The Journal of Experimental Education, *2006*, *74*(3), 267–288

The Search for "Optimal" Cutoff Properties: Fit Index Criteria in Structural Equation Modeling

STEPHEN A. SIVO
University of Central Florida

XITAO FAN
University of Virginia

E. LEA WITTA
University of Central Florida

JOHN T. WILLSE
University of North Carolina–Greensboro

ABSTRACT. This study is a partial replication of L. Hu and P. M. Bentler's (1999) fit criteria work. The purpose of this study was twofold: (a) to determine whether cut-off values vary according to which model is the true population model for a dataset and (b) to identify which of 13 fit indexes behave optimally by retaining all of the correct models while simultaneously rejecting all of the misspecified models in a manner invariant across sample size and data distribution. The authors found that for most indexes the results do not vary depending on which model serves as the correct model. Furthermore, the search for an optimal cut-off value led to a new discovery about the nature of McDonald's measure of centrality and the root mean square error of approximation. Unlike all other indexes considered in this study, the cut-off value of both indexes actually *decreases* for incorrect models as sample size increases. This may suggest that power calculations are more likely to be optimal when based on those indices.

Key words: cut-off values, fit indexes, model fit, structural equation modeling

Address correspondence to: Stephen A. Sivo, College of Education, University of Central Florida, 4000 Central Florida Parkway, Orlando, FL 32816-1250. E-mail: ssivo@mail.ucf.edu

THE APPLICATION OF STRUCTURAL EQUATION MODELING (SEM) requires a researcher to evaluate how well a model fits sample data. Although many fit indexes have been developed (Fan & Sivo, 2005), the indexes are by no means parallel. In fact, the variety of indexes available complicates model evaluation. In addition to not having directly comparable scales or distributions, different fit indexes address different aspects of model appropriateness (e.g., parsimony, sample size effects, comparisons to null models).

The issue of how to determine the propriety of models has not been resolved (Marsh, Hau, & Wen, 2004). Cudeck and Henly (1991) pointed out that assessment of model fit will always have an aspect of statistical decision making that cannot be totally reduced to numbers. Despite this truism, there is still a need for some clear guiding principles regarding model fit assessment.

Much of the research concerning model fit has used Monte Carlo studies in which data are simulated to investigate the performance of fit indexes (e.g., Fan, Thompson, & Wang, 1999; Gerbing & Anderson, 1993; Hu & Bentler, 1999; Marsh, Balla, & McDonald, 1988; McDonald & Marsh, 1990; Mulaik et al., 1989). One goal that Monte Carlo researchers in SEM have sought is the development of well-defined "rules of thumb" for assessing model appropriateness. The lack of comparability of different fit indexes caused by their discrepant functioning (Fan et al.), the effects of sample size on fit index functioning (Marsh et al.), and, to some extent, the inherent inability of a structural model to exactly account for the phenomena it seeks to describe (Tanaka, 1993) make the development of a rule of thumb for dichotomous fit or no-fit decisions difficult. With these problems in mind, Hu and Bentler undertook developing rules of thumb for fit criteria.

Hu and Bentler's (1999) approach to studying the criterion problem was to design a Monte Carlo study considering a variety of indexes under different data conditions (sample size, variate independence, and distributional assumptions). The seven data conditions investigated by Hu and Bentler had been identified in earlier work as affecting the behavior of statistical tests (Hu, Bentler, & Kano, 1992). In addition, Hu and Bentler's examination of two-index strategies for assessing model fit was an important step toward addressing the problem that fit indexes may or may not provide complementary results.

Hu and Bentler's (1999) results led them to make suggestions for using the combination of two indexes to identify misspecified models and about tentative cut-off criteria of fit indexes for assessing model fit. Although Hu and Bentler were able to make several specific suggestions, the extent to which their findings are generalizable beyond the correct and misspecified model conditions used in their study is questionable. As Gerbing and Anderson (1993) have pointed out, the variability of potential structural models makes designing a generalizable Monte Carlo study difficult, but clearly there is a need for such a study, as noted by Marsh et al. (2004) and Fan and Sivo (2005). The possibility that fit index cut-off criteria may vary depending on the specific correct and misspecified models considered in a study needs to be studied. In the search for optimal cut-off criteria of fit indexes, both the cut-off values needed for retaining correct models and those needed for rejecting misspecified models should be considered simultaneously.

The purpose of this study was twofold: (a) to determine whether cut-off values vary depending on which model serves as the correct model, and (b) to identify which of 13 fit indexes behave optimally in terms of both retaining the correct models and simultaneously rejecting the misspecified models in a manner invariant across sample size and data distribution. We pursued four specific research questions in this study. As a partial replication of Hu and Bentler's (1999) study, we considered the original correct models used in that study and some alternative correct models similar to their models. Although we could have considered a set of correct models entirely different from those in Hu and Bentler's study, we chose to examine a family of models relevant to their study. It is our position that this is logically the first necessary step prior to the consideration of alternative models as we intend to focus on the generalizability of Hu and Bentler's conclusions. With this in mind, we considered the following research questions:

Questions for Correct Models

1. Do the fit indexes show consistent behavior across different but correctly specified models?
2. What are the highest possible fit index values (the lowest for the root mean residual [*RMR*], the standardized root mean square residual [*SRMR*], and the root mean square error of approximation [*RMSEA*]) not resulting in the rejection of any correct models, with respect to different sample size and data distribution conditions?
3. Do sample size and data distribution interact to affect the maximum cut-off values possible for correctly specified models?

Question for Misspecified Models

4. What are the lowest possible fit index values (the highest for *RMR*, *SRMR*, and *RMSEA*) resulting in the rejection of all misspecified models, with respect to different sample size and data distribution conditions?

