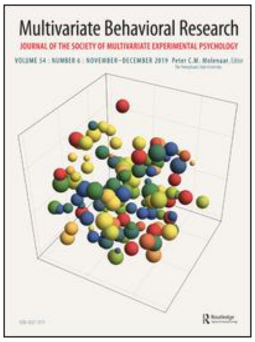


Multivariate Behavioral Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmbr20

Evaluating SEM Model Fit with Small Degrees of Freedom

Dexin Shi, Christine DiStefano, Alberto Maydeu-Olivares & Taehun Lee

MULTIVARIATE BEHAVIORAL RESEARCH
https://doi.org/10.1080/00273171.2020.1868965

Evaluating SEM Model Fit with Small Degrees of Freedom

Dexin Shi[a], Christine DiStefano[b], Alberto Maydeu-Olivares[a,c], and Taehun Lee[d]

[a]Department of Psychology, University of South Carolina; [b]Department of Educational Studies, University of South Carolina; [c]Faculty of Psychology, University of Barcelona; [d]Department of Psychology, Chung-Ang University;

ABSTRACT

Research has revealed that the performance of root mean square error of approximation (RMSEA) in assessing structural equation models with small degrees of freedom (*df*) is suboptimal, often resulting in the rejection of correctly specified or closely fitted models. This study investigates the performance of standardized root mean square residual (SRMR) and comparative fit index (CFI) in small *df* models with various levels of factor loadings, sample sizes, and model misspecifications. We find that, in comparison with RMSEA, population SRMR and CFI are less susceptible to the effects of *df*. In small *df* models, the sample SRMR and CFI could provide more useful information to differentiate models with various levels of misfit. The confidence intervals and *p*-values of a close fit were generally accurate for all three fit indices. We recommend researchers use caution when interpreting RMSEA for models with small *df* and to rely more on SRMR and CFI.

KEYWORDS
SEM; model fit; RMSEA; SRMR; CFI; degrees of freedom

Structural equation modeling (SEM) has remained a popular data analytic technique in psychology, business, education, and other disciplines (Austin & Calderón, 1996; MacCallum & Austin, 2000; Tremblay & Gardner, 1996). Often, the purpose of conducting an SEM study is to evaluate the viability of a hypothesized theoretical structure. In most practical situations, the model under consideration is, to some degree, incorrect or misspecified (Box, 1979; MacCallum, 2003). As a result, it makes sense to test whether the model has a close fit or, put differently, whether any misfit is substantively ignorable (Shi et al., 2018). To evaluate how well the data fit the theory, researchers rely upon fit information to verify that the tested model correctly approximates the theoretical underpinnings hypothesized by the researchers. Fit indices, as well as other information from the results (e.g., parameter estimates), are employed to describe the model's fit and provide support for decisions, such as altering the relationships estimated by the model, or to support removing non-performing items. Thus, researchers rely upon fit indices to provide information about modeled relationships.

Currently, one of the most widely used goodness-of-fit indices is the root mean square error of approximation (RMSEA) (Steiger, 1989, 1990; Steiger & Lind, 1980). RMSEA measures the unstandardized discrepancy between the population and the fitted model, adjusted by its degrees of freedom (*df*). Different proposals have been made as to the correct use of RMSEA.

The most common approach is to calculate and interpret the sample's RMSEA (Hancock & Mueller, 2010; McDonald & Ho, 2002). RMSEA is considered a "badness-of-fit measure," meaning that lower index values represent a better-fitting model. Often, researchers compare the sample's RMSEA with a cutoff value; often, an RMSEA value of less than 0.06 is used to denote an acceptable model (Hu & Bentler, 1999). However, interpreting the sample's RMSEA this way is a heuristic decision, basing the evaluation solely on the value of the sample statistic while ignoring its sampling variability. Often, to help assess the sampling error of the RMSEA, a confidence interval (CI) is provided with the point estimate. As an alternative, formal statistical inferences can be formed by testing the hypothesis that RMSEA $\leq k$, where k is the reference cutoff in the population. However, this (less common) approach requires identifying the

CONTACT Dexin Shi shid@mailbox.sc.edu Department of Psychology, University of South Carolina, Barnwell College, 1512 Pendleton St., Columbia, SC 29208-0001, USA; Taehun Lee lee.taehun@gmail.com Department of Psychology, Chung-Ang University, 84 Heukseok-Ro, Dongjak-Gu, Seoul, 60974, Korea.

Supplemental data for this article is available online at https://doi.org/10.1080/00273171.2020.1868965

population's RMSEA cutoff value (Browne & Cudeck, 1993).

In practice, RMSEA is a very popular fit index; however, a major drawback is that it is an unstandardized effect size. As a result, the population's RMSEA values cannot be substantively interpreted or compared across models. In addition to the extent of model misspecification, the RMSEA value depends on other characteristics of the population model (i.e., incidental factors) (Saris et al., 2009). One important incidental factor is the size of the model[1]. For example, at the population level, the same RMSEA value (e.g., 0.06) may have different meanings if the tested models have different degrees of freedom (Chen et al., 2008; Savalei, 2012).

Methodological studies have shown that RMSEA is unsuitable for assessing models with small degrees of freedom (Kenny et al., 2015; Shi et al., 2019). This situation may arise when the tested model is small, or when there are many estimated paths relative to the information available for analysis (cf. Bollen, 1989a). The RMSEA penalizes model complexity by incorporating *df* in the denominator of its formula. As a result, when the model has very small *df* (e.g., $df = 1$), the population value of RMSEA for a close-fitting model (e.g., omitting a residual correlation of 0.10) may be large, and the conventional cutoff may be misleading at the population level (Shi et al., 2019). In addition, prior studies have pointed out that sample RMSEA does not perform well when it is used to assess models with small *df* using sample data. Given identical sample sizes, the bias observed with a point estimate is larger in models with very small *df* (Kenny et al., 2015). For example, when fitting correctly specified models with very small *df*, Kenny et al. (2015) found that, as *df* decreased, the sampling variability of RMSEA increased, and sample estimates exceeded the cutoff value more frequently. The coverage of RMSEA may be acceptable when considering the CI while fitting models with small *df*, but the CI becomes wider as the *df* decreases, suggesting a greater level of uncertainty in the fit index. Prior research has suggested that RMSEA may not be useful when fitting models with very small *df*. Kenny et al. (2015, p. 486) stated, "We recommend not computing the RMSEA for small *df* models, especially those with small sample sizes."

This finding is informative, but in practice, social science researchers often encounter very small models. For example, a three-wave latent growth model has $df = 1$ (Meredith & Tisak, 1990)[2]. Models with small *df* are often observed in path analysis applications. For example, with three waves of data collected from two variables, the cross-lagged panel model (CLPM) and random intercept cross-lagged panel model (RI-CLPM) have $df = 4$ and 1, respectively (Hamaker et al., 2015). Also, the use of very short scales, to lower the cost of data collection, reduce the response burden and frustration for survey participants, and increase response rates, is commonplace (Maydeu-Olivares & Steenkamp, 2019; Robinson, 2018). As a result, short scales have become popular in many social science fields such as psychology (Ziegler et al., 2014), human resource management (Robinson, 2018), and marketing research (Bruner et al., 2005; de Jong et al., 2009). For example, short scales with four items have been widely utilized to measure global job satisfaction (Price, 1977), organizational citizenship behavior (Lee & Allen, 2002), male sexual health (Rosen et al., 2007), patient decisional conflict (Légaré et al., 2010), and depression and anxiety (Löwe et al., 2010). Unfortunately, fitting a one-factor model with four items results in a tested model with only two degrees of freedom.

Researchers need to explore the plausibility of using other fit indices to assess models with very small *df*. In this study, we focus on two potentially useful indices: the standardized root mean square residual (SRMR) (Bentler, 1995; Jöreskog & Sörbom, 1988) and comparative fit index (CFI) (Bentler, 1990). We selected these indices as they have been routinely reported in most SEM software (e.g., Mplus) and widely used in empirical SEM applications (McDonald & Ho, 2002). In addition, the theoretical sampling distributions of both SRMR and CFI have been derived using asymptotic methods (Lai, 2019a; Maydeu-Olivares, 2017; Ogasawara, 2001a, 2001b, 2007). Researchers may construct the CIs (or conduct significance tests; e.g., SRMR $\leq k$) for population effect sizes of misfit, which makes SRMR and CFI comparable to RMSEA.

Prior model fit studies have shown that compared to the population RMSEA, the population SRMR and

[1]Researchers have used different indices to indicate the size of an SEM model, such as the number of observed variables (*p*), the number of estimated parameters (*q*), the degrees of freedom (*df*), and the number of observed variables per latent factor (*p/f*; Shi, Lee, et al., 2015, 2018). Although the above indicators are different concepts, in many cases, they tend to vary together. For example, a larger number of indicators (*p*) often resulted in a larger *df*. In this study, we refer to small models as models with small *df*.

[2]When fitting a three-wave latent growth model, the number of sample moments = 6 covariances) + 3 (means) = 9, and the number of free parameters = 8 (2 factor variances + 2 factor means + 1 covariance between factors + 3 residual variances). Therefore, $df = 9 - 8 = 1$.

CFI are less susceptible to the influence of model sizes (Kenny & McCoach, 2003; Maydeu-Olivares, 2017; Shi, Maydeu-Olivares, & DiStefano, 2018; Shi et al., 2019). Several simulation studies have investigated the behavior of the sample values of SRMR and CFI (e.g., Fan & Sivo, 2005, 2007; Hu & Bentler, 1999; Sharma et al., 2005) under various levels and types of model (mis)specifications. Concerning the model size, however, most studies focus on models with at least 10–15 observed variables; thus, it is not clear whether the findings can be generalized to models with very small *df*.

In terms of the CIs and close fit tests, Maydeu-Olivares et al. (2018) compared the performance of RMSEA and SRMR under models with *df* ranging from 35 to 1,710. They found that the CIs for SRMR yielded better coverage of its population value when assessing models with large *df* (e.g., $df > 400$). Lai (2019a) investigated the performance of CIs for CFI and found that the CIs were generally accurate for models with *df* ranging from 24 to 119. While these studies have provided information on the viability of SRMR and CFI in large models, details of their performance in very small models remain unknown.

To fill this research gap, we compared the performance of RMSEA, SRMR, and CFI in assessing very small models. We considered various levels of model (mis)specifications (i.e., correctly specified, close fitting, and severely misspecified) and sample sizes. To assist empirical researchers, we focused on population values, sample point estimates, interval estimates (i.e., 90% CIs), and close fit tests for all three fit indices. First, we reviewed the statistical theories underlying RMSEA, SRMR, and CFI. Second, we presented a simulation study that compares the performances of the indices in small *df* models, along with an empirical example. Finally, we discussed the implications and provide recommendations.

The statistical theory underlying RMSEA, SRMR, and CFI

Root mean square error of approximation

The population RMSEA is defined as (Browne & Cudeck, 1993):

$$\mathrm{RMSEA} = \sqrt{\frac{\tilde{F}}{df}} \tag{1}$$

where $\tilde{F}$ denotes the minimum of a discrepancy function between the population covariance matrix, Σ, and the model implied covariance matrix, $\mathbf{\Sigma}_0$, for the postulated model, and *df* denotes the *df* for the tested model. From Equation (1), it is easy to show that for a fixed (nonzero) value of $\tilde{F}$, the population RMSEA increases as *df* decreases. It is also noteworthy that the population value of RMSEA depends on the estimation method employed. When maximum likelihood (ML) is used, the population RMSEA is

$$\mathrm{RMSEA}_{\mathrm{ML}} = \sqrt{\frac{\tilde{F}_{\mathrm{ML}}}{df}}. \tag{2}$$

Assuming no mean structure is present[3], $\tilde{F}_{\mathrm{ML}}$ is obtained by minimizing

$$F_{\mathrm{ML}} = \ln|\mathbf{\Sigma}_0| - \ln|\mathbf{\Sigma}| + \mathrm{tr}\left(\mathbf{\Sigma}\mathbf{\Sigma}_0^{-1}\right) - p, \tag{3}$$

where p denotes the number of observed variables. Given sample data and assuming normality, the population RMSEA can be estimated as

$$\widehat{\mathrm{RMSEA}}_{\mathrm{ML}} = \sqrt{\max\left(\frac{X^2_{\mathrm{ML}} - df}{N \times df}, 0\right)}, \tag{4}$$

where N denotes sample size. X^2_{ML} is the ML-based likelihood-ratio chi-squared test statistics.

A 90% CI for $\mathrm{RMSEA}_{\mathrm{ML}}$ can be constructed as

$$\sqrt{\max\left(0, \frac{L}{N \times df}\right)}, \sqrt{\max\left(0, \frac{U}{N \times df}\right)}, \tag{5}$$

where L and U are the solutions to

$$F_{\chi^2}(X^2_{\mathrm{ML}}; df, L) = .95 \text{ and } F_{\chi^2}(X^2_{\mathrm{ML}}; df, U) = .05 \tag{6}$$

and $F_{\chi^2}(\cdot; df, \lambda)$ denotes the cumulative non-central chi-squared distribution with *df* degrees of freedom and non-centrality parameter λ (Browne & Cudeck, 1993). Finally, the *p*-value for a close fit test with a population cutoff k $(H_0^* : RMSEA \le k)$ can be obtained using

$$1 - F_{\chi^2}\left(X^2_{\mathrm{ML}}; df, N \times df \times k^2\right). \tag{7}$$

All the notations are the same as above. The RMSEA is a "badness" measure of fit: Lower RMSEA values indicate better fit. In practice, RMSEA $\le .06$ is the most commonly used cutoff for acceptable fit (Hu & Bentler, 1999).

Standardized root mean square residual

The SRMR in the population is defined as (Jöreskog & Sörbom, 1988; Maydeu-Olivares, 2017):

[3]We focused on "classical" covariance structure models without mean structures.

$$\text{SRMR} = \sqrt{\frac{\varepsilon'_s\varepsilon_s}{t}} = \sqrt{\frac{1}{t}\sum_{i\leq j}\left[\frac{\left(\boldsymbol{\sigma}_{ij}-\boldsymbol{\sigma}^0_{ij}\right)}{\sqrt{\boldsymbol{\sigma}_{ii}\boldsymbol{\sigma}_{jj}}}\right]^2}, \tag{8}$$

where $\boldsymbol{\sigma}_{ij}$ denotes the unknown population covariance between variables i and j (or the variance, if $i = j$) and $\boldsymbol{\sigma}^0_{ij}$ denotes the population covariance (or variance) under the tested model. ε_s is the vector of the population standardized residual covariances, and $t = p(p+1)/2$ signifies the number of unique elements in the residual covariance matrix (i.e., the difference between the observed and model-implied covariance matrices) where p denotes the number of observed variables. Thus, the population value of SRMR can be approximately interpreted as the average population standardized residual covariance.

In finite samples, let s_{ij} be the sample covariance, $\hat{\boldsymbol{\sigma}}_{ij}$ the model implied covariance, and $\mathbf{e}_s$ the t vector of the standardized residual covariances with elements $\frac{s_{ij}-\hat{\boldsymbol{\sigma}}_{ij}}{\sqrt{s_{ii}s_{jj}}}$. A sample counterpart of the population SRMR in Equation (8) can be estimated using

$$\widehat{\text{SRMR}}_b = \sqrt{\frac{\mathbf{e}_s'\mathbf{e}_s}{t}} = \sqrt{\frac{1}{t}\sum_{i\leq j}\left[\frac{\left(s_{ij}-\hat{\boldsymbol{\sigma}}_{ij}\right)}{\sqrt{s_{ii}s_{jj}}}\right]^2}. \tag{9}$$

The sample SRMR value is the value typically produced in SEM software packages and evaluated by researchers to assess a model's fit[4].

Maydeu-Olivares (2017) showed that the sample estimates for SRMR are biased, and an asymptotically unbiased estimate of the population SRMR can be expressed as follows:

$$\widehat{\text{SRMR}}_u = \hat{k}_s^{-1}\sqrt{\frac{\max\left(\mathbf{e}_s'\mathbf{e}_s - \text{tr}(\hat{\Xi}_s), 0\right)}{t}}, \tag{10}$$

where $\hat{k}_s = 1 - \frac{\text{tr}(\hat{\Xi}_s^2)+2\mathbf{e}'_s\hat{\Xi}_s\mathbf{e}_s}{4(\mathbf{e}'_s\mathbf{e}_s)^2}$ and Ξ_s represents the asymptotic covariance matrix of $\mathbf{e}_s$. This study investigates the performance of both the biased ($\widehat{\text{SRMR}}_b$) and unbiased ($\widehat{\text{SRMR}}u$) estimates of SRMR.

CIs for the SRMR and close fit tests can be obtained using a reference normal distribution. Specifically, with large samples, a (1–α)% CI for the SRMR can be computed as

$$\max\left[0, \left(\widehat{\text{SRMR}}_u - z_{\alpha/2}\text{SE}(\widehat{\text{SRMR}}_u)\right)\right],$$
$$\max\left[0, \left(\widehat{\text{SRMR}}_u + z_{\alpha/2}\text{SE}(\widehat{\text{SRMR}}_u)\right)\right], \tag{11}$$

where $z_{\alpha/2}$ is the (two-tailed) critical z value for the given α level, and SE () denotes asymptotic standard error, given as (Maydeu-Olivares, 2017):

$$SE(\text{SRMR}_u) = \sqrt{k_s^{-2}\frac{\text{tr}(\Xi_s^2)+2\mathbf{e}'_s\Xi_s\mathbf{e}_s}{2t\mathbf{e}'_s\mathbf{e}_s}}. \tag{12}$$

In addition, a statistical test for model close fit can be conducted using the hypotheses H_0 : SRMR $\leq k$ vs. H_1 : SRMR $> k$, where $k > 0$ is a reference cut-off value for close fit at the population level. The p-values are obtained using $p = 1 - \Phi(z)$, where $\Phi()$ denotes a standard normal distribution function and

$$z = \frac{\widehat{\text{SRMR}}_u - k}{\text{SE}(\widehat{\text{SRMR}}_u)}. \tag{13}$$

Similar to RMSEA, SRMR also measures the "badness" of model fit. The most commonly used criterion for a good fit is SRMR $\leq$ .08 (Hu & Bentler, 1999). Recently, Shi et al. (2018) proposed new criteria by considering the impact of the measurement quality. Specifically, the cutoffs for good fit and acceptable fit were SRMR $\leq .05\times\overline{R^2}$ and SRMR $\leq .10\times\overline{R^2}$, respectively. $\overline{R^2}$ is the average communality (i.e., the squared standardized loading) across all items.

Comparative fit index

The population CFI is defined as (Bentler, 1990):

$$\text{CFI} = 1 - \frac{F_m}{F_b} \tag{14}$$

where F_m and F_b represent the minimum of the discrepancy function for the researcher's proposed model and baseline model, respectively. Therefore, the CFI measures the relative improvement in fit going from the baseline model to the postulated model.

The population values of CFI also depend on the model estimation method. The population ML fit function is given in Equation (3). When using ML, the sample CFI can be estimated as[5]:

[4]We focused on models with a saturated mean structure and follow the sample SRMR formula (Equation (9)) computed in many widely used SEM software programs (e.g., LISREL, Jöreskog & Sörbom, 2017; EQS, Bentler, 2004, and *lavaan*, Rosseel, 2012). However, in Mplus (Muthén & Muthén, 2017), by default, the sample SRMR is computed as follows:

$$\widehat{\text{SRMR}}_{\text{Mplus}} = \sqrt{\frac{1}{t+p}\left(\sum_{i\leq j}\left(\hat{\varepsilon}^*_{ij}\right)^2 + \sum_i\left(\hat{\varepsilon}^*_i\right)^2\right)}$$

$$\hat{\varepsilon}^*_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}s_{jj}}} - \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}\hat{\varepsilon}^*_i = \frac{m_i}{\sqrt{s_{ii}}} - \frac{\hat{\mu}_i}{\sqrt{\hat{\sigma}_{ii}}}$$

where m_i and $\hat{\boldsymbol{\mu}}_i$ denote the sample and expected mean of variable i, respectively. Mplus users can estimate the sample SRMR defined in Equation (9) by using "MODEL = NOMEANSTRUCTURE" in the ANALYSIS command (Asparouhov & Muthén, 2018; Pavlov et al., 2020).

