Meta-Psychology, 2024, vol 8, MP.2021.2757
https://doi.org/10.15626/MP.2021.2757
Article type: Original Article

Open data: Not Applicable
Open materials: Yes
Open Code: Yes
Open and reproducible analysis: Yes
Open reviews and editorial process: Yes
Preregistration: No

Edited by: Brenton Wiernik
Reviewed by: Victoria Savalei, Casey Giordano
Analysis reproduced by: Rickard Carlsson
All supplementary files can be accessed at OSF:
https://doi.org/10.17605/OSF.IO/9U3ZC

The Devil is Mainly in the Nuisance Parameters: Performance of Structural Fit Indices Under Misspecified Structural Models in SEM

Moritz Heene[1], Michael Maraun[2], Nadine Juliana Glushko[3], and Sunthud Pornprasertmanit[4]
[1]Department of Psychology, Ludwig-Maximilians-Universität München, Germany
[2]Department of Psychology, Simon Fraser University, Burnaby, B.C., Canada
[3]Independent scholar
[4]Chulalongkorn University, Bangkok, Thailand

To provide researchers with a means of assessing the fit of the structural component of structural equation models, structural fit indices- modifications of the composite fit indices, RMSEA, SRMR, and CFI- have recently been developed. We investigated the performance of four of these structural fit indices- RMSEA-P, RMSEA_S, SRMR_S, and CFI_S-, when paired with widely accepted cutoff values, in the service of detecting structural misspecification. In particular, by way of simulation study, for each of seven fit indices- 3 composite and 4 structural-, and the traditional chi-square test of perfect composite fit, we estimated the following rates: a) Type I error rate (i.e., the probability of (incorrect) rejection of a correctly specified structural component), under each of four degrees of misspecification in the measurement component; and b) Power (i.e., the probability of (correct) rejection of an incorrectly specified structural model), under each condition formed of the pairing of one of three degrees of structural misspecification with one of four degrees of measurement component misspecification. In addition to sample size, the impacts of two model features, incidental to model misspecification- number of manifest variables per latent variable and magnitude of factor loading- were investigated. The results suggested that, although the structural fit indices performed relatively better than the composite fit indices, none of the goodness-of-fit index with a fixed cutoff value pairings was capable of delivering an entirely satisfactory Type I error rate/Power balance, [RMSEA_S,.05] failing entirely in this regard. Of the remaining pairings; a) RMSEA-P and CFI_S suffered from a severely inflated Type I error rate; b) despite the fact that they were designed to pick up on structural features of candidate models, all pairings- and especially, RMSEA-P and CFI_S-manifested sensitivities to model features, incidental to structural misspecification; and c) although, in the main, behaving in a sensible fashion, SRMR_S was only sensitive to structural misspecification when it occurred at a relatively high degree.

Keywords: structural equation modeling, model fit, goodness-of-fit, model misfit, fit indices, cutoff values, model test, theory testing

The remarkable growth in quantitative theory relating to structural equation modeling (SEM) that has occurred during the past four decades, has equipped the researcher with tools of unprecedented power with which to estimate and test an equally remarkable diversity of candidate models. Because parameter estimates—more broadly, research implications—attaching to a particular candidate model can be taken seriously, only if the model is a reasonable approximation to reality, the detection of—and assessment of degree of—model misspecification has become a topic of increasing importance. Letting $\{\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Phi}, \mathbf{\Psi}, \ldots\}_T$ be the numerical values assumed by the parameters of a true population model T, in a particular population, a candidate model C is misspecified *vis à vis* T if it imposes restrictions such that it cannot be the case that $\{\mathbf{B}, \mathbf{\Gamma}, \mathbf{\Phi}, \mathbf{\Psi}, \ldots\}_T = \{\hat{\mathbf{B}}, \hat{\mathbf{\Gamma}}, \hat{\mathbf{\Phi}}, \hat{\mathbf{\Psi}}, \ldots\}_C$. Thus, for example, a model C is misspecified in the event that it restricts to zero, a regression parameter linking two latent variables, when, in fact, the parameter is nonzero under T.

Traditionally, approaches to the detection of misspecification have relied upon a sample-based estimate of the difference between the population covariance matrix Σ and the covariance matrix $\hat{\Sigma}(\mathbf{\Theta})$ implied under a particular candidate model, wherein $\mathbf{\Theta}$ contains model parameters; or, in other words, on a sample-based quantification of the degree of *misfit*.

Evidently, misfit is a function of both misspecification error and sampling error. Yuan (2005) classifies misfit-based approaches to the detection of misspecification into two categories: binary hypothesis tests—notably, the chi-square test of the hypothesis pair [$H_0 : \Sigma = \tilde{\Sigma}(\mathbf{\Theta}), H_1 : \Sigma$ any Gramian matrix]—and goodness-of-fit indices. Jöreskog's (1969) early warnings to the effect that, as sample size becomes large, even trivial degrees of model misspecification will lead to rejection of H_0 by the chi-square test, were highly influential, however, and goodness-of-fit indices have become, nowadays, the chief means by which researchers adjudicate model misspecification. Curiously, perhaps because the practical interpretation of such fit indices is rarely straightforward, it has come to be a commonplace for applied researchers to pair fit indices with cutoff values recommended by methodological experts, Hu and Bentler (1999) work on the latter topic being considerably the most cited. Of course, to pair a goodness-of-fit index with a fixed cutoff value—hereafter, called a *GFICV* pairing or strategy—is to create a *de facto* binary hypothesis testing procedure, and, accordingly, an inferential tool of the very same sort as the heavily criticized chi-square test of the pair [$H_0 : \Sigma = \tilde{\Sigma}(\mathbf{\Theta}), H_1 : \Sigma$ any Gramian matrix].

Not surprisingly, then, the employment of *GFICV* pairings in the service of adjudicating model misspecification has, itself, been heavily criticized. In the first place, it has been pointed out that, although Hu and Bentler explicitly conditioned their recommendations respecting cutoff values on the specific type of confirmatory factor analytic model which featured in their work, researchers seem to have interpreted these values as global standards broadly applicable across the domain of SEM. Jackson et al. (2009), for example, found that, although almost 60% of studies explicitly employed or referenced Hu's and Bentler's recommendations, in few of these studies was there evidence that "...warnings about strict adherence to Hu and Bentler's suggestions were being heeded" (p. 18). In the second place, being that a given *GFICV* pairing is a species of binary inferential decision-making machinery, its decision-making performance can be efficiently characterized in terms of the dual concepts of *Power* (i.e., the probability that it will yield a (correct) rejection of a misspecified candidate model) and *Type I error rate* (i.e., the probability that it will yield an (incorrect) rejection of a (correctly specified model). *GFICV* pairings have, then, been criticized on grounds that the Power they deliver turns out to be a function of not only *sample size* and extant *degree of misspecification*, but, also, manifold model features, incidental to misspecification, among these, *number of indicators per latent variable* and *magnitude of factor loadings* (e.g.,Breivik and Olsson, 2001; Chen et al., 2008; Fan and Sivo, 2007; Heene et al., 2011, 2012; Saris and Satorra, 1993; Saris et al., 2009; Yuan, 2005). It is important to stress that these nuisance parameters have nothing to do with the degree of the misspecification as they are entirely unrelated to incorrectly imposed model restrictions. Their effects on Power are rather complex and can be different for different parameters of the model as the above-mentioned literature has shown and it is therefore essential to investigate their effects in this study.

Even if not employed in conjunction with fixed cutoff values, there exist serious, and still largely unresolved, difficulties attendant to the employment of goodness-of-fit indices in the service of adjudicating *degree of model misspecification*. A general problem is that the relationship between degree of sample-based misfit—i.e., the extent of disagreement between Σ and $\hat{\Sigma}(\theta)$—and candidate model misspecification is exceedingly complex. Hayduk (2014) observed that there is no necessity that a candidate model in which, for example, there is falsely fixed to zero, a single regression coefficient linking two latent variables[1], will yield a $\hat{\Sigma}(\theta)$ which differs appreciably from Σ. To the contrary, it is entirely possible that a misspecification possessing of significant theoretical implications will result in but a small *overall* difference between Σ and $\hat{\Sigma}(\theta)$. And because they are functions of $\Sigma - \hat{\Sigma}(\theta)$, under such a circumstance, and even in conjunction with a respectable sample size, the chi-square test statistic and goodness-of-fit indices, alike, may not depart appreciably from zero (equivalently, may fail to signal the presence of misspecification).

A second difficulty relates to the issue of the *type* of misspecification with respect to which a goodness-of-fit index is sensitive. Traditional goodness-of-fit indices are *composite* or *broad-band*, in the sense that they are functions of $\Sigma - \hat{\Sigma}(\theta)$, the latter of which captures, in a nonparticularized fashion, misspecification occurring in either of the measurement (i.e., the relationships between latent variables and indicators) or structural (i.e., the relationships among latent variables) component of a candidate model. As when a researcher formalizes his or her theory about the interplay of *learned helplessness, anxiety, and social stress* in determining the trajectory of *depression*, in a specification of fixed and free regression parameters linking latent variables, psychological theory is, of course, most immediately instantiated in the *structural* component of a candidate model. Arguably, then, it is the detection of structural misspecification

[1] And which, in consequence, asserts a falsehood which could potentially serve to adversely influence the development of psychological theory as related to the causal interplay of constructs.

that is of paramount importance. If the validity of this line of argumentation is granted, then it must also be granted that it is of critical importance to the progress of psychological science, that researchers have available to them, goodness-of-fit indices, the sensitivities of which are tuned to misspecification occurring within the structural component of the candidate models they test.

