Title
Comparative Fit Indices in Structural Models

Author
Peter Bentler

Publication Date
2011-10-25

Comparative Fit Indexes in Structural Models

P. M. Bentler
University of California, Los Angeles

Normed and nonnormed fit indexes are frequently used as adjuncts to chi-square statistics for evaluating the fit of a structural model. A drawback of existing indexes is that they estimate no known population parameters. A new coefficient is proposed to summarize the relative reduction in the noncentrality parameters of two nested models. Two estimators of the coefficient yield new normed (CFI) and nonnormed (FI) fit indexes. CFI avoids the underestimation of fit often noted in small samples for Bentler and Bonett's (1980) normed fit index (NFI). FI is a linear function of Bentler and Bonett's non-normed fit index (NNFI) that avoids the extreme underestimation and overestimation often found in NNFI. Asymptotically, CFI, FI, NFI, and a new index developed by Bollen are equivalent measures of comparative fit, whereas NNFI measures relative fit by comparing noncentrality per degree of freedom. All of the indexes are generalized to permit use of Wald and Lagrange multiplier statistics. An example illustrates the behavior of these indexes under conditions of correct specification and misspecification. The new fit indexes perform very well at all sample sizes.

As is well known, the goodness-of-fit test statistic T used in evaluating the adequacy of a structural model is typically referred to the chi-square distribution to determine acceptance or rejection of a specific null hypothesis, $\Sigma = \Sigma(\theta)$. In the context of covariance structure analysis, Σ is the population covariance matrix and θ is a vector of more basic parameters, for example, the factor loadings and intercorrelations and unique variances in a confirmatory factor analysis. The statistic T reflects the closeness of $\hat{\Sigma} = \Sigma(\hat{\theta})$, based on the estimator $\hat{\theta}$, to the sample matrix S, the sample covariance matrix in covariance structure analysis, in the chi-square metric. Acceptance or rejection of the null hypothesis via a test based on T may be inappropriate or incomplete in model evaluation for several reasons:

1. Some basic assumptions underlying T may be false and the distribution of the statistic may not be robust to violation of these assumptions.
2. No specific model $\Sigma(\theta)$ may be assumed to exist in the population, and T is intended to provide a summary regarding closeness of $\hat{\Sigma}$ to S, but not necessarily a test of $\Sigma = \Sigma(\theta)$.
3. In small samples, T may not be chi-square distributed; hence, the probability values used to evaluate the null hypothesis may not be correct.
4. In large samples, any a priori hypothesis $\Sigma = \Sigma(\theta)$, although only trivially false, may be rejected.

As a consequence, the statistic T may not be clearly interpretable, and transformations of T designed to map it into a more interpretable 0–1, or approximate 0–1, range have been developed. Those indexes are usually called goodness-of-fit indexes (e.g., Bentler, 1983, p. 507; Jöreskog & Sörbom, 1984, p. I.40). A related class of indexes, here called comparative goodness-of-fit indexes, assess T in relation to the fit of a more restrictive model. These comparative fit indexes, formalized by Bentler and Bonett (1980), are very widely used (Bentler & Bonett, 1987) and are the sole object of this article. Alternative approaches to evaluating model adequacy are reviewed elsewhere (e.g., Bollen & Liang, 1988; Bozdogan, 1987; LaDu & Tanaka, in press; Wheaton, 1987). Although covariance structure analysis is emphasized, the methods developed here hold for any type of structural model, including, for example, mean–covariance structures and log-linear models.

Although more than 30 fit indexes have been reported and their empirical behavior studied (Marsh, Balla, & McDonald, 1988), and although new ones continue to be developed (Bollen, 1989), it is surprising to note that they have been developed as purely descriptive statistics. Apparently, no population parameter has been defined that is being estimated by any of the existing indexes. In this article, I define an explicit population comparative fit coefficient, provide two alternative estimators of the coefficient, and investigate the asymptotic relations between the new and previously defined comparative fit indexes. Furthermore, new indexes based on Wald and Lagrange multiplier statistics are developed.

Nested Models and Comparative Fit

In evaluating comparative model fit, it is helpful to focus on more than one pair of models. Consider a series of nested models,

This research was supported in part by United States Public Health Service Grants DA01070 and DA00017 and is based on a February 1988 technical report and a paper presented at the Psychometric Society meetings, June 1988, Los Angeles.

Helpful discussions with J. de Leeuw, R. I. Jennrich, T. A. B. Snijders, and J. A. Woodward; the computer assistance of Shinn-Tzong Wu; and the production assistance of Julie Speckart are gratefully acknowledged.

Correspondence concerning this article should be addressed to P. M. Bentler, Department of Psychology, University of California, Los Angeles, California 90024-1563.

Psychological Bulletin, 1990, Vol. 107, No. 2, 238–246

$$M_i, \ldots, M_j, \ldots, M_k, \ldots, M_s, \tag{1}$$

beginning with the most restricted model M_i that one might consider and extending to the least restricted or saturated model M_s. The models are assumed to be nested so that a more restricted model is obtained by imposing constraints on a more general model. For example, M_j may be obtained from M_k by fixing a free parameter in M_k to some a priori value. That is, $M_i \subseteq M_j \subseteq M_k \subseteq M_s$. In covariance structure analysis, M_i is typically the baseline model corresponding to uncorrelated measured variables, or a model of modified uncorrelatedness that allows some independent variables to have known nonzero covariances. M_i is sometimes called a null model, indicating no mutual influences among variables. If the measured variables that generate Σ are multivariate normally distributed, then M_i is the independence or modified independence model. M_i is not necessarily the most restricted model that can be considered, but it is intended to be the most restricted one that would reasonably be considered in practice. Thus, a model containing no free parameters would be still more restrictive than the independence model, but such a model would almost never describe data and is thus not considered seriously. At times it may also make sense to have M_i be a more general model than the uncorrelatedness model (Sobel & Bohrnstedt, 1985). At the other end of the continuum, M_s is the saturated model in which there are as many parameters in θ as there are nonredundant elements in Σ. In M_s there is no falsifiable structural hypothesis.