[5]In this study, $\widehat{\text{CFI}}_{\text{ML}}$ (defined in Equation (15)) is the usual point estimator of the population CFI reported in most SEM software when ML is used, which is different from Equation (8) in Lai (2019a).

$$\widehat{\mathrm{CFI}}_{\mathrm{ML}} = \frac{\max({\chi_b}^2 - df_b, 0) - \max({\chi_m}^2 - df_m, 0)}{\max({\chi_b}^2 - df_b, 0)}, \tag{15}$$

where χ_b^2 and df_b denote the likelihood-ratio chi-squared test statistics and the corresponding df for the baseline model, respectively, and χ_m^2 and df_m represent the chi-squared test statistic and df for the proposed model, respectively.

Lai (2019a) proposed two new consistent point estimators for CFI directly based on the fit function ($\hat{F}_m$) and recommended a bias-corrected sample estimate of CFI ($\widehat{\mathrm{CFI}}_{\mathrm{FBC}}$). Let $\mathbf{s}$ and $\boldsymbol{\sigma}$ be the $t \times 1$ vector of the unique elements in $\mathbf{S}$ and $\boldsymbol{\Sigma}$, respectively. $\widehat{\mathrm{CFI}}_{\mathrm{FBC}}$ is expressed as

$$\widehat{\mathrm{CFI}}_{\mathrm{FBC}} = 1 - \frac{\hat{F}^{(m)}_{\mathrm{MLBC}}}{\hat{F}^{(b)}_{\mathrm{MLBC}}}; \tag{16}$$

$$\hat{F}_{\mathrm{MLBC}} = \hat{F} - (2N)^{-1} tr[\ddot{\phi}\hat{\boldsymbol{\Gamma}}], \tag{17}$$

where $\ddot{\phi}$ is the second derivative of ϕ, which is a function of $\mathbf{s}$ (i.e., $\hat{F}_m = \phi(\mathbf{s})$), evaluated at $\mathbf{s} = \boldsymbol{\sigma}$ and $\boldsymbol{\theta} = \boldsymbol{\theta}$ ($\boldsymbol{\theta}^*$ denotes the population model parameter), and $\hat{\boldsymbol{\Gamma}}$ is the sample estimate for the asymptotic covariance matrix of $\sqrt{N}\mathbf{s}$.

Under a normal reference distribution, a (1–α) % CI for the CFI can be computed as

$$\min\left[1, \left(\widehat{\mathrm{CFI}} - z_{\alpha/2}\mathrm{SE}(\widehat{\mathrm{CFI}})\right)\right],$$

$$\min\left[1, \left(\widehat{\mathrm{CFI}} + z_{\alpha/2}\mathrm{SE}(\widehat{\mathrm{CFI}})\right)\right]. \tag{18}$$

In Equation (18), $z_{\alpha/2}$ is the (two-tailed) critical z value for the given α level, $\widehat{\mathrm{CFI}}$ is a sample estimate of CFI, and SE () is asymptotic standard error. Lai (2019a) derived two forms of $\mathrm{SE}(\widehat{\mathrm{CFI}})$ and recommended using

$$\mathrm{SE}(\widehat{\mathrm{CFI}}) = \sqrt{\hat{\mathbf{g}}'\hat{\boldsymbol{\Gamma}}\hat{\mathbf{g}}/N}, \tag{19}$$

where $\hat{\mathbf{g}}$ is computed as the first-order derivative of the function of CFI with respect to $\mathbf{s}$ ($\partial\varphi/\partial\mathbf{s}$) evaluated at $\mathbf{s} = \mathbf{s}$ and $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ ($\hat{\boldsymbol{\theta}}$ denotes the sample estimates of the model parameters).

A statistical test for acceptable fit (with a population cutoff k) can be conducted under the hypotheses H_0 : CFI $\geq k$ vs. H_1 : CFI $< k$. The p values are obtained under a standard normal distribution function, $\boldsymbol{\Phi}()$, as $p = 1 - \boldsymbol{\Phi}(z)$, where z is expressed as

$$z = \frac{\widehat{\mathrm{CFI}} - k}{\mathrm{SE}(\widehat{\mathrm{CFI}})}. \tag{20}$$

All the notations are the same as above. The values of CFI are normed between 0 and 1: higher values of CFI indicate better model fit. In practice, the most widely used criterion for a good fit is CFI $\geq$ .95 (Hu & Bentler, 1999).

We provided the theory underlying RMSEA, SRMR, and CFI to orient readers to the similarities and differences among the indices. All three indices are often employed to evaluate models, and researchers compute both point estimates and CIs for each index. Further study of the three indices can aid applied researchers interested in model testing.

Monte Carlo simulations

We performed simulations to investigate the behaviors of RMSEA, SRMR, and CFI when assessing models with small *df*. We considered scenarios with both correctly specified and misspecified models. In correctly specified models, the population model was a one-factor confirmatory factor analysis (CFA) model and the same model was fit to the data. The population model in misspecified conditions was a two-factor CFA model, but a one-factor model was fit to the data. The population factor variance(s) were set to 1.0. To create the different simulation conditions, we manipulated four variables: model size, size of the factor loadings, magnitude of model (mis)specification, and sample size.

Model size

We manipulated model size by changing p and the number of observed variables (Moshagen, 2012; Shi, Lee, et al., 2015, 2018). In this study, we focused on small models with ps of 4, 8, and 12. For the population model with two factors, each factor had an equal number of items loaded. As we estimated the one-factor model, the *df*s for the fitted model ranged from 2 ($p = 4$) to 54 ($p = 12$).

Size of factor loadings

Three levels of factor loadings were considered: low ($\lambda = .40$), medium ($\lambda = .60$), and high ($\lambda = .80$). The error variances were set as $1 - \lambda^2$ so that the factor loadings were standardized.

Sample size

Sample sizes were 50, 100, 200, 500, and 1,000. These numbers were chosen to represent the range of very small to large samples frequently observed in psychological and behavioral research.

Table 1. Population values and average sample estimates: RMSEA.

			Correctly specified models $\rho = 1.0$			Close fit models $\rho = 0.9$				Severely misspecifed models $\rho = 0.6$			
N	λ	df	pop.	mean	sd	pop.	mean	sd	c.v.	pop	mean	sd	c.v.
50	.4	2	0	0.030	0.058	0.015	0.030	0.056	3.73	0.058	0.028	0.056	0.97
		20	0	0.039	0.043	0.011	0.039	0.043	3.91	0.040	0.043	0.044	1.10
		54	0	0.047	0.036	0.010	0.047	0.035	3.50	0.036	0.054	0.037	1.03
	.6	2	0	0.046	0.072	0.044	0.051	0.075	1.70	0.154	0.099	0.098	0.64
		20	0	0.044	0.045	0.030	0.050	0.046	1.53	0.101	0.097	0.050	0.50
		54	0	0.048	0.036	0.028	0.054	0.036	1.29	0.088	0.096	0.034	0.39
	.8	2	0	0.051	0.075	0.126	0.099	0.101	0.80	0.374	0.325	0.109	0.29
		20	0	0.044	0.045	0.084	0.084	0.051	0.61	0.228	0.221	0.043	0.19
		54	0	0.048	0.036	0.074	0.085	0.035	0.47	0.190	0.191	0.030	0.16
100	.4	2	0	0.026	0.045	0.015	0.025	0.044	2.93	0.058	0.029	0.048	0.83
		20	0	0.025	0.029	0.011	0.026	0.029	2.64	0.040	0.036	0.033	0.83
		54	0	0.025	0.024	0.010	0.026	0.024	2.40	0.036	0.038	0.026	0.72
	.6	2	0	0.033	0.052	0.044	0.041	0.056	1.27	0.154	0.120	0.079	0.51
		20	0	0.026	0.030	0.030	0.034	0.032	1.07	0.101	0.096	0.033	0.33
		54	0	0.025	0.024	0.028	0.033	0.026	0.93	0.088	0.088	0.022	0.25
	.8	2	0	0.033	0.052	0.126	0.102	0.080	0.63	0.374	0.346	0.072	0.19
		20	0	0.026	0.030	0.084	0.079	0.035	0.42	0.228	0.222	0.029	0.13
		54	0	0.025	0.024	0.074	0.074	0.023	0.31	0.190	0.188	0.020	0.11
200	.4	2	0	0.021	0.034	0.015	0.022	0.035	2.33	0.058	0.033	0.042	0.72
		20	0	0.016	0.020	0.011	0.018	0.021	1.91	0.040	0.035	0.025	0.63
		54	0	0.015	0.016	0.010	0.016	0.017	1.70	0.036	0.034	0.018	0.50
	.6	2	0	0.023	0.036	0.044	0.035	0.044	1.00	0.154	0.136	0.056	0.36
		20	0	0.017	0.020	0.030	0.028	0.024	0.80	0.101	0.098	0.021	0.21
		54	0	0.015	0.016	0.028	0.027	0.018	0.64	0.088	0.088	0.013	0.15
	.8	2	0	0.023	0.036	0.126	0.110	0.060	0.48	0.374	0.356	0.050	0.13
		20	0	0.017	0.020	0.084	0.081	0.022	0.26	0.228	0.224	0.020	0.09
		54	0	0.015	0.016	0.074	0.073	0.014	0.19	0.190	0.188	0.014	0.07
500	.4	2	0	0.014	0.023	0.015	0.017	0.025	1.67	0.058	0.044	0.035	0.60
		20	0	0.010	0.013	0.011	0.012	0.014	1.27	0.040	0.037	0.015	0.38
		54	0	0.008	0.010	0.010	0.011	0.011	1.10	0.036	0.035	0.009	0.25
	.6	2	0	0.015	0.023	0.044	0.035	0.033	0.75	0.154	0.148	0.034	0.22
		20	0	0.010	0.013	0.030	0.027	0.016	0.53	0.101	0.100	0.012	0.12
		54	0	0.008	0.010	0.028	0.026	0.011	0.39	0.088	0.088	0.008	0.09
	.8	2	0	0.015	0.023	0.126	0.121	0.035	0.28	0.374	0.365	0.032	0.09
		20	0	0.010	0.013	0.084	0.083	0.012	0.14	0.228	0.226	0.013	0.06
		54	0	0.009	0.010	0.074	0.073	0.008	0.11	0.190	0.189	0.009	0.05
1000	.4	2	0	0.010	0.016	0.015	0.014	0.019	1.27	0.058	0.050	0.026	0.45
		20	0	0.007	0.009	0.011	0.010	0.010	0.91	0.040	0.039	0.009	0.23
		54	0	0.006	0.007	0.010	0.009	0.008	0.80	0.036	0.036	0.006	0.17
	.6	2	0	0.010	0.016	0.044	0.037	0.026	0.59	0.154	0.151	0.024	0.16
		20	0	0.007	0.009	0.030	0.029	0.010	0.33	0.101	0.101	0.008	0.08
		54	0	0.006	0.007	0.028	0.027	0.006	0.21	0.088	0.088	0.005	0.06
	.8	2	0	0.011	0.016	0.126	0.124	0.024	0.19	0.374	0.368	0.023	0.06
		20	0	0.007	0.009	0.084	0.083	0.008	0.10	0.228	0.227	0.009	0.04
		54	0	0.006	0.007	0.074	0.073	0.005	0.07	0.190	0.189	0.006	0.03

Note. N = sample size; λ = standardized factor loadings; *df* = degrees of freedom; ρ = interfactor correlations; pop. = the population values; mean = the average sample estimates across replications; sd = the standard deviations of the sample estimates across replications; c.v. = the coefficient of variation.

Magnitude of model (mis)specification

Since model misspecification was introduced by ignoring the multidimensional structure, the magnitude of model misfit was manipulated using different ρs in the population model. Specifically, under misspecified conditions, the population ρs included 0.60 and 0.90; a smaller correlation coefficient indicated greater misspecification. Note that the population model under the correctly specified conditions was a one-factor model, which is equivalent to a two-factor model with a perfect inter-factor correlation (i.e., $\rho = 1.0$).

We considered a range of model misspecifications: correctly specified ($\rho = 1.0$), minor ($\rho = 0.90$), and severe ($\rho = 0.60$) (Shi, Maydeu-Olivares, & DiStefano, 2018). For example, when fitting a one-factor model to two-factor data with an inter-factor correlation of $\rho = 0.90$, most researchers would consider the misfit ignorable. However, when the population model has two factors with $\rho = 0.60$, the one-factor model cannot be overlooked and should be rejected based on the model fit information.

In summary, we considered 135 simulated conditions: 5 (sample sizes) × 3 (model sizes) × 3 (factor loading levels) × 3 (factor inter-correlations). For each simulated condition, 5,000 replications were generated with the *simsem* package in R

(Pornprasertmanit et al., 2012; R Development Core Team, 2015). The observed data were generated using a multivariate normal distribution.

In each simulation condition, first, we computed the population values for RMSEA, SRMR, and CFI by fitting the one-factor CFA models to the population covariance matrix. We then fit a one-factor model to each simulated data set and computed the sample's RMSEA, SRMR, and CFI. This provided the empirical distributions of the sample fit indices with over 5,000 replications. Under the model misspecification conditions, we computed 90% CIs for RMSEA, SRMR, and CFI. We also computed the p-value of a close fit test (when the RMSEA and SRMR were less than or equal to their population values and the CFI was greater than or equal to its population values). We conducted all data analyses with ML estimation using the *lavaan* package in R (R Development Core Team, 2015; Rosseel, 2012).

For all three fit indices, we reported population values under different model sizes, factor loading levels, and magnitudes of model misspecification. To better summarize the empirical distributions of the sample estimates, for each simulated condition, we computed the empirical means and standard deviations of the sample RMSEA, SRMR, and CFI (across 5,000 replications). We also calculated the proportion of replications for each index that would be rejected based on the conventional cutoff values for acceptable fit (Hu & Bentler, 1999; Shi, Maydeu-Olivares, & DiStefano, 2018). In terms of interval estimates, we computed the average and median width of the 90% CIs for each fit index. The accuracy of these CIs was determined by computing their coverage rates – how often the population value was within the CI. Finally, we evaluated the accuracy of the p-values for the close fit tests by computing the Type I error rates (i.e., $\alpha = 0.05$). Specifically, we calculated the percentage of replications in which the null hypothesis of close fit[6] was rejected.

Results

In our study, not every replication converged. Low convergence rates resulted when both the df and sample size were very small (e.g., $df = 2$ and $N = 50$), factor loadings were low (i.e., $\lambda = .40$), and the model was severely misspecified (i.e., $\rho = 0.60$). For that specific condition, the convergence rate was 57%. The complete results of the convergence rates are provided in the supplementary materials. Only cases that converged were included when calculating the outcome variables. We also conducted analyses of variance (ANOVA) on selected outcome measures. Specifically, an eta-squared (η^2) value above 5% was used to identify the main conditions that contributed to practically sizeable variability in the outcome (Shi, Maydeu-Olivares, & DiStefano, 2018)[7].

Population values

The population values for RMSEA, SRMR, and CFI across simulation conditions are reported in Tables 1–3. Figures 1–3 plot the population RMSEAs, SRMRs, and CFIs against the *df*s across the different magnitudes of model misspecification and factor loading levels. When the model was correctly specified, the population RMSEA and SRMR were both equal to zero and the population CFI was equal to one. Under the misspecified condition, the population RMSEAs ranged from 0.010 to 0.374, the population SRMRs from 0.006 to 0.117, and the population CFIs from 0.748 to 0.996. In general, the population RMSEA and SRMR increased and the population CFI decreased as the level of model misspecification increased, and the magnitude of factor loadings increased, suggesting a worse fit. In addition, the effect of factor loadings was more pronounced when the level of model misfit is more severe.

The ANOVA results showed that the most important sources of variability in the population RMSEA were the levels of factor loadings (λ; $\eta^2 = 0.48$), magnitudes of model misspecification (ρ; $\eta^2 = 0.27$), the interaction between λ and ρ ($\lambda \times \rho$; $\eta^2 = 0.10$), and model size (df; $\eta^2 = 0.07$). For a certain magnitude of model misspecification, the population RMSEA decreased dramatically as the model size increased and level of factor loadings decreased. For example, when fitting a one-factor model to two-factor data with $\rho = 0.60$ and $\lambda = 0.60$, as the df increased from 2 ($p = 4$) to 54 ($p = 12$), the population RMSEA fell from 0.154 to 0.088. Meanwhile, for a fixed level of model misfit and model size (e.g., $\rho = 0.60$ and $df = 2$), as the standardized factor loadings decreased from 0.80 to 0.40, the population RMSEA dropped from 0.374 to 0.058.

The most important factors that affect the population values of SRMR were the magnitudes of model misspecification (ρ; $\eta^2 = 0.50$), levels of factor loadings (λ; $\eta^2 = 0.36$), and their interaction ($\lambda \times \rho$; $\eta^2 =$

[6]The RMSEA and SRMR were less than or equal to their population values and the CFI was greater than or equal to its population value.

[7]The ANOVA tables are also available in the supplementary materials.

Table 2. Population values and average sample estimates: SRMR.