Unfortunately, traditional composite fit indices turn out to be notably *insensitive* to misspecification occurring within the structural component of candidate models. In their investigation of the issue, McDonald and Ho (2002) reviewed 14 SEM studies and found that the measurement components of candidate models were better fitting than their structural counterparts, the relatively smaller contribution to global misfit they made, serving to mask misspecification on the structural side of the coin. As they concluded, “the fit of the composite model can appear satisfactory when the few constraints implied by the path model are not, in fact, correctly specified” (McDonald and Ho, 2002, p. 75). In response to the need for such, a number of authors have attempted to create fit indices possessing of a narrow-band sensitivity to misspecification in the structural components of candidate models. The *structural fit indices* thus far developed—RMSEA-P (McDonald & Ho, 2002; Williams & O'Boyle, 2011), RMSEA$_S$, SRMR$_S$, and CFI$_S$ (the latter three, courtesy of Hancock and Mueller, 2011)—are modifications of the composite fit indices, RMSEA, SRMR, and CFI (hence lowercase ‘s’ denotes “structural”).

In what follows, we investigate the performance of RMSEA-P, RMSEA$_S$, SRMR$_S$, and CFI$_S$, in the service of detecting structural misspecification, when paired with widely accepted cutoff values. In particular, by way of simulation study, for each of seven fit indices—3 composite and 4 structural—and the traditional chi-square test of the pair [$H_0 : \Sigma = \tilde{\Sigma}(\boldsymbol{\Theta}), H_1 :$ Σ any Gramian matrix], we estimate the following rates: a) Type I error rate (i.e., the probability of incorrect rejection of a correctly specified structural model), under condition of each of four degrees of misspecification in the measurement component; b) Power (i.e., the probability of correct rejection of an incorrectly specified structural model), under each condition formed of the pairing of one of three degrees of structural misspecification with one of four degrees of misspecification in the measurement component. In addition to sample size, the impacts of two model features incidental to structural misspecification—*number of manifest variables per latent variable* and *magnitude of factor loadings*—are investigated. We will be particularly interested in comparing the performances of the structural fit indices—these, designed for the express purpose of detecting structural misspecification—to those of the traditional composite indices. A key secondary issue on which we hope to shed some light, is the sensitivity of the performance of structural fit indices to the nuisance factor, misspecification in the measurement components of models.

The organization of the manuscript is as follows: firstly, we present a brief overview of each of the structural fit indices, RMSEA-P, RMSEA$_S$, SRMR$_S$, and CFI$_S$; secondly, we describe the design of the simulation study; and, finally, we report and discuss the results, emphasizing the implications for applied research.

Structural Fit Indices

Let there be a candidate model T featuring m endogenous latent variables with p indicators, and n exogenous latent variables with q indicators [$m+n = K$; $p+q = S$]. In matrix algebra notation, the model implied covariance matrix is given as

$$\Sigma = \Lambda\Omega\Lambda' + \Theta, \Sigma = \Lambda\Omega\Lambda' + \Theta, \Sigma = \Lambda\Omega\Lambda' + \Theta, \Sigma = \Lambda\Omega\Lambda' + \Theta, \quad (1)$$

wherein: Λ is an $S \times K$ matrix containing the factor loadings of the S indicators on the K latent variables; Ω is the $K \times K$ covariance matrix of the latent variables; and Θ is the $S \times S$ covariance matrix of the error terms of the indicators (cf. Bollen, 1989, for a detailed account). The matrix Ω, which is solely a function of the structural component of model T, is given as

$$\Omega = \begin{bmatrix} \Phi & \Phi\Gamma'[(\mathbf{I} - \mathbf{B})^{-1}]' \\ (\mathbf{I} - \mathbf{B})^{-1}\Gamma\Phi & [(\mathbf{I} - \mathbf{B})^{-1}](\Gamma\Phi\Gamma' + \Psi)[(\mathbf{I} - \mathbf{B})^{-1}]' \end{bmatrix} \quad (2)$$

In Ω, Φ is the $n \times n$ covariance matrix of the exogenous latent variables; B is an $m \times m$ matrix, the ij-th element of which is the structural impact (regression coefficient) of the j-th endogenous latent variable on the i-th endogenous latent variable on the i-th endogenous latent variable; Γ is an $m \times n$ matrix, the ij-th element of which is the structural impact (regression coefficient) of the j-th exogenous latent variable on the i-th endogenous latent variable; and Ψ is the $m \times m$ covariance matrix of disturbances (i.e., the residuals attaching to the structural equations of the latent endogenous variables).

RMSEA-P

Let $\chi^2_T \left[d_T = \chi^2_T - df_T\right]$ and $\chi^2_{SS} \left[d_{SS} = \chi^2_{SS} - df_{SS}\right]$ be the chi-square test statistics (noncentrality parameters) for candidate model T and its structurally saturated counterpart (wherein are free to be estimated, all possible directional paths linking latent variables), respectively; $\chi^2_P = \chi^2_T - \chi^2_{SS}$; and df_T, df_{SS}, and df_P be the three associated degrees of freedom.

RMSEA-P (McDonald & Ho, 2002; Williams & O'Boyle, 2011) is, then, given as

$$\sqrt{\max\left(\frac{\chi_P^2 - df_P}{df_P(N-1)}, 0\right)} = \sqrt{\max\left(\frac{d_T - d_{SS}}{(df_T - df_{SS})(N-1)}, 0\right)} \tag{3}$$

Evidently, RMSEA-P is interpretable as the amount of misfit contributed by the structural component of T per degree of freedom. Obviously, smaller values of RMSEA-P correspond to better structural fit.

RMSEA$_S$

Let

$$df_{\text{model}} = \frac{K(K+1)}{2} - t_{\text{model}},$$

wherein t_{model} is the number of free parameters in $\{\Phi, \Gamma, B, \Psi\}$, under candidate model T;

$$\tilde{F}_{\text{model}} = \ln|\tilde{\Omega}_T| + \text{tr}\left(\tilde{\Omega}_{SS}\tilde{\Omega}_T^{-1}\right) - \ln|\tilde{\Omega}_{SS}| - K,$$

wherein $\tilde{\Omega}_T[\tilde{\Omega}_{SS}]$ is the $K \times K$ covariance matrix of the latent variables implied by T (i.e., T's structurally saturated counterpart]; and $\tilde{T}_{\text{model}} = (n-1)\tilde{F}_{\text{model}}$. Evidently, $\tilde{F}_{\text{model}}$ quantifies the discrepancy between T and its structurally saturated counterpart, and $\tilde{T}_{\text{model}}$ is a pseudo test statistic. RMSEA$_S$ (Hancock and Mueller, 2011) is, then, given as

$$\sqrt{\max\left(\frac{\tilde{T}_{\text{model}} - df_{\text{model}}}{df_{\text{model}}(N-1)}, 0\right)} = \sqrt{\max\left(\frac{\tilde{F}_{\text{model}} - 1}{df_{\text{model}}(N-1)}, 0\right)} \tag{4}$$

It is clear, by inspection, that smaller values of RMSEA$_S$ correspond to better structural fit.

Hancock and Mueller (2011) warn against the employment of $\tilde{T}_{\text{model}}$ as a test statistic in formal binary hypothesis testing contexts, and emphasize, instead, its role as a component of descriptive structural fit indices. That being said, the employment of $\tilde{T}_{\text{model}} - df_{\text{model}}$ as an estimator of noncentrality does carry with it a tacit presumption that $\tilde{T}_{\text{model}}$ is at least approximately distributed as a noncentral chi-square variate. It is not yet known whether this is a reasonable presumption. An essential difference between RMSEA-P and RMSEA$_S$ can be expressed as follows. RMSEA-P is a chi-square-based index which captures the difference in composite model fit between candidate model T and its structurally saturated counterpart. Being that it is simply the difference between chi-square statistics deriving from two distinct estimated model implied covariance matrices, it is—in theory, if not in actuality—dependent upon model features, incidental to structural misspecification, such as *magnitude of the factor loadings* and *number of indicators per latent variable* (Hancock & Mueller, 2011; Heene et al., 2011; Saris & Satorra, 1988; Saris et al., 2009; Savalei, 2012). Because its key ingredient is a quantification of the discrepancy between $\tilde{\Omega}_T$ and $\tilde{\Omega}_{SS}$, each but a component element of the model implied covariance matrix, RMSEA$_S$ can be understood as having been constructed explicitly with the aim that it capture misfit solely in the structural component of a candidate model, and, accordingly, that it be free of the effects of incidental model features such as those related to the measurement component (see Hancock and Mueller, 2011, p. 319).

SRMR$_S$

Let $\hat{\Omega}_{SSij}[\hat{\Omega}_{Tij}]$ be the ij-th element of $\hat{\Omega}_{SS}[\hat{\Omega}_T]$ and $\hat{\omega}_{(sat)ij}$ be the ij-th element of $\hat{\Omega}_{sat}$, the saturated (just-identified) covariance matrix of the structural model. SRMR$_S$, recognizable as a structurally oriented specialization of the standardized root mean squared residual, is then defined as

$$\text{SRMR}_S = \sqrt{\frac{2}{K(K+1)}\sum_i z_i\left(\frac{\hat{\omega}_{SSij} - \hat{\omega}_{Tij}}{\sqrt{\hat{\omega}_{Tii}\hat{\omega}_{Tjj}}}\right)^2} \tag{5}$$

The index is bound below by zero, with smaller values indicative of better structural fit.