Corresponding to the sequence of nested models (Equation 1) is a sequence of goodness-of-fit test statistics,

$$T_i, \ldots, T_j, \ldots, T_k, \ldots, T_s, \tag{2}$$

and corresponding degrees of freedom, $d_i, \ldots, d_j, \ldots, d_k, \ldots, d_s$, obtained by optimizing a specific statistical fitting function such as maximum likelihood or generalized least squares using a set of data S and the models (Equation 1). Thus, T_i is the chi-square value based on d_i degrees of freedom obtained by fitting model M_i to S; T_j and d_j are the corresponding values obtained for model M_j; T_k and d_k correspond to M_k; and T_s and d_s correspond to M_s. The saturated model M_s, not necessarily unique, has the characteristic that $T_s = 0$ and $d_s = 0$. Typically, $T = NF$, where N is the sample size (or sample size minus 1) and F is the minimum of some discrepancy function. When alternative models are compared with the same discrepancy function, $T_i \geq T_j \geq T_k \geq T_s = 0$, indicating that the independence model has the worst fit, intermediate models have intermediate degrees of fit, and the saturated model has a perfect fit. Similarly, $d_i > d_j > d_k > d_s$. Corresponding to the models (Equation 1) and test statistics (Equation 2) are the parameter vectors $\theta_i, \ldots, \theta_j, \ldots, \theta_k, \ldots, \theta_s$ and the corresponding model matrices $\Sigma_i = \Sigma(\theta_i), \ldots, \Sigma_j, \ldots, \Sigma_k, \ldots, \Sigma_s$, as well as their estimated values $\hat{\Sigma}_i, \ldots, \hat{\Sigma}_j, \ldots, \hat{\Sigma}_k, \ldots, \hat{\Sigma}_s$. In covariance structure analysis, under the model of uncorrelated variables, generally $\hat{\Sigma}_i = \text{diag}(S)$ and no covariances are accounted for by the model. Intermediate model matrices $\hat{\Sigma}_i$ and $\hat{\Sigma}_j$ account for the off-diagonal elements of S, the covariances, to varying degrees. At the other extreme, $\hat{\Sigma}_s = S$, so that the model perfectly reflects the data.

Comparative fit indexes evaluate the adequacy of a particular model M_k in relation to the endpoint models M_i and M_s on the continuum (Equation 1) of models. In practice this is done by evaluating where T_k falls in relation to T_i and T_s. If T_k is close to T_i, M_k is hardly an improvement over M_i and the fit index is close to 0. If T_k is close to T_s, M_k is almost as good as the saturated model, which corresponds to the data, and the fit index is close to 1. Different fitting functions will, of course, yield somewhat different values of a fit index (LaDu & Tanaka, in press).

In the next section, the comparative fit indexes of Bentler and Bonett (1980) and Bollen (1989) are reviewed, and their limitations are noted. A population fit index designed to overcome these limitations is defined in the subsequent section, and some estimators of the index are developed. Relations among the existing and new indexes are also developed. The various indexes are extended in the following section to include information from Wald and Lagrange multiplier statistics. A sampling study illustrates the behavior of these indexes. Some concluding comments are then offered.

Normed and Non-Normed Fit Indexes

Bentler and Bonett (1980) proposed to evaluate model M_k by comparing T_k with T_i via

$$NFI = \frac{T_i - T_k}{T_i}, \tag{3}$$

which equals 0 when $T_i = T_k$, equals 1.0 when $T_k = 0$, and is in the 0–1 range otherwise, with higher values indicating better fit. Because of the 0–1 range, this index was called the normed fit index. James, Mulaik, and Brett (1982, p. 155) suggested multiplying Equation 3 by d_k/d_i to yield an index to reflect model parsimony. The issue of parsimony, or degrees of freedom used, is not addressed in this analysis (see Bentler & Mooijaart, 1989; Mulaik et al., 1989). An important characteristic of NFI is that the index is additive for nested model comparisons. Thus, if one defines the incremental normed fit index comparing models M_j and M_k as

$$NFI(jk) = \frac{T_j - T_k}{T_i}, \tag{4}$$

it will be obvious that NFI for model M_k is the additive sum of the component fits. For example, $NFI = NFI(ij) + NFI(jk)$. This characteristic permits isolating relative sources of fit or of model misspecification.

A disadvantage of NFI is that it is affected by sample size (Bearden, Sharma, & Teel, 1982). It may not reach 1.0 even when the model is correct, especially in smaller samples. This can occur because the expected value of T_k may be greater than 0, for example, when T_k is a $\chi^2(d_k)$ variate, $E(T_k) = d_k$. This difficulty with range was resolved by the modified index

$$NNFI = \frac{T_i - d_i d_k^{-1} T_k}{T_i - d_i}, \tag{5}$$

called the non-normed fit index. Bentler and Bonett (1980) built this index on one developed by Tucker and Lewis (1973) for evaluating the fit of exploratory factor analysis models estimated by maximum likelihood. The degrees of freedom adjustment in the index was designed to improve its performance near 1.0, not necessarily to permit the index to reflect other model features such as parsimony. When $T_k = E(T_k) = d_k$, the $NNFI$ =

1.0, thus obviating a major difficulty with NFI. However, $NNFI$ can fall outside the 0–1 range. It will be negative when $d_i d_k^{-1} T_k > T_i$, as usually $T_i \gg d_i$. It will exceed 1.0 when $T_k < d_k$. In fact, the index can be anomalously small, especially in small samples, implying a terrible fit when other indexes suggest an acceptable model fit (Anderson & Gerbing, 1984). As a consequence, the variance of $NNFI$ is, in sampling studies, substantially larger than the variance of NFI. This is a negative feature. The comparable incremental fit index,

$$NNFI(jk) = \frac{d_i d_j^{-1} T_j - d_j d_k^{-1} T_k}{T_i - d_i}, \qquad (6)$$

shares the advantages and disadvantages of the basic index.

The nonnormed fit index has the major advantage of reflecting model fit very well at all sample sizes (Anderson & Gerbing, 1984; Marsh et al., 1988; Wheaton, 1987). It would be desirable to modify this index so as to maintain its desirable feature while minimizing its undesirable features. A modification relating to sample size was proposed by Bollen (1986), but it did not solve the major problem of variability in the index. This problem was addressed by Bollen (1989). He defined the incremental fit index as

$$IFI = \frac{T_i - T_k}{T_i - d_k}, \qquad (7)$$

and showed that it behaved like $NNFI$ in a sampling study but had a smaller sampling variance.

Unfortunately, population parameters corresponding to the indexes that have been described have not been given, so it is not clear what quantity or quantities they are estimating. Let me first define a population fit index and two estimators of it, and then return to these indexes.