			Correctly specified models $\rho = 1.0$					Close fit models $\rho = 0.9$							Severely misspecifed models $\rho = 0.6$						
				SRMR.b		SRMR.u			SRMR.b			SRMR.u				SRMR.b			SRMR.u		
N	λ	df	pop.	mean	sd	mean	sd	pop.	mean	sd	c.v.	mean	sd	c.v.	pop.	mean	Sd	c.v.	mean	sd	c.v.
50	.4	2	0	0.042	0.021	0.014	0.025	0.006	0.042	0.021	3.53	0.013	0.025	4.12	0.023	0.042	0.022	0.96	0.013	0.025	1.10
		20	0	0.087	0.014	0.018	0.025	0.007	0.088	0.014	2.01	0.019	0.024	3.49	0.028	0.092	0.015	0.53	0.022	0.027	0.98
		54	0	0.099	0.010	0.018	0.021	0.007	0.100	0.010	1.44	0.018	0.022	3.10	0.029	0.105	0.011	0.38	0.026	0.025	0.88
	.6	2	0	0.038	0.020	0.016	0.025	0.013	0.040	0.021	1.62	0.018	0.026	2.03	0.053	0.059	0.029	0.55	0.040	0.039	0.73
		20	0	0.068	0.013	0.016	0.020	0.016	0.072	0.013	0.83	0.020	0.022	1.38	0.063	0.097	0.021	0.33	0.056	0.033	0.52
		54	0	0.076	0.009	0.014	0.017	0.016	0.080	0.010	0.63	0.019	0.019	1.19	0.066	0.106	0.017	0.26	0.061	0.027	0.42
	.8	2	0	0.022	0.012	0.010	0.015	0.023	0.033	0.018	0.78	0.022	0.023	0.99	0.094	0.102	0.036	0.38	0.099	0.038	0.40
		20	0	0.039	0.009	0.009	0.012	0.028	0.051	0.012	0.44	0.025	0.018	0.65	0.112	0.130	0.034	0.31	0.119	0.036	0.33
		54	0	0.043	0.007	0.008	0.010	0.029	0.056	0.010	0.36	0.027	0.015	0.52	0.117	0.136	0.032	0.28	0.123	0.034	0.29
100	.4	2	0	0.032	0.016	0.012	0.020	0.006	0.031	0.016	2.73	0.012	0.020	3.29	0.023	0.034	0.017	0.75	0.014	0.022	0.95
		20	0	0.062	0.010	0.014	0.018	0.007	0.063	0.010	1.44	0.014	0.018	2.57	0.028	0.069	0.011	0.41	0.023	0.022	0.78
		54	0	0.070	0.007	0.012	0.015	0.007	0.071	0.007	1.01	0.014	0.016	2.26	0.029	0.077	0.008	0.28	0.025	0.019	0.67
	.6	2	0	0.026	0.014	0.011	0.017	0.013	0.029	0.015	1.19	0.015	0.020	1.51	0.053	0.057	0.023	0.44	0.047	0.030	0.57
		20	0	0.048	0.008	0.011	0.014	0.016	0.052	0.009	0.56	0.016	0.016	1.01	0.063	0.081	0.016	0.25	0.060	0.022	0.35
		54	0	0.053	0.006	0.010	0.012	0.016	0.057	0.007	0.42	0.015	0.014	0.88	0.066	0.088	0.013	0.20	0.064	0.018	0.27
	.8	2	0	0.015	0.008	0.006	0.010	0.023	0.028	0.013	0.57	0.022	0.017	0.73	0.094	0.102	0.027	0.28	0.101	0.027	0.29
		20	0	0.027	0.005	0.006	0.008	0.028	0.040	0.009	0.31	0.026	0.013	0.45	0.112	0.123	0.026	0.23	0.118	0.026	0.24
		54	0	0.030	0.004	0.006	0.007	0.029	0.044	0.007	0.25	0.028	0.010	0.35	0.117	0.128	0.024	0.20	0.122	0.024	0.21
200	.4	2	0	0.023	0.012	0.009	0.015	0.006	0.024	0.012	2.06	0.010	0.015	2.57	0.023	0.029	0.014	0.63	0.016	0.019	0.83
		20	0	0.044	0.007	0.010	0.013	0.007	0.045	0.007	1.02	0.011	0.013	1.87	0.028	0.053	0.009	0.32	0.024	0.017	0.61
		54	0	0.049	0.005	0.009	0.011	0.007	0.050	0.005	0.70	0.010	0.011	1.62	0.029	0.059	0.006	0.22	0.026	0.014	0.49
	.6	2	0	0.018	0.010	0.008	0.012	0.013	0.022	0.011	0.88	0.012	0.015	1.15	0.053	0.055	0.018	0.34	0.051	0.021	0.39
		20	0	0.034	0.006	0.008	0.010	0.016	0.038	0.006	0.40	0.014	0.012	0.77	0.063	0.072	0.012	0.19	0.062	0.014	0.23
		54	0	0.038	0.004	0.007	0.008	0.016	0.042	0.005	0.29	0.015	0.010	0.66	0.066	0.078	0.010	0.15	0.065	0.012	0.18
	.8	2	0	0.010	0.005	0.004	0.007	0.023	0.025	0.010	0.42	0.022	0.012	0.51	0.094	0.101	0.020	0.21	0.101	0.020	0.21
		20	0	0.019	0.003	0.004	0.005	0.028	0.035	0.006	0.23	0.027	0.008	0.29	0.112	0.119	0.019	0.17	0.117	0.019	0.17
		54	0	0.021	0.002	0.004	0.005	0.029	0.037	0.005	0.18	0.029	0.006	0.22	0.117	0.124	0.017	0.14	0.121	0.017	0.14
500	.4	2	0	0.015	0.008	0.007	0.010	0.006	0.016	0.008	1.39	0.008	0.011	1.78	0.023	0.026	0.011	0.49	0.020	0.015	0.65
		20	0	0.028	0.004	0.006	0.008	0.007	0.029	0.005	0.66	0.008	0.009	1.25	0.028	0.040	0.006	0.23	0.026	0.010	0.37
		54	0	0.031	0.003	0.005	0.007	0.007	0.032	0.003	0.46	0.007	0.008	1.08	0.029	0.043	0.005	0.16	0.029	0.007	0.26
	.6	2	0	0.011	0.006	0.005	0.008	0.013	0.017	0.008	0.63	0.012	0.011	0.83	0.053	0.054	0.012	0.23	0.053	0.012	0.23
		20	0	0.021	0.003	0.005	0.006	0.016	0.027	0.004	0.28	0.014	0.008	0.51	0.063	0.067	0.008	0.13	0.063	0.009	0.14
		54	0	0.024	0.002	0.004	0.005	0.016	0.029	0.003	0.20	0.016	0.006	0.40	0.066	0.071	0.007	0.10	0.066	0.007	0.11
	.8	2	0	0.006	0.003	0.003	0.004	0.023	0.024	0.006	0.28	0.023	0.007	0.29	0.094	0.100	0.013	0.14	0.100	0.013	0.14
		20	0	0.012	0.002	0.003	0.003	0.028	0.031	0.004	0.15	0.028	0.005	0.17	0.112	0.116	0.011	0.10	0.115	0.012	0.10
		54	0	0.013	0.001	0.002	0.003	0.029	0.033	0.004	0.12	0.029	0.004	0.13	0.117	0.121	0.010	0.09	0.119	0.010	0.09
1000	.4	2	0	0.011	0.006	0.005	0.007	0.006	0.012	0.006	1.03	0.006	0.008	1.34	0.023	0.025	0.009	0.38	0.022	0.011	0.46
		20	0	0.020	0.003	0.004	0.006	0.007	0.021	0.003	0.47	0.007	0.007	0.94	0.028	0.034	0.005	0.17	0.027	0.006	0.22
		54	0	0.022	0.002	0.004	0.005	0.007	0.023	0.002	0.33	0.007	0.006	0.81	0.029	0.037	0.004	0.12	0.029	0.005	0.16
	.6	2	0	0.008	0.004	0.004	0.005	0.013	0.015	0.006	0.48	0.012	0.008	0.62	0.053	0.053	0.009	0.16	0.053	0.009	0.16
		20	0	0.015	0.002	0.003	0.004	0.016	0.022	0.003	0.21	0.015	0.005	0.32	0.063	0.065	0.006	0.09	0.063	0.006	0.09
		54	0	0.017	0.002	0.003	0.004	0.016	0.024	0.002	0.15	0.016	0.004	0.23	0.066	0.068	0.005	0.07	0.066	0.005	0.08
	.8	2	0	0.005	0.002	0.002	0.003	0.023	0.024	0.005	0.20	0.023	0.005	0.20	0.094	0.099	0.010	0.10	0.099	0.010	0.10
		20	0	0.008	0.001	0.002	0.002	0.028	0.029	0.003	0.11	0.028	0.003	0.11	0.112	0.114	0.008	0.07	0.114	0.008	0.07
		54	0	0.009	0.001	0.002	0.002	0.029	0.031	0.003	0.09	0.029	0.003	0.09	0.117	0.119	0.007	0.06	0.118	0.007	0.06

Note. N = sample size; λ = standardized factor loadings; *df* = degrees of freedom; ρ = interfactor correlations; pop. = the population values; mean = the average sample estimates across replications; sd = the standard deviations of the sample estimates across replications; c.v. = the coefficient of variation. SRMR.b = sample SRMR using the biased formula ($\widehat{SRMR}_b$; Equation (9)); SRMR.u = sample SRMR computed using the unbiased formula ($\widehat{SRMR}u$; Equation (10)).

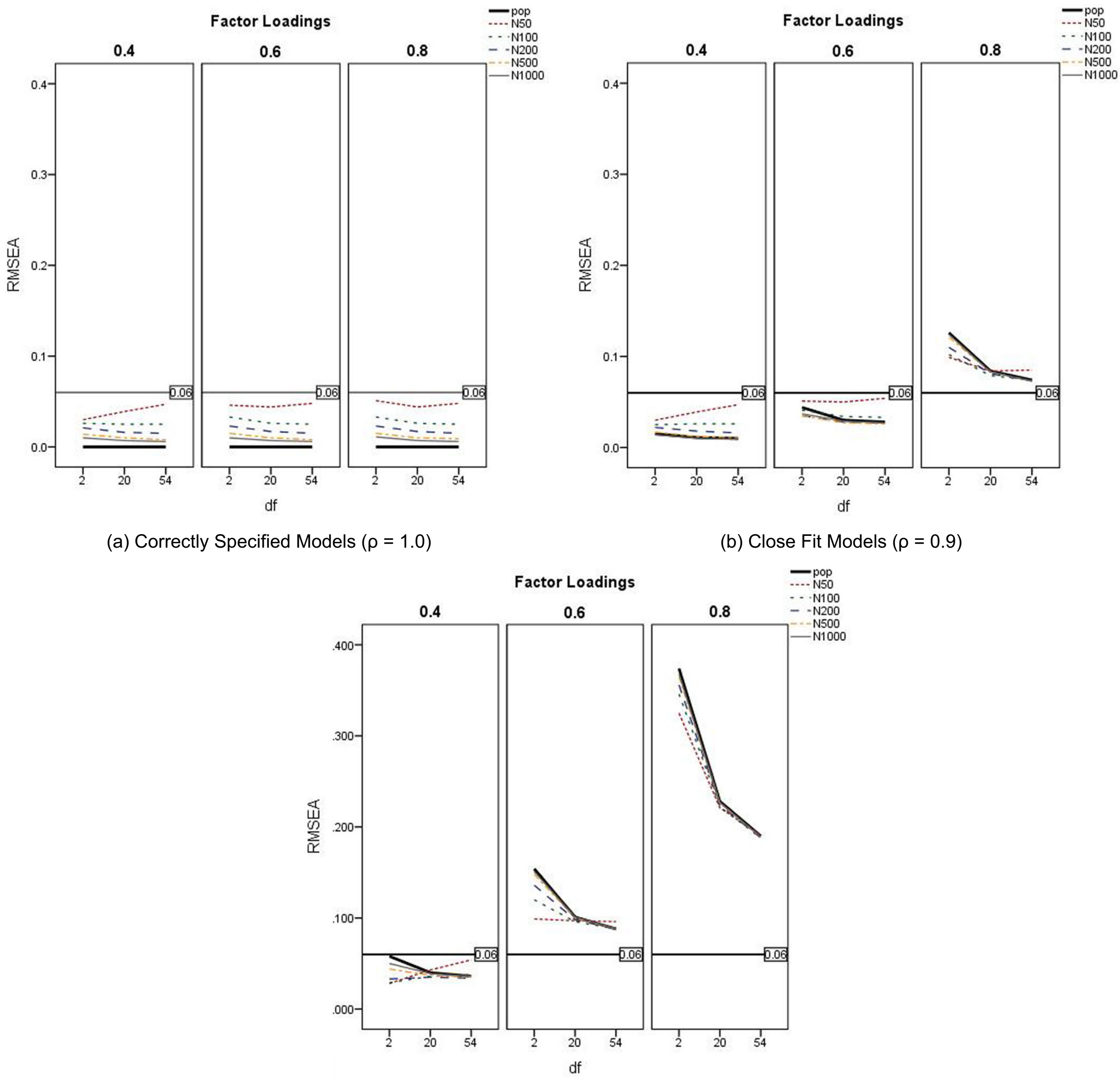
Table 3. Population values and average sample estimates: CFI.

			Correctly specified models $\rho = 1.0$					Close fit models $\rho = 0.9$							Severely misspecifed models $\rho = 0.6$						
				CFI.ML		CFI.FBC			CFI.ML			CFI.FBC				CFI.ML			CFI.FBC		
N	λ	df	pop.	mean	sd	mean	sd	pop.	mean	sd	c.v.	mean	sd	c.v.	pop.	mean	Sd	c.v.	mean	sd	c.v.
50	.4	2	1	0.950	0.130	0.918	0.167	0.996	0.949	0.134	0.13	0.917	0.170	0.17	0.915	0.937	0.165	0.18	0.907	0.196	0.21
		20	1	0.893	0.144	0.829	0.200	0.994	0.887	0.150	0.15	0.822	0.207	0.21	0.903	0.852	0.181	0.20	0.767	0.243	0.27
		54	1	0.845	0.148	0.781	0.180	0.993	0.835	0.157	0.16	0.766	0.190	0.19	0.894	0.767	0.188	0.21	0.675	0.228	0.26
	.6	2	1	0.975	0.053	0.966	0.070	0.993	0.969	0.061	0.06	0.958	0.080	0.08	0.876	0.906	0.127	0.14	0.876	0.162	0.18
		20	1	0.959	0.055	0.950	0.061	0.989	0.948	0.064	0.06	0.937	0.072	0.07	0.854	0.839	0.119	0.14	0.810	0.137	0.16
		54	1	0.946	0.054	0.934	0.058	0.986	0.931	0.064	0.06	0.917	0.069	0.07	0.842	0.802	0.107	0.13	0.777	0.114	0.14
	.8	2	1	0.992	0.017	0.991	0.018	0.982	0.976	0.035	0.04	0.974	0.038	0.04	0.785	0.816	0.106	0.14	0.804	0.113	0.14
		20	1	0.985	0.020	0.983	0.021	0.972	0.961	0.035	0.04	0.957	0.036	0.04	0.759	0.765	0.084	0.11	0.752	0.089	0.12
		54	1	0.979	0.021	0.975	0.022	0.966	0.947	0.034	0.04	0.941	0.035	0.04	0.748	0.742	0.073	0.10	0.731	0.077	0.10
100	.4	2	1	0.963	0.088	0.939	0.128	0.996	0.960	0.096	0.10	0.935	0.137	0.14	0.915	0.936	0.140	0.15	0.902	0.178	0.19
		20	1	0.942	0.086	0.923	0.109	0.994	0.935	0.094	0.09	0.911	0.122	0.12	0.903	0.877	0.139	0.15	0.827	0.182	0.20
		54	1	0.938	0.078	0.922	0.088	0.993	0.928	0.087	0.09	0.910	0.099	0.10	0.894	0.854	0.129	0.14	0.816	0.149	0.17
	.6	2	1	0.987	0.028	0.985	0.032	0.993	0.981	0.035	0.04	0.978	0.040	0.04	0.876	0.893	0.102	0.12	0.875	0.121	0.14
		20	1	0.983	0.025	0.981	0.026	0.989	0.974	0.033	0.03	0.971	0.035	0.04	0.854	0.853	0.082	0.10	0.843	0.088	0.10
		54	1	0.981	0.024	0.978	0.025	0.986	0.970	0.032	0.03	0.966	0.033	0.03	0.842	0.835	0.070	0.08	0.826	0.073	0.09
	.8	2	1	0.996	0.008	0.996	0.008	0.982	0.980	0.023	0.02	0.980	0.024	0.02	0.785	0.805	0.075	0.10	0.797	0.079	0.10
		20	1	0.994	0.008	0.994	0.009	0.972	0.969	0.022	0.02	0.968	0.022	0.02	0.759	0.766	0.059	0.08	0.759	0.062	0.08
		54	1	0.993	0.009	0.992	0.009	0.966	0.961	0.020	0.02	0.960	0.021	0.02	0.748	0.751	0.052	0.07	0.745	0.055	0.07
200	.4	2	1	0.976	0.054	0.966	0.075	0.996	0.971	0.060	0.06	0.958	0.087	0.09	0.915	0.932	0.113	0.12	0.903	0.153	0.17
		20	1	0.972	0.043	0.968	0.047	0.994	0.966	0.051	0.05	0.961	0.056	0.06	0.903	0.895	0.099	0.11	0.877	0.115	0.13
		54	1	0.973	0.037	0.970	0.039	0.993	0.967	0.043	0.04	0.962	0.046	0.05	0.894	0.886	0.084	0.09	0.874	0.089	0.10
	.6	2	1	0.994	0.013	0.993	0.014	0.993	0.988	0.021	0.02	0.987	0.022	0.02	0.876	0.885	0.077	0.09	0.875	0.086	0.10
		20	1	0.992	0.012	0.992	0.012	0.989	0.984	0.018	0.02	0.983	0.019	0.02	0.854	0.855	0.055	0.06	0.851	0.058	0.07
		54	1	0.992	0.011	0.992	0.011	0.986	0.982	0.017	0.02	0.981	0.018	0.02	0.842	0.840	0.046	0.05	0.837	0.047	0.06
	.8	2	1	0.998	0.004	0.998	0.004	0.982	0.982	0.016	0.02	0.981	0.016	0.02	0.785	0.799	0.054	0.07	0.793	0.056	0.07
		20	1	0.997	0.004	0.997	0.004	0.972	0.971	0.014	0.01	0.971	0.014	0.01	0.759	0.763	0.042	0.06	0.759	0.044	0.06
		54	1	0.997	0.004	0.997	0.004	0.966	0.964	0.013	0.01	0.964	0.013	0.01	0.748	0.750	0.037	0.05	0.747	0.038	0.05
500	.4	2	1	0.989	0.024	0.987	0.027	0.996	0.984	0.031	0.03	0.982	0.036	0.04	0.915	0.922	0.086	0.09	0.909	0.102	0.11
		20	1	0.989	0.018	0.989	0.018	0.994	0.985	0.022	0.02	0.984	0.023	0.02	0.903	0.904	0.060	0.07	0.899	0.062	0.07
		54	1	0.990	0.015	0.990	0.015	0.993	0.985	0.019	0.02	0.985	0.019	0.02	0.894	0.893	0.047	0.05	0.891	0.048	0.05
	.6	2	1	0.998	0.005	0.997	0.005	0.993	0.991	0.012	0.01	0.991	0.012	0.01	0.876	0.879	0.051	0.06	0.874	0.054	0.06
		20	1	0.997	0.005	0.997	0.005	0.989	0.988	0.010	0.01	0.988	0.010	0.01	0.854	0.854	0.034	0.04	0.853	0.034	0.04
		54	1	0.997	0.004	0.997	0.004	0.986	0.986	0.009	0.01	0.986	0.009	0.01	0.842	0.841	0.028	0.03	0.840	0.028	0.03
	.8	2	1	0.999	0.002	0.999	0.002	0.982	0.982	0.010	0.01	0.981	0.010	0.01	0.785	0.793	0.035	0.04	0.789	0.036	0.05
		20	1	0.999	0.002	0.999	0.002	0.972	0.972	0.008	0.01	0.972	0.008	0.01	0.759	0.761	0.027	0.04	0.759	0.028	0.04
		54	1	0.999	0.001	0.999	0.002	0.966	0.965	0.007	0.01	0.965	0.007	0.01	0.748	0.748	0.024	0.03	0.747	0.024	0.03
1000	.4	2	1	0.994	0.012	0.994	0.013	0.996	0.990	0.018	0.02	0.990	0.019	0.02	0.915	0.919	0.065	0.07	0.913	0.073	0.08
		20	1	0.995	0.009	0.995	0.009	0.994	0.990	0.013	0.01	0.990	0.013	0.01	0.903	0.903	0.039	0.04	0.902	0.040	0.04
		54	1	0.995	0.007	0.995	0.007	0.993	0.991	0.011	0.01	0.990	0.011	0.01	0.894	0.895	0.030	0.03	0.894	0.030	0.03
	.6	2	1	0.999	0.003	0.999	0.003	0.993	0.992	0.008	0.01	0.992	0.008	0.01	0.876	0.877	0.037	0.04	0.875	0.038	0.04
		20	1	0.999	0.002	0.999	0.002	0.989	0.989	0.006	0.01	0.989	0.006	0.01	0.854	0.854	0.023	0.03	0.853	0.023	0.03
		54	1	0.999	0.002	0.999	0.002	0.986	0.986	0.006	0.01	0.986	0.006	0.01	0.842	0.842	0.019	0.02	0.841	0.019	0.02
	.8	2	1	1.000	0.001	1.000	0.001	0.982	0.982	0.007	0.01	0.982	0.007	0.01	0.785	0.790	0.025	0.03	0.788	0.026	0.03
		20	1	1.000	0.001	1.000	0.001	0.972	0.972	0.006	0.01	0.972	0.006	0.01	0.759	0.760	0.019	0.03	0.758	0.019	0.03
		54	1	1.000	0.001	1.000	0.001	0.966	0.966	0.005	0.01	0.965	0.005	0.01	0.748	0.748	0.017	0.02	0.747	0.020	0.03

Note. *N* = sample size; λ = standardized factor loadings; *df* = degrees of freedom; ρ = interfactor correlations; pop. = the population values; mean = the average sample estimates across replications; sd = the standard deviations of the sample estimates across replications; c.v. = the coefficient of variation. CFI.ML = sample CFI computed using $\widehat{\mathrm{CFI}}_{ML}$ (Equation (15)); CFI.FBC = sample CFI computed using $\widehat{\mathrm{CFI}}_{FBC}$ (Equation (16)).

0.13). When fitting a one-factor model to two-factor data with $\rho = 0.60$ and $df = 2$, as λ decreased from 0.80 to 0.40, the population SRMR dropped dramatically from 0.094 to 0.023. However, under the models considered in this study, the population SRMR was less susceptible to the influence of model size (df; $\eta^2 < 0.01$) compared to the population RMSEA. Keeping the level of model misfit and factor loadings the same (e.g., $\rho = 0.60$ and $\lambda = .60$), as the *df* increased from 2 (p = 4) to 54 (p = 12), the population SRMR increased slightly, from 0.053 ($df = 2$) to 0.066 ($df = 54$).

Finally, for population CFI, variability is mostly attributed to the magnitudes of model misspecification (ρ; $\eta^2 = 0.74$), levels of factor loadings (λ; $\eta^2 = 0.16$), and their interaction ($\lambda \times \rho$; $\eta^2 = 0.09$). For the models considered in this study, the population CFI was not sensitive to the impact of model size (df; $\eta^2 < 0.01$). Taking the same conditions as in the example above, as λ decreased from 0.80 to 0.40 ($\rho = 0.60$ and $df = 2$), the population CFI increased from 0.748 to 0.915. As the *df* increased from 2 to 54 ($\rho = 0.60$ and $\lambda = .60$), the population CFI was relatively stable and slightly decreased from 0.876 to 0.842.