Method

We retained the basic model structure in Hu and Bentler's (1999) study in the current study. As such, we investigated the behaviors of the fit indexes of interest across two categories of confirmatory factor analysis models. Hu and Bentler defined the first category as the *simple* model, in which each manifest variable served as an indicator of only 1 latent factor. The correct simple model had 3 correlated latent factors and 15 manifest variables. Each latent factor had 5 manifest variables as its indicators. Two misspecified simple models were created by specifying as zeros 1 or 2 covariances among the latent factors.

Hu and Bentler (1999) defined the second category as the *complex* model, which had 3 correlated latent factors and 15 manifest variables. In the correct complex model, however, 2 manifest variables served as indicators for 2 latent factors simultaneously (i.e., "double loading"). Two misspecified complex models were created by specifying 1 or 2 double-loading indicators to be single-loading indicators.

In the current study, we used the same six models (i.e., three simple and three complex models) defined in Hu and Bentler (1999); however, we employed a fully crossed design in which each of the three simple and three complex models served as the correct model, and the other models were used as misspecified models (as long as the condition of underparameterization was satisfied). In this way, the number of correct and misspecified models fitted and analyzed was greatly increased.

Models

Both the simple and complex models contained 15 manifest variables and 3 latent factors. The factor pattern matrix (see Thompson, 2004) for the three simple models was the same, as shown (transposed):

$$\begin{bmatrix} .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .00 & .00 & .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .70 & .70 & .75 & .80 & .80 \end{bmatrix}.$$

The residual variances of all manifest variables in the simple models were specified in such a way so that all the manifest variables had unit variance under normality conditions (see Hu & Bentler, 1999, and Hu et al., 1992, for more details about the models and data conditions). The three simple models differed in their respective factor covariance matrices (i.e., ϕ matrix) shown as follows:

$$\text{Simple 1: } \begin{bmatrix} 1 & & \\ .50 & 1 & \\ .40 & .30 & 1 \end{bmatrix};$$

$$\text{Simple 2: } \begin{bmatrix} 1 & & \\ .00 & 1 & \\ .40 & .30 & 1 \end{bmatrix};$$

$$\text{Simple 3:} \begin{bmatrix} 1 & & \\ .00 & 1 & \\ .00 & .30 & 1 \end{bmatrix}.$$

The complex models had the same factor covariance matrix as Simple Model 1, but the complex models differed from each other in their factor pattern matrices. The transposed factor pattern matrices for the three complex models were as follows:

$$\text{Complex 1:} \begin{bmatrix} .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .70 & .00 & .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 \\ .70 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .70 & .00 & .70 & .70 & .75 & .80 & .80 \end{bmatrix}.$$

Complex Models 2 and 3 differed by the number of factors for which the first and fourth manifest variables were specified as indicators, as follows:

$$\text{Complex 2:} \begin{bmatrix} .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .70 & .00 & .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .70 & .00 & .70 & .70 & .75 & .80 & .80 \end{bmatrix}$$

and

$$\text{Complex 3:} \begin{bmatrix} .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .00 & .00 & .70 & .70 & .75 & .80 & .80 & .00 & .00 & .00 & .00 & .00 \\ .00 & .00 & .00 & .00 & .00 & .00 & .00 & .00 & .70 & .00 & .70 & .70 & .75 & .80 & .80 \end{bmatrix}.$$

In the three complex models, the residual variances were specified in such a way so that all but three manifest variables had unit variances under normality conditions. The first, fourth, and ninth variables (i.e., the variables with multiple pattern coefficients) were given unique variances of .51, .36, and .36, respectively.

The four questions were intended to focus on the most restrictive index cut-off limits for correct models and on the least restrictive index cut-off limits for misspecified models. The emphasis on index cut-off limits is meant to delimit this study. Questions concerning the empirical distributions of these indexes are outside the purview of this inquiry. Identification of cut-off limits under the conditions of this study requires the single-minded focus on the absolute limits. Therefore, in this study, we did not consider other relevant issues, such as setting the cut-off boundary so that, say, 5% of the correct models would be rejected. Moreover, setting the cut-off boundary to 5% creates the illusion that 5% of the correct models are actually being rejected when, in fact, a 5% cut-off boundary is an artifact of the parametric conditions chosen for the design of this particular simulation study.

It was desirable to find the highest possible cut-off value (the lowest for *RMR*, *SRMR*, and *RMSEA*) not resulting in the rejection of any of the correct models.

These cut-off values are best disposed to reject as many incorrect models as possible without rejecting any correct models. Conversely, it was also desirable to find the lowest possible cut-off value (the highest for *RMR*, *SRMR*, and *RMSEA*) resulting in the rejection of all incorrect (misspecified) models. These cut-off values are best disposed to reject as few correct models as possible while rejecting all incorrect models.

In this study, underparameterized models are nested under correct models, and correct models may be nested under overparameterized models. Hu and Bentler (1999) noted that overparameterized models have zero population noncentrality. For this reason, we excluded overparameterized models from any of the correct model conditions. Table 1 shows the different models considered in this study.

To fit a correct model, sample data were generated based on the model parameters of a model in Column 1 (e.g., Simple Model 1; S1), and then the same model (i.e., S1) was fitted to its own sample data. To fit a misspecified (underparameterized) model, sample data were generated based on the model parameters of a model in Column 1 (e.g., Simple Model 1; S1), and the corresponding underparameterized models listed in Column 2 (i.e., Simple Models 2 and 3) were fitted to the sample data generated from Simple Model 1. The overparameterized models in Column 3, however, were not used in our study.