Fit Indexes and Noncentrality

Suppose that the distribution of each of the test statistics T given in Equation 2 can be approximated in large samples by the noncentral chi-square distribution with given degrees of freedom. This is a reasonable assumption for the true model and for small model misspecifications; that is, if systematic errors due to discrepancy between the true population covariance matrix, say Σ^o, and the population model matrix, say $\Sigma(\theta^o)$, are not large relative to the sampling errors in the matrix S (see, e.g., Satorra, 1989). If the mean or variance of the distribution of T substantially differs from the corresponding reference noncentral chi-square distribution, T can be scaled or adjusted to more closely achieve this result (Satorra & Bentler, 1988). Thus, the reference distribution for T_k is the noncentral $\chi^2(d_k)$ distribution with parameter λ_k, known as the noncentrality parameter. Asymptotically, $\lambda_k = T_k^o = NF_k^o$, where T_k^o is the value of T_k obtained when Σ^o substitutes for S in the discrepancy function F used, and F_k^o is the corresponding minimum of F under M_k obtained when $\Sigma(\theta^o)$ is fitted to Σ^o. If M_k is the true model, $F_k^o = 0$ and asymptotically T_k is distributed as a central $\chi^2(d_k)$ variate with $\lambda_k = 0$. Hence, the size of λ_k can be taken as a population indicator of model misspecification, with larger values of λ_k reflecting greater misspecification. The relative size of the noncentrality parameters associated with Equation 2,

$$\lambda_i \geq \lambda_j \geq \lambda_k \geq \lambda_s = 0, \qquad (8a)$$

will reflect the degree of model misspecification. In view of the fact that the models are nested, the standardized noncentrality parameters are also ordered

$$F_i^o \geq F_j^o \geq F_k^o \geq F_s^o = 0. \qquad (8b)$$

The relations in Equation 8 permit defining a population measure of comparative model misspecification, that is, a comparative fit index.

The fit index is built as follows. Let λ_k be the measure of misspecification of model M_k. The corresponding misspecification for model M_i is λ_i. In general, one hopes that λ_k is small and expects that λ_i is large. The smaller the ratio λ_k/λ_i, the greater the information provided by model M_k as compared with M_i. Hence,

$$\Delta = 1 - \lambda_k/\lambda_i \qquad (9a)$$

would equal 1.0 if λ_k is 0, and would be close to 0 if $\lambda_k \cong \lambda_i$. In view of Equation 8, Δ is naturally a normed coefficient having a 0–1 range. For a fixed null model misspecification λ_i, decreases in misspecification yield increasing values of Δ. Thus, Δ is a measure of comparative fit. This index can equivalently be written as

$$\Delta = \frac{\lambda_i - \lambda_k}{\lambda_i}, \qquad (9b)$$

showing that the index Δ measures the relative improvement in noncentrality in going from model M_i to M_k. The corresponding incremental fit index comparing models M_j and M_k is

$$\Delta_{jk} = \frac{\lambda_j - \lambda_k}{\lambda_i}. \qquad (10)$$

It is easy to verify that $\Delta = \Delta_{ij} + \Delta_{jk}$, that is, that the increments are additive.

It is apparent that Δ and Δ_{jk} are invariant to a rescaling of the noncentrality parameters by a nonzero constant, for example, if for some c, $\lambda_k \rightarrow c\lambda_k$ and $\lambda_i \rightarrow c\lambda_i$, Δ is unchanged. This invariance is critical to the definition of comparative fit via noncentrality because the noncentrality parameters depend on sample size. The ordering of Equation 8a only makes sense when all noncentrality parameters are based on the same sample size. As a consequence, it is assumed that all nested models and fit statistics (Equation 2) being compared are based on the same sample size, as in fact model comparisons are essentially always implemented. In the unusual situation that M_i is evaluated on a sample of size N_i and M_k on a sample of size N_k, for example, Equations 9a–9b would need to be replaced by

$$\Delta = 1 - F_k^o/F_i^o, \qquad (11)$$

using the standardized noncentrality parameters (Equation 8b). Of course, asymptotically Equations 11 and 9a are equal when N is equal for all models. Differential sample size would not affect Equation 11, although it could affect Equation 9a. Modifications parallel to Equation 11 would be made to Equation 10 in this atypical situation.

These indexes (Equations 9a, 9b, 10, and 11) are population quantities. To implement them in practice, estimators of the noncentrality parameters (Equation 8) must be available. A variety of estimators can be obtained, but let me concentrate on two. Let $\hat{\lambda}_k = T_k - d_k = N\hat{F}_k^o$, $\hat{\lambda}_i = T_i - d_i = N\hat{F}_i^o$, and correspondingly for other noncentrality parameters. Then

$$\hat{\Delta} = FI = 1 - \hat{\lambda}_k/\hat{\lambda}_i = 1 - \hat{F}_k^o/\hat{F}_i^o \quad (12)$$

is a natural index of comparative fit. Apparently, like the *NNFI*, *FI* can be outside the 0–1 range. The alternative estimator

$$\bar{\Delta} = CFI = 1 - \bar{\lambda}_k/\bar{\lambda}_i, \quad (13)$$

based on $\bar{\lambda}_i = \max(\hat{\lambda}_i, \hat{\lambda}_k, 0)$ and $\bar{\lambda}_k = \max(\hat{\lambda}_k, 0)$, is a normed comparative fit index. Because $\bar{\lambda}_i \geq \bar{\lambda}_k \geq 0$, Equation 13 must lie in the 0–1 interval. Saxena and Alam (1982) noted that $\bar{\lambda}_k$ dominates the maximum likelihood estimator of λ_k with squared error as the loss function. Estimators of Δ_{jk} are built analogously. In the unnatural situation that sample size is not constant for all models, $\hat{F}_k^o = (T_k - d_k)/N_k$ and $\hat{F}_i^o = (T_i - d_i)/N_i$ would be used in Equation 12 and similarly applied to obtain Equation 13.

The estimator *FI* is a consistent estimator of Δ. As $N \to \infty$, $\hat{F}_k^o = (T_k - d_k)/N$ and $\hat{F}_i^o = (T_i - d_i)/N$ converge in probability to the constants

$$\hat{F}_k^o \xrightarrow{p} F_k^o \quad \text{and} \quad \hat{F}_i^o \xrightarrow{p} F_i^o. \quad (14)$$

Assuming $F_i^o > 0$, *FI* converges in probability to Δ. Similarly, *CFI* is a consistent estimator of Δ. In view of Equations 14 and 8b, and the definition of $\bar{\Delta}$, in the limit $\hat{\Delta}$ and $\bar{\Delta}$ are equal, and equal to Δ. This means that *FI* behaves as a normed fit index asymptotically. Thus, although *FI* can fall outside the 0–1 range, such behavior would be a small sample effect.

