Borrower: SYB

Lending String: TXI,*NNM,MNU,ZYU,ZYU

Patron: Prince, Mark

Structural equation modeling ; a multidisciplinary journal.

Volume: 3 **Issue:** 4
Month/Year: 1996
Pages: 369-379

Article Author: Rigdon, Edward

Article Title: CFI versus RMSEA; A comparison of two fit indexes for structural equation modeling

ILL Number: 104401933

Call #:

Location:

ODYSSEY ENABLED
Charge
Maxcost: 30.00IFM

Shipping Address:
Syracuse Univ.Library
ILL
222 Waverly Ave.
Syracuse, N.Y.13244-2010

EMAIL: ILL@SYR.EDU
email: ill@syr.edu

Billing Notes; IDS, Shares

STRUCTURAL EQUATION MODELING, *3*(4), 369–379

CFI Versus RMSEA: A Comparison of Two Fit Indexes for Structural Equation Modeling

Edward E. Rigdon
Marketing Department
Georgia State University

This article compares two structural equation modeling fit indexes—Bentler's (1990; Bentler & Bonett, 1980) Confirmatory Fit Index (CFI) and Steiger and Lind's (1980; Browne & Cudeck, 1993) Root Mean Square Error of Approximation (RMSEA). These two fit indexes are both conceptually linked to the noncentral chi-square distribution, but CFI has seen much wider use in applied research, whereas RMSEA has only recently been gaining attention. The article suggests that use of CFI is problematic because of its baseline model. CFI seems to be appropriate in more exploratory contexts, whereas RMSEA is appropriate in more confirmatory contexts. On the other hand, CFI does have an established parsimony adjustment, although the adjustment included in RMSEA may be inadequate.

By now, most structural equation modeling (SEM) researchers are familiar with the limitations of the χ^2 fit statistic. Interest centers, instead, on the variety of alternative fit indexes that are available. One family of fit indexes, developed by Bentler and Bonett (1980) and reconceptualized by Bentler (1990), involve comparing the fit of the model of interest against a baseline model. These "comparative fit indexes" have proven very popular. Another fit index, the Root Mean Square Error of Approximation (RMSEA) was also introduced in 1980 (Steiger & Lind, 1980). Given its current name by Browne and Cudeck (1993), RMSEA has received considerably less attention until recently, even in the SEM literature. Recently, however, Sugawara and MacCallum (1993) found that "incremental fit indexes,"

Requests for reprints should be sent to Edward Rigdon, Department of Marketing, Georgia State University, University Plaza, Atlanta, GA 30303. E-mail: mkteer@langate.gsu.edu.

like CFI, were less stable across different estimation methods than "nonincremental fit indexes," including RMSEA, in analyses of three real-world data sets. This suggests that researchers using the same data but different estimation methods would be more likely to reach different conclusions about models if they relied on CFI instead of RMSEA. So it appears that this is a good time to take a comparative look at these two fit indexes.

BACKGROUND

In SEM, researchers typically compare $\hat{\Sigma}$, the covariance structure implied by some model, with an empirical covariance matrix, S. Using Bentler's (1990) symbols and approach, for a given set of manifest variables and a given S, researchers who wish to compute comparative fit indexes must specify at least two models. Model M_i is a highly restricted baseline model, whereas model M_k is the model implied by the theory under investigation. Two test statistic values, T_i and T_k, result from fitting M_i and M_k to S. Let the test statistic be the χ^2 statistic, with degrees of freedom d_i and d_k for the two models.

These test statistics will follow χ^2 distributions with appropriate degrees of freedom only if the two models M_i and M_k are exactly correct in the population. Beginning with a rather overlooked conference presentation by Steiger and Lind (1980), researchers argued that if the two models are not exactly correct in the population, but the degree of discrepancy is limited; the test statistic will instead follow a noncentral χ^2 distribution, with noncentrality parameter λ. Kendall and Stuart (1979) designated this distribution χ'^2. Applying this concept, Bentler (1990; see also Steiger, Shapiro, & Browne, 1985) reconceptualized the family of comparative fit indexes as functions of the estimated noncentrality parameters, λ_i and λ_k, of models M_i and M_k. For example, Bentler (1990) defined the Comparative Fit Index (CFI) as:

$$\mathrm{CFI} = 1 - (\lambda_k / \lambda_i) \tag{1}$$

The value of λ for a given model can be estimated as $T - d$. To give his normed CFI a (0, 1) range, Bentler (1990) further constrained the estimation of λ_k and λ_i as follows:

$$\lambda_k = \max(T_k - d_k,\ 0) \tag{2}$$

$$\lambda_i = \max(T_i - d_i,\ T_k - d_k,\ 0) \tag{3}$$

Going back to original work by Tucker and Lewis (1973), the most widely used baseline model has been a null or independence model, which posits that all manifest variables in the model are mutually uncorrelated. Bentler and Bonett (1980) arrived at the same baseline model by generalizing the concept of nested model comparisons, and this use was endorsed explicitly by Mulaik et al. (1989, p. 343) and implicitly by Bentler (1990). Based on an informal survey of recent literature, it appears that applied researchers adopt the null baseline model as a matter of course.

RMSEA is computed based on sample size (N) and the noncentrality parameter and degrees of freedom for the target model, as (Browne & Cudeck, 1993; Steiger, 1990):

$$\text{RMSEA} = \sqrt{\frac{\hat{\lambda}}{(N-1)\times d}} \tag{4}$$

The rationale for this index begins with Browne's (1982) development of the generic discrepancy function:

$$\hat{F} = (\mathrm{s}-\hat{\sigma})' * W^{-1} * (\mathrm{s}-\hat{\sigma}) \tag{5}$$

where s and $\hat{\sigma}$ are vectors formed from the nonredundant elements of the empirical (S) and model-implied ($\hat{\Sigma}$) covariance matrices, respectively, and W is a weight matrix, generally taken to be an asymptotically unbiased estimate of the covariance matrix of the elements of s. So $\hat{F}$ is a weighted sum of squared deviations. The use of W^{-1} effectively standardizes the $\mathrm{s}-\hat{\sigma}$ residuals, making $\hat{F}$ scale-free (Steiger, message to SEMNET, May 8, 1995; see also Steiger & Lind, 1980, p. 6). Given that S is an unbiased estimator of the population covariance matrix, Σ, then $\hat{F}$ is an asymptotically unbiased estimator of F, the population value. With finite sample size, however, $\hat{F}$ is biased because it reflects both random sampling error and systematic specification error, whereas the population value naturally includes no random sampling error.

The χ^2 statistic, computed as $(N-1) * \hat{F}$ (where N is sample size), provides a mechanism for eliminating the random error component of this estimate of F. The χ^2 statistic's degrees of freedom parameter measures the expected quantity of random error in the statistic. When we estimate $\hat{\lambda}$ as $\chi^2 - d$ (with a minimum of zero), we eliminate the random error component, and $\hat{\lambda}/(N-1)$ becomes an unbiased estimator of F in the population. This process of removing both the random error component and sample size from the computed χ^2 is also supposed to make the estimate of RMSEA largely insensitive to changes in sample size, especially when sample size is large.

Dividing this estimator of F by d produces a measure of misspecification per degree of freedom in the population. Also, because adding free parameters to the model reduces the degrees of freedom, this division may serve as a parsimony adjustment. Finally, taking the square root returns this function of squared deviations to the metric of the standardized discrepancies between s and $\hat{\sigma}$.

RMSEA has a minimum of 0 which implies perfect fit. Browne and Cudeck (1993) reluctantly proposed this rule of thumb for interpreting larger values:

> Practical experience has made us feel that a value of the RMSEA of about 0.05 or less would indicate a close fit of the model in relation to the degrees of freedom. ... We are also of the opinion that a value of about 0.08 or less ... would indicate a reasonable error of approximation and would not want to employ a model with a RMSEA greater than 0.1. (p. 144)

CFI VERSUS RMSEA: A COMPARISON

Noncentral Chi-Square Distribution

Although the χ'^2 distribution is common to both indexes, the ways that CFI and RMSEA relate to this distribution highlight several key differences between the indexes. RMSEA is calculated only from the χ^2 statistic for the target model. The CFI computation involves the χ^2 statistics for both the target model and the baseline model. Given the reference to the χ'^2 distribution, then, both statistics carry the implicit assumption that the target model is approximately correct in the population, but CFI carries the additional assumption that the baseline model is also approximately correct.