(a) Correctly Specified Models (ρ = 1.0)

(b) Close Fit Models (ρ = 0.9)

(c) Severely Misspecifed Models (ρ = 0.6)

Figure 1. Population values and average sample estimates: RMSEA.
Note. $N =$ sample size; df $=$ degrees of freedom. The horizontal lines indicate Hu & Bentler (1999)'s conventional cutoff for sample RMSEA (i.e., 0.06).

Average sample estimates

Table 1 summarizes the empirical distributions of the sample RMSEAs, SRMRs, and CFIs across replications. Specifically, it reports the means and standard deviations of the empirical distributions. We reported sample SRMRs using both $\widehat{\mathrm{SRMR}}_b$(Equation (9)) and $\widehat{\mathrm{SRMR}}u$(Equation (10)) and reported sample CFIs using both $\widehat{\mathrm{CFI}}_{\mathrm{ML}}$(Equation (15)) and $\widehat{\mathrm{CFI}}_{\mathrm{FBC}}$(Equation (16)) as presented earlier.

The accuracy (bias) of the point estimates was determined by comparing the average sample estimates with the corresponding population values. For each fit index, ANOVA were conducted by using the relative bias of average point estimates as the outcome variable[8]. Figures 1–3 plot the average sample estimates of RMSEA,$\widehat{\mathrm{SRMR}}_b$, $\widehat{\mathrm{SRMR}}u$, $\widehat{\mathrm{CFI}}_{\mathrm{ML}}$, and $\widehat{\mathrm{CFI}}_{\mathrm{FBC}}$ by different sample sizes across simulation conditions. Not surprisingly, the average sample

[8]The relative bias was computed as $\frac{\bar{\theta}_{\mathrm{est}}-\theta_{\mathrm{pop}}}{\theta_{\mathrm{pop}}}$, where $\bar{\theta}_{\mathrm{est}}$ represents the average sample estimate of the fit indices across all replications, and θ_{pop} indicates the population value of fit indices. For RMSEA and SRMR, the relative bias was not computed under correctly specified models as the population value (the denominator) equals zero.

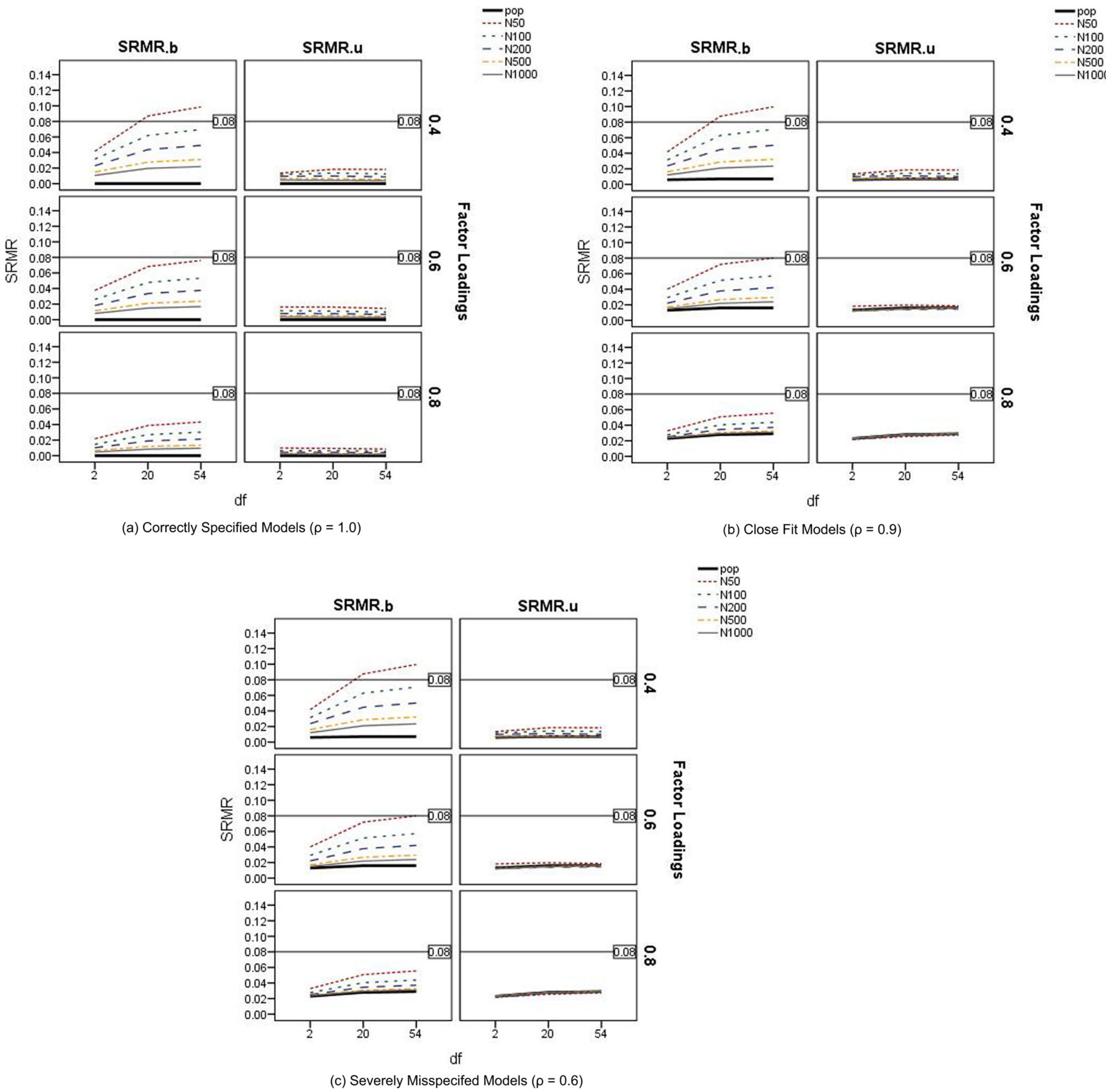


Figure 2. Population values and average sample estimates: SRMR.
Note. N = sample size; df = degrees of freedom. SRMR.b = sample SRMR using the biased formula ($\widehat{\mathrm{SRMR}}_b$; Equation 9); SRMR.u = sample SRMR computed using the unbiased formula ($\widehat{\mathrm{SRMR}}u$; Equation (10)); the horizontal lines indicate Hu & Bentler's (1999) conventional cutoff for sample SRMR (i.e., 0.08).

estimates of all three fit indices approached the population values as the sample size (N) increased. Other than the sample size (N; $\eta^2 = 0.11$), the important sources of (relative) biases in estimating the population RMSEA include the level of factor loadings (λ; $\eta^2 = 0.10$), the magnitudes of model misspecification (ρ; $\eta^2 = 0.10$), and three interaction terms (i.e., $\lambda \times \rho$; $\eta^2 = 0.14$; $N \times \lambda$; $\eta^2 = 0.13$; $N \times \rho$; $\eta^2 = 0.12$). Generally, the relative bias in estimating the population RMSEA decreases as the level of factor loadings increase, and the level of model misfit increases. The effects of the magnitude of factor loadings and the level of model misfit were more pronounced as the sample size became smaller.

The most important sources of (relative) biases in estimating the population SRMR was the choice of formula ($\eta^2 = 0.12$). The effect of the choice of formula was also moderated by the magnitude of factor loadings ($\eta^2 = 0.10$) and the level of model misfit ($\eta^2 = 0.06$). Other noticeable factors included the sample size (N; $\eta^2 = 0.07$), the level of factor loadings (λ; η^2

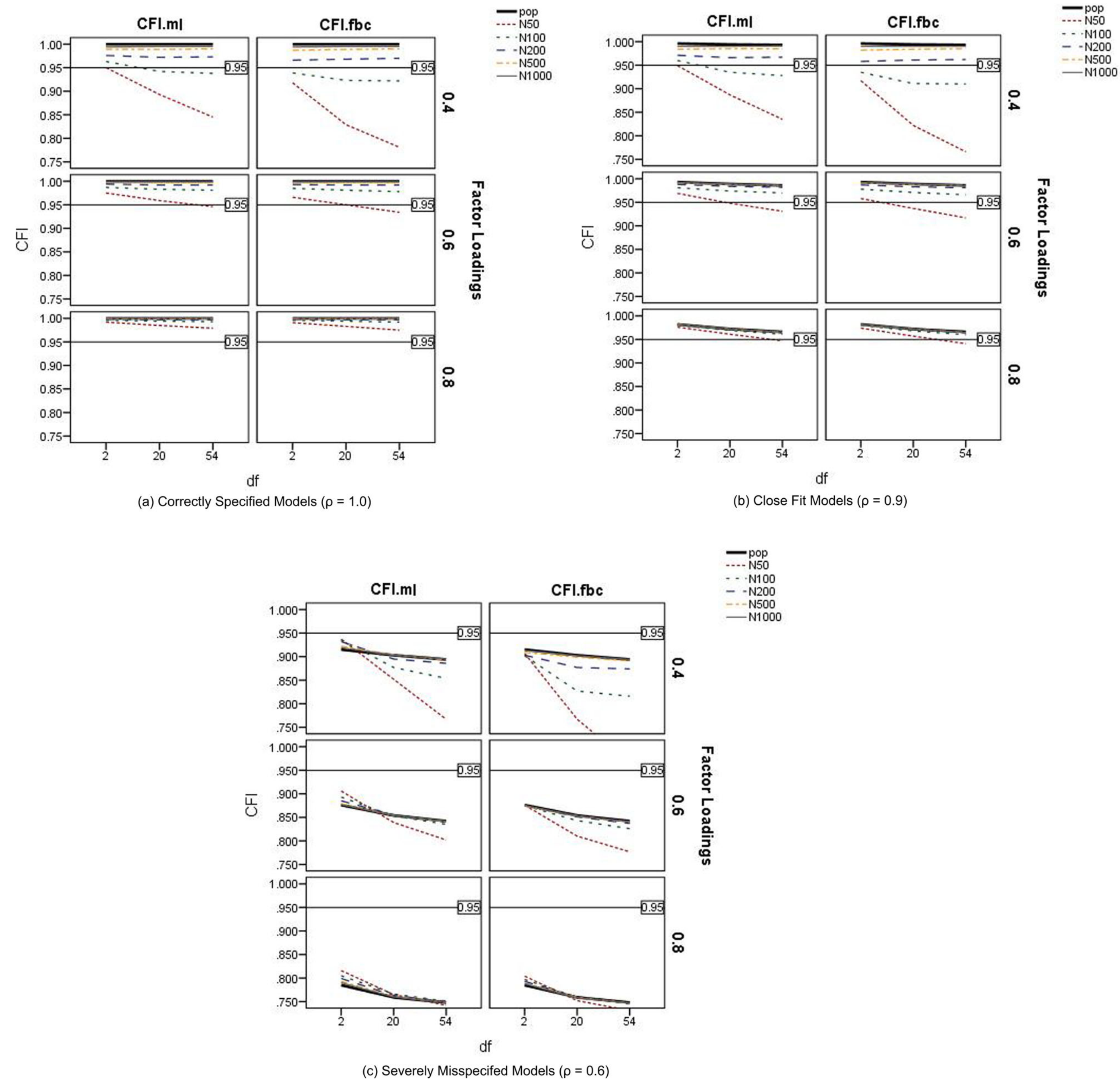


Figure 3. Population values and average sample estimates: CFI.
Note. N = sample size; df = degrees of freedom. CFI.ml = sample CFI computed using $\widehat{\mathrm{CFI}}_{\mathrm{ML}}$ (Equation (15)); CFI.fbc = sample CFI computed using $\widehat{\mathrm{CFI}}_{\mathrm{FBC}}$ (Equation 16); the horizontal lines indicate Hu & Bentler (1999)'s conventional cutoff for sample CFI (i.e., 0.95).

$= 0.14$), the magnitudes of model misspecification (ρ; $\eta^2 = 0.09$), and their interaction ($\lambda \times \rho$; $\eta^2 = 0.09$).

Specifically, the average sample SRMR using the biased formula ($\widehat{\mathrm{SRMR}}_b$) was more sensitive to the impact of sample size (N), model size (df), and the level of factor loadings (λ). As shown in the figure, the bias in $\widehat{\mathrm{SRMR}}_b$ could increase dramatically as N decreases and λ and df increase. Given the same level of df and λ, the sample SRMR computed using the unbiased formula ($\widehat{\mathrm{SRMR}}u$) converged on its population value faster. Furthermore, when the model was correctly specified (i.e., population RMSEA and SRMR were both equal to zero), the sample estimates of $\widehat{\mathrm{SRMR}}u$ tended to be more accurate, on average, than those calculated from RMSEA, especially when the df was very small. For example, when $\lambda = 0.60$, $df = 2$, and $N = 50$, the average of the sample RMSEA and $\widehat{\mathrm{SRMR}}u$ for the correctly specified model was 0.050 and 0.020, respectively.

For estimating CFI, according to the ANOVA results, the impact of the choice between $\widehat{\mathrm{CFI}}_{\mathrm{ML}}$ and $\widehat{\mathrm{CFI}}_{\mathrm{FBC}}$ was not practically important ($\eta^2 = 0.02$).

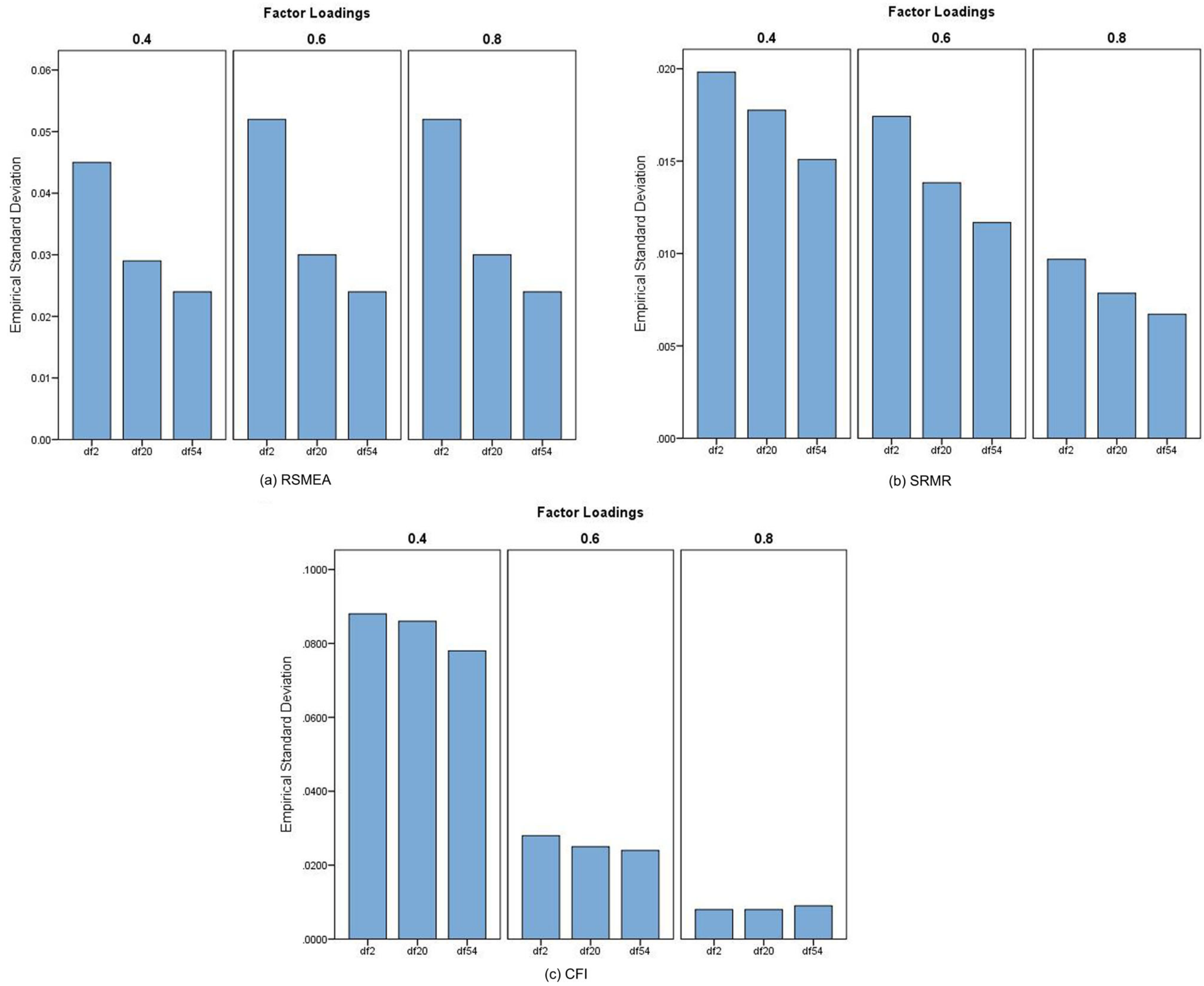


Figure 4. Effect of *df* on the empirical standard deviations of RMSEA, SRMR, and CFI ($\rho = 1.0$; $N = 100$).
Note. N = sample size; df = degrees of freedom; ρ = interfactor correlation; the results for SRMR and CFI were based on $\widehat{\text{SRMR}}u$ and $\widehat{\text{CFI}}_{\text{ML}}$.

However, as shown in the figure, $\widehat{\text{CFI}}_{\text{ML}}$ generally converged on its population value faster than $\widehat{\text{CFI}}_{\text{FBC}}$, especially for models with lower factor loadings (λ) and larger *df*. For both $\widehat{\text{CFI}}_{\text{ML}}$ and $\widehat{\text{CFI}}_{\text{FBC}}$, the bias in estimating the population value decreased as sample size (N; $\eta^2 = 0.28$) increased, and the magnitude of factor loadings increased (λ; $\eta^2 = 0.20$). The effect of sample size was more noticeable as λ decreased ($N \times \lambda$; $\eta^2 = 0.19$) and/or *df* increased ($N \times df$; $\eta^2 = 0.09$). For example, when fitting correctly specified models (i.e., population CFI = 1.0) with $N = 100$, $\lambda = 0.80$, and $df = 2$, the average $\widehat{\text{CFI}}_{\text{ML}}$ and $\widehat{\text{CFI}}_{\text{FBC}}$ were both 0.996. Keeping all other conditions the same, as λ decreased to 0.40 and *df* increased to 54, the average $\widehat{\text{CFI}}_{\text{ML}}$ and $\widehat{\text{CFI}}_{\text{FBC}}$ dropped to 0.938 and 0.922, respectively.

Standard deviations of point estimates

Next, we checked the variability of the point estimates. For SRMR and CFI, we focused on the empirical variability of $\widehat{\text{SRMR}}u$ and $\widehat{\text{CFI}}_{\text{ML}}$ as they yield more accurate point estimates compared with $\widehat{\text{SRMR}}b$ and $\widehat{\text{CFI}}_{\text{FBC}}$. It is noteworthy that using asymptotic methods, the empirical standard deviations for $\widehat{\text{SRMR}}u$ and $\widehat{\text{CFI}}_{\text{FBC}}$ tended to be larger than those obtained from $\widehat{\text{SRMR}}b$ and $\widehat{\text{CFI}}_{\text{ML}}$, especially when the sample size was small. We first summarized the results under correctly specified models as the population values of RMSEA, SRMR, and CFI were constant across conditions. Not surprisingly, for all three fit indices, as sample size (N) increased, the empirical standard deviations of the sample estimates decreased.

Table 4. Empirical rejection rates using conventional cutoffs.