CFI$_S$

Let

$$\tilde{F}_{\text{null}} = \ln|\hat{\Omega}_{\text{null}}| + \text{tr}\left(\hat{\Omega}_{SS}\hat{\Omega}_{\text{null}}^{-1}\right) - \ln|\hat{\Omega}_{SS}| - K,$$

wherein $\hat{\Omega}_{\text{null}}$ is the $K \times K$ covariance matrix of the latent variables implied by T's structurally null counterpart (wherein are fixed to zero, all possible directional paths linking latent variables); $\tilde{T}_{\text{null}} = (N-1)\tilde{F}_{\text{null}}$; and

$$df_{\text{null}} = \frac{K(K+1)}{2} - K = \frac{K(K-1)}{2}.$$

CFI$_S$, a structural variant of the CFI composite fit index which quantifies the proportional improvement in structural fit achieved by the candidate model relative to its structurally null counterpart, is then defined as

$$\begin{aligned}\text{CFI}_S &= 1 - \frac{\max\left[(\tilde{T}_{\text{model}} - df_{\text{model}}), 0\right]}{\max\left[(\tilde{T}_{\text{model}} - df_{\text{model}}), (\tilde{T}_{\text{null}} - df_{\text{null}}), 0\right]} \\ &= 1 - \frac{\max\left[(n-1)\tilde{F}_{\text{model}} - df_{\text{model}}, 0\right]}{\max\left[(n-1)\tilde{F}_{\text{model}} - df_{\text{model}}, (n-1)\tilde{F}_{\text{null}} - df_{\text{null}}, 0\right]}\end{aligned} \tag{6}$$

CFI_S is bounded above by unity, with large values indicative of better structural fit.

Design of the Simulation Study

True Population Models

To arrive at true population models of the sort encountered in social research, we relied upon both a Google Images search of the phrase, "structural equation model", and our combined modeling experience. In particular, we determined, a) both the *number of latent variables* and *number of indicators per latent variable*, b) the complexity of the structural component (i.e., the number of pathways linking latent variables) rather typical of published social research in which SEM is employed. A reviewer commented that the Google Images search we undertook is likely to be inadequate to the task of identifying the sorts of structural models commonly arising within the social sciences. In response, we ran a Google Scholar search using the phrase "double mediation model" because a double latent mediation model (as described in the next paragraph) was the model that appeared most often in our Google Images web search. The Google Scholar search yielded 585 hits per June 05, 2022, showing that a double latent mediation model is indeed commonly applied in the social sciences. On the other hand, we stress the point that our chosen model structure should not be regarded as *prototypical* of psychological research in the sense that it represents a core model of applied SEM enabling us to extrapolate our results to various other model structures. Hence, the actual reason that we chose this kind of model structure was to investigate how structural fit indices and their cutoff values perform under those circumstances wherein they encounter a commonly, hence, typically highly-parametrized model.

Structural Component.

Although confirmatory factor analytic models are the typical focus in SEM simulation studies, complex SEMs, possessing of structural components, are the norm in applied research. The structural component herein created as typical of those arising in applied research, was a double latent mediation model with six latent variables (3 exogenous, 3 endogenous) linked by eight pathways.[2] Standardized regression coefficients were selected with the guidance of the study by Peterson and Brown (2005), in which is presented the distribution of 1,504 standardized regression weights appearing in published articles in social and behavioral journals. Because misspecification relating to even a single path in the structural component of a candidate model may constitute a significant substantive violation of the theory instantiated in a true population model, our approach was to select large standardized regression coefficients.

Peterson and Brown (2005) reported that approximately 97% of published standardized regression coefficients lie within the interval [-.5,.5] units (M = .06, SD = .21). Accordingly, we selected an arbitrary set of eight standardized regression coefficients lying within the interval [.30,.57]. The lower and upper bound of this interval corresponds to the 87$^{\text{th}}$ and 99$^{\text{th}}$ percentile, respectively, of a random variable following a truncated normal distribution $X \sim N_{\text{trunc}}(.06, .21, -.82, .82)$. The rationale for selecting regression coefficients from this percentile interval was to define nontrivial (i.e., strong) misspecifications in terms of the difference between population regression weights and their counterparts of the misspecified models being incorrectly set to zero and to assess the power of fit index cutoff values under such nontrivial misspecifications.

Correlations among the three exogenous latent variables were set to $\rho_{\eta_1\eta_2} = .30$, $\rho_{\eta_1\eta_3} = .50$, and $\rho_{\eta_2\eta_3} = .40$, and, accordingly, were comparable to those reported in Hu and Bentler (1999).

Figure 1 depicts the structural component of the true population models employed in the simulation study.

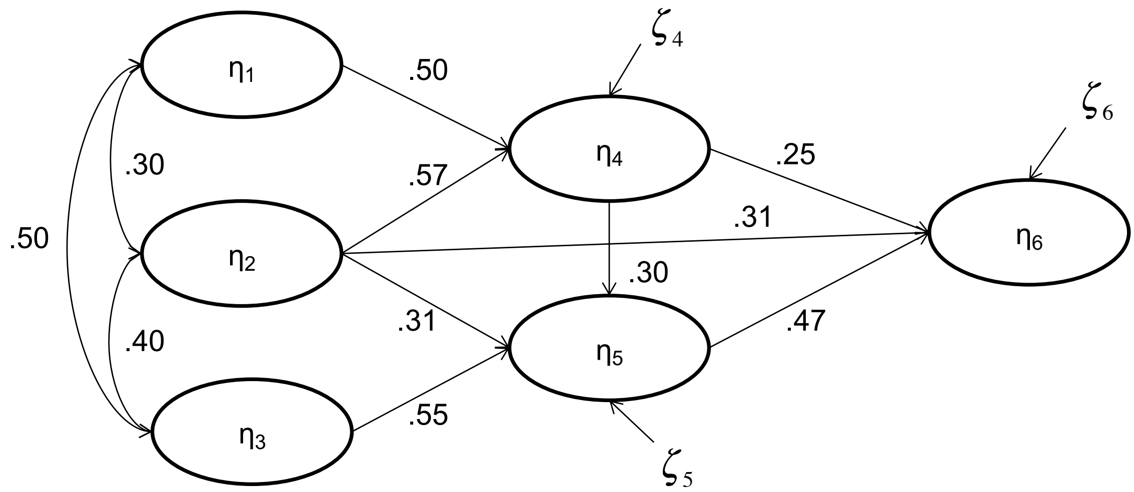
Measurement Component.

Misspecification in the measurement component of a candidate model, the magnitudes of the factor loadings of the true population model (cf. Browne et al., 2002; Hancock and Mueller, 2011; Heene et al., 2011; Schonemann, 1981), and the number of indicators per latent variable (Breivik & Olsson, 2001; Heene et al., 2011; Kenny & McCoach, 2003), are model features, incidental to structural misspecification, which have been shown to have an impact upon the performance of fit indices. Factor loading sizes were based on the results concerning typical factor loadings of psychological questionnaires Peterson (2000) and cognitive performance tests (Carroll, 1993, pp. 592–593). To allow for an assessment of their relative impacts on the performance, as detectors of structural misspecification, of the fit indices, herein, under investigation, variants of the true population model were created. In particular, the double mediation structural core was paired with each of the measurement components yielded by crossing the following factors:

[2] Though their treatment is both beyond the scope of, and irrelevant to the concerns of, the present research, we acknowledge the deeper scientific difficulties attendant to the employment of mediation models in social research (see, e.g., Grice et al., 2015; Kline, 2015; Tate, 2015).

Figure 1

Structural component of true population models.



- Number of indicators per latent variable: 5 or 10
- Magnitude of factor loadings: small (randomly sampled from the [.4,.6] interval) or large (randomly sampled from the [.6,.8] interval)
- Magnitude of measurement error correlations: zero (no correlated measurement errors), small (randomly sampled from the [-.1,.1] interval), medium (randomly sampled from the [-.2,.2] interval), or large (randomly sampled from the [-.3,.3] interval).

All told, then, data was simulated under each of sixteen true population models.

Candidate Models

Structural Component.

Candidate models were based on one of four distinct structural components, each representing a particular degree of structural misspecification. In particular, either zero [no structural misspecification], one, two, or three of the eight nonzero structural pathways of the true population models were, in each candidate model, set to zero. Figures 2 to 4 depict the three misspecified structural components.

Measurement Component.

Each of the four candidate structural components was paired with a measurement component wherein all measurement error correlations were restricted to be zero. Accordingly, measurement model misspecification was induced under all instances in which a candidate model happened to be paired with a true population model wherein were present non-zero measurement error correlations. More specifically, the latter three levels of the *magnitude of measurement error correlations* factor—i.e., small, medium, and large—can be taken as corresponding to three (ascending) degrees of measurement model misspecification.[3]

Notational Convention

Each of the sixty-four distinct pairings of candidate model with true population model is uniquely representable in terms of an S-M notational scheme (employed as a labeling device in the results section), wherein

[3]Measurement component misspecification was operationalized in terms of measurement error correlations for the reason that the latter represent a common phenomenon in the social sciences. Cole et al. (2007) found that 26.6% to 31.9% of articles published in five prominent APA journals involved the estimation of measurement error correlations.