The asymptotic definitions of *NFI*, *NNFI*, and *IFI*, say, NFI^o, $NNFI^o$, and IFI^o, are obtained similarly. Specifically, as T_k/N and T_i/N have the same probability limits as given in Equation 14,

$$NFI^o = IFI^o = \Delta, \quad (15)$$

so that these indexes have the same limit in very large samples. Thus, asymptotically *NFI* and *IFI* also can be related to the comparative reduction in noncentrality as proposed here. On the other hand,

$$NNFI^o = 1 - \frac{d_i F_k^o}{d_k F_i^o} = 1 - \frac{d_i \lambda_k}{d_k \lambda_i} \quad (16)$$

does not have the same limiting definition as the other indexes. This result is consistent with Bollen's (1989) conclusions regarding the similar asymptotic limits of *NFI* and *IFI* and their differences from *NNFI* and Bollen's (1986) index. *NNFI* does not have an interpretation as a comparative reduction in noncentrality, but it can be interpreted as a relative reduction in noncentrality per degree of freedom. Thus, it does appear to have a parsimony rationale.

There is an interesting relation between Equations 12 and 5. Let $\beta = d_k/d_i$ and $\alpha = 1 - \beta$. Then $FI = \alpha + \beta(NNFI)$. As a result, *FI* will behave better than *NNFI*. Although both *NNFI* and *FI* are not restricted to the 0–1 range, *FI* will not be negative as frequently as *NNFI*. This can be seen as follows. Suppose that *NNFI* is negative. Then, as long as $\alpha > \beta|NNFI|$, *FI* remains positive. *FI* also behaves better than *NNFI* at the upper end. Suppose that *NNFI* is greater than 1.0. Then, although it will also exceed 1.0, $FI < NNFI$; that is, it will exceed 1.0 by a smaller amount. Another consequence of the relation between *FI* and *NNFI* is that the standard error of *FI* will be smaller than the standard error of *NNFI* by the factor β. Stated differently, $\text{var}(FI) = \beta^2\text{var}(NNFI) < \text{var}(NNFI)$ because $0 < \beta < 1.0$. Thus, *FI* is a more precise measure of fit than *NNFI*. This effect is illustrated in the sampling study described later. If $\beta = d_k/d_i$ is small, that is, if M_k has many parameters and hence few degrees of freedom, d_k, the reduction in variance possible by using *FI* rather than *NNFI* can be quite substantial. Finally, it will be apparent that as $CFI = FI$ when $0 \leq FI \leq 1$, $CFI > FI$ when $FI < 0$, and $CFI < FI$ when $FI > 1$, the variability of *CFI* will always be less than the variability of *FI*.

All of the previously defined fit indexes, including the new coefficients *FI* and *CFI* introduced here, are based on comparative model fit as measured by the fit T_i and T_k of two nested models. In effect, they are based on a rationale involving difference tests. However, in view of the basic definition (Equations 9a, 9b, and 11), this is not the only way such coefficients need to be stated. Consequently, some new comparative fit indexes based on a different rationale are also introduced.

Fit Indexes for Wald and Lagrange Multiplier Tests

In recent years, Wald and Lagrange multiplier (or score) tests have been introduced into structural modeling (Bentler & Dijkstra, 1985; Dijkstra, 1981, Lee, 1985; Lee & Bentler, 1980). They are routinely available in a public computer program (Bentler, 1986, 1989) and are typically applied to compare nested submodels. These tests provide fit information from the perspective of the less restricted model M_k (Wald) or the more restrictive model M_i (Lagrange multiplier). That is, when estimating M_k and obtaining T_k, one can calculate a Wald statistic W_{ik} at M_k that evaluates the hypothesis that the parameters that differentiate models M_i and M_k are 0. When estimating M_i and obtaining T_i, one can calculate a Lagrange multiplier statistic L_{ik} at M_i that evaluates this same hypothesis. Under standard regularity conditions, W_{ik} and L_{ik} behave as asymptotic noncentral chi-square statistics with $d_{ik} = d_i - d_k$ degrees of freedom and noncentrality parameter $\lambda_{ik} = \lambda_i - \lambda_k$ (Davidson & MacKinnon, 1987; Satorra, 1989). The well-known difference test $D_{ik} = T_i - T_k$ also has the same distribution, although it requires estimating the two models M_i and M_k rather than only one of them (Steiger, Shapiro, & Browne, 1985). In general, these statistics can be used interchangeably in large samples, that is, asymptotically $W_{ik} = L_{ik} = D_{ik} = T_i - T_k$. Thus, this equivalence can be used to form goodness-of-fit indexes to assess model misspecification.

As $\lambda_{ik} = \lambda_i - \lambda_k$, Equations 9a, 9b, and 11 may be equivalently written as

$$\Delta = \lambda_{ik}/\lambda_i = \lambda_{ik}/(\lambda_{ik} + \lambda_k), \quad (17)$$

and estimators of λ_{ik} obtained from W_{ik} and L_{ik} tests can be used to implement Equation 17. Using $\hat{\lambda}_{ik} = W_{ik} - d_{ik}$ or $\hat{\lambda}_{ik} = L_{ik} - d_{ik}$ along with $\hat{\lambda}_i$ and $\hat{\lambda}_k$ as previously defined yields

$$\tilde{\Delta}_W = FIW = \frac{W_{ik} - d_{ik}}{W_{ik} + T_k - d_{ik} - d_k}$$

$$\tilde{\Delta}_L = FIL = \frac{L_{ik} - d_{ik}}{T_i - d_i}. \quad (18)$$

Estimators out of the 0–1 range can be avoided by using $\hat{\lambda}_{ik} = \max(\tilde{\lambda}_{ik}, 0)$ along with $\hat{\lambda}_i$ and $\hat{\lambda}_k$ previously given via

$$\hat{\Delta}_W = CFIW = \hat{\lambda}_{ik}/(\hat{\lambda}_{ik} + \hat{\lambda}_k) \quad (19a)$$

$$\hat{\Delta}_L = CFIL = \hat{\lambda}_{ik}/\hat{\lambda}_i, \quad (19b)$$

where $\hat{\lambda}_{ik}$ is based on W_{ik} in Equation 19a and on L_{ik} in Equation 19b. The practical importance of these indexes (Equations 18, 19a, and 19b) is that they can be implemented when the standard indexes cannot be used: FI and CFI require estimates of both models M_i and M_k, FIW and $CFIW$ require estimating only model M_k, and FIL and $CFIL$ require estimating only model M_i.