			Correctly specified models $\rho = 1.0$				Close fit models $\rho = 0.9$				Severely misspecifed models $\rho = 0.6$			
N	λ	df	RMSEA > 0.06	SRMR > 0.08	SRMR $> 0.10\times R^2$	CFI < 0.95	RMSEA > 0.06	SRMR > 0.08	SRMR $> 0.10\times R^2$	CFI < 0.95	RMSEA > 0.06	SRMR > 0.08	SRMR $> 0.10\times R^2$	CFI < 0.95
50	.4	2	0.22	0.02	0.25	0.21	0.22	0.02	0.26	0.21	0.20	0.03	0.24	0.21
		20	0.34	0.01	0.40	0.46	0.34	0.01	0.41	0.47	0.37	0.03	0.45	0.52
		54	0.41	0.00	0.45	0.66	0.41	0.00	0.45	0.66	0.50	0.02	0.56	0.74
	.6	2	0.31	0.02	0.23	0.17	0.34	0.03	0.26	0.21	0.57	0.16	0.52	0.48
		20	0.38	0.00	0.20	0.32	0.44	0.01	0.27	0.40	0.79	0.24	0.75	0.79
		54	0.43	0.00	0.14	0.46	0.49	0.00	0.22	0.55	0.87	0.24	0.84	0.92
	.8	2	0.34	0.00	0.00	0.04	0.55	0.02	0.05	0.17	0.98	0.69	0.83	0.91
		20	0.39	0.00	0.00	0.06	0.71	0.00	0.02	0.33	1.00	0.86	0.95	0.99
		54	0.43	0.00	0.00	0.11	0.80	0.00	0.01	0.50	1.00	0.91	0.98	1.00
100	.4	2	0.21	0.00	0.29	0.20	0.20	0.00	0.28	0.21	0.24	0.01	0.33	0.27
		20	0.16	0.00	0.38	0.36	0.16	0.00	0.40	0.39	0.28	0.00	0.55	0.55
		54	0.08	0.00	0.38	0.43	0.09	0.00	0.40	0.46	0.21	0.00	0.64	0.70
	.6	2	0.26	0.00	0.12	0.09	0.32	0.00	0.18	0.14	0.75	0.14	0.68	0.64
		20	0.17	0.00	0.06	0.11	0.26	0.00	0.13	0.20	0.88	0.18	0.87	0.89
		54	0.08	0.00	0.02	0.12	0.17	0.00	0.08	0.25	0.90	0.17	0.94	0.96
	.8	2	0.26	0.00	0.00	0.00	0.66	0.00	0.01	0.11	1.00	0.78	0.92	0.99
		20	0.17	0.00	0.00	0.00	0.74	0.00	0.00	0.18	1.00	0.94	0.99	1.00
		54	0.09	0.00	0.00	0.00	0.77	0.00	0.00	0.28	1.00	0.98	1.00	1.00
200	.4	2	0.16	0.00	0.28	0.17	0.17	0.00	0.30	0.19	0.27	0.00	0.43	0.36
		20	0.02	0.00	0.31	0.23	0.03	0.00	0.35	0.27	0.16	0.00	0.67	0.62
		54	0.00	0.00	0.28	0.23	0.00	0.00	0.34	0.29	0.05	0.00	0.78	0.75
	.6	2	0.18	0.00	0.03	0.02	0.28	0.00	0.08	0.07	0.91	0.07	0.79	0.78
		20	0.03	0.00	0.00	0.01	0.10	0.00	0.04	0.06	0.97	0.10	0.97	0.97
		54	0.00	0.00	0.00	0.00	0.02	0.00	0.01	0.05	0.98	0.10	1.00	1.00
	.8	2	0.18	0.00	0.00	0.00	0.80	0.00	0.00	0.04	1.00	0.85	0.97	1.00
		20	0.03	0.00	0.00	0.00	0.84	0.00	0.00	0.07	1.00	0.99	1.00	1.00
		54	0.00	0.00	0.00	0.00	0.84	0.00	0.00	0.13	1.00	1.00	1.00	1.00
500	.4	2	0.06	0.00	0.21	0.08	0.08	0.00	0.25	0.11	0.34	0.00	0.63	0.51
		20	0.00	0.00	0.16	0.04	0.00	0.00	0.22	0.09	0.04	0.00	0.85	0.76
		54	0.00	0.00	0.10	0.03	0.00	0.00	0.16	0.06	0.00	0.00	0.95	0.89
	.6	2	0.06	0.00	0.00	0.00	0.25	0.00	0.01	0.01	0.99	0.02	0.92	0.94
		20	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	1.00	0.02	1.00	1.00
		54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.03	1.00	1.00
	.8	2	0.06	0.00	0.00	0.00	0.95	0.00	0.00	0.01	1.00	0.94	1.00	1.00
		20	0.00	0.00	0.00	0.00	0.97	0.00	0.00	0.01	1.00	1.00	1.00	1.00
		54	0.00	0.00	0.00	0.00	0.96	0.00	0.00	0.03	1.00	1.00	1.00	1.00
1000	.4	2	0.01	0.00	0.10	0.02	0.03	0.00	0.16	0.04	0.37	0.00	0.75	0.63
		20	0.00	0.00	0.04	0.00	0.00	0.00	0.10	0.01	0.01	0.00	0.96	0.89
		54	0.00	0.00	0.01	0.00	0.00	0.00	0.04	0.00	0.00	0.00	1.00	0.98
	.6	2	0.01	0.00	0.00	0.00	0.20	0.00	0.00	0.00	1.00	0.00	0.98	0.99
		20	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00
		54	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	1.00	1.00
	.8	2	0.01	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	0.98	1.00	1.00
		20	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00
		54	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	1.00	1.00	1.00	1.00

Note. N = sample size; λ = standardized factor loadings; *df* = degrees of freedom; ρ = interfactor correlation; the results for SRMR and CFI were based on $\widehat{\text{SRMR}}u$ and $\widehat{\text{CFI}}_{\text{ML}}$.

The ANOVA results also showed that the model size (*df*) was an important source to explain the sampling variability of RMSEA ($\eta^2 = 0.35$). For SRMR, the important sources of sampling variability are the level of model misfit (ρ; $\eta^2 = 0.16$) and the interaction between ρ and *df* ($\rho \times df$; $\eta^2 = 0.06$). In general, the empirical standard deviations of SRMR decreased as *df* decreased, especially as the model misfit level decreased. However, the empirical standard deviation of the sample CFI was not sensitive to the model size ($\eta^2 = .01$). Instead, the levels of model misfit (ρ; $\eta^2 = 0.19$) and magnitudes of factor loadings (λ; $\eta^2 = 0.25$) played an important role in explaining the empirical standard deviation of the sample CFI, especially when the sample size was small ($N \times \lambda$; $\eta^2 = 0.11$).

To better demonstrate the patterns, we plotted the empirical standard deviations of the three fit indices against the *df*s (e.g., $N = 100$; $\rho = 1.0$) in Figure 4. As shown in this figure and in Tables 1–3, for both RMSEA and SRMR, their empirical standard deviation increased as *df* decreased. The empirical standard deviation of the sample CFI was generally stable as *df* decreased, but a larger standard deviation of CFI was associated with lower factor loadings (λ). The standard deviation of the sample SRMR also tended to increase as the level of factor loadings increased;

however, these changes were not as pronounced as those observed from the sample CFI.

Regarding the empirical standard deviations, similar behaviors were observed under misspecified conditions. When models were misspecified, the population values of the fit indices varied by model size and level of factor loadings. As a result, the raw standard deviations may not be comparable across conditions. To account for the differences in the population values, we also computed the coefficient of variation (CV; Everitt, 1998) as the ratio of the empirical standard deviation divided by the population fit indices. The values of the CVs are reported in Tables 1–3. Similar patterns were observed in terms of the relationships between *df* and CV.

By comparing the empirical standard deviations among the three fit indices, we found that under correctly specified models (where both population RMSEA and SRMR are equal to zero), the empirical standard deviations of SRMR were noticeably smaller than the standard deviations of RMSEA, especially when fitting small *df* models. For example, when $N = 100$, $\lambda = 0.60$, and $df = 2$, the standard deviations of RMSEA and SRMR were 0.052 and 0.018, respectively. Under misspecified models, the standard deviations of SRMR were smaller than the standard deviations of RMSEA under most simulated conditions. However, the CVs for RMSEA were generally smaller than those obtained from SRMR under small *df* models. RMSEA yielded larger standard deviations but smaller CVs because the population RMSEA increased dramatically as *df* decreased. Meanwhile, the scale of the population CFI was much larger. Consequently, the CVs for CFI were noticeably smaller than those calculated from RMSEA and SRMR.

Empirical rejection rates using conventional cutoffs

Following prior research, models are considered to be adequately fitted if their RMSEA values are below 0.06 or CFI values above 0.95 (Hu & Bentler, 1999). For SRMR, we applied two cutoffs for acceptable fit: SRMR values below 0.08 (Hu & Bentler, 1999) or SRMR values below $R^2 \times .10$ (Shi, Maydeu-Olivares, & DiStefano, 2018). Table 4 reports the percentage of replications with sample RMSEAs and SRMRs exceeding the cutoffs and CFIs falling below the conventional cutoffs (suggesting that the model does not fit adequately).

We focused on the results from $\widehat{\text{SRMR}}u$ and $\widehat{\text{CFI}}_{\text{ML}}$ as they produced less biased point estimates (compared to $\widehat{\text{SRMR}}b$ and $\widehat{\text{CFI}}_{\text{FBC}}$). Results showed that applying the conventional cutoff of SRMR $\leq .08$ to $\widehat{\text{SRMR}}u$ could yield misleading conclusions as it often retains severely misspecified models unless the level of factor loadings is very high ($\lambda = 0.80$). Therefore, for SRMR, we applied the cutoff of SRMR $\leq .10 \times R^2$ and compared the results with those obtained from RMSEA and CFI. We plotted the empirical rejection rates for RMSEA ($\widehat{\text{RMSEA}}_{\text{ML}} \leq .06$), SRMR ($\widehat{\text{SRMR}}u \leq 10 \times \text{R}^2$), and CFI ($\widehat{\text{CFI}}_{\text{ML}} \geq 0.95$) across all simulated conditions in Figure 5.

As shown, for correctly specified models ($\rho = 1.00$) or models with minor misspecification ($\rho = 0.90$), the observed sample RMSEAs tended to exceed the cutoff of 0.06 when the sample sizes decreased, especially when *df* was very small and the level of factor loadings (λ) was high. For example, in models with minor misspecifications ($\rho = .90$) where $N = 200$, $\lambda = 0.80$, and $df = 2$, sample RMSEA values were greater than 0.06 in 80% of the replications. In addition, the sample RMSEA did not have enough power to reject severely misspecified models ($\rho = 0.60$) when the level of factor loadings was very low ($\lambda = 0.40$), even when the sample size (N) reached 1,000. For example, when $\rho = 0.60$, $N = 1{,}000$, $\lambda = 0.40$, and $df = 2$, only 37% of the sample RMSEA exceeded 0.06.

As sample sizes decreased, the sample cutoff for SRMR and CFI rejected greater proportions of the correctly specified or slightly misspecified models, especially when the levels of factor loadings were very low (i.e., $\lambda = 0.40$). The rejection rates for SRMR generally increased as *df* decreased, while the rejection rates for CFI tended to decrease as *df* decreased. For example, when fitting correctly specified models with $N = 500$ and $\lambda = 0.40$, as *df* decreased from 54 to 2, the rejection rates for the sample SRMR increased from 10% to 21%, whereas the rejection rates for the sample CFI decreased from 8% to 3%. In addition, for both SRMR and CFI, lower levels of factor loadings were associated with lower power rates to reject the severely misspecifed models.

In general, with medium or high levels of factor loadings ($\lambda \geq .60$), as the sample size reaches 200, using $\widehat{\text{SRMR}}u$ and $\widehat{\text{CFI}}_{\text{ML}}$ with the conventional cutoffs, researchers reject severely misspecified models and retain correctly specified or close-fitting models (power greater than 80% and Type I error rates of less than 10%) even when *df* is very small. For example, when $df = 2$, $\lambda = 0.80$, and $N = 200$, the proportion of rejections based on $\widehat{\text{SRMR}}u$ were 0%, 0%, and 97%,

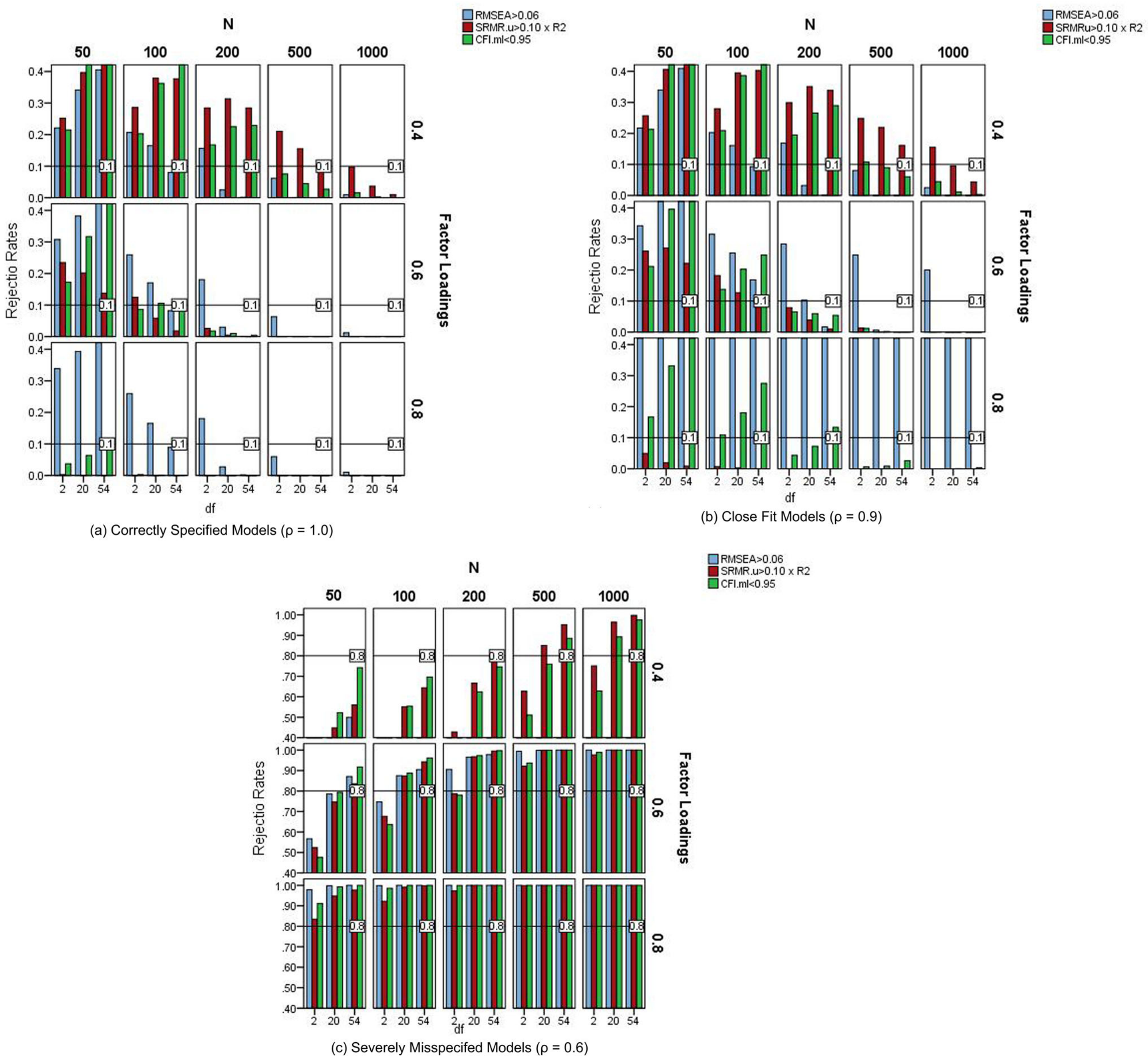


Figure 5. Empirical rejection rates using conventional cutoffs.
Note. N = sample size; df = degrees of freedom; the results for SRMR and CFI are based on $\widehat{\mathrm{SRMR}}u$ and $\widehat{\mathrm{CFI}}_{\mathrm{ML}}$; reference lines of 0.10 and 0.80 are included for correctly specified models/models with minor misspecification and severely misspecified models, respectively.

and the proportion of rejections based on $\widehat{\mathrm{CFI}}_{\mathrm{ML}}$ were 0%, 4%, and 100% for correctly specified, slightly misspecified, and severely misspecified models, respectively. However, under the same conditions, the rejection rates based on the sample $\widehat{\mathrm{RMSEA}}_{\mathrm{ML}}$ were 17%, 80%, and 100%, respectively.

Interval estimates

Under model misspecifications for all three fit indices, we examined the accuracy of the CIs and p values for the close fit tests. Specifically, the CIs and close fit tests for SRMR were computed based on the unbiased sample estimates ($\widehat{\mathrm{SRMR}}u$). For CFI, the CIs and close fit tests were computed using $\widehat{\mathrm{CFI}}_{\mathrm{ML}}$ in Equations (18) and (20)[9]. Table 5 reports the coverage rates for the 90% CIs, where coverage rates between 85% and 95% (90% ± 5%) were considered acceptable (Maydeu-Olivares et al., 2018). These cases are highlighted in bold in the table. Figure 6 provides the boxplots of the coverage rates for the 90% CIs of the three fit indices

[9]We also computed CIs and conducted close fit tests using $\widehat{\mathrm{CFI}}_{\mathrm{FBC}}$ in Equations (18) and (20), and the results were similar to those obtained using $\widehat{\mathrm{CFI}}_{\mathrm{ML}}$.

across all simulated conditions. As shown in Table 5 and Figure 6, in small *df* models, the 90% CIs for RMSEA, SRMR, and CFI were accurate across most simulated conditions. The CIs for the RMSEA generally performed better than the CIs for SRMR and CFI when there were minor model misspecifications ($\rho = 0.90$) and/or when the level of factor loadings was low (e.g., $\lambda = 0.40$). However, the CIs for SRMR and CFI were generally more accurate when the model misspecification was more severe (ρ = .60) and/or when the level of factor loadings was high (e.g., $\lambda = 0.80$).

We also computed the widths of the CIs across replications. Table 5 presents the average and median widths of the CIs. Figures 7–9 plot the average and median widths of the CIs for RMSEA, SRMR, and CFI against model sizes (*df*s) across various sample sizes (i.e., $N = 50$, 100, and 200), magnitudes of model misspecification, and levels of factor loadings. The ANOVA results showed that for all three fit indices, sizeable variability in the average width of the CIs could be explained by sample size (N). Not surprisingly, as sample size increased, the CIs for all three fit indices became narrower.

Model size (*df*) was an important factor in predicting the average width of the CIs for RMSEA (η^2 = 0.46) and SRMR (η^2 = 0.28). As shown in Table 5 and Figures 6–7, for both RMSEA and SRMR, the average width of the CIs increased as *df* decreased, especially when sample size was small. When fitting models with very small *df*s (e.g., $df = 2$) and small sample sizes (e.g., $N \leq 100$), the average widths of the CIs for both RMSEA and SRMR could be very large. For example, when $\rho = 0.90$, $N = 100$, $\lambda = 0.60$, and $df = 2$, the average width of the CIs for RMSEA and SRMR was both 0.18. Such wide intervals reflected high uncertainty and could not provide useful information in practice. For SRMR, we also observed a noticeable interaction effect between standardized factor loadings and the model size ($\lambda \times df$; η^2 = 0.09). As indicated in Figures 7 and 8, the average widths of the CIs for SRMR decreased noticeably as the level of factor loadings increased, especially when *df* was small. In addition, for SRMR, the median widths of the CIs were much narrower than their mean widths. For example, when $\rho = 0.90$, $N = 100$, $\lambda = 0.80$, and $df = 2$, the average and median width of CIs for SRMR was 0.13 and 0.05, respectively. Such patterns were not observed for the CIs for RMSEA.

Meanwhile, the widths of the CIs for CFI were less sensitive to the model size (η^2 = 0.02). Instead, sizable variability in the widths of the CIs for CFI can be explained by the size of model misfit (ρ; η^2 = 0.23) and level of factor loadings (λ; η^2 =0.20). As shown in Figure 9, the widths of the CIs for CFI became narrower as the level of factor loadings (λ) increased and model misfit became less severe (e.g., $\rho = 0.90$). The average and median widths of the CIs for CFI yielded similar values.

Close fit tests

Finally, we examined the accuracy of the p values for close fit tests. Table 6 depicts the empirical rejection rates (Type I error rates) at the 5% significance level when testing whether the RMSEAs and SRMRs were less than or equal to and the CFIs were greater than or equal to their population values. We considered a 5% rejection rate range between 2% and 8% to be reasonably accurate (5% ± 3%; Bradley, 1978); thus, those cases are highlighted in bold.

The results showed that, in general, the empirical rejection rates for all three indices were close to their nominal levels under the examined conditions, even when the model size (*df*) was very small. Figure 10 provides the boxplots of the Type I error rates for close fit tests based on the three indices across levels of factor loadings (λ) and model misfit (ρ). As shown in Table 7 and Figure 10, the Type I error rates for RMSEA were generally more accurate when the level of factor loadings was low (e.g., $\lambda = 0.40$) and/or the level of model misfit was minor (e.g., $\rho = 0.90$). However, for SRMR and CFI, Type I error rates tended to be more accurate under models with more severe levels of misfit (e.g., $\rho = 0.60$) and/or higher levels of factor loadings (e.g., $\lambda = 0.80$).

