows; (b) all five manifest variables, V1 Time 1–V1 Time 5, load on the Intercept factor with fixed loadings of 1.0; and (c) the mean, α_1, and variance, Ψ_{11}, on the latent variable and the residual variances of the five manifest variables, θ_{11}–θ_{55}, are the parameters to be estimated in the model. Thus, the Intercept-only model has only seven parameter estimates. Because the standard null model has 10 parameter estimates, it is an unacceptable null model for this application because the standard null model cannot be nested within this more restricted Intercept-only model. If the Intercept-only model were the most restricted substantive model for the data, then an acceptable null model might have six parameter estimates: constrained equal

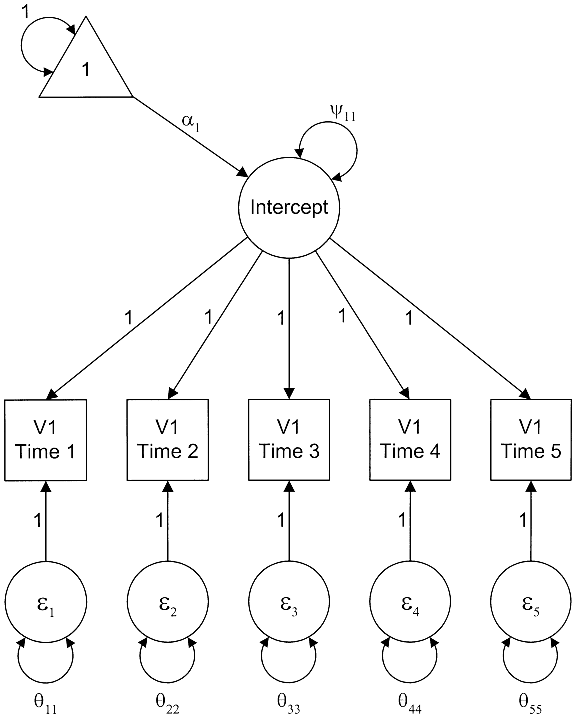

Figure 2. An Intercept-only growth model for a single manifest variable, V1, measured at five times of measurement. △ stands for unit constant, ○ for latent variable, □ for manifest variable, → for directed path, and ↔ for nondirected path. Paths with coefficients fixed at unity are shown with unit coefficients; each path with a parameter to be estimated is labeled with a symbol for the parameter estimate.

[4] This research application was suggested by one of the reviewers of a previous version of this article, and our presentation of this application borrows heavily from the reviewer's description in the review.

means for the five manifest variables and separate estimates of the residual variance of each manifest variable. This acceptable null model could be specified by fixing to zero the variance on the Intercept factor, Ψ_{11}, and leaving intact the remainder of the model specification shown in Figure 2.

Furthermore, a "constrained Intercept-only" model is a second substantive model that could be considered. This model would be specified by constraining to equality the residual variances of the manifest variables, θ_{11}–θ_{55}, consistent with homogeneity of variance assumptions in many statistical models. This constrained Intercept-only model has only three parameter estimates: α_1, Ψ_{11}, and the constrained-equal θ_{11}–θ_{55}. Clearly, neither the standard null model with 10 parameter estimates nor the preceding null model with 6 parameter estimates could be nested within this more constrained substantive model. An acceptable null model for this parameterization could be specified by fixing the variance on the Intercept factor, Ψ_{11}, to zero, leaving two parameter estimates: the mean on the Intercept factor, α_1, and the constrained-equal residual variances of the manifest variables, θ_{11}–θ_{55}. Because of the form of the latent growth model shown in Figure 2, the parameter estimate reproducing the means of manifest variables is estimated in the $\boldsymbol{\alpha}$ vector (cf. Equation 2), rather than in the $\boldsymbol{\tau}$ vector. However, the two acceptable null models described above yield nonzero estimates of the mean and variance on manifest variables and constrain covariances among manifest variables to be null or zero, satisfying key stipulations for an acceptable null model. Because an acceptable null model must be nested within the most restricted substantive model fit to the data, the nature of the most restricted substantive model would determine which of the two modified null models described above would be the acceptable null model for computing incremental fit indices.