Data Generation

Each condition in this study was replicated 200 times and under 6 sample sizes (150, 250, 500, 1,000, 2,500, and 5,000). In addition, data were generat-

TABLE 1. Correct, Underparameterized (Misspecified), and Overparameterized Models

If the correctly specified (true) model is	then the underparameterized (misspecified) models are	and the overparameterized models are
Simple Model 1 (S1)	Models S2 and S3	Models C1, C2, and C3
Simple Model 2 (S2)	Model S3	Models S1, C1, C2, and C3
Simple Model 3 (S3)	No models	Models S1, S2, C1, C2, and C3
Complex Model 1 (C1)	Models S1, S2, S3, C2, and C3	No models
Complex Model 2 (C2)	Models S1, S2, S3, and C3	Model C1
Complex Model 3 (C3)	Models S1, S2, and S3	Models C1 and C2

ed under seven conditions of data distribution (SAS Institute, 1991). In Condition 1, factor scores and errors were distributed multivariate normal. Conditions 2 and 3 required the factors to deviate from normal kurtosis, whereas in Condition 3 the nonnormal factors and errors were independent. The kurtosis of the three factors in Conditions 2 and 3 was –1.0, 2.0, and 5.0. The kurtosis of the unique variates was adjusted to –1.0, 0.5, 2.5, 4.5, 6.5, –1.0, 1.0, 3.0, 5.0, 7.0, –0.5, 1.5, 3.5, 5.5, and 7.5 to help create Conditions 2 through 4. Conditions 5 through 7 required that the factors and error variates be divided by a random variable equal to $[\chi^2(5)]^{0.5}/3^{0.5}$. In Condition 5, the data were distributed elliptically with the factors and errors being uncorrelated but dependent. In Condition 6, the errors alone were distributed elliptically with the factors and errors being uncorrelated but dependent. In Condition 7, the factors and errors were multivariate normal, whereas the factors and errors were uncorrelated but dependent (see Hu et al., 1992). For a better understanding of the procedures used for data generation, see Fan, Felsovalyi, Sivo, and Keenan (2002).

The sample data were generated with each of the 6 models (Column 1 in Table 1) used as a correct model. Both the correct model (i.e., the model used for data generation) and the corresponding (underparameterized) misspecified models (Column 2 in Table 1) were then fitted to the sample data. This resulted in obtaining fit statistics for 6 correct models (Column 1 in Table 1) and 15 misspecified models (Column 2 in Table 1). This methodology resulted in 176,400 replications [21 (6 correct + 15 misspecified) models × 6 sample sizes × 7 data distribution conditions × 200 replications per cell].

Fit Statistics

We examined how 13 popular fit indexes responded to the conditions described. Several of these indexes have been studied ad nauseum, but they have not been studied within the context similar to the current design, nor for the reasons motivating this investigation. The indexes considered include the goodness-of-fit index (GFI), adjusted goodness-of fit-index (AGFI), comparative fit index (CFI), Tucker–Lewis Index (TLI) or non-normed index (NNFI; Bentler & Bonett, 1980; Tucker & Lewis, 1973), normed fit index (NFI), Bollen's Normed Index Rho1 (Bollen, 1986), Bollen's Non-normed Index Delta2 (Bollen, 1989), McDonald's Measure of Centrality (Mc; McDonald & Hartman, 1990; McDonald & Marsh, 1992), parsimonious goodness-of-fit index (PGFI), parsimonious normed fit index (PNFI), *RMR*, *SRMR*, and *RMSEA*.

Many readily available publications describe the calculations and considerations for each of the indexes used in this study. We assume that the reader has foreknowledge of these indexes. A quick overview of many of the indexes considered in this study may be found in Marsh et al. (1988).

Results

The results are organized according to the four research questions.

Questions for Correct Models

Question 1: Do the fit indexes show consistent behavior across different but correctly specified models? We conducted a 6 × 6 × 7 (Correct Model Type × Sample Size × Data Condition) factorial analysis of variance (ANOVA) to examine the sensitivity of fit indexes to different factors (model types, sample size, data conditions) for correctly specified models. For each fit index, with 200 replications per cell condition, a total of 50,400 index values were available in each factorial ANOVA.

The ANOVAs showed that, when there was no model misspecification (i.e., correct models alone considered), fit index values were statistically different (at $\alpha = .05$) across model types for the following fit indexes: GFI, AGFI, CFI, NFI, *RMR*, *SRMR*, Bollen's Rho1, PGFI, and PNFI. Nevertheless, the effect size measures (η^2 in Table 2) reveal that these statistical differences across model types are largely the result of statistical power. The actual fit index value differences across model types on inspection were negligible, except for the two parsimonious indexes (PGFI and PNFI), because the proportions of variation of fit index values contributed by model types were minimal (close to 0%). Indeed, inspection of the mean fit index values (not presented) suggests that differences occur across model types only to the thousandth place. It should be noted that differences in *SRMR* values may not be negligible, with 4% of its variation explained by model types. Notably, the PGFI and PNFI fit values were influenced by model types, accounting for 36% of the variation in PGFI values and 27% of the variation in PNFI values. This suggests that the penalty function in these parsimonious fit indexes was not consistent across the models examined in the study.

Question 2: What are the highest possible fit index values (the lowest for RMR*,* SRMR*, and* RMSEA*) not resulting in the rejection of any correct models, with respect to different sample size and data distribution conditions?* For correctly specified models, the cut-off point considered as optimal for fit indexes are the highest values (the lowest for *RMR*, *SRMR*, and *RMSEA*) at which 100% of the correct models will be retained (i.e., no Type I error). For misspecified models, these values maximize the chances of rejecting the misspecified models (i.e., minimizing Type II error). Analyses (Table 2) showed that data distribution conditions had minimal and negligible effect on the values of fit indexes. For this reason, we will ignore data distribution conditions and will focus strictly on sample size conditions.