This is problematic, given the predominant use of the null model as the baseline. Rarely will the null model be approximately correct in the population. More likely, a model that posits all manifest variables as uncorrelated will be tremendously incorrect. The null model cannot be approximately correct unless the parameters in the target model that are typically of interest (parameters indicating correlations between variables) are all approximately zero. Even leaving the target model aside, variables in nonexperimental research do tend to be correlated, whether due to general background correlation (Meehl, 1990; Standing, Sproule, & Khouzam, 1991), or due to "method factors" (Bagozzi, 1984) or "halo effects" (Cooper, 1981). Therefore, it seems unlikely that the χ'^2 distribution provides a sound basis for making inferences about the behavior of CFI.

By contrast, the χ'^2 distribution does provide a basis for describing the statistical behavior of RMSEA, to the extent of allowing hypothesis tests and the construction of confidence intervals around particular point estimates (Browne & Cudeck, 1993). Besides being more informative than point estimates, confidence intervals may also help researchers to resist the temptation to free additional parameters in a model for the sake of marginal improvements in fit.

Calculation

On the other hand, the genuine usefulness of the χ'^2 distribution in connection with RMSEA means that the researchers using RMSEA must actually work with the χ'^2 distribution, and this is no small task. With a continuous noncentrality parameter, in addition to a degrees of freedom parameter, the χ'^2 distribution is not well tabled—at least, not in the ranges of interest to most SEM researchers—nor is it a well-established part of widely distributed analytical programs. For example, Wolfram's (1991) *Mathematica* symbolic math package does include a NoncentralChisquareDistribution function, but the function tends to break down when the parameters become large. The distribution has become a staple of SEM packages, but researchers who rely entirely on those packages are limited in the types of analyses that they can perform, and have no way to verify their results. Approximations to the normal distribution exist, such as the Moschopoulos (1984) approximation recommended by Browne and Cudeck (1993), but such methods tend to be either simple or accurate, but not both. Furthermore, the χ'^2 distribution has a fatter, more convex upper tail than the normal, so approximations to the normal are likely to be inaccurate in this area. Unfortunately, this is the area that will often be of greatest interest to researchers.

However, RMSEA will be more generally useful in the task of reevaluating earlier research because computing this index requires only the χ^2 statistic, degrees of freedom, and sample size for the target model. Computing CFI requires χ^2 statistics and degrees of freedom for both the target model and the baseline model. If the report of the earlier research did not include statistics for a baseline model, and if a covariance matrix is not available for reanalysis, computing CFI is impossible.

Exploratory Versus Confirmatory

SEM is most commonly considered a confirmatory methodology, yet few research projects are purely confirmatory or exploratory. Projects that involve low sample sizes or that represent early efforts in a field are bound to have an exploratory flavor, even when they employ SEM methods. Arguably, the exploratory/confirmatory distinction marks another point of difference between CFI and RMSEA.

Bentler and Bonett (1980) clearly stated that "The statistical value of the null model seems to be more important in small samples" (p. 604). This emphasis on small samples may well be taken to suggest an exploratory context. Remarking on the use of a null model as baseline, Sobel and Bohrnstedt (1985) stated, "Only in the purely exploratory case is M_0 an appropriate baseline model" (p. 161). If there is prior knowledge on the phenomena in question, they noted, then using a simple null model as a baseline is inefficient because, in a typical confirmatory context, that same null model has already been rejected. As argued previously, the χ^2 statistic for the null model is unlikely to follow the χ^{-2} distribution except in exploratory contexts. In confirmatory settings, the χ^2 statistic for the null model is likely to be so large that even rather poor target models look quite good by comparison. Recognizing this, Mulaik has recently suggested (Mulaik, message to SEMNET, February 24, 1995; see also Carlson & Mulaik, 1993) raising the rule-of-thumb minimum standard for the CFI from 0.90 to 0.95, in order to reduce the number of severely misspecified models that achieve acceptable values on this criterion. It may be that the problem lies not in the value adopted for the rule of thumb, but rather in the research context in which the rule is applied.