Multiple-Group Models

Factor analysis models. When multiple-group modeling was first introduced (Jöreskog, 1971), analyses were typically based only on covariance matrices for each sample. Recent advances (e.g., Meredith, 1993) require the use of information about manifest variable means in addition to their covariances to test interesting models, such as models with varying forms of factorial invariance and the informative substantive questions associated with them. These models can be represented, by modifying Equation 2, as

$$\boldsymbol{\Sigma}_g^* = (\boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\alpha}_g)(\boldsymbol{\tau}_g + \boldsymbol{\Lambda}_g \boldsymbol{\alpha}_g)' + \boldsymbol{\Lambda}_g \boldsymbol{\Psi}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g, \quad (14)$$

where all symbols are as defined above, and the g subscript on each matrix indicates that the equation is written for the gth group ($g = 1, \ldots, G$).

Based on Thurstone (1947), Horn, McArdle, and Mason (1983), and Meredith (1993), Widaman and Reise (1997) identified four levels of factorial invariance, represented as a nested sequence of models: (a) the *configural invariance* model has the same pattern of fixed and free loadings in $\boldsymbol{\Lambda}$ in each group but no cross-group invariance constraints in any matrices and is thus the least constrained model; (b) the *weak factorial invariance* model is identical to the configural invariance model, but invokes cross-group invariance constraints on factor loadings in $\boldsymbol{\Lambda}$; (c) the *strong factorial invariance* model adds to the weak factorial invariance model cross-group invariance constraints on intercept terms in $\boldsymbol{\tau}$; and (d) the *strict factorial invariance* model adds to the strong factorial invariance model cross-group constraints on unique factor variances in $\boldsymbol{\Theta}$.

The tree structure of models shown in Figure 1 can be used to portray alternative multiple-group factor analysis models if the constraints on parameter matrices, which represented within-group invariance constraints for Example 1, now represent cross-group constraints. Thus, under Structure 0, all factor loadings in all groups are constrained to zero; under Structure 1, estimates of factor loadings are constrained to invariance across groups; and under Structure 2, factor loadings are freely estimated within each group. Similarly, under Structure A, intercept estimates in $\boldsymbol{\tau}$ and unique variances in $\boldsymbol{\Theta}$ are constrained to invariance across groups; under Structures B and B′, cross-group invariance constraints are relaxed on the $\boldsymbol{\Theta}$ and $\boldsymbol{\tau}$ matrices, respectively; and under Structure C, intercept and unique variance parameters are freely estimated within each group. Based on these stipulations, Model 2C represents the configural invariance model, Model 1C represents the weak factorial invariance model, Model 1B represents the strong factorial invariance model, and Model 1A represents the strict factorial invariance model.

The standard null model, Model 0C, is once again inappropriate for calculating either statistical difference tests or incremental fit indices for the multiple-sample models outlined above because the standard null model is not nested within the most constrained of the competing models. Using covariances and means in multiple groups, Model 0C has $2pG$ parameter estimates, or two parameter estimates (i.e., mean and variance) for each indicator within each of the G groups. However, under the strict factorial invariance model, intercepts and unique factor variances for indicators are constrained to invariance across groups. Thus, the acceptable null model should include cross-group invariance constraints on both unique factor variances and intercepts for all manifest variables. The verbal statement for this null model, with reference to Equation 14, is as follows: Each manifest variable has positive variance, and these variances in $\boldsymbol{\Theta}$ are invariant across groups; each manifest variable has an unknown mean, and these means in $\boldsymbol{\tau}$ are invariant across groups; and the manifest variables are mutually uncorrelated (i.e., all parameters in $\boldsymbol{\Lambda}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\Psi}$ are constrained to zero in all groups). Therefore, the optimal acceptable null model, Model 0A, has $2p$ parameter estimates, rather than the $2pG$ parameter estimates automatically made under the inappropriate, standard null model.

Example 2: Data from Reise, Widaman, and Pugh (*1993*). To illustrate a multiple-group model, we reanalyzed data from Reise et al., who modeled means on and covariances among five negative mood items in samples of college students from China ($N = 598$) and the US ($N = 540$). The means and variances on and correlations among the five items are shown in Table 4.

Using Equation 14, Reise et al. (1993) specified a one-factor model on which all five items loaded and justified model identification in detail. Briefly, if the subscript 1 stands for the U.S. sample and subscript 2 for the Chinese, the factor variance in $\boldsymbol{\Psi}$ was fixed to unity in the U.S. sample (i.e., $\boldsymbol{\Psi}_1 = 1.0$) and estimated in the Chinese sample (i.e., $\boldsymbol{\Psi}_2 =$ free), the first factor loading in $\boldsymbol{\Lambda}$ and the first intercept term in $\boldsymbol{\tau}$ were constrained to invariance, and the factor mean in $\boldsymbol{\alpha}$ was fixed at zero in the U.S. sample (i.e., $\boldsymbol{\alpha}_1 = 0$) and estimated in the Chinese sample (i.e., $\boldsymbol{\alpha}_2 =$ free). Six models from the tree structure in Figure 1 were fit to the data, specified in the following fashion:

Model 2C: $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2; \boldsymbol{\Lambda}_1, \boldsymbol{\Lambda}_2; \boldsymbol{\alpha}_1 = 0, \boldsymbol{\alpha}_2; \boldsymbol{\Psi}_1 = 1.0, \boldsymbol{\Psi}_2; \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2;$

Model 1C: $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2; \boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2; \boldsymbol{\alpha}_1 = 0, \boldsymbol{\alpha}_2; \boldsymbol{\Psi}_1 = 1.0, \boldsymbol{\Psi}_2; \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2;$

Model 1B: $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2; \boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2; \boldsymbol{\alpha}_1 = 0, \boldsymbol{\alpha}_2; \boldsymbol{\Psi}_1 = 1.0, \boldsymbol{\Psi}_2; \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2;$

Model 1A: $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2; \boldsymbol{\Lambda}_1 = \boldsymbol{\Lambda}_2; \boldsymbol{\alpha}_1 = 0, \boldsymbol{\alpha}_2; \boldsymbol{\Psi}_1 = 1.0, \boldsymbol{\Psi}_2; \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2;$

Model 0C: $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2; \boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2;$

Model 0A: $\boldsymbol{\tau}_1 = \boldsymbol{\tau}_2; \boldsymbol{\Theta}_1 = \boldsymbol{\Theta}_2.$

Table 4
Example 2: Correlations Among Five Manifest Variables, With Variable Means and Standard Deviations, in Two Samples, From Reise, Widaman, and Pugh (1993)

	Variable				
Variable and sample	Nervous	Worried	Jittery	Tense	Distressed
		Correlations			
Nervous	—	.282	.320	.405	.386
Worried	.577	—	.223	.269	.411
Jittery	.537	.411	—	.191	.372
Tense	.550	.533	.486	—	.448
Distressed	.471	.551	.417	.623	—
		Variable means and standard deviations			
MN					
M	2.170	2.520	2.010	2.350	2.290
SD	1.120	1.220	1.090	1.190	1.250
Nanj					
M	1.890	2.090	1.600	2.150	1.930
SD	0.930	1.110	1.010	1.040	1.120

Note. Correlations below the diagonal are for the Minnesota (MN) sample ($N = 540$), and correlations above the diagonal are for the Nanjing (Nanj) sample ($N = 598$). In the Reise et al. (1993) article, covariances among the five items were incorrectly identified as correlations. In the present table, the correlations among items are correctly reported, so all SEM results reported by Reise et al. and in Tables 5 and 6 of the present article can be reproduced from values in this table.

Moving from least to most restricted, Models 2C, 1C, 1B, and 1A represented, respectively, the configural invariance and weak, strong, and strict factorial invariance models. Given their patterns of constraint, Models 2C, 1C, 1B, and 1A had 30, 26, 22, and 17 parameter estimates, respectively. Model 0C was the standard null model, with $\mathbf{\Lambda}$, $\boldsymbol{\alpha}$, and $\mathbf{\Psi}$ fixed to zero in both groups and a total of 20 parameter estimates: 5 in the $\boldsymbol{\tau}$ vector and 5 in the diagonal $\mathbf{\Theta}$ matrix in each group. Model 0A, the optimal acceptable null model, had only 10 parameter estimates, invoking cross-group equality constraints on elements in $\boldsymbol{\tau}$ and elements on the diagonal of $\mathbf{\Theta}$ because these constraints were invoked in Model 1A.

The chi-square values and fit statistics for each of the models described above are shown in Table 5. In the top half of Table 5, all values were calculated using the standard null model, Model 0C. The most restricted model, Model 1A, had fewer parameter estimates than the standard null model and also contained cross-group equality constraints on elements in the $\boldsymbol{\tau}$ and $\mathbf{\Theta}$ matrices that were relaxed in the standard null model. Thus, Model 0C was an improper comparison model because it could not be nested within Model 1A. As a result, the incremental fit index values shown in the top half of Table 5 have no straightforward interpretation, are invalid, and should be disregarded but are presented to allow comparison with values computed using an acceptable null model.