Table 3 presents the highest values of the fit indexes (the lowest for *RMR*, *SRMR*, and *RMSEA*) that would not result in the rejection of any correct models.

TABLE 2. Sources of Variation (as Indexed by η^2) of Fit Indexes for the Correct Models

	Fit index												
Source	GFI	AGFI	CFI	NNFI	NFI	Mc	Rho1	Delta2	PGFI	PNFI	*RMR*	*SRMR*	*RMSEA*
Model types (M)	.00	.00	.00	.00	.01	.00	.01	.00	.36	.27	.00	.04	.00
Sample sze (S)	.96	.96	.23	.04	.93	.04	.93	.04	.62	.68	.27	.87	.22
Data conditions (D)	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
M × S	.00	.00	.00	.00	.01	.00	.01	.00	.00	.01	.00	.01	.00
M × D	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
S × D	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00
M × S × D	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Notes. η^2 quantifies the proportion of the total variation in a fit index contributed by a factor: $\eta^2 = SS_{source}/SS_{total}$. GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald’s Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

TABLE 3. Optimal Index Value Obtained Without Rejecting Any Correct Models

	Sample size					
Fit index	150	250	500	1,000	2,500	5,000
GFI	.89	.93	.96	.98	.99	.99
AGFI	.87	.91	.95	.97	.99	.99
CFI	.95	.97	.98	.99	.99	.99
NNFI	.95	.97	.98	.99	.99	.99
NFI	.88	.92	.96	.98	.99	.99
Mc	.87	.92	.96	.98	.99	.99
Rho1	.87	.91	.95	.97	.99	.99
Delta2	.96	.97	.98	.99	.99	.99
PGFI	.72	.75	.77	.78	.79	.79
PNFI	.72	.75	.77	.78	.79	.79
RMR	.14	.12	.11	.11	.07	.05
SRMR	.12	.10	.07	.05	.03	.03
RMSEA	.06	.05	.03	.03	.02	.01

Notes. The results displayed are across all six models fitted to their own data 200 times under each condition. High values indicate better model fit for all except the 3 indexes at the bottom (*RMR*, *SRMR*, and *RMSEA*). GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald's Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

The findings suggest that optimal cut-off values (for correct models only) may vary considerably depending on sample size, with smaller sample size resulting in lower optimal cut-off values. For example, at the lower end of sample size ($N = 150$), the optimal value of GFI for not rejecting correct models is about .90, and it may increase to .99 for the very large sample size condition (e.g., $N \geq 2{,}500$). For *RMSEA*, the optimal value is around .06 for small sample size ($N = 150$) and may decrease to .02 for very large sample size condition ($N = 2{,}500$).

Table 3 also shows that the optimal values should probably be different for some indexes even though they may appear to be on a comparable scale. For example, CFI, NNFI, and Bollen's Delta2 values for correct models are around .95 for the very small sample condition ($N = 150$), higher than some other indexes such as Rho1, Mc, and NFI.

Question 3: Do sample size and data distribution interact to affect the maximum cut-off values possible for correctly specified models? We examined the interaction effects obtained from the previous factorial ANOVA (see Table 2). Before re-

TABLE 4. Mean Fit Index Values for Correct Models, by Sample Size

	Sample size					
Fit index	150	250	500	1,000	2,500	5,000
GFI	.928	.955	.977	.988	.995	.997
AGFI	.900	.938	.968	.984	.993	.996
CFI	.993	.996	.998	.999	.999	.999
NNFI	.995	.998	.999	.999	.999	.999
NFI	.923	.953	.976	.988	.995	.997
Mc	.988	.995	.998	.999	.999	.999
Rho1	.907	.943	.971	.985	.994	.997
Delta2	.996	.998	.999	.999	.999	.999
PGFI	.764	.787	.805	.814	.820	.821
PNFI	.760	.785	.804	.813	.819	.821
RMR	.051	.039	.027	.019	.012	.008
SRMR	.070	.054	.038	.027	.017	.012
RMSEA	.017	.011	.007	.005	.003	.002

Notes. The results displayed are across all six models fitted to their own data 200 times under each condition. High values indicate better model fit for all except the 3 indexes at the bottom (*RMR*, *SRMR*, and *RMSEA*). GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald’s Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

porting any of the interaction effect results, however, it is useful to note for which fit indexes the effect of sample size was found, because the effect of sample size was by far the most prominent contributor to the variation of various fit indexes (see Table 2) for the correctly specified models. The results for all fit indexes were affected by sample size to a statistically significant degree ($p < .0001$), although the effect sizes varied greatly. Based on the η^2 results alone, NNFI, Mc, and Bollen’s Delta2 were the least affected, with sample size accounting for only 4% of the variation in the respective fit index values. Results for CFI suggest that 23% of the variation in CFI values may be explained by the sample size conditions considered. Although this percentage is sizeable, consultation of the actual mean values by sample size presented in Table 4 is more informative. It may be seen that the differences in CFI values by sample size are found in the thousandth place, so sample size influences are in reality negligible. Inspection of the means in Table 4 also suggests that sample size has a negligible impact on CFI and *RMSEA* in addition to the previously mentioned NNFI, Mc, and Bollen’s Delta2.

Table 2 presents the four interaction results for the fit indexes: (a) Model ×

Sample Size, (b) Model × Data Condition, (c) Sample Size × Data Condition, and (d) Model × Sample Size × Data Condition. The results reveal that, in general, the interactions explain a miniscule and negligible amount of variation in fit indexes, even though some of them are statistically significant.