Figure 2

Structural component of candidate models containing one misspecification

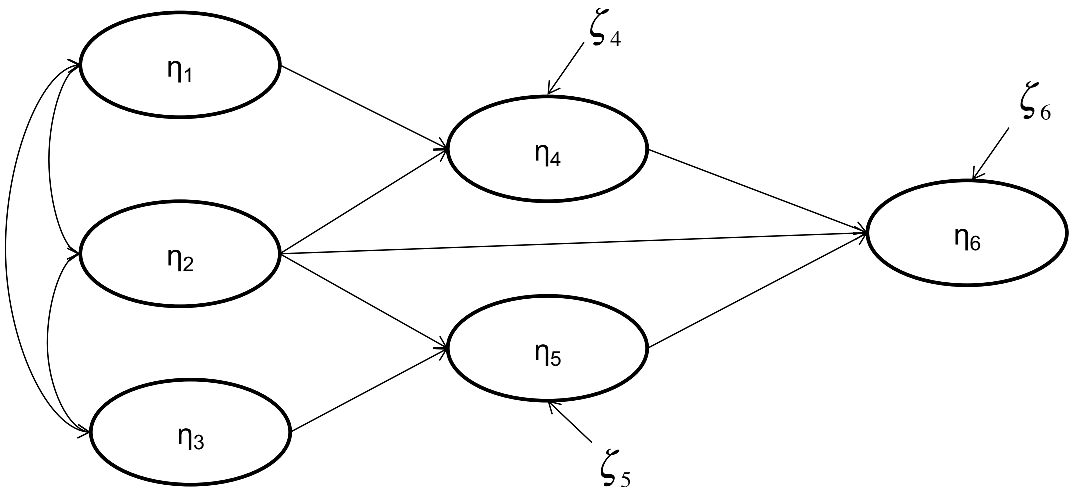


Figure 3

Structural component of candidate models containing two misspecifications

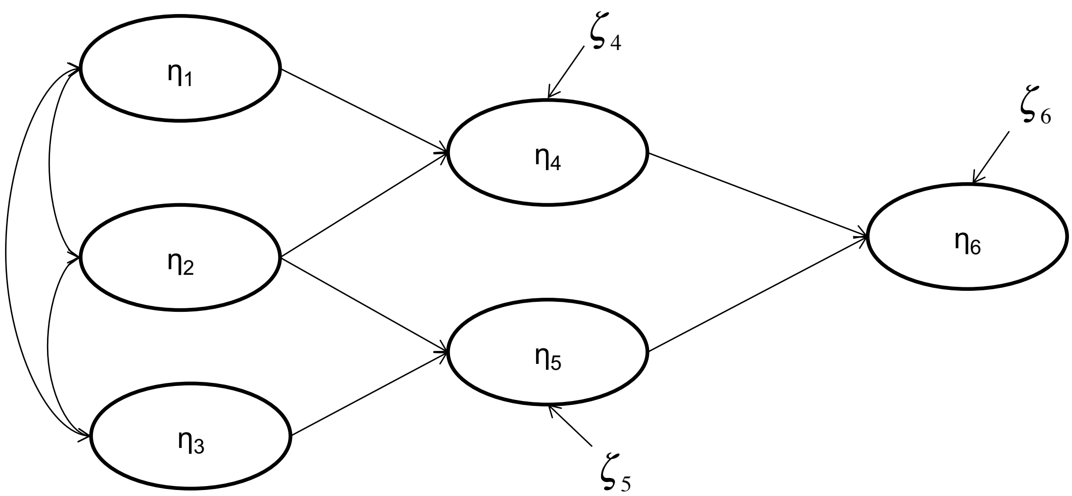
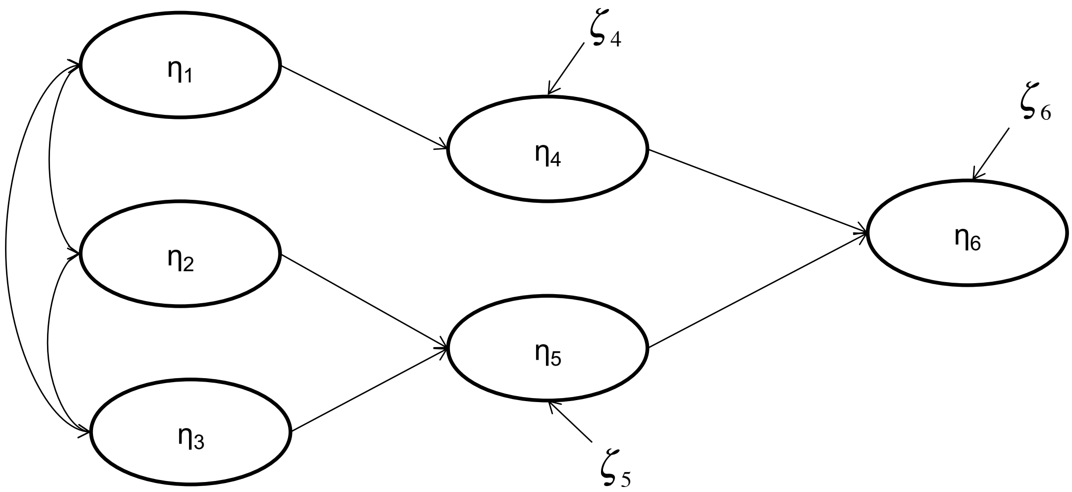


Figure 4

Structural component of candidate models containing three misspecification



S = *degree of structural component misspecification* (0 [no structural component misspecifications], 1 [1 omitted nonzero pathway], 2 [2 omitted nonzero pathways], 3 [3 omitted nonzero pathways]) and M = *degree of measurement component misspecification* (0 [no measurement component misspecification], 1 [omission of measurement error correlations, when, in the true population model, they are of small magnitude], 2 [omission of measurement error correlations, when, in the true population model, they are of medium magnitude], 3 [omission of measurement error correlations, when, in the true population model, they are of large magnitude]). Thus, the notation 0-1 indicates the pairing of a candidate model with a true population model, wherein there is no structural misspecification, but a mild degree of measurement component misspecification; 2-3, a pairing under which there is both structural misspecification (the candidate model omits 2 nonzero pathways) and the greatest degree of measurement component misspecification.

Cutoff values

For each of the composite fit indices (and its structural counterpart(s)), cutoff values were as suggested by Hu and Bentler (1999) or Hancock and Mueller (2011) and Williams and O'Boyle (2011), respectively. Accordingly, each of RMSEA (RMSEA-P, RMSEA_S) < .05, SRMR (SRMR_S) < .08, and CFI (CFI_S) > .95, returned the decision that the structural component of a candidate model was correctly specified. The nominal Type I error rate of the chi-square test of the hypothesis pair [$H_0 : \Sigma = \tilde{\Sigma}(\boldsymbol{\Theta}), H_1 : \Sigma$ any Gramian matrix] was set to .05.

General Details

Typically, Power is a function of sample size. In the service of elucidating the relation between sample size and the Power to detect structural misspecification yielded by each of the seven GFICV pairings, and the chi-square test, three different levels of sample size ($N = 150, 250, \text{and} 1000$) were selected. The two smaller sample size conditions, especially the one with $N = 250$, reflect rather typical sample size conditions in applied SEM (MacCallum & Austin, 2000; Shah & Goldstein, 2006; Westland, 2010). These typical sample sizes may be too small to yield sufficient Power, which is why we also chose the sample size condition of 1000 to ensure greater Power. Candidate models were fit by maximum likelihood, under condition of the multivariate normality of the indicators.[4] For each of

[4]Though normally distributed data can seldom be found in the social sciences (Micceri, 1989), in the service of maintaining comparability with other simulation studies devoted to the investigation of the performance of GFICV pairings, maximum likelihood estimation under multivariate normality was

the 4 [degree of structural component misspecification] $\times$ 4 [degree of measurement component misspecification] $\times$ 2 [number of indicators per latent variable] $\times$ 2 [magnitude of factor loadings] $\times$ 3 [sample size] = 192 conditions, 1000 simulated samples were generated. Equivalently and, perhaps, more transparently: a) 16 populations were defined under the crossing of the factors number of indicators per latent variable, magnitude of factor loadings, and magnitude of measurement error correlations; b) for each population, there was a fixed set of parameter values of each of the structural and measurement variety; c) the structural parameter values were constant over population; d) the measurement parameter values were unique to each population; e) under each population—i.e., on the basis of a single fixed set of values—and for each of the sample sizes $n = \{150, 250, 1000\}$, 1000 simulated data sets were generated. The simulation was programmed in R (R Core Team, 2015) using the package simsem, version .5-11 (Pornprasertmanit et al., 2015). The simulation code can be downloaded from https://osf.io/4by5m/.

In simulation designs wherein certain of the conditions correspond to the circumstance of factor loadings of a small magnitude and/or small sample size, improper solutions must be expected (Driel, 1978). Herein, simulated samples yielding of improper solutions were replaced by fresh simulated samples, until the number of proper solutions was equal to the prespecified 1000 (see Curran et al., 2002 for a similar approach). Under each of the 192 conditions, for each of the 7 GFICV pairings, and the chi-square test, we present—over the 1000 simulations—both the Type I error rate (for the case of no misspecification in the structural component) or Power (for the case of misspecification in the structural component), and the mean and standard deviation of the quantity (fit index or chi-square test statistic) in question. For purposes of exposition, a Type I rate of .05 [Power of .8] was viewed as constituting acceptable performance.

Results

With the aim of improving stringency and clarity, results are displayed separately for conditions wherein the candidate model does, and does not, feature misspecification in the measurement component. Specifically, conditions featuring a pairing of candidate and true model of type S-0 (i.e., in which there is present, no misspecification in the measurement component) are presented first. Though we discuss, in the body of the text, results relating to the composite fit indices, for only the chi-square test and GFICV pairings involving structural fit indices, do we present the corresponding tables. Tables for composite fit indices—which, along with the chi-square test, in each subsection, are addressed first—can be found in the Appendix. In the shaded main-panels of Table 3, for each GFICV pairing, and under each of the two conditions of no misspecification in the measurement component, and misspecification in the measurement component, we offer the reader synoptic recapitulations of the results discussed, herein.

Candidate models absent misspecification in the measurement component

Chi-square test: Type I error rate and Power.

Table 4 contains the results for the chi-square test paired with a nominal Type I error rate of .05.

With respect to Type I error rate, key findings were as follows: i) irrespective of condition, the true Type I error rate exceeded the nominal counterpart; and ii) the degree of inflation in Type I error rate was a decreasing [increasing] function of N [*number of indicators per latent variable*]. The explanation for the first of these relations– which, at various points in our discussion of the results, we will have occasion to invoke–, resides in the fact that, for small sample sizes, the distribution of the chi-square test statistic departs appreciably from its theoretical (null) sampling distribution[5] (see Bentler & Yuan, 1999, for a similar conclusion).