In the bottom half of Table 5, all fit index values were based on the acceptable null model, Model 0A, which was nested within all four substantive models. Both the chi-square and estimated NCP for Model 0A were about 11% higher than those for Model 0C, so NFI, IFI, RNI, and CFI values were all higher under the acceptable null model than under the standard null model, as predicted. In contrast, the χ^2/df ratio for Model 0A (57.81) was smaller than that for Model 0C (77.59). As a result, TLI and RFI values tended to be lower using the acceptable null model. Indeed, all eight TLI and RFI values for Models 2C through 1A fell above the frequently used critical cutoff of .90 under the incorrect standard null model, but fully half of these values failed to reach the .90 cutoff when the acceptable null model was used. Thus, relying on the TLI and RFI, use of the improper standard null model implies that one can choose among several models (2C–1A), each of which had adequate fit, whereas use of the acceptable null model signals that several of the substantive models fail to demonstrate minimally adequate fit to the data.

The differences in incremental fit index values between models are shown in Table 6, with differences based on the improper standard null model shown in the top half of the table and those based on the acceptable null model shown in the bottom of the table (cf. Footnote 3). Because the fit index values for the NFI, IFI, RNI, and CFI were slightly higher under the acceptable null model than under the standard null model, the differences in fit for competing models were either unchanged or slightly smaller under the acceptable null model. The opposite pattern held for

Table 5
Example 2: Fit Indices for Structural Models Fit to the Reise, Widaman, and Pugh (1993) Data

Model	No. est.	*df*	χ^2	χ^2/df	NCP	TLI	RFI	NFI	IFI	RNI and CFI[a]
				Standard null, Model 0C						
0C: Standard null	20	20	1,551.82	77.59	1,531.82	0	0	0	0	0
2C: Configural inv.	30	10	75.38	7.53	65.38	.915	.903	.951	.958	.957
1C: $\mathbf{\Lambda}$ inv.	26	14	89.99	6.42	75.99	.929	.917	.942	.951	.950
1B: $\mathbf{\Lambda}$ & $\boldsymbol{\tau}$ inv.	22	18	111.24	6.18	93.24	.932	.920	.928	.939	.939
1A: $\mathbf{\Lambda}$, $\boldsymbol{\tau}$, & $\mathbf{\Theta}$ inv.	17	23	123.12	5.35	100.12	.943	.931	.921	.934	.935
				Acceptable null, Model 0A						
0A: Acceptable null	10	30	1,734.40	57.81	1,704.40	0	0	0	0	0
2C: Configural inv.	30	10	75.38	7.53	65.38	.885	.870	.957	.962	.961
1C: $\mathbf{\Lambda}$ inv.	26	14	89.99	6.42	75.99	.904	.889	.948	.956	.955
1B: $\mathbf{\Lambda}$ & $\boldsymbol{\tau}$ inv.	22	18	111.24	6.18	93.24	.909	.893	.936	.946	.945
1A: $\mathbf{\Lambda}$, $\boldsymbol{\tau}$, & $\mathbf{\Theta}$ inv.	17	23	123.12	5.35	100.12	.923	.907	.929	.942	.941

Note. No. est. = the number of parameter estimates in the model; NCP = estimated noncentrality parameter; TLI = Tucker–Lewis Index; RFI = Bollen (1986) relative fit index; NFI = normed fit index; IFI = Bollen (1989a) incremental fit index; RNI = relative noncentrality index; CFI = comparative fit index; inv. = invariance.
[a] When RNI values fall between 0 and 1, RNI = CFI. This occurred for all models in this table; to conserve space, only a single value for the RNI and CFI is reported.

Table 6
Example 2: Indices of Difference in Fit For Alternative Structural Models Fit to the Reise, Widaman, and Pugh (1993) Data

Model comparison	Δdf	$\Delta\chi^2$	ΔTLI	ΔRFI	ΔNFI	ΔIFI	ΔRNI and ΔCFI[a]
Standard null, Model 0C							
Model 2C vs. 1C: $\mathbf{\Lambda}$ invariance	4	38.00	.014	.014	−.009	−.007	−.007
Model 1C vs. 1B: $\boldsymbol{\tau}$ invariance	4	21.25	.003	.003	−.014	−.012	−.011
Model 1B vs. 1A: $\mathbf{\Theta}$ invariance	5	11.88	.011	.011	−.007	−.005	−.004
Acceptable null, Model 0A							
Model 2C vs. 1C: $\mathbf{\Lambda}$ invariance	4	38.00	.019	.019	−.009	−.006	−.006
Model 1C vs. 1B: $\boldsymbol{\tau}$ invariance	4	21.25	.005	.004	−.012	−.010	−.010
Model 1B vs. 1A: $\mathbf{\Theta}$ invariance	5	11.88	.014	.014	−.007	−.004	−.004