By contrast, RMSEA may be better suited to more confirmatory contexts, and may behave in unintended ways when, for example, sample size is small. First, recall that RMSEA attempts to estimate a population value. It is logical that the quality of this estimate would be particularly good when sample size is large, and poor when sample size is low. This becomes clearer with another look at the RMSEA formula. Recast the formula this way:

$$\text{RMSEA} = \sqrt{\frac{\chi^2 - d}{(N-1)\times d}} \tag{6}$$

$$\text{RMSEA} = \sqrt{\frac{\chi^2}{(N-1)\times d} - \frac{d}{(N-1)\times d}} \tag{7}$$

$$\text{RMSEA} = \sqrt{\frac{(N-1)\times \hat{F}}{(N-1)\times d} - \frac{1}{(N-1)}} \tag{8}$$

$$\text{RMSEA} = \sqrt{\frac{\hat{F}}{d} - \frac{1}{(N-1)}} \tag{9}$$

Equation 9 indicates that the RMSEA value derives from two key elements— $\hat{F}$ / d and 1 / (N – 1). When sample size is large, the second term approaches 0 asymptotically, and even its contribution to the overall square root will be small. When sample size is small, however, the second term, and particularly its contribution to the square root, can be substantial, especially in light of the values in Browne and Cudeck's (1993) rule of thumb (see Figure 1). When the difference between .05 and .08 is the difference between a model that fits well and one that doesn't, the square root of the second term may be tremendously influential. For instance, when $N = 50$, the square root of the second term by itself is almost 0.15, and when N is 100, the value is approximately 0.10.

Another way to look at this problem is to examine critical values for the (central) χ^2 distribution, and to observe how these values translate into RMSEA values. For instance, suppose a certain model yields a χ^2 statistic with $p = .05$. Most researchers would interpret a χ^2 statistic like this as indicating good fit. Does RMSEA suggest the same thing? The answer depends on both the degrees of freedom and the sample size for the model. For example, with 10 degrees of freedom, the critical value for the χ^2 distribution is 18.307. With $N = 100$, this translates into RMSEA = 0.092, a value indicating relatively poor fit according to Browne and Cudeck's (1993) criteria. When $N = 300$, however, RMSEA = 0.053, and when $N = 500$, RMSEA = 0.042.—values that suggest relatively good fit. Thus, when sample size is low, RMSEA may suggest rejecting a model that otherwise would be accepted. (Argu-