The second relation is consonant with the results of Moshagen (2012), who demonstrated that, under the state of H_0 true, the departure of the distribution of the chi-square test statistic from its theoretical counterpart is exacerbated by increments to the order of the input covariance matrix (see Yuan, 2005; Yuan et al., 2017 for approaches to the amelioration of the effect).

With respect to Power, key findings were as follows: i) as one would expect, holding constant all other factors, Power was an increasing function of *degree of structural misspecification*; ii) Power ranged over the interval [.216, 1], and reached acceptable levels (> .80) under 64% of the conditions; iii) Power was an increasing function of both *number of indicators per latent variable* (cf. Saris and Satorra, 1988, who make the same observation) and *magnitude of factor loadings* (see, also, Hancock and Mueller, 2011; Heene et al., 2011; Schonemann, 1981 for a theoretical explanation of the effect,

employed (e.g., Browne et al., 2002; Curran et al., 2002; Hu & Bentler, 1999; Marsh et al., 1988).

[5]In particular, whereas, in the case of a variate X distributed as central χ^2, $\mu_X = df$ and

$$\sigma_X = \sqrt{2 \cdot df} = \sqrt{2 \cdot \mu_X},$$

for the mean, standard deviation pairings displayed in Table 4, $\hat{\sigma}_X \ll \sqrt{2 \cdot \hat{\mu}_X}$.

see Heene et al., 2011; Schonemann, 1981); iv) although Power standardly increased with N, when *degree of structural misspecification* was either small or moderate, and under condition of five indicators per latent variable, Power yielded by $N = 250$ was *lower* than that yielded by $N = 150$. This counterintuitive effect can be explained, once again, with reference to the fact of the poor correspondence, at small N, between the empirical and theoretical sampling distributions of the chi-square test statistic; in particular, to the findings of Curran et al. (2002), who demonstrated that, under condition of small N, the mean of the chi-square test statistic exceeds that of the corresponding non-central χ^2 distribution. As is evident from the 3-0 column of Table 4, the effect is mitigated under increasing *degree of structural misspecification*; v) when *degree of structural misspecification* was either small or moderate, under condition of small *magnitude of factor loadings* and five *indicators per latent variable*, Power was uncharacteristically low.

RMSEA < .05: Type I error rate and Power.

For [RMSEA < .05], results are presented in Appendix 1. As can be seen, therein, under all conditions, the Type I error rate and Power were identically zero. This state of affairs is consonant with the {mean, standard deviation} pairings which accompany these rates, within which the mean is always less than .05, and the standard deviation, small. Based on these results, it can be concluded that, although the researcher employing this particular GFICV pairing will never risk falsely rejecting a structurally correct candidate model, they, on the other hand, will never correctly reject a false candidate model. Even in the extreme case wherein the candidate model is most severely misspecified (i.e., wherein there are three omitted pathways), the population factor loadings are of small magnitude, and there are five indicators per latent variable, the population value of RMSEA is only (approximately) .022[6]. This result once again highlights the aforementioned distinction between model misspecification and model misfit, in as much as that, it demonstrates that a seriously misspecified candidate model may yet yield a model implied covariance matrix that is close to the true population counterpart.

SRMR < .08: Type I error rate and Power.

For SRMR paired with the usual cutoff value of .08, values for the Type I error rate and Power are presented in Appendix 2. Under all conditions, the Type I error rate was equal to zero. Acceptable Power (> .80) –Power greater than zero, moreover– was achieved only under condition of highest level of *degree of structural misspecification* combined with large *magnitude of factor loadings*. When *degree of structural misspecification* was at its highest level, and, in consequence, Power was nonzero, Power was a decreasing [increasing] function of N [*number of indicators per latent variable, magnitude of factor loadings*]. The simulated population value of SRMR, under condition of greatest degree of misspecification, loadings of small magnitude, and five indicators per latent variable, was .031, a value considerably beneath the .08 cutoff level. The fact that, as is evident from Appendix 2, the mean of the sampling distribution of SRMR tends to *decrease* with increasing N (see, also, Marsh, Hau, and Wen, 2004), implies that the problem cannot be remediated by recourse to the usual tactic of increasing sample size.

CFI > .95: Type I error rate and Power.

Results for CFI paired with a cutoff value of .95 are presented in Appendix 3. With respect to Type I error rate, key findings were as follows: i) in contradistinction to the results for the GFICV pairings involving SRMR and RMSEA, the Type I error rate yielded by CFI varied widely –from 0 to .999– over conditions; ii) was a decreasing function of N; and iii) for $N < 1000$, was a decreasing [increasing] function of *magnitude of factor loadings* [*number of indicators per latent variable*].

With respect to Power, the key findings were: i) Power reached acceptable levels (> .80) under only 22% of conditions; ii) for $N < 1000$, and holding constant all other factors, Power was an increasing function of *degree of structural misspecification*; iii) Power was a decreasing function of N, with rates essentially zero under condition that $N = 1000$[7]; iv) Power was a decreasing [strongly increasing] function of *magnitude of factor loadings*[8] [*number of indicators per latent variable*].

[6]We used a finite population of 1,000,000 subjects to approximate the population RMSEA, SRMR, and CFI. The R syntax can be downloaded from https://osf.io/wurt6/.

[7]For this effect, the explanation is analogous to that introduced in the case of the chi-square test statistic; namely that, under condition of small N, the mean of the chi-square test statistic on which CFI is based, exceeds that of the corresponding (theoretical) non-central χ^2 distribution (see Curran et al., 2002). In fact, under small N and greatest degree of structural misspecification, even when the loadings are of small magnitude, and there is only five indicators per latent variable, the simulated population mean of CFI (.956) still exceeded the .95 cutoff value.

[8]The explanation for the effect residing in the fact that, all things being equal, the magnitude of the covariance between a pair of indicators is an increasing function of the magnitude of the factor loadings of each. Accordingly, even in the presence of structural misspecification, loadings of smaller magnitude will yield a relatively small separation between the model im-

RMSEA-P < .05: Type I error rate and Power.

Results for the RMSEA-P are displayed in Table 5.

The Type I error rate: i) ranged over the interval [0, .213] and, overall, was mildly inflated under 58% of conditions; and ii) was a decreasing function of N (at $N = 1000$, brought to levels beneath .05). Power: i) reached acceptable levels (> .80) under 90% of conditions; ii) was an increasing function of *degree of structural misspecification*; and iii) in cases in which *degree of structural misspecification* was either small or moderate, was strongly increasing in both *magnitude of factor loadings* and *number of indicators per latent variable*.

RMSEA$_S$ < .05: Type I error rate and Power.

Table 6 presents the results for the pairing of RMSEA$_S$ and a cutoff value of .05.

As can be seen, over all conditions, as the case may be, the Type I error rate was egregiously high and Power, essentially, unity. In light of the surprisingly large means and standard deviations appearing throughout Table 6, both findings call for an explanation in terms of a poor correspondence between the theoretical and empirical distribution of RMSEA$_S$; in particular, that both the mean and standard deviation of the empirical distributions exceed those of their theoretical counterparts (see Curran et al., 2002; Yuan, 2005; see Appendix 4 for a technical elucidation of the phenomenon). For example, the simulated population RMSEA$_S$, under condition of five indicators per latent variable, loadings of small magnitude, and one (three) misspecifications, turns out to be .278 (.459), which lies beneath each and every one of the means of Table 6. The standard deviation of the theoretical counterpart, under condition of misspecification, is given by

$$SD(\text{RMSEA}_S) = \frac{\text{Var}\left(\chi^2\right)}{\sqrt{[df(N-1)]^2}} = \frac{\sqrt{2(df + 2\lambda)}}{\sqrt{[df(N-1)]^2}}, \tag{7}$$

with λ denoting the noncentrality parameter of a noncentral chi-square distribution and degrees of freedom df.

Under condition of five indicators per latent variable, loadings of small magnitude, $N = 250$, and one (three) misspecifications, (7) yields a standard deviation equal to .011 (.021)[9], a value which is, indeed, considerably much smaller than that – .162 (.124)– which appears in Table 6. This inflation of mean and standard deviation can be shown to be exacerbated by the circumstance of the true (population) model possessing of factor loadings of small magnitude[10]; indicating, of course, that the distribution of RMSEA$_S$ is sensitive to model features, incidental to the structural component.

SRMR$_S$ < .08: Type I error rate and Power.

As is evident from Table 7, the results bearing on the GFICV pairing of SRMR$_S$ and cutoff value of .08 are relatively straightforward.

Under all conditions, the Type I error rate was essentially zero. Power, on the other hand, was near zero under all conditions, excepting that of highest *degree of misspecification*, under which, irrespective of the levels of other factors, was essentially unity.

CFI$_S$ > .95: Type I error rate and Power.

Table 8 presents results for the GFICV pairing of CFI$_S$ and a cutoff value of .95.

The Type I error rate: i) exceeded .05 under 58% of conditions, but was, in the main, brought to acceptable levels at $N = 1000$; ii) was a decreasing function of each of N and *number of indicators per latent variable*, and was strongly decreasing in *magnitude of factor loadings*.

Power: i) ranged between .775 and unity and, accordingly, under all conditions, reached an acceptable level; ii) was an increasing function of *degree of structural misspecification*, reaching unity when encountering structural misspecification of the highest degree; and iii) when *degree of structural misspecification* was either small or moderate, was a decreasing function of *magnitude of factor loadings*, and only weakly related to N and *number of indicators per latent variable*.