Note. The $\Delta\chi^2$, distributed as a chi-square variate with Δdf, enables a statistical test of the difference in fit of the models compared. All differences in practical fit index values were calculated to indicate the worsening of fit levels for the more restricted model. Thus, a negative value indicates that the more restricted model had worse fit relative to the less restricted model, and a positive value indicates that the more restricted model had better fit than did the less restricted model. TLI = Tucker–Lewis Index; RFI = Bollen (1986) relative fit index; NFI = normed fit index; IFI = Bollen (1989a) incremental fit index; RNI = relative noncentrality index; CFI = comparative fit index.
[a] When RNI values fall between 0 and 1, RNI = CFI, which occurred for all models. In such cases, ΔRNI = ΔCFI. To conserve space, only a single value for the ΔRNI and ΔCFI is reported.

the TLI and RFI. For these fit indices, use of an acceptable null model led to somewhat lower overall levels of fit and therefore to modestly larger differences between competing models. Despite these changes, most differences shown in Table 6 were affected little by choice of null model.

How Often Is a More Restricted Null Model Needed?

Assuming we have made the case that a more restricted null model is needed in certain research situations, one may wonder how often these situations arise in practice. To answer this question, we reviewed all studies published during the year 2000 in three selected American Psychological Association (APA) journals to determine whether the issues raised in this article would affect calculation of fit index values. A total of 28 articles in the *Journal of Personality and Social Psychology* used SEM; of these, 11 used multiple-sample analyses, 8 presented analyses of longitudinal data, and due to overlap, 13 of 28, or 46%, used multiple-sample and/or longitudinal analyses. In *Developmental Psychology,* only 4 articles used SEM; 1 of these articles, or 25%, presented analyses of longitudinal data. A total of 7 articles published in *Psychological Methods* used SEM; of these, 4 analyzed longitudinal data (one in a multiple-group context), and 1 other article concerned psychometric models. So, 5 of the 7 SEM articles published in *Psychological Methods,* or 71%, contained analyses that would be influenced by the need for an acceptable null model. Admittedly, some articles identified as potentially influenced by our considerations did not report incremental fit index values, and others did not use detailed constraints on means and/or residual variances in substantive models. Thus, these articles might not, strictly speaking, require a revised null model that is a restricted version of the standard null model. However, as researchers become more aware of the theoretical advantages of certain kinds of models (e.g., growth curve models, multiple-group models), particularly advantages associated with highly constrained models (e.g., strict factorial invariance), the need for more restricted null models will only increase. At present, given our review of articles published during the year 2000, at least one-third to one-half of the articles using SEM that are published in APA journals are likely to be affected by the choice of an acceptable null model if maximally informative, highly constrained substantive models are used.

Discussion

Incremental fit indices are often reported in empirical articles that present results analyzed using SEM. Indeed, incremental fit indices are often given at least as much weight as, or even more weight than, the chi-square index of fit when evaluating the fit of a structural model to data. This increased reliance on incremental fit indices is justified on at least two grounds. First, the chi-square index has increasing sensitivity to minor levels of model misfit as sample size increases, supporting the rejection of models that provide close, but not perfect, fit to the data (Browne & Cudeck, 1993). Second, Monte Carlo investigations have demonstrated that incremental fit indices exhibit strong properties, including relative independence from sample size, supporting their use as indices of practical fit of models to data (e.g., Hu & Bentler, 1998).

The major aims of this article were to review critically the nature of the null model that is a crucial component in the calculation of incremental fit indices and to investigate the consequences of using an acceptably specified null model. For their calculation, incremental fit indices require a null model to be evaluated, so acceptable specification of the null model is an important, yet often disregarded, issue. Our consideration of the nature of the null model led us to conclude that research situations in which the standard null model is an inappropriate comparison model are not uncommon. One incontrovertible characteristic of an acceptable null model is that it be nested within all competing substantive models, as only then will incremental fit indices be interpretable as reflecting improvement in fit over that of a null model. However, the standard null model is not nested within common psychometric models, such as models for parallel or tau-equivalent tests, nor is it nested within many state-of-the-art multiple-sample factor analysis models. Moreover, we discussed in detail only two problematic situations; additional examples could easily be generated. If the standard null model is not nested within competing substantive models, incremental fit indices computed using the standard null model will be incorrect, have no useful interpretation, and should be disregarded. In such situations, researchers must specify and estimate an acceptable null model for the data and then use the fit statistics (e.g., chi-square and degrees of freedom) for this null model to calculate values for incremental fit indices. (See the Appendix, which is available in the online version of this article in PsycARTICLES, for SAS code to calculate incremental fit indices.)