Numerical example: open-book closed-book data

In this section, we present a numeric example to compare the sample estimates and CIs for SRMR and RMSEA when assessing a factor analysis model with small *df*. The R code and data used in this example are included as supplementary materials. For this, we used the open-book closed-book (OBCB) data set first introduced in Mardia, et al. (1979). The OBCB data consisted of test scores of five topics (i.e., $p = 5$): mechanics, vectors, algebra, analysis, and statistics. The first two tests are from closed-book tests, while the other three are from open-book tests. The five tests (variables) were measured on a scale of 0–100. Table 7 summarizes the descriptive statistics of the measured variables. The sample size is $N = 88$. The

Table 5. Coverage rates, average, and median widths of the 90% confidence intervals (CI).

			90% CI coverage rates						Average CI widths						Median CI widths					
			$\rho = .90$			$\rho = .60$			$\rho = .90$			$\rho = .60$			$\rho = .90$			$\rho = .60$		
N	λ	df	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI
50	.4	2	**0.92**	**0.90**	0.97	**0.91**	0.80	**0.93**	0.22	0.46	0.34	0.22	0.58	0.39	0.23	0.14	0.29	0.23	0.14	0.34
		20	**0.89**	**0.90**	0.81	**0.89**	0.97	0.83	0.12	0.07	0.33	0.12	0.07	0.37	0.13	0.06	0.31	0.13	0.06	0.34
		54	0.82	0.83	0.59	0.83	0.95	0.61	0.10	0.05	0.29	0.10	0.06	0.31	0.10	0.04	0.30	0.10	0.06	0.32
	.6	2	**0.91**	**0.87**	0.99	**0.91**	**0.87**	0.79	0.25	0.48	0.17	0.29	0.31	0.30	0.26	0.12	0.13	0.31	0.14	0.26
		20	**0.87**	**0.94**	**0.94**	**0.88**	0.83	**0.85**	0.12	0.06	0.18	0.14	0.09	0.30	0.13	0.05	0.17	0.14	0.09	0.30
		54	0.80	**0.91**	**0.85**	0.82	0.84	0.79	0.10	0.04	0.18	0.10	0.07	0.27	0.10	0.04	0.17	0.09	0.08	0.27
	.8	2	**0.90**	**0.88**	**0.93**	**0.89**	**0.94**	0.78	0.28	0.17	0.08	0.34	0.13	0.29	0.31	0.08	0.07	0.34	0.13	0.30
		20	**0.86**	0.80	0.97	0.83	**0.88**	0.83	0.14	0.04	0.10	0.12	0.11	0.24	0.14	0.05	0.10	0.11	0.11	0.24
		54	0.81	0.81	**0.93**	0.78	**0.88**	**0.85**	0.10	0.04	0.11	0.07	0.10	0.22	0.10	0.04	0.11	0.07	0.10	0.21
100	.4	2	**0.91**	**0.88**	0.98	**0.92**	0.83	**0.92**	0.16	0.27	0.26	0.17	0.25	0.34	0.17	0.10	0.21	0.18	0.11	0.29
		20	**0.89**	**0.87**	**0.91**	**0.90**	0.96	**0.89**	0.08	0.04	0.25	0.09	0.05	0.32	0.09	0.04	0.23	0.09	0.05	0.30
		54	**0.88**	0.80	0.83	**0.88**	0.66	0.82	0.06	0.03	0.21	0.07	0.04	0.26	0.07	0.03	0.19	0.07	0.04	0.28
	.6	2	**0.90**	**0.85**	0.99	**0.92**	**0.94**	0.79	0.18	0.18	0.10	0.23	0.14	0.27	0.19	0.08	0.07	0.24	0.10	0.25
		20	**0.89**	**0.93**	0.98	**0.87**	**0.86**	**0.86**	0.09	0.04	0.11	0.09	0.06	0.24	0.09	0.04	0.09	0.09	0.06	0.24
		54	**0.87**	**0.88**	0.96	0.84	**0.86**	**0.86**	0.07	0.03	0.10	0.06	0.05	0.21	0.07	0.03	0.09	0.06	0.05	0.21
	.8	2	**0.90**	**0.91**	**0.87**	**0.90**	**0.90**	0.83	0.22	0.13	0.06	0.24	0.09	0.24	0.24	0.05	0.05	0.24	0.09	0.23
		20	**0.87**	0.83	**0.93**	0.82	**0.86**	**0.86**	0.10	0.03	0.07	0.08	0.08	0.18	0.10	0.03	0.07	0.08	0.08	0.18
		54	0.84	0.83	**0.93**	0.76	**0.86**	**0.86**	0.07	0.03	0.07	0.05	0.07	0.16	0.06	0.03	0.07	0.05	0.07	0.16
200	.4	2	**0.91**	**0.87**	0.99	**0.92**	0.84	**0.89**	0.12	0.16	0.18	0.13	0.17	0.28	0.12	0.07	0.14	0.14	0.08	0.23
		20	**0.90**	**0.87**	0.96	**0.89**	0.71	**0.93**	0.06	0.03	0.16	0.07	0.04	0.25	0.06	0.03	0.14	0.07	0.04	0.25
		54	**0.89**	0.80	**0.95**	**0.89**	0.75	**0.90**	0.04	0.02	0.14	0.05	0.03	0.21	0.05	0.02	0.12	0.05	0.03	0.22
	.6	2	**0.90**	0.83	0.98	**0.90**	0.96	0.82	0.13	0.11	0.06	0.17	0.07	0.22	0.14	0.06	0.04	0.17	0.07	0.22
		20	**0.89**	**0.92**	0.99	**0.85**	**0.88**	**0.87**	0.06	0.03	0.06	0.06	0.04	0.18	0.07	0.03	0.05	0.06	0.04	0.18
		54	**0.89**	0.66	0.98	0.83	**0.88**	**0.89**	0.05	0.02	0.06	0.04	0.04	0.15	0.05	0.02	0.05	0.04	0.04	0.15
	.8	2	**0.89**	**0.94**	0.83	**0.89**	**0.86**	**0.86**	0.17	0.04	0.04	0.17	0.06	0.18	0.17	0.04	0.04	0.17	0.06	0.18
		20	**0.86**	**0.86**	**0.90**	0.80	0.84	**0.87**	0.06	0.02	0.05	0.05	0.05	0.13	0.06	0.02	0.05	0.05	0.05	0.13
		54	0.84	**0.85**	**0.93**	0.74	**0.87**	**0.88**	0.04	0.02	0.05	0.03	0.05	0.12	0.04	0.02	0.04	0.03	0.05	0.12
500	.4	2	**0.90**	**0.85**	0.99	**0.91**	**0.91**	0.82	0.08	0.10	0.09	0.10	0.08	0.22	0.08	0.05	0.07	0.11	0.05	0.20
		20	**0.90**	**0.88**	0.99	**0.89**	0.84	**0.88**	0.04	0.02	0.08	0.04	0.03	0.18	0.04	0.02	0.07	0.04	0.03	0.18
		54	**0.89**	0.83	0.98	**0.88**	0.84	**0.89**	0.03	0.01	0.07	0.03	0.02	0.15	0.03	0.02	0.06	0.03	0.02	0.15
	.6	2	**0.90**	**0.87**	**0.93**	**0.89**	**0.91**	**0.86**	0.09	0.06	0.03	0.11	0.04	0.16	0.10	0.04	0.02	0.11	0.04	0.16
		20	**0.89**	0.77	0.98	0.84	**0.90**	**0.88**	0.04	0.02	0.03	0.03	0.03	0.11	0.04	0.02	0.03	0.03	0.03	0.11
		54	**0.89**	0.78	0.97	0.81	**0.90**	**0.90**	0.03	0.01	0.03	0.02	0.02	0.09	0.03	0.01	0.03	0.02	0.02	0.09
	.8	2	**0.89**	**0.93**	**0.87**	**0.89**	0.80	**0.88**	0.11	0.02	0.03	0.10	0.04	0.12	0.11	0.02	0.03	0.10	0.04	0.12
		20	**0.86**	**0.88**	**0.90**	0.81	0.85	**0.89**	0.04	0.01	0.03	0.03	0.03	0.09	0.04	0.01	0.03	0.03	0.03	0.09
		54	0.84	**0.88**	**0.92**	0.74	**0.87**	**0.90**	0.02	0.01	0.03	0.02	0.03	0.08	0.02	0.01	0.03	0.02	0.03	0.08
1000	.4	2	**0.89**	0.83	0.99	**0.90**	**0.94**	0.82	0.06	0.10	0.05	0.07	0.04	0.18	0.06	0.03	0.04	0.08	0.03	0.17
		20	**0.90**	**0.90**	0.99	**0.88**	**0.88**	**0.90**	0.03	0.01	0.04	0.03	0.02	0.13	0.03	0.02	0.04	0.03	0.02	0.13
		54	**0.90**	0.54	0.99	**0.87**	**0.87**	**0.90**	0.02	0.01	0.04	0.02	0.01	0.10	0.02	0.01	0.03	0.02	0.01	0.10
	.6	2	**0.90**	**0.91**	**0.87**	**0.89**	**0.90**	**0.88**	0.07	0.03	0.02	0.07	0.03	0.12	0.08	0.03	0.02	0.07	0.03	0.12
		20	**0.89**	**0.85**	**0.93**	0.85	**0.91**	**0.89**	0.03	0.01	0.02	0.02	0.02	0.08	0.03	0.01	0.02	0.02	0.02	0.08
		54	**0.88**	0.84	**0.93**	0.82	**0.90**	**0.90**	0.02	0.01	0.02	0.01	0.02	0.06	0.02	0.01	0.02	0.01	0.02	0.06
	.8	2	**0.89**	**0.90**	**0.89**	**0.89**	0.76	**0.89**	0.07	0.02	0.02	0.07	0.03	0.08	0.07	0.02	0.02	0.07	0.03	0.08
		20	**0.86**	**0.89**	**0.90**	0.80	**0.86**	**0.89**	0.02	0.01	0.02	0.02	0.02	0.06	0.02	0.01	0.02	0.02	0.02	0.06
		54	0.83	**0.89**	**0.91**	0.73	**0.88**	**0.90**	0.01	0.01	0.02	0.01	0.02	0.06	0.01	0.01	0.02	0.01	0.02	0.06

Note. N = sample size; λ = standardized factor loadings; df = degrees of freedom; ρ = interfactor correlation; the coverage rates between 0.85 to 0.95 are highlighted in bold.

OBCB data closely matched the conditions considered in our simulation study (i.e., small p, small N, and approximately normally distributed data).

We first fit a one-factor CFA model with no mean structure to the OBCB data using ML estimation ($df = 5$). The standardized factor loadings under the one-factor model ranged from 0.60 to 0.92. The sample RMSEA was 0.096, with a 90% CI between 0 and 0.195. The sample SRMRs using the biased (SRMR.b) and unbiased (SRMR.u) formulas were 0.048 and 0.039, respectively. The 90% CI using the unbiased formula SRMR was [0.004, 0.074]. In terms of CFI, the sample CFI_{ML} and CFI_{FBC} were 0.979 and 0.976, respectively, while their corresponding 90% CIs were [0.923, 1] and [0.918, 1], respectively. Using conventional cutoffs, the sample RMSEA suggested that the one-factor model fit the data poorly. The sample SRMR and CFI met the cutoffs for close fit; however, their CIs indicated that with a 90% level of confidence, and it is not clear whether the population values met the conventional cutoffs.

Based on previous studies (e.g., Cai & Lee, 2009), we fit a simple-structure CFA model with two correlated factors (i.e., closed-book tests and open-book tests). The df for the two-factor model was four, and the standardized factor loadings ranged from 0.70 to

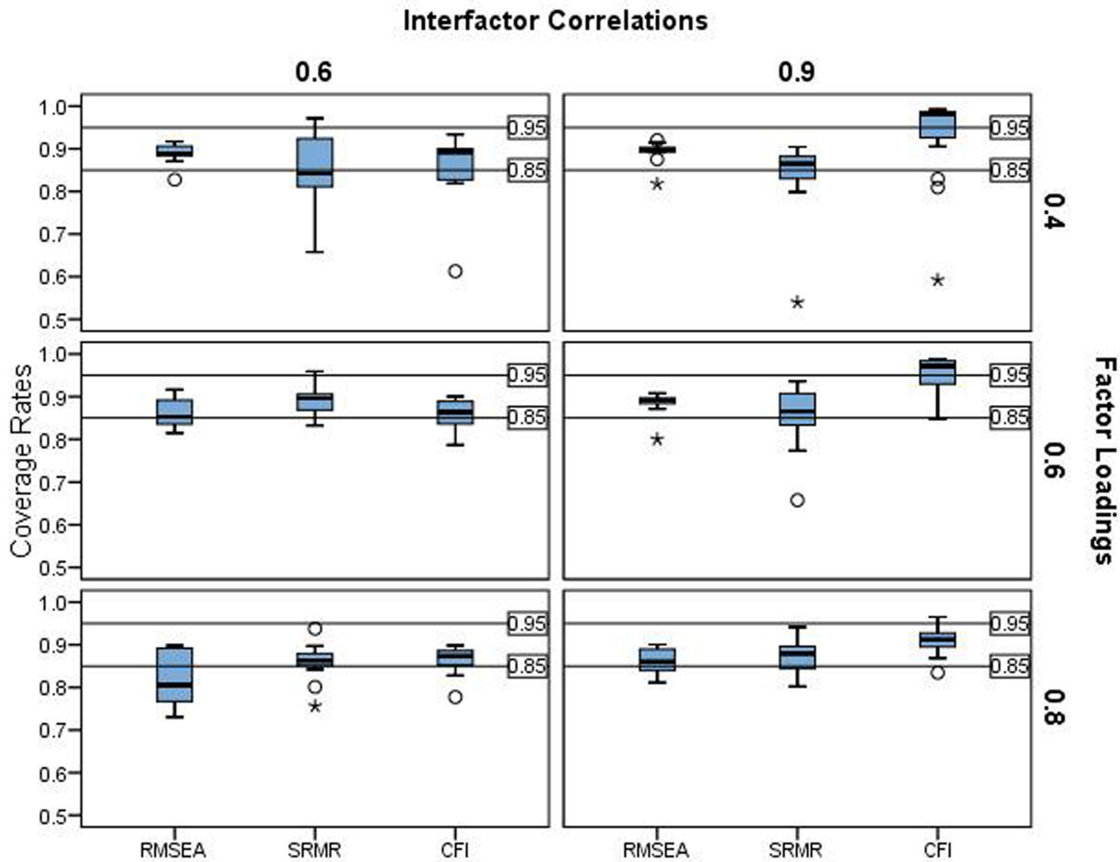


Figure 6. Coverage rates of the 90% confidence intervals.
Note. Reference lines indicate the range of acceptable coverage rates of the 90% confidence intervals (i.e., between 0.85 and 0.95).

0.93. The sample RMSEA for the two-factor model was 0; however, the 90% CI for the RMSEA was [0, 0.118], suggesting high uncertainty in estimating the RMSEA. The sample SRMRs using the biased (SRMR.b) and unbiased (SRMR.u) formulas were 0.019 and 0, respectively. The 90% CI for the unbiased formula SRMR was [0, 0.061]. For CFI, both the sample CFI_{ML} and CFI_{FBC} were equal to 1.000. The 90% CIs based on CFI_{ML} and CFI_{FBC} were [0.976, 1] and [0.967, 1], respectively. The 90% CIs for SRMR and CFI are noticeably narrower than those obtained from RMSEA. Using the recommended cutoffs[10], the point estimates and 90% CIs of the SRMR and CFI suggested that the proposed two-factor model fit well.

Discussion and conclusion

This study compared the suitability of RMSEA, SRMR, and CFI in assessing factor analysis models with small *df*s. The theoretical presentation of the formulas for RMSEA, SRMR, and CFI provided a foundation for understanding the indices and how their point estimates and CIs were calculated. At the population level, the results showed that given a fixed magnitude of model misspecification, the population RMSEA (i.e., the true value) increased noticeably as *df* decreased. This finding is not surprising. By definition, RMSEA penalizes more complex models by including their *df* in the denominator of its formula. Therefore, keeping the misspecified parameter(s) fixed, the population RMSEA is expected to increase as *df* decreases.

On the other hand, for a fixed magnitude of misspecification (i.e., the inter-factor correlation, ρ), population SRMR and CFI are less susceptible to the effect of model size (*df*). In addition, as the levels of factor loadings (λ) increased, the population RMSEA and SRMR increased and population CFI decreased, suggesting a worse fit. Our findings regarding the effects of model size (*df*) and level of factor loadings (λ) on the population fit indices are consistent with those of previous studies (Hancock & Mueller, 2011;

[10]For the acceptable fit cutoff, we used CFI $\geq$ 0.95 and SRMR $\leq R^2 \times 0.10$. For the numerical example, the average communality of the manifest variables was $R^2 = 0.620$. The cutoff value for the SRMR was $0.620 \times 0.10 = 0.062$.

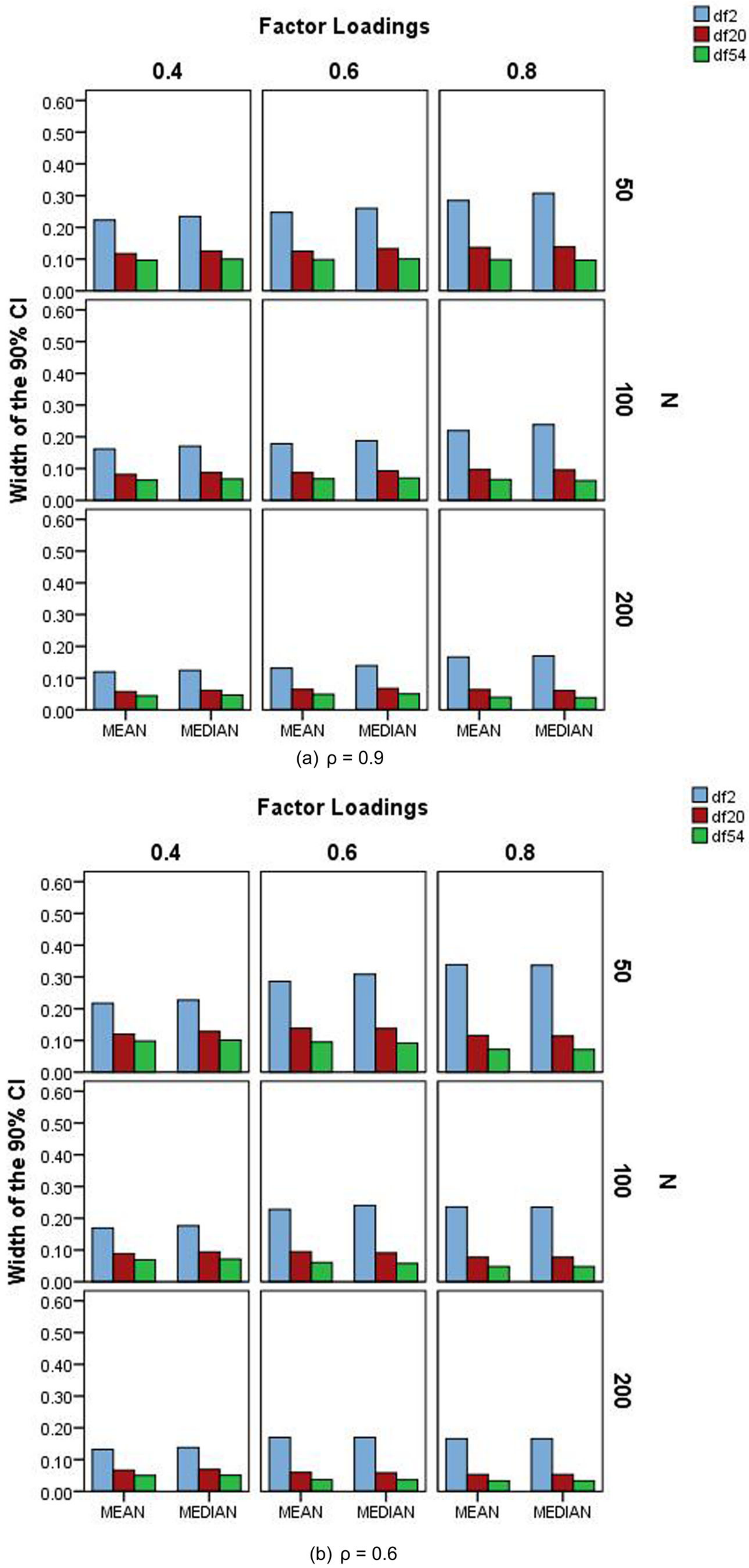


Figure 7. The average and median widths of the 90% confidence intervals: RMSEA.
Note. N = sample size; df = degrees of freedom