When considering the mean index values for correct models (shown in Table 4), in each of the five model comparison scenarios, the average fit index value for the correct model was consistently larger than those of the misspecified models considered as rival hypotheses. An optimal finding for any index, in this case, would be one in which the mean fit index value of the correct model is much more discrepant from the mean fit index value of the misspecified models. This information is presented in Table 5. A review of discrepancies between the correct and misspecified model index mean values suggests that more obvious discrepancy is observed for Mc, S*RMR*, and *RMSEA*, suggesting that these indexes may do a better job at both retaining the true model and rejecting the misspecified models.

Question for Misspecified Models

Question 4: What are the lowest possible fit index values (the highest for RMR*,* SRMR*, and* RMSEA*) resulting in the rejection of all misspecified models, with respect to different sample size and data distribution conditions?* Optimal cut-off values for fit indexes are the lowest (highest for *RMR*, *SRMR*, and *RMSEA*) values needed to reject 100% of all misspecified models (no Type II error), and such optimal values minimize the chances of rejecting correct models (i.e., minimizing Type I error). Table 6 presents these values for fitting misspecified models (S1 and S3; see Table 1) to data generated from Simple Model 1.

A review of these values in Table 6 is not very consoling. Sample size affects all index cut-off values and notably not in the same way across all fit indexes. Sample size raises the optimal cut-off value in some cases and lowers it in other cases, depending on the index under consideration. Moreover, indexes that were shown to be affected minimally by sample size require a cut-off value so high that their purportedly impervious stance in the face of sample size (NNFI, Mc, Delta2, CFI; see Table 2) may be nothing more than a ceiling effect; sample size cannot affect the cut-off value when the fit value has to be so high to reject all incorrect models in the first place. McDonald's Measure of Centrality is useful in rejecting incorrect models when the sample size is set to 500 or greater. Conversely, for GFI, AGFI, NFI, and Rho1, the required cut-off values for rejecting all incorrect models are lower for smaller sample size conditions. These results suggest that sample size has a differential effect on optimal cut-off values of fit indexes for rejecting misspecified models. Because fit indexes generally do not exceed 1.00, cut-off values near 1.00 for rejecting a misspecified model are disconcerting. The .95 cut-off for Mc is sufficient when the sample size is set to 500. Considering that the lowest possible cut-off value without rejecting a correct

TABLE 5. Mean Fit Index Values for Correct and Misspecified Model Combinations

	Correct model vs. corresponding misspecified (underparameterized) models																			
	Simple 1 (TS_1) vs. incorrect S2 and S3			Simple 2 (T_S2) vs. incorrect S3		Complex 1 (T_C1) vs. incorrect S1, S2, S3, C2, and C3						Complex 2 (T_C2) vs. incorrect S1, S2, S3, and C3					Complex 3 (T_C3) vs. incorrect S1, S2, and S3			
Fit index	T_S1	S2	S3	T_S2	S3	T_C1	C2	C3	S1	S2	S3	T_C2	C3	S1	S2	S3	T_C3	S1	S2	S3
GFI	974	951	944	973	958	974	928	891	823	818	806	974	923	869	864	855	974	918	899	890
AGFI	963	933	925	964	944	964	899	849	756	751	739	964	893	820	814	804	964	886	862	852
CFI	997	969	959	998	979	998	950	908	848	792	782	998	952	887	832	828	998	930	897	892
NNFI	999	963	952	999	976	999	938	888	816	752	743	999	941	864	800	797	999	915	877	872
NFI	971	942	931	970	951	977	929	888	829	774	764	975	928	866	812	808	973	906	873	868
Mc	997	901	869	997	934	997	811	681	529	419	401	997	827	642	517	509	997	775	686	673
Rho1	965	931	919	964	942	971	912	863	793	731	722	969	913	838	776	773	967	886	848	844
Delta2	999	970	959	999	980	999	950	909	849	793	783	999	952	888	833	829	999	930	897	892
PGFI	807	797	801	816	812	780	751	730	683	682	685	789	756	720	724	725	798	760	753	755
PNFI	804	789	789	813	806	781	752	727	687	649	648	789	760	717	680	685	797	750	732	736
RMR	025	139	167	029	106	025	084	098	126	295	326	025	059	097	258	273	025	082	186	202
SRMR	035	194	233	041	150	032	090	106	139	296	337	033	071	113	270	297	034	096	226	258
RMSEA	008	048	056	008	039	008	070	095	121	141	144	008	066	101	123	123	008	077	093	094

Notes. All mean fit values are presented to the thousandth place, without a decimal. The prefix “T_” indicates that it is from a true model (i.e., correctly specified model). GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald’s Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

TABLE 6. Optimal Index Value for Rejecting All Misspecified Models Fitted to Simple Model 1 Data

	Sample size					
Fit index	150	250	500	1,000	2,500	5,000
GFI	.94	.96	.97	.98	.98	.98
AGFI	.92	.94	.96	.97	.97	.97
CFI	1.0	1.0	.99	.98	.98	.98
NNFI	1.0	1.0	.99	.98	.98	.97
NFI	.93	.95	.96	.97	.98	.98
Mc	1.0	.98	.95	.94	.93	.92
Rho1	.92	.94	.96	.97	.97	.97
Delta2	1.0	1.0	.99	.98	.98	.98
PGFI	.79	.81	.82	.82	.83	.83
PNFI	.78	.80	.81	.82	.82	.82
RMR	.02	.03	.03	.02	.02	.02
SRMR	.12	.14	.15	.16	.16	.17
RMSEA	.00	.02	.03	.04	.04	.04

Notes. High values indicate better model fit for all except the 3 indexes at the bottom (*RMR*, *SRMR*, and *RMSEA*). GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald's Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

model (see Table 3; with cut-offs increasing as sample size increases) was the same for Mc and other indexes, it becomes clear that Mc performs best for larger sample size conditions (at least 500 in this study).