Fortunately, determining that a null model is nested within competing substantive models is often fairly straightforward. For analyses based on covariance matrices, an acceptable null model usually contains parameter estimates only in the $\boldsymbol{\Theta}$ matrix; if both means and covariances are used, an acceptable null model will typically contain parameter estimates only in the $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ matrices. In either case, the researcher must verify that the most restrictive within-group and/or between-group constraints on means and variances made in any competing substantive model are also invoked when specifying the null model. If this is done and the null model represents manifest variables as mutually uncorrelated, the resulting model will be an acceptable null model.

The tree structure in Figure 1 highlights the presence of multiple potential null models (e.g. all four models in the first column of the tree structure). Depending on the most restricted substantive model considered for a set of data, a single, optimal null model should be selected. In our opinion, the optimal acceptable null model is the least restricted null model that is nested within the most restricted substantive model considered for the data. Using this optimal null model to compute fit indices for all competing substantive models will ensure that the 0–1 continuum on which incremental fit index values fall is defined in a uniform, globally consistent fashion.

Figure 1 also illustrates nesting relations based on patterns of constraint among a common subset of structural equation models. Of course, the tree structure in Figure 1 will not be adequate for all research situations. Indeed, formulating a simple, understandable tree structure may be difficult for rather complex models, such as multiple-group models for longitudinal data with parallel test constraints for some indicators. However, to the extent that it is possible, arraying models in a tree structure is often a helpful guide for ensuring that all necessary, and no unnecessary, within-group and/or between-group constraints in $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ are placed in the null model.

In some research applications, nesting relations among models will be impossible to derive. For example, neither Model 2A nor Model 1B in Figure 1 is nested within the other, so computing differences between these models in their levels of statistical or practical fit is not justified. However, if Model 0A were used as the acceptable null model, then incremental fit index values for the two models would fall on a comparable metric, enabling comparisons between models with regard to their fit to the data. As another example, consider the analysis of multitrait–multimethod data. One set of nested models, from the taxonomy developed by Widaman (1985), could be fit using the linear factor analysis model, and a different set of nested models could be fit using the multiplicative model proposed by Browne (1984). Here, the linear factor analysis models might constitute one major branch of a tree structure, and the multiplicative models would constitute a second major branch. Although nesting relations could be derived for alternative models within each major branch, establishing nesting relations among models from separate branches might be impossible. If both of these major branches were formulated as emanating from a common, shared acceptable null model, then incremental fit index values would be on a comparable metric and could be used to evaluate relative fit of models across the two major branches, even if models had no nesting relations.

In situations in which the standard null model is an inappropriate comparison model, an acceptable null model will be more restrictive, requiring within-group and/or between-group constraints on elements in the $\boldsymbol{\tau}$ and $\boldsymbol{\Theta}$ matrices. As a result, an acceptable null model will have fewer parameter estimates and therefore more degrees of freedom than the standard null model. Given this, we highlighted the effects on fit index values if the overall chi-square, the χ^2/df ratio, or the estimated NCP for an acceptable null model differed substantially from comparable values for the standard null model. In some research contexts (e.g., our Example 2), changes in fit index values arising from use of an acceptable null model may be rather small; in other research contexts (e.g., our Example 1), changes may be pronounced. At least two related issues deserve attention. First, changes in absolute levels of incremental fit index values under an acceptable null model might necessitate revision of recommendations concerning fit index cutoffs that reflect adequate model fit (e.g., .90 vs. .95; Hu & Bentler, 1999). Second, recommendations might be different for fit indices with a strong correction for model complexity (i.e., the TLI and RFI) than for fit indices with no such correction (i.e., the NFI, IFI, RNI, and CFI) because these two classes of fit indices behaved rather differently under altered specification of the null model. Regardless of the magnitude of changes accompanying the use of an acceptable null model, incremental fit index values have an interpretation only if an acceptable null model is used in calculations. Furthermore, the quality of research using SEM and the quality of thinking about structural models will only increase if researchers are held responsible for the specification of an acceptable null model as well as for specification of each competing substantive model considered.