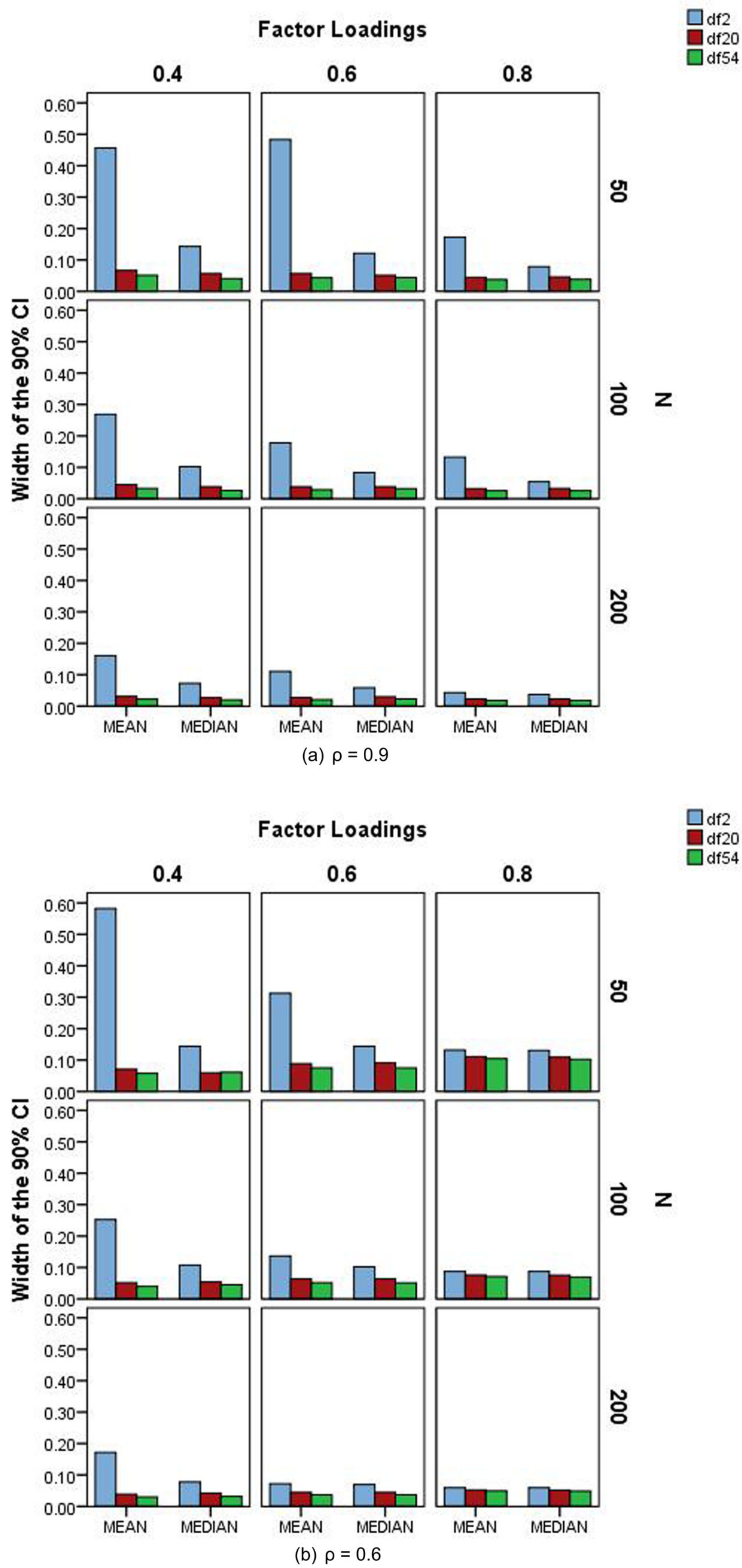
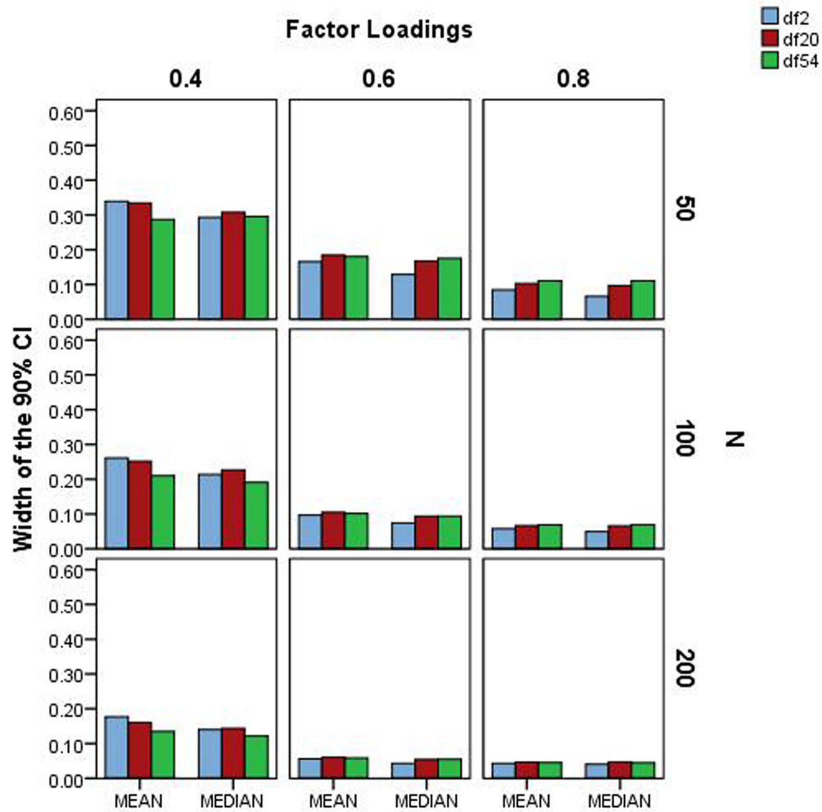
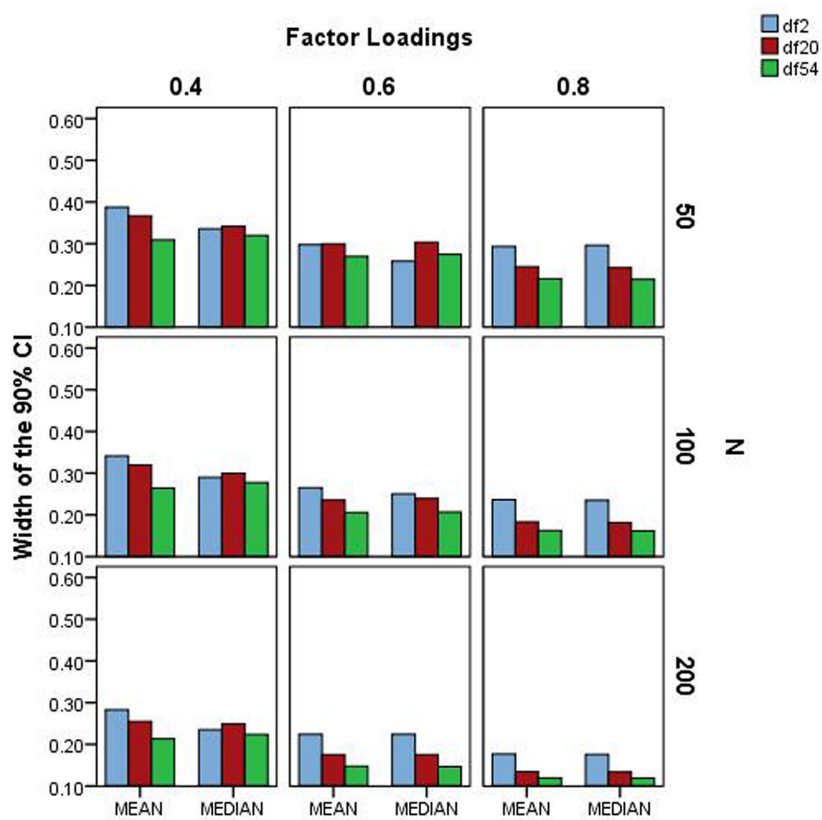


Figure 8. The average and median widths of the 90% confidence intervals: SRMR.
Note. N = sample size; df = degrees of freedom



(a) ρ = 0.9



(b) ρ = 0.6

Figure 9. The average and median widths of the 90% confidence intervals: CFI.
Note. N = sample size; df = degrees of freedom

Kenny & McCoach, 2003; Maydeu-Olivares, 2017; McNeish et al., 2018; Rigdon, 1996; Savalei, 2012; Shi, Maydeu-Olivares, & DiStefano, 2018; Shi et al., 2019).

The results of this study provide a better understanding of the behavior of sample RMSEA, SRMR, and CFI when assessing models with small *df*. For all three fit indices, the biases for the sample estimates decreased as the levels of factor loadings (λ) and sample size (N) increased. Compared to the biased sample SRMR ($\widehat{\text{SRMR}}_b$; Equation (9)), the sample SRMR computed using the unbiased formula ($\widehat{\text{SRMR}}u$; Equation (10)) performed better with small sample sizes, converging to its population value faster. We also compared sample CFIs computed using two formulas. The CFI computed under $\widehat{\text{CFI}}_{\text{ML}}$ (Equation (15)) could converge to its population value faster than $\widehat{\text{CFI}}_{\text{FBC}}$ (Equation (16)) under the conditions considered in the current study, especially for models with a low level of factor loadings (λ) and large *df*. Under both formulas, noticeable biases in the sample CFIs were observed when fitting large models with low factor loadings and very small sample size.

When fitting models with smaller *df*, the variability (i.e., standard deviation) of the sample RMSEA and SRMR tended to increase, especially when the sample size (N) was small. However, the standard deviation of the sample estimates of CFI was less sensitive to model size (*df*). We provided explanations of the impact of model size (*df*) on the standard deviations of the three fit indices. First, we showed, in the Appendix, that under a null hypothesis of close fit, the asymptotic variance of the sample-squared RMSEA has variance $\frac{2}{N^2 \times df} + \frac{4F_0}{N \times df^2}$. This generalizes the result of Rigdon (1996) and Kenny et al. (2015), who showed that in the special case of exact model fit, $F_0 = 0$, this variance is $\frac{2}{N^2 \times df}$. As a result, holding the level of population model misspecification F_0 constant, the variance of the sample RMSEA increases as *df* decreases, leading to higher uncertainty when estimating this population parameter. For the unbiased SRMR, under the normal reference distribution, the expected standard error for SRMR is approximated asymptotically as given in Equation (12). We see in this equation that when holding the degree of misfit constant, as $t = p(p+1)/2$ decreases (in small models), the variability of SRMR_u increases. Finally, in small models, both the independence baseline model and fitted model will involve a small *df*. Since the CFI is a comparative index, the impact of *df* on the variability of the chi-squared test statistics could be canceled out to some degree. Additional studies are needed to further investigate this issue.

From empirical users' perspective, we also examined the performance of applying the conventional cutoffs to the evaluation of small *df* models. When the *df* was very small ($df = 2$), sample RMSEA tended to reject close-fitting models ($\rho = .90$) often (20%), even as the sample size reached 1,000. The sample SRMR and CFI performed better with small *df* models. Generally, as the sample size reaches 200, the unbiased sample SRMR ($\widehat{\text{SRMR}u}$) and CFI ($\widehat{\text{CFI}}_{\text{ML}}$) can be used even for very small models (e.g., $df = 2$), except when the level of factor loadings is very low. Researchers should also be cautious when applying the conventional cutoffs for SRMR or CFI with low factor loadings. When the standardized factor loadings are very low (i.e., $\lambda \leq 0.4$), using sample SRMR or CFI with a conventional cutoff often leads to rejection of correctly specified or close-fitting models, except when the sample size is very large (e.g., $N > 1{,}000$).

The three fit indices performed similarly in terms of the accuracy of their p values in close fit tests and CI coverage rates. RMSEA performed better when the level of factor loadings was low (e.g., $\lambda = 0.40$) and/or the level of model misfit was minor (e.g., $\rho = 0.90$), whereas SRMR and CFI were superior to RMSEA in models with more severe levels of misfit (e.g., $\rho = 0.60$) and/or higher levels of factor loadings (e.g., $\lambda = 0.80$). It is notable that the performance of CIs for SRMR and CFI became worse as the level of model misfit decreased. The plausible explanations are as follows. The values of SRMR and CFI are bounded by zero and one, respectively. As a result, a normal approximation may fail to represent the observations at the tail of the distribution, especially when the degree of model misfit is small (i.e., producing more 0 s for SRMR and more 1 s for CFI).

In addition, the widths of the CIs for CFI are not sensitive to the model size (df). The average widths of the CIs for RMSEA and SRMR can be noticeably wide under models with very small *df* (e.g., $df = 2$), especially when the sample size is small. Under the above conditions, however, the median widths of the CIs for SRMR were much narrower than the average widths; this pattern was not observed for the CIs of RMSEA.

Why were the average and median CI sizes for SRMR so different? We examined the empirical distributions of the size of the CIs for both RMSEA and SRMR under a specific condition (i.e., $\lambda = 0.60$, $\rho = 0.90$, $N = 100$, and $df = 2$). For RMSEA, the size of the CIs was generally large across most replications: 4,478 of 5,000 (89.6%) of RMSEA CIs were larger than 0.10. However, the CIs for SRMR were generally

Table 6. Empirical rejection rates of the close fit test ($\alpha = .05$).

			$\rho = .90$			$\rho = .60$		
N	λ	df	RMSEA	SRMR	CFI	RMSEA	SRMR	CFI
50	.4	2	0.02	0.01	**0.03**	0.01	0.01	**0.03**
		20	**0.07**	0.10	0.19	**0.06**	**0.03**	0.17
		54	0.16	0.17	0.41	0.15	**0.05**	0.39
	.6	2	**0.04**	**0.02**	0.01	**0.02**	0.01	**0.02**
		20	0.09	**0.06**	**0.06**	**0.07**	0.01	0.09
		54	0.18	0.09	0.15	0.15	0.01	0.15
	.8	2	**0.05**	**0.02**	0.01	0.02	**0.02**	**0.03**
		20	0.09	**0.02**	**0.03**	**0.07**	**0.03**	**0.05**
		54	0.16	**0.03**	**0.07**	0.13	**0.03**	**0.07**
100	.4	2	**0.02**	**0.02**	0.02	0.02	0.01	**0.02**
		20	**0.06**	0.13	0.09	**0.05**	**0.04**	0.11
		54	0.10	0.20	0.17	0.08	**0.06**	0.18
	.6	2	**0.04**	**0.03**	0.01	**0.02**	0.02	0.02
		20	**0.07**	**0.07**	**0.02**	**0.06**	**0.02**	**0.05**
		54	0.10	0.10	**0.04**	0.09	0.02	**0.07**
	.8	2	**0.05**	**0.03**	0.01	**0.02**	**0.06**	**0.03**
		20	**0.07**	**0.03**	**0.02**	**0.06**	**0.07**	**0.03**
		54	0.10	**0.04**	**0.03**	0.10	**0.06**	**0.04**
200	.4	2	**0.04**	**0.03**	0.01	0.02	0.01	0.01
		20	**0.05**	0.13	**0.04**	**0.05**	**0.05**	**0.06**
		54	**0.07**	0.20	**0.05**	**0.07**	**0.06**	0.08
	.6	2	**0.05**	**0.03**	0.00	**0.04**	**0.03**	0.02
		20	**0.07**	**0.07**	0.01	**0.07**	**0.03**	**0.04**
		54	**0.07**	0.09	0.02	0.09	**0.03**	**0.04**
	.8	2	**0.05**	**0.04**	0.01	**0.02**	0.09	**0.02**
		20	**0.07**	**0.04**	**0.02**	**0.07**	0.09	**0.04**
		54	0.09	**0.05**	**0.03**	0.11	**0.08**	**0.04**
500	.4	2	**0.05**	**0.04**	0.01	**0.03**	**0.03**	0.02
		20	**0.05**	0.12	0.01	**0.06**	**0.05**	**0.04**
		54	**0.06**	0.17	0.02	**0.07**	**0.06**	**0.04**
	.6	2	**0.05**	**0.04**	0.01	**0.04**	**0.04**	**0.02**
		20	**0.06**	**0.05**	0.01	**0.07**	**0.03**	**0.04**
		54	**0.06**	0.09	0.01	0.10	**0.03**	**0.04**
	.8	2	**0.05**	**0.05**	**0.02**	**0.03**	0.16	**0.03**
		20	**0.06**	**0.04**	**0.03**	**0.08**	0.10	**0.04**
		54	**0.08**	**0.06**	**0.03**	0.11	0.09	**0.04**
1000	.4	2	**0.05**	**0.04**	0.00	**0.04**	**0.04**	0.02
		20	**0.05**	0.10	0.01	**0.06**	**0.04**	**0.03**
		54	**0.05**	0.14	0.01	**0.07**	**0.06**	**0.03**
	.6	2	**0.05**	**0.04**	0.01	**0.05**	**0.04**	**0.03**
		20	**0.06**	**0.05**	0.01	**0.08**	**0.04**	**0.04**
		54	**0.06**	0.08	0.02	0.09	**0.04**	**0.04**
	.8	2	**0.05**	**0.05**	**0.03**	**0.03**	0.21	**0.02**
		20	**0.07**	**0.04**	**0.03**	0.09	0.09	**0.04**
		54	**0.07**	**0.06**	**0.04**	0.11	**0.08**	**0.04**

Note. N = sample size; λ = standardized factor loadings; *df* = degrees of freedom; ρ = interfactor correlation; empirical rejection rates between 0.02 and 0.08 are highlighted in bold.

much narrower than the RMSEA CIs in the majority of replications: Only 1,192 of 5,000 (23.8%) SRMR CIs were larger than 0.10. Nevertheless, a small proportion of SRMR CIs were extremely large: 276 out of 5,000 (5.5%) were larger than 0.30. Due to this small number of extreme values, the average size of SRMR CIs was inflated and their median size was noticeably smaller than average. In summary, in assessing very small models with small sample sizes (e.g., $df = 2$ and $N < 100$), SRMR CIs were mostly narrower than RMSEA CIs. Nevertheless, there is a slight chance that unreasonably wide CIs may be observed using SRMR. Future studies should explore this phenomenon further.

This study is not without limitations. First, its findings on sample and interval estimates are based on multivariate normally distributed data, and many conclusions are conditional on the assumption of multivariate normality. The assumption of normality can be violated in many applications (Micceri, 1989). Statistical theories and formulas to estimate RMSEA, SRMR, CFI, and close fit tests under non-normal data are available (Brosseau-Liard et al., 2012; Brosseau-Liard & Savalei, 2014; Gao et al., 2020; Lai, 2019a, 2019b, 2020; Maydeu-Olivares, 2017; Savalei, 2018; Shi et al., 2020). Based on previous findings (e.g., Gao et al., 2020; Lai, 2019a, 2019b, 2020; Maydeu-Olivares et al., 2018), we expected the formulas to show

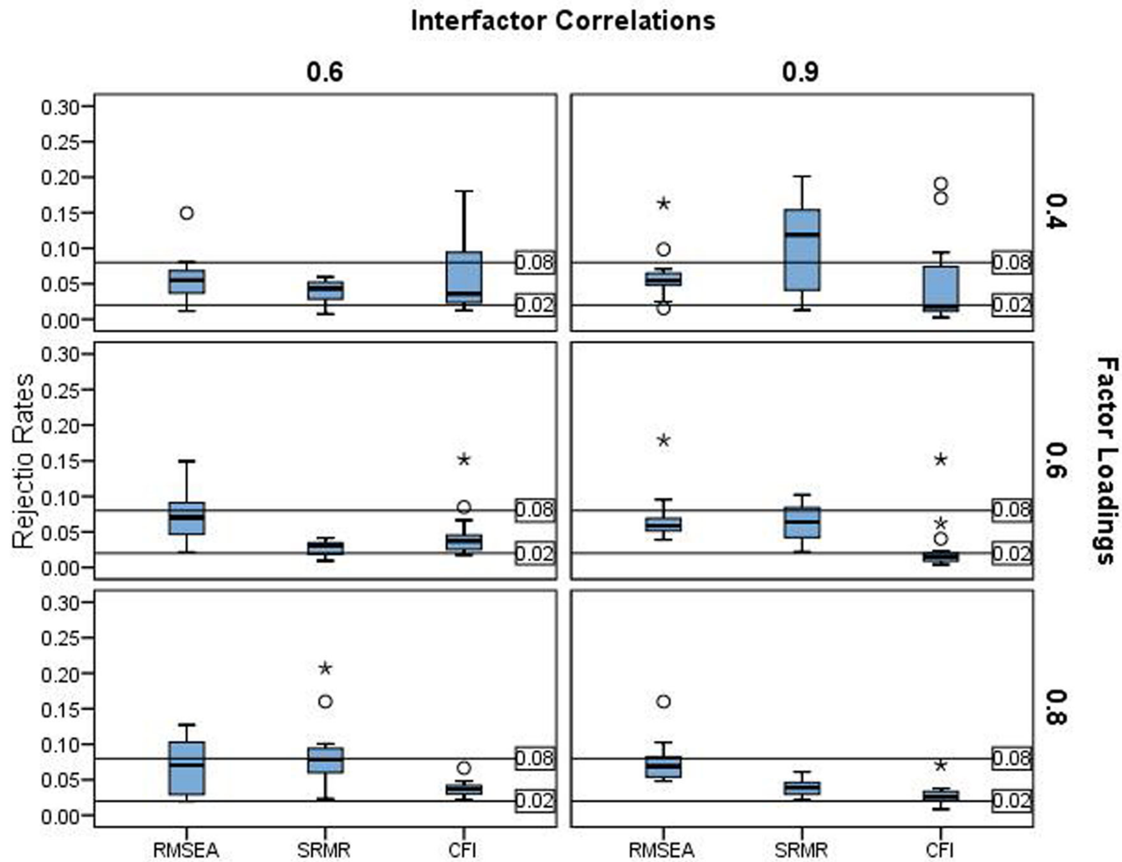


Figure 10. Empirical rejection rates of the close fit test (α =.05).
Note. Reference lines indicate the range of acceptable rejection rates (i.e., between 0.02 and 0.08).