Are the results for other correct model conditions (Simple Model 2, Complex Models 1, 2, and 3; see Table 1) similar to those in Table 6? Tables 7 through 10 show these results for the misspecified models fitted to data generated from other correct models (Simple Model 2, Complex Models 1, 2, and 3; see Table 1). A close look at Tables 7 through 10 shows that, in terms of the differential effect of sample size on the choice of cut-off values, the results are similar to those observed in Table 6. In terms of the possible cut-off value needed to reject all misspecified models, the cut-off value varies depending on which model's data are of concern.

For the three indexes for which a lower value indicates better model fit, all model data indicated that for *RMSEA* and *RMR* the needed cut-off value would be very low (see Tables 6 and 7). *SRMR*, on the other hand, is capable of ruling out misspecified models with a rather high cut-off value (higher than the conventional .05) for the underparameterized simple models (those misspecified

TABLE 7. Optimal Index Value for Rejecting All Misspecified Models Fitted to Simple Model 2 Data

	Sample size					
Fit index	150	250	500	1,000	2,500	5,000
GFI	.95	.97	.98	.98	.99	.99
AGFI	.93	.95	.97	.98	.98	.98
CFI	1.0	1.0	1.0	.99	.99	.99
NNFI	1.0	1.0	1.0	.99	.99	.99
NFI	.94	.96	.98	.98	.99	.99
Mc	1.0	1.0	.99	.97	.96	.96
Rho1	.93	.95	.96	.98	.98	.98
Delta2	1.0	1.0	1.0	.99	.99	.99
PGFI	.80	.82	.83	.83	.84	.84
PNFI	.80	.81	.83	.83	.84	.84
RMR	.01	.02	.01	.01	.01	.02
SRMR	.08	.09	.10	.11	.12	.12
RMSEA	.00	.00	.01	.02	.03	.03

Notes. High values indicate better model fit for all except the 3 indexes at the bottom (*RMR*, *SRMR*, and *RMSEA*). GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald’s Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

models fitted to data from Simple Models 1 and 2). This suggests that the *SRMR* is sensitive to misspecification in the measurement model components (misspecification in factor pattern matrix), a result different from the conclusion by Hu and Bentler (1999) that *SRMR* was more sensitive to misspecification in structural model components (e.g., misspecification in factor covariance matrix). Moreover, the .05 criterion for the *SRMR* is sufficient across sample size conditions, unlike other fit indexes for which a higher value indicates better model fit.

Discussion

Hu and Bentler (1999) conducted a Monte Carlo study to find the optimal criteria for fit indexes used in applied SEM research. The generalizability of Hu and Bentler’s findings, however, may be uncertain without replication involving more model conditions. Our study, for this reason, included more model conditions by (a) allowing all of Hu and Bentler’s misspecified models to have a turn as a correct model (in addition to their correct models) and (b) fitting Hu and Bentler’s complex models to all simple model datasets and vice versa (Hu and Bentler did

TABLE 8. Optimal Index Value for Rejecting All Misspecified Models Fitted to Complex Model 1 Data

	Sample size					
Fit index	150	250	500	1,000	2,500	5,000
GFI	.92	.94	.95	.95	.96	.96
AGFI	.87	.91	.92	.93	.94	.94
CFI	.98	.98	.97	.97	.96	.96
NNFI	.98	.97	.96	.96	.95	.95
NFI	.92	.94	.95	.96	.96	.96
Mc	.92	.89	.87	.85	.84	.83
Rho1	.90	.92	.94	.94	.95	.95
Delta2	.98	.98	.97	.97	.96	.96
PGFI	.74	.76	.77	.77	.78	.78
PNFI	.75	.76	.77	.77	.78	.78
RMR	.02	.02	.02	.02	.02	.02
SRMR	.07	.07	.07	.07	.07	.07
RMSEA	.04	.05	.05	.06	.06	.06

Notes. High values indicate better model fit for all except the 3 indexes at the bottom (*RMR*, *SRMR*, and *RMSEA*). GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald's Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

not make these comparisons). In the end, the purpose of this study was twofold: (a) to determine whether optimal fit index cut-off values vary according to which model serves as the correct model, and (b) to identify which fit indexes behave optimally by both retaining all correct models while simultaneously rejecting all misspecified models in a manner invariant across sample size and data distribution. To accomplish both purposes, we generated data conforming to all parametric and model conditions heretofore mentioned, and then we endeavored to answer four research questions instrumental in focusing our analysis on both purposes.

Model Invariance for Correct Models

The analytical results obtained for Question 1 supported our first purpose: determining whether optimal fit index cut-off values vary according to which model serves as the correct model. For this analysis only, we included the data generated for correct (true) models in the analysis. If the indexes perform in a model invariant manner, then fit results would be the same regardless of which model is the cor-

TABLE 9. Optimal Index Value for Rejecting All Misspecified Models Fitted to Complex Model 2 Data

	Sample size					
Fit index	150	250	500	1,000	2,500	5,000
GFI	.92	.94	.95	.95	.95	.95
AGFI	.89	.91	.93	.93	.93	.93
CFI	.99	.98	.97	.97	.96	.96
NNFI	.99	.97	.97	.96	.96	.95
NFI	.93	.94	.95	.96	.96	.96
Mc	.95	.91	.89	.87	.86	.85
Rho1	.91	.93	.94	.95	.95	.95
Delta2	.99	.98	.97	.97	.96	.96
PGFI	.75	.77	.78	.78	.78	.78
PNFI	.76	.77	.78	.79	.79	.79
RMR	.01	.01	.01	.01	.01	.01
SRMR	.06	.06	.05	.05	.05	.05
RMSEA	.03	.04	.05	.05	.06	.06

Notes. High values indicate better model fit for all except the 3 indexes at the bottom (*RMR*, *SRMR*, and *RMSEA*). GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald's Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

rect model. Our results favorably confirmed that the actual fit index value differences across model types were negligible, except for two parsimonious indexes (PGFI and PNFI) and perhaps *SRMR*. PGFI and PNFI values were influenced by model types, accounting for 36% of the variation in PGFI values and 27% of the variation in PNFI values. This suggests that the penalty function in these parsimonious fit indexes favor some correct models more than other correct models.