Candidate models containing misspecification in the measurement component

In what follows, we present results for candidate models containing misspecification in both the structural and measurement components; i.e., when confronted by pairings of types S-1, S-2, and S-3. As will be recalled, measurement misspecification was induced by pairing a candidate model absent of measurement error correlations, with a true population model in which were present measurement error correlations, the latter of a magnitude either small, medium, or large.

plied covariance matrices of each of a candidate model and null model (under which the manifest variables are uncorrelated) (see Heene et al., 2011).

[9] The associated syntax files can be found on https://osf.io/k3n4x/ and https://osf.io/8wa24/.

[10] An R syntax illustrating this phenomenon can be downloaded from https://osf.io/pmfu8/ and https://osf.io/hzxjs/

Chi-square test: Type I error rate and Power.

As is patent in Table 9, under condition of misspecification in the measurement component, the results for the chi-square test paired with a nominal Type I error rate of .05 are particularly simple, in that, for all conditions under which it was defined—i.e., 0-M, M = 1, 2, 3—, the true Type I error rate was equal to unity, and, for all conditions under which it was defined—S-M, S = 1, 2, 3—, power was equal to unity.

RMSEA < .05: Type I error rate and Power.

For RMSEA paired with the usual cutoff value of .05, results are presented in Appendix 5. As can be seen, therein, under condition of misspecification in the measurement component, the Type I error rate yielded by this pairing: i) ranged from .006 to unity, was egregiously high under 94.4% of conditions, and, in fact, was equal to unity under 83.3% of conditions; ii) was an increasing function of each of *degree of misspecification in the measurement component* and *number of indicators per latent variable*; and iii) was in the ballpark of reasonable only under condition of lowest *degree of misspecification in the measurement component*, *number of indicators per latent variable* equal to 5, and $N = 1000$. The Power yielded: i) exceeded the .8 threshold of acceptability under 88% of conditions, with the only problematically low values occurring under S-1 conditions (i.e., those wherein *misspecification in the measurement component* was of lowest degree); and ii) under condition that *number of indicators per latent variable* was equal to 5, was a decreasing function of N.

SRMR < .08: Type I error rate and Power.

For the pair [SRMR, .08], values for the Type I error rate and Power are presented in Appendix 6. Evidently, under condition of misspecification in the measurement component, the Type I error rate yielded by this pairing: i) was either zero (72% of conditions) or essentially unity (28%); ii) was a strongly increasing function of each of *degree of misspecification in the measurement component* and *number of indicators per latent variable*; iii) was a decreasing function of *magnitude of factor loadings* and N; and iv) approached unity under condition of small factor loadings and at least moderate *degree of misspecification in the measurement component*. Power yielded by this pairing: i) reached the .8 threshold of acceptability under 47% of conditions, of these, the vast majority of the 3-M type (3-M conditions wherein power was *inadequate*, those for which *number of indicators per latent variable* was equal to 5); ii) in general, was an increasing function of both *degree of misspecification in the measurement component* and *misspecification in the structural component*, with the former appearing to be the more decisive determiner; and iii) was largely insensitive to N.

CFI > .95: Type I error rate and Power.

The situation respecting the pairing of CFI and a cutoff value of .95—presented in Appendix 7—is, in its simplicity, similar to that of the pairing of the chi-square test. Specifically, over all conditions under which it was defined, the Type I error rate was near, or equal to, unity; and under all conditions under which it was defined, Power was equal to unity.

RMSEA-P < .05: Type I error rate and Power.

In light of their complexity, the results for [RMSEA-P, .05] are depicted in both Table 10 and Figure 5.

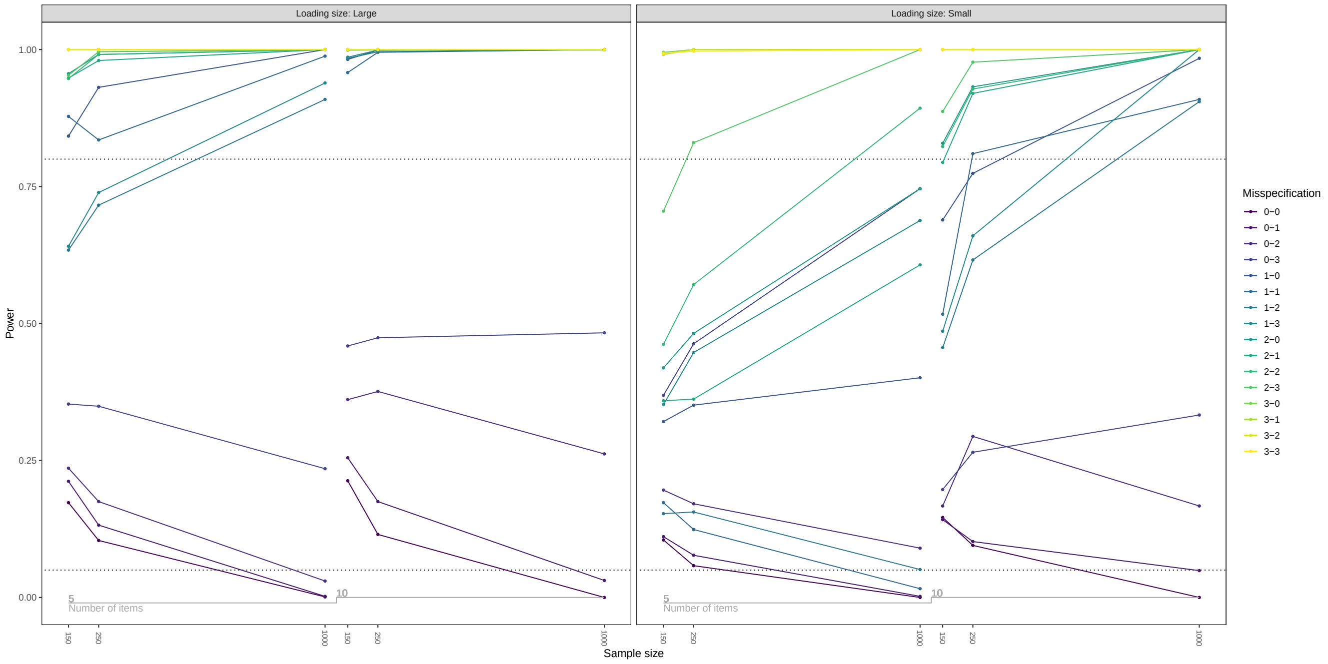
With respect to Type I error rate, key findings were as follows: i) it ranged within the interval [.002, .746] and, under 86% of conditions, exceeded the .05 level; ii) it was an increasing function of *degree of misspecification in the measurement component*, this fact indicative of the sensitivity of [RMSEA-P, .05] to model features, incidental to structural misspecification; iii) it manifested complex, conditional dependencies on N, *magnitude of factor loadings*, and *number of indicators per latent variable*. With respect to Power, key findings were as follows: i) under 79% of conditions, it reached acceptable (> .80) levels; ii) as with the Type I error rate, it was, under many combinations of factor levels, an increasing function of *degree of misspecification in the measurement component*, implying, once again, the pairing's sensitivity to model features incidental to structural misspecification; iii) when *degree of structural misspecification* was at its highest level, an acceptable level of Power was achieved, irrespective of *degree of misspecification in the measurement component* (i.e., for all of pairings 3-1 to 3-3); iv) irrespective of *degree of misspecification in the measurement component*, Power did not reach acceptable levels when *magnitude of factor loadings* was small; v) when $N \leq 250$ and *degree of structural misspecification* was either small or moderate (i.e., for pairings 1-1 to 2-3), Power was a complex function of *magnitude of factor loadings* and *number of indicators per latent variable*; vi) under condition of five indicators per latent variable and factor loadings of small magnitude, the Power yielded for pairings 1-1 to 2-2 was often egregiously low.

Random Forest Analysis.

With the aim of coming to a finer-grained understanding of the dependency of Power on S-M pairing,

Figure 5

Rejection Rates for Misspecified Measurement and Structural Models for the RMSEA-P



magnitude of factor loading, *number of indicators per latent variable*, and N, a random forest analysis was carried out using the "caret" package (Wing et al., 2018). Notice that this analysis was carried out only for the *RMSEA-P* and further below for the *CFIs* because Power rates of the other GFICVs varied so little that such an analysis would have been uninformative. Because the simulation results being analyzed here were already based on resampled data, cross-validations were not employed. Scaled conditional variable importance (VIMP) measures—ranging between 0 and 100—were calculated using the permutation accuracy importance mode (e.g., Strobl et al., 2008)[11]. The random forest regression explained 64.5% of the variance in Power. Scaled VIMP measures are reported in Table 1.

As one would hope, *degree of structural misspecification* had the greatest impact upon Power yielded by the GFICV pairing of RMSEA-P and .05. *Magnitude of loadings* turned out to be a stronger predictor than *number of indicators per latent variable*, and N had negligible impact.

$RMSEA_S$ < .05: Type I error rate and Power.

Results for the GFICV pairing of $RMSEA_S$ and a cutoff value of .05 are presented in Table 11.

Key findings for the Type I error rate are as follows: i) under all conditions, it vastly exceeded the .05 level, approaching unity under 78% of conditions; ii) it was sensitive to the degree of misspecification in the measurement component. With respect to Power, in light of the Type I error rates observed, it was no surprise that, under all conditions, it exceeded the .8 threshold, and, in fact, diverged, only marginally, from unity.

$SRMR_S$ < .08: Type I error rate and Power.

In Table 12 are presented results for [$SRMR_S$, .08].