Although I prefer the new indexes because of their clear rationale, the NFI, $NNFI$, and IFI indexes can also be modified to yield information from the W_{ik} and L_{ik} tests. Thus,

$$NFIW = \frac{W_{ik}}{W_{ik} + T_k}, \quad NFIL = \frac{L_{ik}}{T_i}. \quad (20)$$

$$NNFIW = \frac{W_{ik} - d_{ik}d_k^{-1}T_k}{W_{ik} + T_k - d_{ik} - d_k},$$

$$NNFIL = \frac{d_iL_{ik} - d_{ik}T_i}{(d_i - d_{ik})(T_i - d_i)}, \quad (21)$$

and

$$IFIW = W_{ik}/(W_{ik} + T_k - d_k)$$

$$IFIL = L_{ik}/(T_i - d_i + d_{ik}). \quad (22)$$

Modifications to ensure that Equations 20–22 have a 0–1 range are obvious. The simplest implementation is to pull estimates outside the 0–1 range to the 0 or 1 endpoints.

The estimators (Equations 18–22) converge in probability as $N \rightarrow \infty$ to interpretable constants. Let $F_{ik}^o = W_{ik}^o/N$ (or $= L_{ik}^o/N$). Then

$$FIW^o = CFIW^o = NFIW^o = IFIW^o = F_{ik}^o/(F_{ik}^o + F_k^o) \quad (23)$$

$$FIL^o = CFIL^o = NFIL^o = IFIL^o = F_{ik}^o/F_k^o \quad (24)$$

and are equal if $F_i^o = F_{ik}^o + F_k^o$ as assumed. Furthermore, both equal Equation 15 under this circumstance, indicating that the W-based and L-based indexes are equivalent to the traditional indexes. However,

$$NNFIW^o = \frac{F_{ik}^o - d_{ik}d_k^{-1}F_k^o}{F_{ik}^o + F_k^o} \quad (25)$$

does not equal Equation 23 but does equal Equation 16 if the assumptions are met. Similarly,

$$NNFIL^o = \frac{d_iF_{ik}^o - d_{ik}F_i^o}{(d_i - d_{ik})F_i^o}, \quad (26)$$

which equals Equations 16 and 25 but not Equation 24, under the assumed conditions.

Finally, although new W- and L-based versions of fit indexes have been presented, the L-based indexes may fail to be defined meaningfully in an important practical circumstance. The indexes are defined when the corresponding Lagrange multiplier statistic L_{ik} or L_{jk} is defined. In the standard application of L_{jk} statistics, there is rarely a problem. However, when the baseline model M_i is the model of uncorrelated variables, the statistic L_{ik} may be 0. For example, if M_k: $\Sigma = \Lambda\Phi\Lambda' + \Psi$ and M_i: $\Sigma = \Psi$, the derivatives of Σ with respect to elements of Λ and Φ under model M_i will be 0. Yet these are key components involved in computing L_{ik}. It is apparent that the asymptotic equivalence of W_{ik}, L_{ik}, and D_{ik} breaks down in this situation. For these reasons, the L-based indexes are not recommended for application in the context of the independence model. However, they will be applicable when other baseline models (Sobel & Bohrnstedt, 1985) are used.

Two Sampling Studies

An example was created to illustrate some of the indexes described earlier and their characteristics. A population model M_k based on the stability of alienation model (see, e.g., Bentler & Bonett, 1980, p. 601) was created. This model contained six measured variables, three factors each with two (mutually exclusive) indicators, and regressions among the factors. M_k contains 15 parameters that require estimation in a sample, with $d_k = 6$. The baseline model M_i, which is false, is the independence model with 6 parameters and $d_i = 15$. A multivariate normal sample of a given size was drawn from this population, and T_i, T_k, W_{ik}, and the corresponding degrees of freedom were pulled from a standard EQS (Bentler, 1989) maximum likelihood estimation run. These statistics were transformed to calculate several of the fit indexes described earlier. This process was repeated until 200 samples of the given size were drawn and the corresponding indexes were obtained. The resulting distributions of indexes form the basic data to be described. This process was repeated at the sample sizes 50, 100, 200, 400, 800, and 1,600.

Summary statistics for the performance of the fit indices at the various sample sizes are shown in Table 1. The simulation for $N = 50$ is presented first. The 10 indexes that were computed are listed along the left, with the 5 indexes based on traditional fit information presented first and the 5 corresponding new Wald-based indexes presented next. A row summarizes the performance of a given index across the 200 replications: the mean value of the index, its standard deviation, its minimum value, and its maximum. As the model is correct, the means should be close to 1.0.

The first 5 rows give fit indexes based on fit of two nested models. The problem of underestimation via small samples known for NFI is evident in the first row of the table. Its mean of .921 is substantially below the means of all other indexes, perhaps inappropriately leading one to question whether the model is correct. The $NNFI$ performs much better on the average, with a mean of 0.998. However, its range of 0.570 to 1.355 is so large that in many samples one would suspect model incorrectness and, in many other samples, overfitting. This increase in variability is also seen in the standard deviations: The $NNFI$ has a threefold increase in standard deviation as compared with