Results for the parsimony indexes are perhaps more academic, however, because these models are seldom used in applied SEM research. With respect to *SRMR* fit values, discrepancies among results for correct models may need to be studied further given that *SRMR* is widely used and that 4% of the variation in *SRMR* is explained in our study by which model happened to be designated as correct.

Optimal Cut-Off Values for Correct Models

The analytical results obtained for Questions 2, 3, and 4 supported our second purpose: identifying which fit indexes behave optimally in terms of both retaining

TABLE 10. Optimal Index Value for Rejecting All Misspecified Models Fitted to Complex Model 3 Data

	Sample size					
Fit index	150	250	500	1,000	2,500	5,000
GFI	.91	.93	.94	.95	.95	.95
AGFI	.87	.90	.92	.92	.93	.93
CFI	.97	.96	.95	.95	.94	.94
NNFI	.97	.95	.94	.94	.93	.93
NFI	.93	.94	.95	.96	.96	.96
Mc	.90	.87	.84	.82	.81	.80
Rho1	.87	.90	.92	.92	.93	.93
Delta2	.97	.96	.96	.95	.94	.94
PGFI	.76	.77	.78	.78	.79	.79
PNFI	.75	.76	.77	.78	.78	.78
RMR	.02	.02	.01	.01	.02	.02
SRMR	.08	.08	.07	.07	.08	.08
RMSEA	.05	.05	.06	.06	.07	.07

Notes. High values indicate better model fit for all except the 3 indexes at the bottom (*RMR*, *SRMR*, and *RMSEA*). GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; CFI = comparative fit index: NNFI = non-normed fit index; NFI = normed fit index; Mc = McDonald’s Measure of Centrality; Rho1 = Bollen's Normed Index; Delta2 = Bollen's Non-normed Index; PGFI = parsimonious goodness-of-fit index; PNFI = parsimonious normed fit index; *RMR* = root mean square residual; *SRMR* = standardized *RMR*; *RMSEA* = root mean square error of approximation.

all of the correct models while simultaneously rejecting all of the misspecified models in a manner that is invariant across both sample size and data distribution.

The results for Question 2 were used to evaluate the most stringent cut-offs possible for the indexes without rejecting any of the correct models across sample size and data distribution conditions. Taken into consideration along with Question 3, five conclusions may be drawn from the results of both questions.

Sample size. First, just as fit indexes are affected by sample size, optimal cut-off values (for correct models only) vary considerably depending on sample size, with smaller sample sizes resulting in lower optimal cut-off values. If our interest is to retain all correct models (i.e., no Type I error) while maximizing the chances of rejecting misspecified models (i.e., minimizing Type II error), then the cut-off values for all indexes could become more stringent as sample size increases from 150 to 5,000 (see Table 3). This result suggests that larger sample sizes offer more precision in identifying the correct (i.e., true) model. This finding suggests that, regardless of which index is under consideration, the cut-off values may need to become less rigorous as sample size decreases, so that we could retain all correct models while maximizing the chance of rejecting the incorrect models as rival hypotheses. Although this may seem obvious, it is important to address because the propriety of conventional cut-off values (either .90 or .95) irrespective of the fit index considered is questionable at certain sample sizes (see Table 3). To determine whether it may be meritorious to use a higher cut-off value under the condition of a larger sample size, or a lower cut-off value under the condition of a smaller sample size, it is important to consider jointly the results for the misspecified model (Question 4).

Second, a corollary conclusion based on the findings reported in Table 3 would be that cut-off values for indexes, otherwise appearing to be on a comparable scale, may well require very different optimal cut-off values. A .95 cut-off may be too high for some indexes when sample size is 150, yet it may be appropriate for other indexes. Therefore, the recommendation of .95 for any class of indexes may be inappropriate, ignoring the issue of sample size.

Third, a review of the mean fit index results for correct models in Table 4 reveals that, at least for the conditions considered in this study, certain indexes have the attribute of allowing correct models to be retained under very stringent cut-off values even when sample size is small (N = 150): CFI, NNFI, Mc, *RMSEA*, and Delta2.

Parsimonious indexes. A fourth conclusion is that the concept of optimal values may not apply to the parsimonious indexes (PGFI and PNFI), because their values are primarily influenced by model degrees of freedom. If this is the case, then no standard for these indexes exists, rendering these indexes as not interpretable for applied research. It may be argued that the applied researcher's interpretation of an index must have some standard by which to gauge the magnitude of that index. Consider just how wildly the PGFI and PNFI fit values varied depending on which model happened to be correct (see Table 2).

Interaction. Fifth, the results reveal that, in general, the interactions explain a miniscule and negligible amount of variation in fit indexes, even though some of them are statistically significant. Table 2 presents the four interaction results for the fit indexes: (a) Model × Sample Size, (b) Model × Data Condition, (c) Sample Size × Data Condition, and (d) Model × Sample Size × Data Condition.

Optimal Cut-Off Values for Incorrect Models

The results for Question 4 presented in Tables 6 through 10 were used to examine the least stringent needed cut-off values to reject all incorrect (misspecified) models. The objective was to find the least stringent cut-off values possible by which all misspecified models could be rejected (no Type II error, while minimizing Type I error).