In regards the Type I error rate: i) As was the case under condition of no misspecification in the measurement component, it was less than .05—and, in fact, near zero—under virtually all conditions (the exception being those cases in which $N = 150$, *magnitude of factor loadings* was small, and there were five *indicators per latent variable*, wherein it was still modest); and ii) under condition wherein it was nonzero (i.e., those conditions in which $N = 150$, *magnitude of factor loadings* was small, and there were five *indicators per latent variable*), it was sensitive to the *degree of misspecification in the measurement component*. Respecting Power: i) it reached acceptable (> .80) levels only when de-

[11] It should be stressed that a VIMP for a particular predictor captures all effects—direct impact and interactions of all orders—involving the predictor. Accordingly, though a VIMP captures the overall predictive efficacy of a given predictor, it implies nothing about the partitioning of this efficacy into direct and interaction effects.

gree of structural misspecification was at its highest level (wherein it was uniformly near unity).

CFI_S < .95: Type I error rate and Power.

Results for GFICV pairing [CFI_S, .95] are displayed in Table 13.

Respecting the Type I error rate: i) it ranged over the entire interval [0, 1], exceeding the .05 rate under 81% of conditions, and, in many instances, doing so egregiously; ii) it was an increasing function of *degree of misspecification in the measurement component*; iii) it was strongly decreasing in *magnitude of factor loadings*; iv) excepting the condition wherein *degree of misspecification in the measurement component* was at its highest level, was, in the main, a decreasing function of N; and v) was satisfactory at $N = 1000$, save for the condition under which the *magnitude of factor loadings* was small. As for Power: i) under 87% of conditions, it reached acceptable (> .80) levels, in many instances, approaching unity; ii) it manifested complex dependencies on *degree of misspecification in measurement component*; iii) it was, under a number of factor combinations, a decreasing function of N; and iv) when *degree of structural misspecification* was either small or moderate, was a decreasing [increasing] function of *magnitude of factor loadings* [*number of indicators per latent variable*].

Random Forest Analysis.

A random forest analysis undertaken to assess the dependency of Power on the condition-defining factors, explained 88.95% of the variance. Scaled VIMP measures are reported in Table 2.

Table 1

Scaled Variable Importance Measures of the Random Forest Analysis for the CFIs > .95

	Variable Importance
Misspecification condition	100.00
Loading size	65.84
Number of items per latent variable	9.06
Sample size	0.00

As with RMSEA-P, the *degree of structural misspecification* was the best predictor, followed by *magnitude of factor loadings*, and (with a considerably lower impact) *number of indicators per latent variable*. Sample size, once again, had a negligible impact.

Table 2

Scaled Variable Importance Measures of the Random Forest Analysis for the RMSEA-P < .05

	Variable Importance
Misspecification condition	100.00
Loading size	45.42
Number of items per latent variable	30.41
Sample size	0.00

Discussion

In this study, we investigated the performance of eight GFICV pairings in detecting structural misspecification in candidate models. Four pairings featured a classical composite fit index designed with the aim that it picks up on any type of misspecification, be it located in the structural component or measurement component, and four, a structural fit index invented expressly with the aim that its sensitivity be focused exclusively on structural misspecification.

More particularly, for each pairing, the Type I error rate (under condition of no misspecification in the structural component of the candidate model) and Power (under condition of each of three degrees of structural misspecification) was empirically estimated under a range of conditions formed by crossing: a) three levels of sample size; and b) the two, two, and three levels of three model features, incidental to structural misspecification, namely, *magnitude of population factor loadings*, *number of indicators per latent variable*, and *degree of misspecification in the measurement component* of the candidate model.

Being as they are binary inferential decision-making instruments, the performance of each GFICV pairing can be captured with reference to the distributions of Type I error rates and values of Power, in conjunction with the usual optimality criteria. Generally speaking, the performance of a binary decision-making instrument is deemed acceptable if: a) there is a means by which it can be set to an acceptably low value, the Type I error rate; b) [conditional on satisfaction of a)] the Power it yields is an increasing function of departure from H_0 (in this case, the *degree of structural misspecification* inherent to a candidate model), and can be made acceptably high by recourse to selection of sample size.

With respect to the GFICV pairings which, herein, are the focus, a third, but related, desideratum is that its performance should not be sensitive to features of candidate models, incidental to structural misspecification. To this requirement, we add two qualifications: a) in light of the fact that the sole purpose for having added them to the not-insubstantial complement of fit indices, was to equip the researcher with a means of shedding light on the quality of the *structural* claims made by models, it's especially important that those pairings featuring structural indices be insensitive to misspecifications in the measurement component; b) it is more problematic for a pairing to be sensitive to factors, incidental to structural misspecification, and, additionally, not under control of the researcher, than incidental, and under control of the researcher. Thus, for example, we would view to be a greater strike against a pairing, its sensitivity to one or both of the *magnitude of factor loadings* or *misspecification in the measurement component*, than its sensitivity to the *number of indicators per latent variable* (the researcher able to antecedently select a value of the factor, number of indicators per latent variable, which will ensure an optimal performance from the pairing).

All told, these considerations suggest that, within the present context, the issue of performance can be seen as reducing to the issue of whether an acceptable Type I error rate/Power balance—conventionally, the former in the vicinity of .05 and the latter in excess of, say, .80—can be achieved through the antecedent selection of levels of factors which are under control of the researcher. Given that Power is coded to a specific degree of structural misspecification, under those circumstances wherein an acceptable balance can be achieved, it will be essential to note the corresponding *degree of structural misspecification* for which this holds. In what remains, we will endeavour to provide a global assessment of the performance of each of the individual GFICV pairings. We address those pairings featuring a structural fit index, and composite fit index, respectively, in turn, at the end of each section, offer a general recapitulation, highlighting general themes inherent to the performances of the pairings within each set. Finally, we address the issue of whether the former perform better than the latter, and, accordingly, succeed in making a useful contribution to the fit assessment resources available to the scientist. For easy reference, for each pairing, the global assessments we formulate are echoed in the unshaded sections of Table 3.

GFICV Pairings Featuring a Structural Fit Index

RMSEA-P, .05.

This pairing was sensitive to misspecification in the measurement component and, under such misspecification, manifested complex and conditional dependencies on both the controllable factors, sample size and number of indicators per latent variable, and noncontrollable factor, magnitude of factor loadings. Though, in broad stroke, it behaves in a manner consonant with its intended purpose, it suffers from the defect of rejecting many structurally correct models, merely because they happen to contain misspecification in the measurement component.

RMSEA$_S$, .05.

Because it yields a Type I error rate and Power essentially equal to unity under all conditions for which each is defined, it is not possible to achieve an acceptable Type I error rate/Power balance with this pairing. Employment of [RMSEA$_S$, .05] is logically equivalent to rejecting the structural component of each and every candidate model. That is to say, it will reject both every structurally correct candidate model, and every structurally misspecified model.

SRMR$_S$, .08.

With this pairing, the aim of creating a tool of detection that is largely insensitive to misspecification in the measurement component is realized. The single draw-back is that it will signal misspecification in the structural component, only when this misspecification is relatively pronounced. All told, a well-behaved detector of structural misspecification, as long as the latter is present at a relatively high degree.

CFI$_S$, .95.

This pairing manifested a sensitivity to both *degree of misspecification in the measurement component* and incidental and non-controllable factor—*magnitude of factor loadings*. Generally speaking, it matches the high Type I error rate of [RMSEA-P, .05], while delivering slightly better Power over all degrees of misspecification in the structural component.

General recapitulation.

All told, none of the GFICV pairings was capable of delivering an entirely satisfactory Type I error rate/Power balance, [RMSEA$_S$, .05] failing entirely in this regard. Of the remaining pairings: a) [RMSEA-P, .05] and [CFI$_S$, .95] suffered from a severely inflated Type I error rate; b) despite the fact that they were designed with the aim that their sensitivity be focused on structural features of candidate models, all pairings—and especially, [RMSEA-P, .05] and [CFI$_S$, .95]—were, in fact, sensitive to model features, incidental to structural misspecification; c) [RMSEA-P, .05] and [CFI$_S$, .95] were sensitive to misspecification in the measurement component; and d) as would be expected from an instrument which delivers a Type I error rate close to zero, [SRMR_S, .08] was only sensitive to structural misspecification when it occurred at a relatively high degree.

GFICV Pairings Featuring a Composite Fit Index

Chi-square, .05.

This pairing was sensitive to the incidental, but controllable, factor, *number of indicators per latent variable*. Although, under condition of no misspecification in the measurement component, it has decent power (and an inflated Type I error rate), its sensitivity is tuned in favor of the detection of measurement component misspecification, in that it will falsely reject any structurally correct model that happens to feature misspecification in the measurement component.

RMSEA_S, .05.

The sensitivity of this pair is tuned exclusively towards the detection of misspecification in the measurement component. Only when a model contains misspecification in the measurement component is Power nonzero (and, in those cases, is always near unity). On the other hand, it rejects virtually all structurally correct models that happen to feature misspecification in the measurement component.

SRMR_S, .08.

The sensitivity of this pairing is somewhat less highly tuned to misspecification in the measurement component than in the case of other composite indices. It is, however, possessing of a complex, highly conditional, relationship to the degree of misspecification in the structural component, and the Power it delivers is dependent upon the interaction of the incidental, non-controllable, factors, *magnitude of factor loadings* and presence/absence of misspecification in the measurement component. Even in the presence of misspecification in the measurement component, a reasonable Type I error rate/Power balance can *almost* be assured, through appropriate selection of factor levels—specifically, the combination of $N = 1000$ and at least 10 indicators per latent variable—for high degree of misspecification in the structural component. We should note that Maydeu-Olivares (2017) showed that the SRMR is a positively biased estimator and developed an asymptotically unbiased estimator of the SRMR. Our rather pessimistic overall results concerning Type I and Power rates of the biased, yet still widely used SRMR could be improved in future simulation studies by using its asymptotically unbiased counterpart.

CFI_S, .95.