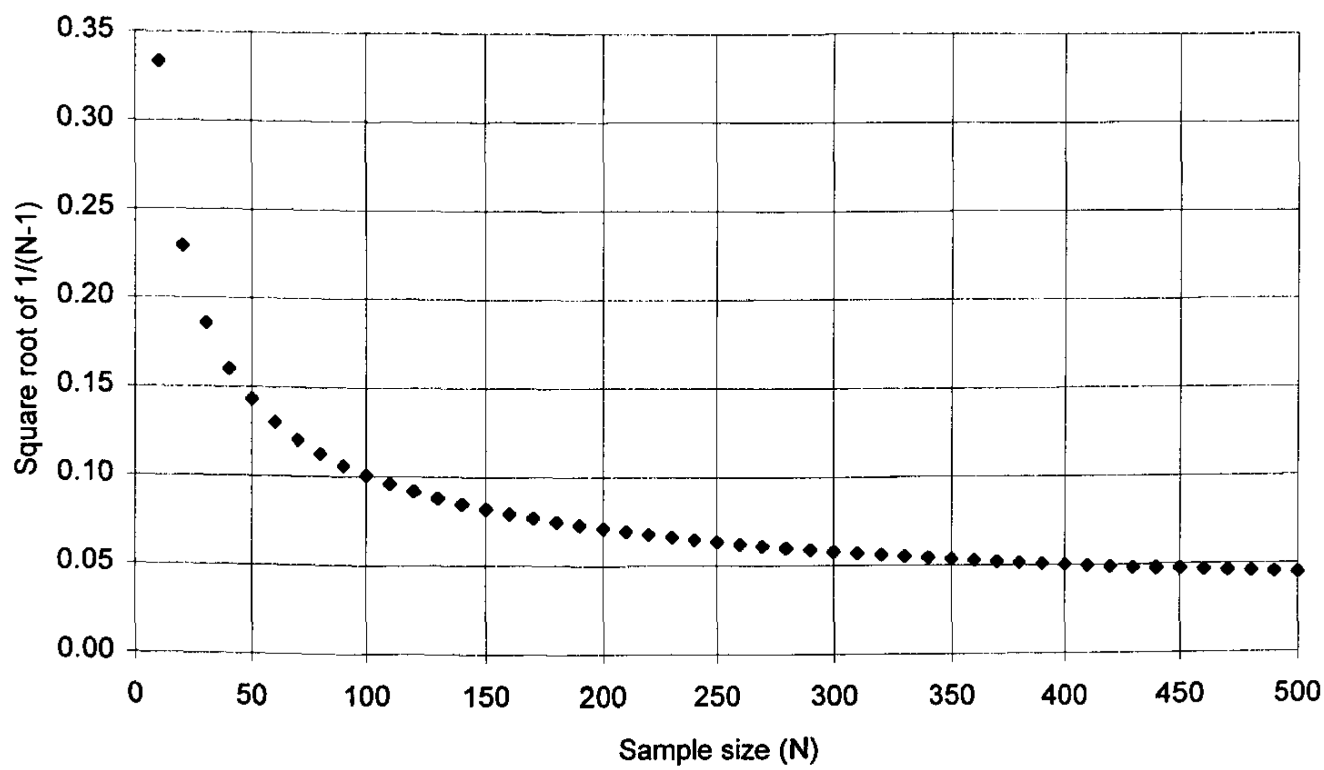

FIGURE 1 The value of the square root of 1 / (N – 1) as a function of N (sample size).

ably, the problem lies in the rule of thumb for interpreting RMSEA; this is a weakness shared by both indexes.) Thus, CFI and RMSEA appear to be complementary indexes, with CFI better suited to more exploratory, small sample cases, and RMSEA better suited to more confirmatory, large sample situations.

Parsimony Adjustment

Concerns about parsimony lie at the heart of the philosophy of SEM. It is easy to achieve good fit for a model if parsimony is neglected. Consequently, researchers prefer fit indexes that take account of the relative parsimony of the models being evaluated. Here again, there is a point of difference between CFI and RMSEA.

CFI itself does not inherently include a parsimony adjustment, except in that, as with RMSEA, the noncentrality parameter is computed as the difference between the χ^2 statistic and the degrees of freedom. Mulaik et al. (1989) proposed a parsimony adjustment for indexes of this type:

$$p\text{CFI} = \text{CFI} \times \frac{d_k}{d_i} \tag{10}$$

As restrictions in the target model are relaxed, its degrees of freedom decline, as does the parsimony-adjusted index. Mulaik (message to SEMNET, February 24, 1995) suggested that a value for pCFI of 0.75 or greater, combined with a CFI value of 0.95 or greater, indicates a parsimonious model that fits well.

It has already been suggested that a focus on RMSEA confidence intervals, rather than point estimates, can help researchers to resist the temptation to allow more free parameters. Further, it can be argued that RMSEA already includes a parsimony adjustment, because the statistic represents misspecification per degree of freedom. But, lacking a highly restricted baseline model as a standard of comparison, RMSEA may leave the researcher without a clear sense of just how many potential degrees of freedom have been sacrificed to achieve the current level of fit. Consequently, researchers who are concerned about parsimony may have reason to favor parsimony-adjusted CFI. Perhaps a new rule of thumb could be developed for a parsimony-adjusted RMSEA, but this would not be a trivial exercise.

ALTERNATIVES TO THE NULL BASELINE MODEL

Many of the objectionable aspects of CFI spring from the use of the null model as baseline. Bentler and Bonett (1980) recognized that "the incremental fit indexes depend critically on the availability of a suitably framed [baseline] model" (p. 604). Are other baseline models available that might resolve these problems?

Bentler and Bonett (1980) endorsed a "model of modified independence" (p. 604), where the baseline model includes the same fixed non-zero covariances that are included in the target model. Bentler and Bonett (1980, p. 604) considered this an excellent choice for a baseline model. Unless a model includes a number of these fixed nonzero covariances, however, it will have roughly the same fit as the standard null model, and will share the same basic problems.