Our discussion of psychometric and multisample analyses demonstrated that the user of SEM software must be aware of the nature of the null model used in automatic calculation of incremental fit indices and often must beware of the values automatically printed out by SEM software. The researcher should not presume that SEM software will always use an acceptable null model when calculating incremental fit indices. SEM programs attempt to provide user-friendly procedures that allow easy fitting of standard models to data. However, developers of SEM software cannot be expected to anticipate every research context or to ensure that their programs can adjust automatically and appropriately to every one of the complex situations in which these models are applied. Several modifications to current software might be considered. For example, software developers could follow the lead of the Mx program (Neale, 1999). Mx is perhaps unique in requiring the user to supply the chi-square and degrees of freedom of the null model; if these are not supplied by the user, the Mx program will not provide values for any indices of practical fit. Alternatively, programs could be written to be sensitive to information provided by the user, allowing the user either to specify the form of an acceptable null model or to supply the chi-square and degrees of freedom of an acceptable null model. If the user supplied either of these types of information, the program could use the altered specification of the null model or its fit statistics when computing indices of practical fit, overriding the default use of the standard null model. Any such changes would place explicit responsibility for specifying an acceptable null model where it rightfully belongs—in the hands of the researcher who specifies the competing substantive models for the data.

The central message of our article is that the investigator must verify whether the standard null model is appropriate for a particular research application. If the standard null model is appropriate, as it will be for many single-group analyses, the fit index values automatically reported by SEM software will be accurate. However, if the standard null model is not an appropriate comparison model, then incremental fit index values reported automatically will be incorrect, and an acceptable null model must be fit to the data to enable the researcher to calculate incremental fit index values correctly. We trust we have sensitized researchers to the importance of this issue, provided tools for determining the form of an acceptable null model, and described clearly how a researcher must proceed if an acceptable null model for a set of data does not correspond to the standard null model.

References

Arbuckle, J. L., & Wothke, W. (1999). *AMOS 4.0 user's guide* [Computer software manual]. Chicago: Smallwaters.

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

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

Bentler, P. M., & Wu, E. J. C. (1995). *EQS for Windows user's guide* [Computer software manual]. Encino, CA: Multivariate Software.

Bollen, K. A. (1986). Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika, 51,* 375–377.

Bollen, K. A. (1989a). A new incremental fit index for general structural equation models. *Sociological Methods and Research, 17,* 303–316.

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

Browne, M. W. (1984). The decomposition of multitrait–multimethod matrices. *British Journal of Mathematical and Statistical Psychology, 37,* 1–21.

Browne, M. W., & Arminger, G. (1995). Specification and estimation of mean- and covariance-structure models. In G. Arminger, C. C. Clogg, & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 185–249). New York: Plenum Press.

Browne, 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.

Browne, M. W., & Mels, G. (1990). *RAMONA user's guide* [Computer software manual]. Unpublished report, Columbus, Ohio State University.

Byrne, B. M. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic concepts, applications, and programming.* Mahwah, NJ: Erlbaum.

Cudeck, R., & Henly, S. (1991). Model selection in covariance structures analysis and the "problem" of sample size: A clarification. *Psychological Bulletin, 109,* 512–519.

Horn, J. L., McArdle, J. J., & Mason, R. (1983). When is invariance not invariant: A practical scientist's look at the ethereal concept of factor invariance. *Southern Psychologist, 1,* 179–188.

Hu, L.-T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3,* 424–453.

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika, 36,* 409–426.

Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology: Measurement, psychophysics, and neural information processing* (Vol. 2, pp. 1–56). San Francisco: Freeman.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1996). *LISREL 8: User's reference guide* [Computer software manual]. Chicago: Scientific Software International.

Marsh, H. W. (1998). The equal correlation baseline model: Comment and constructive alternatives. *Structural Equation Modeling, 5,* 78–86.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification, 6,* 97–103.

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin, 107,* 247–255.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543.

Neale, M. C. (1999). *Mx: Statistical modeling* (5th ed.) [Computer software manual]. Richmond: Virginia Commonwealth University, Department of Psychiatry.

Reise, S. P., Widaman, K. F., & Pugh, R. H. (1993). Confirmatory factor analysis and item response theory: Two approaches for exploring measurement invariance. *Psychological Bulletin, 114,* 552–566.

Rigdon, E. E. (1998a). The equal correlation baseline model for comparative fit assessment in structural equation modeling. *Structural Equation Modeling, 5,* 63–77.

Rigdon, E. E. (1998b). The equal correlation baseline model: A reply to Marsh. *Structural Equation Modeling, 5,* 87–94.

Sobel, M. E., & Bohrnstedt, G. W. (1985). Use of null models in evaluating the fit of covariance structure models. In N. B. Tuma (Ed.), *Sociological methodology 1985* (pp. 152–178). San Francisco: Jossey-Bass/Pfeiffer.

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

Thurstone, L. L. (1947). *Multiple factor analysis.* Chicago: University of Chicago Press.

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10.