Table 1
Statistics From Sampling Study With 200 Replications

Index	*M*	*SD*	Minimum	Maximum	Index	*M*	*SD*	Minimum	Maximum
		Sample size = 50					Sample size = 400		
NFI	0.921	0.042	.781	0.992	NFI	0.989	0.007	.960	0.999
NNFI	0.998	0.131	.570	1.355	NNFI	1.000	0.019	.927	1.030
FI	0.999	0.053	.828	1.142	FI	1.000	0.007	.971	1.012
CFI	0.980	0.034	.828	1.000	CFI	0.997	0.005	.971	1.000
IFI	0.999	0.045	.849	1.110	IFI	1.000	0.007	.971	1.012
NFIW	0.974	0.016	.914	0.998	NFIW	0.996	0.002	.987	1.000
NNFIW	0.998	0.038	.849	1.092	NNFIW	1.000	0.006	.976	1.009
FIW	0.999	0.015	.940	1.037	FIW	1.000	0.002	.990	1.004
CFIW	0.994	0.011	.940	1.000	CFIW	0.999	0.002	.990	1.000
IFIW	0.999	0.015	.943	1.034	IFIW	1.000	0.002	.990	1.003
		Sample size = 100					Sample size = 800		
NFI	0.957	0.023	.869	0.994	NFI	0.994	0.003	.985	1.000
NNFI	1.001	0.066	.735	1.132	NNFI	0.999	0.009	.976	1.015
FI	1.001	0.027	.894	1.053	FI	1.000	0.003	.990	1.006
CFI	0.990	0.017	.894	1.000	CFI	0.999	0.002	.990	1.000
IFI	1.000	0.025	.900	1.048	IFI	1.000	0.003	.990	1.006
NFIW	0.985	0.008	.948	0.998	NFIW	0.998	0.001	.995	1.000
NNFIW	1.000	0.021	.907	1.036	NNFIW	1.000	0.003	.992	1.005
FIW	1.000	0.008	.963	1.014	FIW	1.000	0.001	.997	1.002
CFIW	0.997	0.006	.963	1.000	CFIW	1.000	0.001	.997	1.000
IFIW	1.000	0.008	.964	1.014	IFIW	1.000	0.001	.997	1.002
		Sample size = 200					Sample size = 1,600		
NFI	0.978	0.013	.932	0.998	NFI	0.997	0.002	.990	1.000
NNFI	1.001	0.034	.882	1.061	NNFI	1.000	0.004	.983	1.007
FI	1.001	0.013	.953	1.024	FI	1.000	0.002	.993	1.003
CFI	0.995	0.009	.953	1.000	CFI	0.999	0.001	.993	1.000
IFI	1.001	0.013	.954	1.023	IFI	1.000	0.002	.993	1.003
NFIW	0.993	0.004	.977	0.999	NFIW	0.999	0.001	.997	1.000
NNFIW	1.000	0.011	.962	1.018	NNFIW	1.000	0.001	.994	1.002
FIW	1.000	0.004	.985	1.007	FIW	1.000	0.001	.998	1.001
CFIW	0.998	0.003	.985	1.000	CFIW	1.000	0.000	.998	1.000
IFIW	1.000	0.004	.985	1.007	IFIW	1.000	0.001	.998	1.001

Note. NFI = normed fit index; NNFI = non-normed fit index; FI = non-normed comparative fit index; CFI = normed comparative fit index; IFI = incremental fix index; W = Wald.

NFI. The newly proposed *FI* fares much better. As expected, its mean is almost perfect (0.999) and its range is more circumscribed around 1.0, with a minimum of 0.828 and a maximum of 1.142. The standard deviation is substantially smaller than that shown by *NNFI*, indeed, being only marginally larger than that of *NFI*. The normed index *CFI* has a mean of 0.980, somewhat below that of *FI* due to pulling-in values of *FI* greater than 1.0, and its standard error is even smaller than that shown by *NFI*. Bollen's (1989) *IFI* has a mean of 0.999 and a standard error estimate between those of *FI* and *CFI*. Like *NNFI* and *FI*, its maximum exceeds 1.0.

All of the new W_{ik}-based fit indexes seem to perform very well. The lowest mean value, 0.974, was obtained by *NFIW*, substantially above the 0.921 shown by the traditional *NFI*. The most remarkable feature of all of these indexes is their low ranges and estimated standard errors. The largest standard error is .038 for *NNFIW*, which is only slightly worse than that shown by the best non-W_{ik}-based index, namely *CFI*. All other indexes have standard errors that are smaller by a factor of 2. Not shown in Table 1 is the fact that in 7 of the 200 replications at sample size 50, the information matrix contained linear dependencies and the *W*-test was based on 8 rather than 9 degrees of freedom. Results for the 193 samples with $d_{ik} = 9$ were virtually identical to those for the 200 samples shown in Table 1.

The trends shown at $N = 50$ are visible at all sample sizes, although the effects are less strong. The minimum fit indexes are all much more reasonable at $N = 100$, with all values being substantially higher than the value of .735 obtained with *NNFI*. Although *NFI* is still 0.043 below 1.0 on the average, and *NFIW* is 0.015 below 1.0, the other indexes have means almost on top of 1.0. Again, the *W*-based indexes have the smallest standard deviations and *NNFI* has the largest. At $N = 200$, *FI* and *IFI* are virtually on target, having a mean of 1.001 and a range of 0.95–1.02. In contrast, *NNFI* has a larger range of 0.88–1.06. Again, the *W*-based indexes perform very well. These trends continue at larger sample sizes. The underestimation of perfect fit by *NFI*, which is evident at the smaller sample sizes, becomes trivially small at the largest sample sizes, as noted previously (Bearden et al., 1982; LaDu & Tanaka, in press) and as expected by Equation 15.

Table 2
Statistics From Sampling Study With Misspecified Model

Index	*M*	*SD*	Minimum	Maximum	Index	*M*	*SD*	Minimum	Maximum
Sample size = 50, replications = 176					Sample size = 400, replications = 200				
NFI	0.874	0.058	.679	0.976	NFI	0.938	0.019	.865	.980
NNFI	0.920	0.155	.415	1.261	NNFI	0.894	0.042	.731	.984
FI	0.963	0.072	.727	1.122	FI	0.950	0.020	.875	.993
CFI	0.950	0.058	.727	1.000	CFI	0.950	0.020	.875	.993
IFI	0.967	0.063	.761	1.101	IFI	0.951	0.019	.876	.993
NFIW	0.956	0.025	.855	0.994	NFIW	0.981	0.006	.950	.994
NNFIW	0.973	0.051	.767	1.067	NNFIW	0.968	0.013	.902	.995
FIW	0.987	0.024	.891	1.031	FIW	0.985	0.006	.954	.998
CFIW	0.984	0.021	.891	1.000	CFIW	0.985	0.006	.954	.998
IFIW	0.988	0.023	.896	1.029	IFIW	0.985	0.006	.954	.998
Sample size = 100, replications = 196					Sample size = 800, replications = 200				
NFI	0.907	0.043	.762	0.991	NFI	0.944	0.015	.897	.974
NNFI	0.898	0.099	.553	1.115	NNFI	0.893	0.032	.793	.958
FI	0.952	0.046	.791	1.054	FI	0.950	0.015	.903	.980
CFI	0.950	0.043	.791	1.000	CFI	0.950	0.015	.903	.980
IFI	0.955	0.043	.805	1.050	IFI	0.950	0.015	.904	.980
NFIW	0.969	0.018	.856	0.998	NFIW	0.983	0.005	.968	.993
NNFIW	0.968	0.035	.769	1.028	NNFIW	0.968	0.010	.937	.988
FIW	0.985	0.016	.892	1.013	FIW	0.985	0.005	.971	.994
CFIW	0.984	0.016	.892	1.000	CFIW	0.985	0.005	.971	.994
IFIW	0.985	0.016	.898	1.013	IFIW	0.985	0.005	.971	994
Sample size = 200, replications = 200					Sample size = 1,600, replications = 200				
NFI	0.929	0.033	.826	0.991	NFI	0.947	0.010	.922	.970
NNFI	0.896	0.074	.665	1.029	NNFI	0.892	0.021	.838	.943
FI	0.952	0.034	.844	1.013	FI	0.949	0.010	.925	.973
CFI	0.951	0.034	.844	1.000	CFI	0.949	0.010	.925	.973
IFI	0.953	0.033	.848	1.013	IFI	0.950	0.010	.925	.973
NFIW	0.978	0.012	.925	0.997	NFIW	0.984	0.003	.974	.991
NNFIW	0.968	0.024	.861	1.009	NNFIW	0.968	0.007	.946	.983
FIW	0.985	0.011	.935	1.004	FIW	0.985	0.003	.975	.992
CFIW	0.985	0.011	.935	1.000	CFIW	0.985	0.003	.975	.992
IFIW	0.985	0.011	.935	1.004	IFIW	0.985	0.003	.975	.992