An interesting and previously unreported finding for index behavior emerged in this study, revealing itself across Tables 6 through 10 on careful review of the optimal cut-off values for rejecting incorrect models. To appreciate this finding, it first must be recalled that one of the fundamental limitations of fit indexes is their reported sensitivity to sample size, where, as sample size increases, the index values increase as well. Many fit indexes developed as a solution to this problem have evidenced sensitivity to sample size. In this study, this was indeed true when the results for the correct models were examined in response to Question 1.

Sample Size: New and Encouraging Results for McDonald's Measure of Centrality and RMSEA

It turns out, however, that when incorrect (misspecified) models were fitted to another model's data, 3 of the 13 indexes showed an advantageous pattern of change as sample size increased, a favorable sensitivity to sample size. As sample size increased, the needed cut-off values to reject all incorrect (misspecified) models considered in this study actually decreased for 3 indexes. The cut-off values for all other indexes predictably increased with sample size, whereas Mc, *SRMR*, and *RMSEA* decreased. Unlike other indexes that would require more stringent cut-off values for rejecting all misspecified models as sample size increased, the Mc, *SRMR*, and *RMSEA* required less stringent cut-off values as sample size increased, an obviously positive finding for these 3 indexes.

For example, whereas GFI, AGFI, CFI, NNFI, NFI, Rho1, and Delta2 changed very little or in some cases increased as sample size increased, the cut-off value for Mc rapidly drops in Tables 6 through 10, suggesting that it is getting easier to reject all misspecified models. Recall that in Table 3, for retaining all correct models, the needed cut-off value gets larger with the increase of sample size. This is a desirable finding because it is more likely to locate a realistic cut-off value that performs well in both retaining the correct models and rejecting the incorrect models. A similar pattern is found for both *SRMR* and *RMSEA* when Tables 6 and 7 are examined (keep in mind that these two indexes have the opposite direction compared with Mc) and in Tables 8 through 10 (Hu and Bentler's complex models) for *RMSEA*. Curiously, *SRMR*, however, discontinues this desirable performance under the condition of structural misspecification (see Tables 8–10 for complex models).

Taking this finding into consideration, it is also encouraging that the same three indexes (Mc, *SRMR*, *RMSEA*) showed a more obvious mean index value discrepancy between correct and incorrect models (see Table 5), suggesting that these indexes may do the best job in aiding an applied researcher in distinguishing correct and incorrect (misspecified) models.

A related inference may be that the results for Mc and *RMSEA* suggest that power calculations are more likely to be optimal when based on these indexes. Consider that Kim (2005) derived power estimates using these indexes in addition to CFI and *SRMR*. This will require further investigation.

Recommendations for Future Research

No one investigation into fit indexes can be designed to handle all legitimate considerations simultaneously. In this way, the current study is no exception. This study was designed to be a partial replication of Hu and Bentler's (1999) study but, as such, does not address a number of other worthwhile issues. One plausible concern of the current investigation is whether the models chosen represent models in typical SEM applications. On the one hand, the models considered are not much different from confirmatory models in any one of a number of investigations. On the other hand, there is such a wide array of models found in SEM applications that it would be a mistake to rush to the conclusion that the current findings are generalizable to all SEM applications. Although this study is limited by the models we considered, it represents a necessary first step toward answering the question of whether different results are obtained for different models as motivated by Hu and Bentler's findings. As a partial replication, it was important to consider the models Hu and Bentler used, although alternative models must be considered in future follow-up studies.

In this study, we assume operationally that there are "correct" models (where exact fit is possible) because, as a partial replication, we were required to share certain assumptions made in Hu and Bentler's (1999) study. For the purpose of germane comparison, it was necessary for us to have "correct" models and "misspecified" models. However, SEM researchers usually consider models as useful approximations, and this standpoint should motivate future studies in which all models are misspecified but represent varying approximations of the "true" process.

Another issue that must be considered is the situation in which two or more models all fit equally well because they are mathematically equivalent. Clearly, standard fit indexes, such as those considered in this study, are inadequate in these cases because they are not designed to distinguish mathematically equivalent models. Williams, Bozdogan, and Aiman-Smith (1996) discussed how indexes such as the information complexity index are capable of distinguishing mathematically equivalent models. It turns out, even when models are mathematically equivalent, such indexes are capable of determining the degree to which the equivalent models differ in terms of how information rich they are. The area needs to be studied further as well.

REFERENCES

Bentler, P. M. (1989). *EQS, structural equations, program manual: Version 3.0*. Los Angeles: BMDP Statistical Software, Inc.

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. (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 and Research*, *17*, 303–316.

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

Fan, X., Felsovalyi, A., Sivo, S. A., & Keenan, S. (2002). *SAS for Monte Carlo studies: A guide for quantitative researchers*. Cary, NC: SAS Institute.

Fan, X., & Sivo, S. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling*, *12*, 343–367.

Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation, methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling*, *6*, 56–83.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indexes for structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 40–65). Newbury Park, CA: Sage.

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

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted. *Psychological Bulletin*, *112*, 351–362.

Kim, K. H. (2005). The relation among fit indexes, power, and sample size in structural equation modeling. *Structural Equation Modeling*, *12*, 368–390.

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

Marsh, H. W., Hau, K. T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341.

McDonald, R. P., & Hartmann, W. (1992). A procedure for obtaining initial values of parameters in the RAM model. *Multivariate Behavioral Research, 27,* 57–176.

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

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.

SAS Institute. (1996). *SAS System* (Version 6.12) [computer software]. Cary, NC: Author.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10–39). Newbury Park, CA: Sage.

Thompson, B. (2004). *Exploratory and confirmatory factor analysis: Understanding concepts and applications.* Washington, DC: American Psychological Association.

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

Williams, L. J., Bozdogan, H., & Aiman-Smith, L. (1996). Inference problems with equivalent models. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 279–314). Hillsdale, NJ: Erlbaum.