Table 7. Descriptive statistics for the open-book closed-book data.

Items	Means	Standard deviations	Skewness	Kurtosis
1. Mechanics	38.95	17.49	−0.33	−0.46
2. Vectors	50.59	13.15	−0.23	0.32
3. Algebra	50.60	10.62	−0.32	1.30
4. Analysis	46.68	14.85	−0.64	−0.46
5. Statistics	42.31	17.26	0.47	−0.31

acceptable performance under non-normal data. However, future studies should verify the performance of RMSEA, SRMR, and CFI in assessing small *df* models with non-normal data. Moreover, this study only considered one type of model misspecification (misspecified dimensionality under factor analysis models). In future studies, additional types of models (e.g., path analysis models) and model misspecifications (e.g., omitted cross-loading) should be investigated.

In summary, our findings support the idea that the behaviors of fit indices rely not only on the model fit or misfit, but also on the context of the model, such as the model size (*df*). The population RMSEA is heavily influenced by the model size (*df*). When *df* was very small, the sample RMSEA often erroneously rejected correctly specified or close-fitting models. The CIs of RMSEA also tended to be very wide, suggesting high uncertainty regarding the size of the model misfit. In comparison with RMSEA, the population SRMR and CFI are less susceptible to the effects of changes in *df*. The sample SRMR and CFI could provide more useful information in assessing models with very small *df*. A sample of $N = 200$ observations is generally adequate for interpreting the sample SRMR ($\widehat{\mathrm{SRMR}}u$) or CFI ($\widehat{\mathrm{CFI}}_{\mathrm{ML}}$) in extremely small models (e.g., $df = 2$), unless the level of factor loadings is very low ($\lambda \leq 0.40$). In general, the 90% CIs and p values for close fit tests under SRMR and CFI were accurate. The CIs for SRMR can be fairly wide as the *df* is very small, especially when the sample size (N) is small and factor loadings are very low. The widths of CIs for CFI are less sensitive to the impact of *df*.

Based on the major findings, we provide the following concluding remarks. *When assessing very small models (e.g., $df = 2$), researchers should be cautious in interpreting RMSEA and should rely more on SRMR and CFI. In addition, researchers should pay close attention when interpreting fit indices of models with very small df* *and low factor loadings, especially when the sample size is small.* It is worth noting that based on classic psychometric theory (Lord & Novick, 1968; McDonald, 1999), short scales with low (standardized) factor loadings are not recommended for use in practice as they could generate unreliable test scores. In addition, considering construct representation, short scales are generally suitable for "narrow" measures that have "content homogeneous indicators." However, researchers should be cautious in applying short scales for conceptually broader constructs that have more diverse item content (Cattell, 1966; Reise et al., 2007).

We acknowledge and emphasize that evaluating model fit is a critical but difficult task and that there are no "golden rules" for assessing any model. It is generally recommended that model fit be evaluated based on more than one index. For instance, Hu and Bentler (1999) recommended a two-index strategy using a combination of SRMR and one supplementary index[11]. Our recommendations are consistent with this two-index strategy in that the SRMR and CFI may be used to assess models with small *df*. We hope that the findings from this study inform researchers in psychological, behavioral, and other social science fields, who work with small *df* models.

Article information

Conflict of Interest Disclosures: Each author signed a form for disclosure of potential conflicts of interest. No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: This work was supported in part by the National Science Foundation through Grant No. SES-1659936 to A. Maydeu-Olivares, and the National Research Foundation of Korea through Grant No. 2017R1C1B2012424 and No. 2020R1H1A1102581 to T. Lee.

[11]The supplementary indices included the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973), Bollen's (1989b) fit index (BL89), relative noncentrality index (RNI; McDonald & Marsh, 1990), CFI, Gamma hat (Steiger, 1989), McDonald's centrality index (Mc; McDonald, 1989), and RMSEA.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments: The authors would like to thank Dr. Peter C. M. Molenaar and the anonymous reviewers for their comments on prior versions of this manuscript. The authors are grateful to Dr. Keke Lai for valuable discussions on estimating CFI. The ideas and opinions expressed herein are those of the authors alone, and endorsement by the author's institutions or the National Science Foundation is not intended and should not be inferred.

ORCID

Dexin Shi http://orcid.org/0000-0002-4120-6756
Christine DiStefano http://orcid.org/0000-0001-7504-6554
Alberto Maydeu-Olivares http://orcid.org/0000-0001-5790-392X

References

Asparouhov, T., & Muthén, B. (2018). SRMR in Mplus. Retrieved May 2, 2018, from www.statmodel.com/download/SRMR2.pd

Austin, J. T., & Calderón, R. F. (1996). Theoretical and technical contributions to structural equation modeling: An updated annotated bibliography. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(2), 105–175. https://doi.org/10.1080/10705519609540036

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M. (1995). *EQS 5 [Computer Program]*. Multivariate Software Inc.

Bentler, P. M. (2004). *EQS 6 [Computer Program]*. Multivariate Software Inc.

Bollen, K. A. (1989a). *Structural equations with latent variables*. Wiley.

Bollen, K. A. (1989b). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*(3), 303–316. https://doi.org/10.1177/0049124189017003004

Box, G. E. P. (1979). Some problems of statistics and everyday life. *Journal of the American Statistical Association*, *74*(365), 1–4. https://doi.org/10.1080/01621459.1979.10481600

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology*, *31*(2), 144–152. https://doi.org/10.1111/j.2044-8317.1978.tb00581.x

Brosseau-Liard, P. E., & Savalei, V. (2014). Adjusting incremental fit indices for nonnormality. *Multivariate Behavioral Research*, *49*(5), 460–470. https://doi.org/10.1080/00273171.2014.933697

Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An investigation of the sample performance of two nonnormality corrections for RMSEA. *Multivariate Behavioral Research*, *47*(6), 904–930. https://doi.org/10.1080/00273171.2012.715252

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. s. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Bruner, G. C., Hensel, P. J., & James, K. (2005). *Marketing scales handbook: A compilation of multi-item measures for consumer behavior and advertising* (Vol. IV). GCBII Productions.

Cai, L., & Lee, T. (2009). Covariance structure model fit testing under missing data: An application of the supplemented EM algorithm. *Multivariate Behavioral Research*, *44*(2), 281–304. https://doi.org/10.1080/00273170902794255

Cattell, R. B. (1966). Psychological theory and scientific method. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 1–18). Rand McNally.

Chen, F., Curran, P. J., Bollen, K. A., Kirby, J., & Paxton, P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research*, *36*(4), 462–494. https://doi.org/10.1177/0049124108314720

de Jong, M. G., Steenkamp, J.-B E. M., & Veldkamp, B. P. (2009). A model for the construction of country-specific, yet internationally comparable short-form marketing scales. *Marketing Science*, *28*(4), 674–689. https://doi.org/10.1287/mksc.1080.0439

Everitt, B. (1998). *The Cambridge dictionary of statistics*. Cambridge University Press.

Fan, X., & Sivo, S. A. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rationale of two-index strategy revisited. *Structural Equation Modeling: A Multidisciplinary Journal*, *12*(3), 343–367. https://doi.org/10.1207/s15328007sem1203_1

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research*, *42*(3), 509–529. https://doi.org/10.1080/00273170701382864

Gao, C., Shi, D., & Maydeu-Olivares, A. (2020). Estimating the maximum likelihood root mean square error of approximation (RMSEA) with non-normal data: A Monte-Carlo study. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(2), 192–201. https://doi.org/10.1080/10705511.2019.1637741

Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, *20*(1), 102–116. https://doi.org/10.1037/a0038889

Hancock, G. R., & Mueller, R. O. (2010). *The reviewer's guide to quantitative methods in the social sciences*. Routledge.

Hancock, G. R., & Mueller, R. O. (2011). The reliability paradox in assessing structural relations within covariance structure models. *Educational and Psychological Measurement*, *71*(2), 306–324. https://doi.org/10.1177/0013164410384856

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Lai, K. (2019a). A simple analytic confidence interval for CFI given nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *26*(5), 757–777. https://doi.org/10.1080/10705511.2018.1562351

Lai, K. (2019b). Correct point estimator and confidence interval for RMSEA given categorical data. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(5), 678–695. https://doi.org/10.1080/10705511.2019.1687302

Lai, K. (2020). Better confidence intervals for RMSEA in growth models given nonnormal data. *Structural Equation Modeling: A Multidisciplinary Journal*, *27*(2), 255–274. https://doi.org/10.1080/10705511.2019.1643246

Lee, K., & Allen, N. J. (2002). Organizational citizenship behavior and workplace deviance: The role of affect and cognitions. *The Journal of Applied Psychology*, *87*(1), 131–142. https://doi.org/10.1037/0021-9010.87.1.131

Légaré, F., Kearing, S., Clay, K., Gagnon, S., D'Amours, D., Rousseau, M., & O'Connor, A. (2010). Are you SURE?: Assessing patient decisional conflict with a 4-item screening test. *Canadian Family Physician*, *56*(8), e308–e314.

Lord, F. M., & Novick, M. R. (1968). Statistical Theories of Mental Test Scores. Menlo Park: Addison-Wesley.

Löwe, B., Wahl, I., Rose, M., Spitzer, C., Glaesmer, H., Wingenfeld, K., Schneider, A., & Brähler, E. (2010). A 4-item measure of depression and anxiety: Validation and standardization of the Patient Health Questionnaire-4 (PHQ-4) in the general population. *Journal of Affective Disorders*, *122*(1–2), 86–95. https://doi.org/10.1016/j.jad.2009.06.019

Jöreskog, K. G., & Sörbom, D. (1988). *LISREL 7. A guide to the program and applications* (2nd ed.). International Education Services.

Jöreskog, K. G., & Sörbom, D. (2017). *LISREL (Version 9.3) [Computer program]*. Scientific Software International.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, *44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *10*(3), 333–351. https://doi.org/10.1207/S15328007SEM1003_1

MacCallum, R. C. (2003). 2001 Presidential address: Working with imperfect models. *Multivariate Behavioral Research*, *38*(1), 113–139. https://doi.org/10.1207/S15327906MBR3801_5

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*(1), 201–226. https://doi.org/10.1146/annurev.psych.51.1.201

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1979). *Multivariate analysis*. Academic Press.

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*, *82*(3), 533–558. https://doi.org/10.1007/s11336-016-9552-7

Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing fit in structural equation models: A Monte-Carlo evaluation of RMSEA versus SRMR confidence intervals and tests of close fit. *Structural Equation*

Modeling: A Multidisciplinary Journal, *25*(3), 389–402. https://doi.org/10.1080/10705511.2017.1389611

Maydeu-Olivares, A., & Steenkamp, J. E. M. (2019). An integrated procedure to control for common method variance in survey data using random intercept factor analysis models. *Manuscript Submitted for Publication*.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*(1), 97–103. https://doi.org/10.1007/BF01908590

McDonald, R. P. (1999). *Test theory: A unified approach*. Erlbaum.

McDonald, R. P., & Ho, M.-H R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, *7*(1), 64–82. https://doi.org/10.1037/1082-989x.7.1.64

McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness of fit. *Psychological Bulletin*, *107*(2), 247–255. https://doi.org/10.1037/0033-2909.107.2.247

McNeish, D., An, J., & Hancock, G. R. (2018). The thorny relation between measurement quality and fit index cutoffs in latent variable models. *Journal of Personality Assessment*, *100*(1), 43–52. https://doi.org/10.1080/00223891.2017.1281286

Meredith, W., & Tisak, J. (1990). Latent curve analysis. *Psychometrika*, *55*(1), 107–122. https://doi.org/10.1007/BF02294746

Micceri, T. (1989). The unicorn, the normal curve, and other improbable creatures. *Psychological Bulletin*, *105*(1), 156–166. https://doi.org/10.1037/0033-2909.105.1.156

Moshagen, M. (2012). The model size effect in SEM: Inflated goodness-of-fit statistics are due to the size of the covariance matrix. *Structural Equation Modeling: A Multidisciplinary Journal*, *19*(1), 86–98. https://doi.org/10.1080/10705511.2012.634724

Muthén, L. K., & Muthén, B. O. (2017). *MPLUS 8*. Muthén & Muthén.

Ogasawara, H. (2001a). Approximations to the distributions of fit indexes for misspecified structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal*, *8*(4), 556–574. https://doi.org/10.1207/S15328007SEM0804_03

Ogasawara, H. (2001b). Standard errors of fit indices using residuals in structural equation modeling. *Psychometrika*, *66*(3), 421–436. https://doi.org/10.1007/BF02294443

Ogasawara, H. (2007). Higher-order approximations to the distributions of fit indexes under fixed alternatives in structural equation models. *Psychometrika*, *72*(2), 227–243. https://doi.org/10.1007/s11336-004-1206-5

Pavlov, G., Maydeu-Olivares., & Shi, D. (2020). Using the standardized root mean squared residual (SRMR) to assess exact fit in structural equation models. *Educational and Psychological Measurement*, Manuscript in press.

Pornprasertmanit, S., Miller, P., Schoemann, A. M. (2012). R packagesimsem: SIMulated structural equation modeling. Available from the Comprehensive R Archive Network: http://cran.r-project.org

Price, J. L. (1977). *The study of turnover*. Iowa State University Press.

R Development Core Team. (2015). *R: A language and environment for statistical computing*. R Development Core Team.

Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research*, *16*(S1), 19–31. https://doi.org/10.1007/s11136-007-9183-7

Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(4), 369–379. https://doi.org/10.1080/10705519609540052

Robinson, M. A. (2018). Using multi-item psychometric scales for research and practice in human resource management. *Human Resource Management*, *57*(3), 739–750. https://doi.org/10.1002/hrm.21852

Rosen, R. C., Catania, J. A., Althof, S. E., Pollack, L. M., O'Leary, M., Seftel, A. D., & Coon, D. W. (2007). Development and validation of four-item version of Male Sexual Health Questionnaire to assess ejaculatory dysfunction. *Urology*, *69*(5), 805–809. https://doi.org/10.1016/j.urology.2007.02.036

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(4), 561–582. https://doi.org/10.1080/10705510903203433

Savalei, V. (2012). The relationship between root mean square error of approximation and model misspecification in confirmatory factor analysis models. *Educational and Psychological Measurement*, *72*(6), 910–932. https://doi.org/10.1177/0013164412452564

Savalei, V. (2018). On the computation of the RMSEA and CFI from the mean-and-variance corrected test statistic with nonnormal data in SEM. *Multivariate Behavioral Research*, *53*(3), 419–429. https://doi.org/10.1080/00273171.2018.1455142

Sharma, S., Mukherjee, S., Kumar, A., & Dillon, W. R. (2005). A simulation study to investigate the use of cutoff values for assessing model fit in covariance structure models. *Journal of Business Research*, *58*(7), 935–943. https://doi.org/10.1016/j.jbusres.2003.10.007

Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the model size effect on SEM fit indices. *Educational and Psychological Measurement*, *79*(2), 310–334. https://doi.org/10.1177/0013164418783530

Shi, D., Lee, T., & Terry, R. A. (2015). Abstract: Revisiting the model size effect in structural equation modeling (SEM). *Multivariate Behavioral Research*, *50*(1), 142–142. https://doi.org/10.1080/00273171.2014.989012

Shi, D., Lee, T., & Terry, R. A. (2018). Revisiting the model size effect in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *25*(1), 21–40. https://doi.org/10.1080/10705511.2017.1369088

Shi, D., Maydeu-Olivares, A., & DiStefano, C. (2018). The relationship between the standardized root mean square residual and model misspecification in factor analysis models. *Multivariate Behavioral Research*, *53*(5), 676–694. https://doi.org/10.1080/00273171.2018.1476221

Shi, D., Maydeu-Olivares, A., & Rosseel, Y. (2020). Assessing fit in ordinal factor analysis models: SRMR vs. RMSEA. *Structural Equation Modeling: A*

Multidisciplinary Journal, *27*(1), 1–15. https://doi.org/10.1080/10705511.2019.1611434

Steiger, J. H. (1989). *EzPATH: A supplementary module for SYSTAT and SYGRAPH*. Systat, Inc.

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, *25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Steiger, J., & Lind, J. C. (1980). *Statistically based tests for the number of common factors [Paper presentation]*. Paper Presented at the Annual Meeting of the Annual Spring Meeting of the Psychometric Society, Iowa City.

Tremblay, P. F., & Gardner, R. C. (1996). On the growth of structural equation modeling in psychological journals. *Structural Equation Modeling: A Multidisciplinary Journal*, *3*(2), 93–104. https://doi.org/10.1080/10705519609540035

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales—Five misunderstandings and ways to overcome them. *Journal of Individual Differences*, *35*(4), 185–189. https://doi.org/10.1027/1614-0001/a000148

Appendix

Asymptotic mean and the variance of the sample-squared RMSEA in close fitting models

Under parameter drift assumptions, the asymptotic distribution of the likelihood-ratio test statistic can be approximated by a non-central chi-squared distribution with df degrees of freedom and noncentrality parameter $\boldsymbol{\lambda} = NF_0$, where F_0 denotes the population discrepancy between the data generating process and the fitted model (Browne & Cudeck, 1993; Steiger & Lind, 1980). Since the mean and variance of a non-central chi-square distribution with df degrees of freedom and noncentrality parameter $\boldsymbol{\lambda}$ are $df + \boldsymbol{\lambda}$ and $2df + 4\boldsymbol{\lambda}$, respectively, it follows that under the null hypothesis of close model fit $H_0^* : RMSEA \le k$,

$$E\left(\widehat{\mathrm{RMSEA}}_{ML}^{2}\right) = E\left(\frac{X_{\mathrm{ML}}^2 - df}{N \times df}\right) = \frac{E\left(X_{\mathrm{ML}}^2\right) - df}{N \times df} = \frac{\boldsymbol{\lambda}}{N \times df} = \frac{F_0}{df}$$

$$\mathrm{var}\left(\widehat{\mathrm{RMSEA}}_{ML}^{2}\right) = \mathrm{var}\left(\frac{X_{ML}^2 - df}{N \times df}\right) = \frac{\mathrm{var}\left(X_{ML}^2\right)}{N^2 \times df^2} = \frac{\boldsymbol{2df + 4\lambda}}{N^2 \times df^2} = \frac{2}{N^2 \times df} + \frac{4F_0}{N \times df^2}$$

In the special case where the model fits exactly, i.e., H_0^* : RMSEA $= 0$, $F_0 = 0$, and these expressions reduce to those reported in Kenny et al. (2015, p. 492): $E(\widehat{\mathrm{RMSEA}}_{\mathrm{ML}}^2) = 0$ and $\mathrm{var}(\widehat{\mathrm{RMSEA}}_{\mathrm{ML}}^2) = \frac{2}{N^2 \times df}$.