Note. NFI = normed fit index; NNFI = non-normed fit index; FI = non-normed comparative fit index; CFI = normed comparative fit index; IFI = incremental fit index; W = Wald.

The correlational similarity among the various indexes summarizes a different aspect of their performance, and the trends are clearest at N = 1,600. In spite of the severe restriction in range, essentially all of the correlations are close to 1.0, except that *CFI* and *CFIW* (which correlate ≈ 1.0) correlate about .91 with the remaining indexes. This different behavior of *CFI* and *CFIW* appears to mirror the fact that these indexes correlate only −.91 with T_k, whereas all other indexes correlate below −.995 with T_k. *NNFI*, *FI*, and *IFI* are essentially perfectly correlated, as are the corresponding *W*-based indexes; the correlation between these two classes of indexes is .9996. At smaller sample sizes, these trends are visible but less extreme. At N = 50, *CFI* correlates about .86–.88 with the other indexes and −.88 with T_k. *NNFI* and *FI* correlate 1.0 as expected, and both correlate .999 with *IFI*; these indexes also correlate about −.96 with T_k. The *W*-based indexes show a similar pattern.

A second simulation study was run with a misspecified model, using the design parameters given for the previous study. Although the true model was the same as before, the model that was analyzed omitted the stability path for the repeated latent common factor. The results are shown in Table 2. Turning to the last part of the table, with a sample size of 1,600, one sees that the relative degree of misspecification was on the order of .05 when assessed by all of the fit indexes based on difference tests, except that the *NNFI* indicated a greater degree of misspecification. All of these indexes averaged close to .95, whereas *NNFI* averaged .892. Differences in definition of coefficients are hardly apparent, although *NNFI* has a limit that is not the same as the limit shown by other indexes. These means mirror the asymptotic limits (Equations 15 and 16) quite well. On the other hand, the degree of fit evidenced by the *W*-based indexes is substantially more optimistic than is evidenced by the difference-based indexes. Because of the magnitude of the omitted path, it would seem that the *W*-based indexes provide an unduly optimistic picture of model fit. Apparently, the equivalence $W_{ik} = T_i - T_k$ is not yet true at this sample size, although the asymptotic means of the various *W*-based indexes except *NNFIW* are equal, as expected from Equation 23.

The statistics on the fit indexes shown in the remainder of Table 2 are consistent with the general conclusions that have been mentioned. At sample sizes 50 and 100, there were some problems of nonconvergence as well as singularity of the information matrices that made calculation of the W_{ik} statistics impossible (the ratio of convergence failure to information matrix singularity was 1:3). As a consequence, results from fewer than 200 replications were tabled. The main results to be noted in the other panels involve the comparison to values obtained at N = 1,600: the means of the indexes *FI*, *CFI*, and *IFI* at virtually all sample sizes mirror the large sample results extremely well; *NFI* underestimates the large sample value by a substantial amount in the smaller samples; and the *W*-based indexes seem to be inflated at all sample sizes. Correlational similarity among the indexes mirrored the previously mentioned results under the correct model, with some exceptions. In the misspecified models, the *CFI* indexes were much more similar to all other indexes than they were in the correctly specified model, with the lowest correlation between *CFI* and any other index being .981 at N = 1,600. In fact, the lowest correlation among any pair of indexes was .98, and the correlation between the indexes and T_k was on the order of −.96 for all indexes. At N = 50, the lowest correlation among indexes was .84, and the highest correlation, as expected, was between *NNFI* and *FI* (1.0), with *IFI* correlating .999 with these.

Conclusion

Normed and nonnormed fit indexes are very popular adjuncts to more traditional statistics in structural equation modeling to help assess the quality of a model (Bentler & Bonett, 1987). In spite of their popularity, nothing has been known about the population quantities that these indexes are intended to assess. This is also true of Bollen's (1989) *IFI*, although Bollen showed that *IFI* and *NFI* have the same asymptotic limits and *NNFI* has a different limit. It is apparent from Equation 15 that in large samples, *NFI* and *IFI* will reflect a relative drop in noncentrality, that is, they will mirror the comparative fit indexes *FI* and *CFI* introduced here. Thus, these indexes are equivalent asymptotically and they can be used interchangeably. In small samples, however, this equivalence is less certain and the indexes do not estimate the same quantity. *FI* and *IFI* seem to behave quite similarly, but both can exceed 1.0. *CFI* seems to be the best index: Like the population coefficient Δ, it has a 0–1 range, has small sampling variability, and estimates the relative difference in noncentrality of interest. However, these advantages are obtained at the expense of some downward bias. This bias is quite small, and is certainly much less than the bias of the *NFI*. In fact, there was virtually no bias in the simulation with the misspecified model.

The index *NNFI* seems to have a rationale that is different from the other indexes just mentioned. As seen in Equation 16, *NNFI* can be interpreted in large samples as assessing the relative drop in noncentrality per degree of freedom. In contrast, the new index Δ and its estimators assess the difference in noncentrality on an absolute basis. Thus, *NNFI* should be interpreted differently from Δ and its estimators. However, *NNFI* and *FI* are perfectly correlated as they are linearly related when comparing models with the same degrees of freedom, so that alternative models would be ranked equivalently by these two indexes. The critical issue in the use of *NNFI* as compared with other indexes is its absolute value. Rules of thumb or other more precise decision rules for model acceptability may have to be somewhat different for *NNFI* as compared with the other difference-based indexes considered in this article.