James, Mulaik, and Brett (1982) suggested using a model that allows relations between latent constructs and measures, but constrains the constructs to be mutually uncorrelated. This model could be considered a highly restricted endpoint for a series of nested structural models, as in Anderson and Gerbing's (1988) two-step approach to model testing. This "structural null" model will be problematic in some cases because, in general, measurement models with uncorrelated constructs will only be identified if each construct has at least three measures (Bollen, 1989). Imposing additional restrictions on the individual factor models may eliminate the ambiguity.

Rigdon (1995) proposed an equal correlation (EC) baseline model. Although the null model constrains all correlations between measures to be equal to zero, the EC baseline model only constrains the correlations to be equal to one another. (The EC baseline model allows covariances to be unequal if variances are unequal.) Using the EC baseline model thus amounts to focusing on explaining the pattern of correlations between measures, ignoring the absolute level of correlation. The difference, which involves estimating only one additional model parameter, sometimes accounts for half of the lack of fit in the null model. This, in turn, makes it more likely that χ^2 statistics from the EC baseline model will actually follow the appropriate χ'^2 distribution.

However, the use of any alternative baseline model leads to a major problem of interpretability (Marsh, Balla, & McDonald, 1988). The existing rules of thumb for CFI relate only to indexes computed on the basis of the null model. Certainly, without population distributions for the resulting indexes, researchers who adopt alternative baseline models will be forced to develop and defend criteria for evaluating the index values that result. Consequently, it is unlikely that there will be a movement toward an alternate baseline model anytime soon.

SUMMARY

This comparison suggests that CFI and RMSEA each have comparative strengths and weaknesses. We believe that Sobel and Bohrnstedt (1985) were correct when they suggested limiting the use of the null model to exploratory contexts. Given the nature of the data that researchers often must utilize in exploratory settings, the null model may well represent a viable baseline. It is tempting, then, to call for limiting the use of CFI to similar circumstances, where there is likely to be doubt about whether supposed "good fit" is not actually due either to uniformly weak relations between variables or to very low sample size. But in confirmatory contexts, when researchers wish to determine whether a given model fits well enough to yield interpretable parameters and to provide a basis for further theory development, RMSEA appears to be a better choice. However, RMSEA may not include an adequate parsimony correction, and this may lead researchers to favor more saturated models. Thus, researchers may still feel the need for an alternative fit index that does include a meaningful parsimony penalty.

REFERENCES

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–423.

Bagozzi, R. P. (1984). A prospectus for theory construction in marketing. *Journal of Marketing, 48,* 11–29.

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.

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

Browne, M. W. (1982). Covariance structures. In D. M. Hawkins (Ed.), *Topics in applied multivariate analysis* (pp. 72–141). Cambridge, England: Cambridge University.

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. 136–62), Newbury Park, CA: Sage.

Carlson, M., & Mulaik, S. (1993). Trait ratings from descriptions of behavior as mediated by components of meaning. *Multivariate Behavioral Research, 28,* 111–159.

Cooper, W. H. (1981). Ubiquitous halo. *Psychological Bulletin, 90,* 218–244.

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

Kendall, M., & Stuart, A. (1979). *The advanced theory of statistics* (Vol. 2). London: Charles Griffin.

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.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often uninterpretable. *Psychological Reports, 66,* 195–244.

Moschopoulos, P. G. (1984). On a new transformation to normality. *Communications in Statistics: Theory and Methods, 12,* 1873–1878.

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

Rigdon, E. E. (1995). The equal correlation baseline model for comparative fit assessment in structural equation modeling. In D. W. Stewart & N. J. Vilcassim (Eds.), *Proceedings of the 1995 AMA Winter Educators' Conference* (pp. 175–176). Chicago: American Marketing Association.

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* (pp. 152–78). San Francisco: Jossey-Bass.

Standing, L., Sproule, R., & Khouzam, N. (1991). Empirical statistics: IV. Illustrating Meehl's sixth law of soft psychology: Everything correlates with everything. *Psychological Reports, 69,* 123–126.

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

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically-based tests for the number of common factors.* Handout for a presentation delivered at the meeting of the Psychometric Society, Iowa City, IA.

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

Sugawara, H. M., & MacCallum, R. C. (1993). Effect of estimation method on incremental fit indexes for covariance structure models. *Applied Psychological Measurement, 17,* 365–377.

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

Wolfram, S. (1991). *Mathematica: A system for doing mathematics by computer.* Champaign, IL: Wolfram Research.