Votaw, D. F., Jr. (1948). Testing compound symmetry in a normal multivariate distribution. *Annals of Mathematical Statistics, 19,* 447–473.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9,* 1–26.

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.

Received January 11, 2001
Revision received September 3, 2002
Accepted November 1, 2002 ■

Appendix

SAS Code to Calculate Incremental Fit Indices

Illustrated Using Data From Votaw (1948)

```
options ls=96 pageno=1;
data orig;
input model $ 1-8 snullchi snulldf anullchi anulldf Mkchi Mkdf ;
*********************************************************************************;
*  Results from SEM analyses of a given set of data are entered into an SAS data ;
*  set in a fixed order for each model considered (see below following the       ;
*  “cards” statement). We illustrate this using analyses of the data from Votaw  ;
*  (1948), using information from Table 2 of Widaman & Thompson (2003)           ;
*  For example, for Model 1A, enter the following:                               ;
*     Model1A (model name or identifier, using no more than 8 spaces)            ;
*     272.49  (chi-square for the standard null model, or snullchi)              ;
*     6       (degrees of freedom for the standard null model, or snulldf)       ;
*     277.83  (chi-square for the acceptable null model, or anullchi)            ;
*     12      (degrees of freedom for the acceptable null model, or anulldf)     ;
*     115.27  (chi-square for the current substantive model, or Mkchi)           ;
*     11      (degrees of freedom for the current substantive model, or Mkdf)    ;
*********************************************************************************;
cards;
Model1A   272.49  6  277.83  12  115.27  11
Model1B   272.49  6  277.83  12   44.97   8
Model1C   272.49  6  277.83  12   40.42   5
Model2C   272.49  6  277.83  12    2.28   2
```

```
;
data first;  set orig;                          **************************************;
  NCPk = Mkchi - Mkdf;                          * estimated NCP for Model k           ;
  NCPs = snullchi - snulldf;                    * estimated NCP for STD null model    ;
  Q0   = snullchi / snulldf;                    * STD null model: ratio of chi-sq /df ;
  Qk   = Mkchi / Mkdf;                          * Substan. model: ratio of chi-sq /df ;
  TLI  = (Q0 - Qk)/(Q0 - 1);                    * Tucker-Lewis Index (TLI), or NNFI   ;
  RFI  = (Q0 - Qk)/(Q0);                        * Relative Fit Index (RFI)            ;
  NFI  = (snullchi - Mkchi)/(snullchi);         * Normed Fit Index (NFI)              ;
  IFI  = (snullchi - Mkchi)/(snullchi - Mkdf); * Incremental Fit Index (IFI)         ;
  RNI  = 1 - (NCPk / NCPs);                     * Relative Noncentrality Index (RNI)  ;
  CFI  = 1 - (max(NCPk,0)/max(NCPk,NCPs,0));   * Comparative Fit Index (CFI)         ;
                                                **************************************;
proc print uniform;  var model Mkchi Mkdf NCPk NCPs Q0 Qk;
title4  'VOTAW (1948) data:  Fit Indices based on STANDARD NULL MODEL';
proc print uniform;  var model TLI RFI NFI IFI RNI CFI;

data second;  set orig;                         **************************************;
  NCPk = Mkchi - Mkdf;                          * estimated NCP for Model k           ;
  NCPa = anullchi - anulldf;                    * estimated NCP for ACC null model    ;
  Q0   = anullchi / anulldf;                    * ACC null model: ratio of chi-sq /df ;
  Qk   = Mkchi / Mkdf;                          * Substan. model: ratio of chi-sq /df ;
  TLI  = (Q0 - Qk)/(Q0 - 1);                    * Tucker-Lewis Index (TLI), or NNFI   ;
  RFI  = (Q0 - Qk)/(Q0);                        * Relative Fit Index (RFI)            ;
  NFI  = (anullchi - Mkchi)/(anullchi);         * Normed Fit Index (NFI)              ;
  IFI  = (anullchi - Mkchi)/(anullchi - Mkdf); * Incremental Fit Index (IFI)         ;
  RNI  = 1 - (NCPk / NCPa);                     * Relative Noncentrality Index (RNI)  ;
  CFI  = 1 - (max(NCPk,0)/max(NCPk,NCPa,0));   * Comparative Fit Index (CFI)         ;
```

```
                                                *************************************;
proc print uniform; var model Mkchi Mkdf NCPk NCPa Q0 Qk;
title4  'VOTAW (1948) data:  Fit Indices based on AN ACCEPTABLE NULL MODEL';
proc print uniform; var model TLI RFI NFI IFI RNI CFI;
run;
quit;
```