A class of indexes that use Wald or Lagrange multiplier statistics in addition to fit information on a given model were also developed. The Lagrange multiplier-based indexes appear not to work in a common application where relative fit is assessed by comparing a substantive model to the model of uncorrelated or independent variables. On the other hand, the Wald-based indexes were always implementable and appeared, in the example, to perform quite well, with generally small sampling variability; however, they showed an upward bias. This bias may originate from the fact that equivalence between Wald- and difference-based statistics was not complete at N = 1,600 with the number of replications considered: Asymptotic equivalence would imply that the sampling means of the statistics should approximately relate as $\bar{T}_i = \bar{T}_k + \bar{W}_{ik}$ and that T_i and $(T_k + W_{ik})$ should correlate close to 1.0 across the 200 replications. In fact, $(\bar{T}_k + \bar{W}_{ik})$ exceeded $\bar{T}_i$ by a factor of about 3 in both studies, and the correlations were only .73–.83. Thus, it is possible that even N = 1,600 is too small for asymptotic theory to apply accurately enough, or that the noncentral chi-square distribution is not a totally appropriate reference distribution for T_i. There was some degradation in performance of all indexes at N = 50, where occasional linear dependencies among parameter estimates and less than full-rank Wald tests were observed. At large sample sizes, differences between all fit indexes became quite small. At N = 1,600 the correlations among all indexes studied, computed across 200 sampling replications, were close to 1.0, indicating that correlational performance differences between the indexes become trivial in very large samples. The exception was *CFI* (and *CFIW*), which correlated only .91 with the remaining indexes in the correct model analysis; however, in the misspecified model analysis, these indexes correlated as highly with others as did any other pair of indexes with each other.

The performance of the various fit indexes, summarized in Tables 1 and 2, was studied under a limited range of modeling situations. Further research, made easier by the simulation feature of EQS (Bentler, 1989), is needed. EQS automatically computes CFI, NFI, and NNFI, but the other indexes would need to be computed from the EQS output.

Fit indexes as currently used are primarily descriptive statistics. This article has developed a population index that provides a more fundamental rationale for assessment of comparative fit than has previously been available. Yet essentially nothing is known about the theoretical sampling distribution of the various estimators. A purely theoretical approach will no doubt be difficult as fit indexes are intended to be applied in circumstances not covered by current theory in structural modeling, for example, in small samples, when both models M_i and M_k may be false, and when M_k may be true but M_i is far away from it. Research should also address the use of other reference distributions besides the noncentral chi-square distribution used here for defining comparative fit. It is possible that the null model of independence may be so different from the true model that another distribution could be more appropriate at times (see, e.g., Satorra, Saris, & de Pijper, 1987). Sample fit indexes would then need to be redefined as well. Of course, statistical theory may be limited in its relevance to assessing fit, and the descriptive character of the indexes may continue to be a major feature, as one important application will continue to be in very large samples where almost any a priori model $\Sigma = \Sigma(\theta)$ will be false, that is, where any a priori model will have a comparative fit that is statistically less than a perfect 1.0.

References

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

Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample size effects on chi square and other statistics used in evaluating causal models. *Journal of Marketing Research, 19*, 425–430.

Bentler, P. M. (1983). Some contributions to efficient statistics for structural models: Specification and estimation of moment structures. *Psychometrika, 48*, 493–517.

Bentler, P. M. (1986). *Lagrange Multiplier and Wald tests for EQS and EQS/PC.* Los Angeles: BMDP Statistical Software.

Bentler, P. M. (1989). *EQS structural equations program manual.* Los Angeles: BDMP Statistical Software.

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., & Bonett, D. G. (1987). This week's Citation Classic. *Current Contents S&BS, 19*(37), 16.

Bentler, P. M., & Dijkstra, T. (1985). Efficient estimation via linearization in structural models. In P. R. Krishnaiah (Ed.), *Multivariate analysis VI* (pp. 9–42). Amsterdam: Elsevier.

Bentler, P. M., & Mooijaart, A. (1989). Choice of structural model via parsimony: A rationale based on precision. *Psychological Bulletin, 106*, 315–317.

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

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

Bollen, K. A., & Liang, J. (1988). Some properties of Hoelter's CN. *Sociological Methods & Research, 16*, 492–503.

Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika, 52*, 345–370.

Davidson, R., & MacKinnon, J. G. (1987). Implicit alternatives and the local power of test statistics. *Econometrica, 55*, 1305–1329.

Dijkstra, T. K. (1981). *Latent variables in linear stochastic models.* Groningen, The Netherlands: Rijksuniversiteit te Groningen.

James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models, and data.* Beverly Hills, CA: Sage.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI user's guide.* Mooresville, IN: Scientific Software.

LaDu, T. J., & Tanaka, J. S. (in press). The influence of sample size, estimation method, and model specification on goodness-of-fit assessments in structural equation models. *Journal of Applied Psychology.*

Lee, S. Y. (1985). On testing functional constraints in structural equation modeling. *Biometrika, 72*, 125–131.

Lee, S. Y., & Bentler, P. M. (1980). Some asymptotic properties of constrained generalized least squares estimation in covariance structure models. *South African Statistical Journal, 14*, 121–136.

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.

Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). An evaluation of goodness of fit indices for structural equation models. *Psychological Bulletin, 105*, 430–445.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*, 131–151.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *ASA Proceedings, Business and Economic Statistics Section* (pp. 308–313). Washington, DC: American Statistical Association.

Satorra, A., Saris, W. E., & de Pijper, M. (1987). A comparison of several approximations to the power function of the likelihood ratio test in covariance structure analysis. *ASA Proceedings, Business and Economic Statistics Section* (pp. 393–398). Washington, DC: American Statistical Association.

Saxena, K. M. L., & Alam, K. (1982). Estimation of the non-centrality parameter of a chi-squared distribution. *Annals of Statistics, 10*, 1012–1016.

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.

Steiger, J. H., Shapiro, A., & Browne, M. W. (1985). On the multivariate asymptotic distribution of sequential chi-square statistics. *Psychometrika, 50*, 253–264.

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

Wheaton, B. (1987). Assessment of fit in overidentified models with latent variables. *Sociological Methods & Research, 16*, 118–154.

Received February 25, 1988
Revision received October 12, 1988
Accepted May 16, 1989 ■