The sensitivity of this pairing is tuned almost exclusively towards the detection of misspecification in the measurement component. It will reliably reject only if it is present, misspecification in the measurement component and, on the other hand, will reject a large proportion of structurally correct models that happen to feature misspecification in the measurement component.

General recapitulation.

All told, save for [SRMR, .08], the behavior of which, as a detector of structural misspecification, was disadvantageously complex, the sensitivity of these pairings was tuned emphatically towards detection of misspecification in the measurement component. In particular: a) [RMSEA, .05], on the one hand, under condition of no misspecification in the measurement component, was wholly absent sensitivity to structural misspecification, and, on the other, under condition of misspecification in the measurement component, rejected the vast majority of structurally correct models; b) though possessing of non-zero sensitivity to structural misspecification, irrespective of the presence or absence of misspecification in the measurement component, the chi-square test and [CFI_S, .95] can be expected to reject a very high proportion of structurally correct models.

Are the Structural Indices Superior to the Composite Indices?

As we have seen, composite fit indices are not well suited to the task of detecting structural misspecification, either because they are absent sensitivity to structural misspecification—[RMSEA, .05]—or because they will reject structurally correct models that happen to feature misspecification in the measurement component; [Chi-square, .05] and [CFI_S, .95][12]. Though it is closer to what is needed, the overall complexity of the behaviour of [SRMR_S, .08] renders it, yet, an undesirable option, when the aim is the detection of structural misspecification.

Consequently, the question is not really whether the structural fit indices are *superior* to the composite fit indices, but whether any of the new structural indices perform well enough as detectors of structural misspecification, to warrant their employment in applied research

[12] If correct, McDonald and Ho (2002) observation that the measurement components of the models fit by social scientists, tend to be better fitting than their structural counterparts, counts, here, as an exacerbating factor with respect to the expected performances of these composite fit index-based pairings.

contexts. Unfortunately, the answer, here, also, is somewhat equivocal. The pairing [RMSEA_S, .05] is of no use; despite its billing as a structural fit index, each of [RMSEA_S, .05] and [CFI_S, .95] retains a band sensitivity still too broad, each sensitive to misspecification in the measurement component; and although behaving in a manner consonant with its intended employment, [SRMR_S, .08] will pick up on structural misspecification, only when it occurs at a relatively high degree.

Limitations

To protect against the making of overgeneralizations based on the results of our study, we must keep in mind that our results are tied to a specific model architecture and to specific numerical sets of model parameters. It is thus likely that rejection rates are different for different models, and different numerical values of model parameters will yield distinct results. Particularly, our choice of standardized regression coefficients from the distribution of coefficients coming from articles that used multiple linear regression models may also substantially deviate from the actual distribution of such coefficients from structural equation models. Further simulation studies could take on the herculean task to extract and use latent regression coefficient values from the literature for their simulations to yield simulation designs that are even closer to reality in that regard than ours. Furthermore, we did not investigate the impact of deviations of the data from a multivariate normal distribution. As most data in psychology are categorical, results might differ when a different estimator is used. It should nevertheless be said that applying the commonly suggested fit index cutoff values to categorical data tend to not discover data-model misfit (Xia & Yang, 2018). Hence, the results presented here are fairly optimistic.

Table 3

Summary of Performance of each of Eight GFICV Pairings

<table>
<tr><th>GFICV Pairing</th><th>Type I Error Rate</th><th>Power</th></tr>
<tr><td>Chi-square, .05</td><td>No misspec. in meas. comp.
- Inflated [.078,.999], especially under large [incidental, controllable] number of indicators per latent variable
- Controllable by increasing N
Misspec. in meas. comp.
- Uniformly unity</td><td>No misspec. in meas. comp.
- Ranged over [.216,1], and acceptable (> .80) under 64% of conditions
- Though an increasing function of degree of structural misspecification, not so of N
- Increasing function of both [incidental, controllable] number of indicators per latent variable and [incidental, non-controllable] magnitude of factor loadings
Misspec. in meas. comp.
- Uniformly unity</td></tr>
<tr><td colspan="3">Global Assessment:
No misspec. in meas. comp.: An acceptable Type I error rate/Power balance can be achieved for all degrees of structural misspecification, by ensuring that both N and number of indicators per latent variable are large.
Misspec. in meas. comp.: Model will always be rejected, even if containing no misspecification in structural component.
Overall: Although, under condition of no misspecification in the measurement component, has decent power (and an inflated Type I error rate), its sensitivity is tuned in favor of the detection of measurement component misspecification, in that it will falsely reject any structurally correct model that happens to feature misspecification in the measurement component.</td></tr>
<tr><td>RMSEA, .05</td><td>No misspec. in meas. comp.
- Uniformly near zero
Misspec. in meas. comp.
- Ranged over [.006,1]; equal to unity under 83.3% of conditions</td><td>No misspec. in meas. comp.
- Uniformly near zero
Misspec. in meas. comp.
- Acceptable (> .80) under 86% of conditions</td></tr>
</table>

Continued on next page

Table 3 – *Continued from previous page*

GFICV Pairing	Type I Error Rate	Power
Global Assessment: **No misspec. in meas. comp.**: It will neither falsely reject a structurally correct candidate model, nor reject a structurally misspecified model. **Misspec. in meas. comp.**: Model will most always be rejected, even if containing no misspecification in structural component. **Overall**: Sensitivity tuned exclusively towards the detection of misspecification in the measurement component. Only when a model contains misspecification in the measurement component is Power nonzero (and, then, is always near unity). On the other hand, rejects virtually all structurally correct models that happens to feature misspecification in the measurement component.		
SRMR, .08	**No misspec. in meas. comp.** - Uniformly near zero	**No misspec. in meas. comp.** - Acceptable (> .80) only when degree of structural misspecification at highest level and large [incidental, non-controllable] magnitude of factor loadings (near zero, otherwise) - When degree of structural misspecification at highest level: decreasing function of N; increasing function of both [incidental, controllable] # of indicators per latent variable and [incidental, non-controllable] magnitude of factor loadings
	Misspec. in meas. comp. - Large [approaching unity] under 31% [28%] of conditions - Increasing function of both degree of misspecification in measurement component and [incidental, controllable] # of indicators per latent variable - Decreasing function of both N and [incidental, non-controllable] magnitude of factor loadings	**Misspec. in meas. comp.** - Acceptable (> .80) under 47% of conditions (of these, the vast majority of 3-M type and S-2 and S-3 types in conjunction with small [incidental, non-controllable] magnitude of factor loadings) - In general, an increasing function of each degree of misspecification in measurement component and degree of misspecification in structural component; a decreasing function of [incidental, non-controllable] magnitude of factor loadings - Largely insensitive to N

Continued on next page

Table 3 – *Continued from previous page*

GFICV Pairing	Type I Error Rate	Power
Global Assessment:		
No misspec. in meas. comp.: Acceptable Type I error rate/Power balance achievable only for highest degree of structural misspecification and under condition that [incidental, non-controllable] magnitude of factor loadings happens to assume a large value.		
Misspec. in meas. comp.: Unfavorable Type I error rate/Power balance.		
Overall: Sensitivity somewhat less highly tuned to misspecification in the measurement component than in the case of other composite indices. However, possessing of a complex, highly conditional relationship to the degree of misspecification in structural component; the Power it delivers, dependent upon the interaction of the [incidental, non-controllable] factors, magnitude of factor loadings and presence/absence of misspecification in the measurement component. Even in the presence of misspecification in the measurement component, a reasonable Type I error rate/Power balance can almost be assured, through appropriate selection of factor levels, for high degree of misspecification in the structural component.		
CFI, .95	**No misspec. in meas. comp.** - Varied widely [0, .999] over conditions - For N < 1000, a decreasing [increasing] function of [incidental, non-controllable] *magnitude of factor loadings* [[incidental,controllable] *number of indicators per latent variable*] - Controllable by increasing *N*	**No misspec. in meas. comp.** - Acceptable (> .80) only when degree of structural misspecification at highest level and large [incidental, non-controllable] magnitude of factor loadings (near zero, otherwise) - When degree of structural misspecification at highest level: decreasing function of *N*; increasing function of both [incidental, controllable] # of indicators per latent variable and [incidental, non-controllable] magnitude of factor loadings
	Misspec. in meas. comp. - Under all conditions unity or near unity	**Misspec. in meas. comp.** - Under all conditions, equal to unity
Global Assessment:		
No misspec. in meas. comp.: The large value of *N* necessary to bring Type I error under control, *reduces* Power- for all degrees of structural misspecification- to unacceptably low levels.		
Misspec. in meas. comp.: Model will most always be rejected, even if containing no misspecification in structural component.		
Overall: Sensitivity is tuned almost exclusively towards the detection of misspecification in the measurement component. It will reliably reject only if is present, misspecification in the measurement component and, on the other hand, will reject the vast majority of structurally correct models that happen to feature misspecification in the measurement component.		

Continued on next page

Table 3 – *Continued from previous page*

GFICV Pairing	Type I Error Rate	Power
RMSEA-P, .05	**No misspec. in meas. comp.** - Mildly inflated under 58% of conditions, ranging over [0,.213] - Decreasing function of N (at N=1000, brought to levels beneath .05)	**No misspec. in meas. comp.** - Acceptable (> .80) under 90% of conditions - Increasing function of *degree of structural misspecification* - Under condition of small or moderate *degree of structural misspecification*, strongly increasing in each of [incidental, non-controllable] magnitude of factor loadings and [incidental, controllable] *number of indicators per latent variable*.
	Misspec. in meas. comp. - Ranging over [.002,.746], exceeded .05 level under 86% of conditions - Increasing function of [incidental, non-controllable] *degree of misspecification in measurement component* - Complex conditional dependency on N, [incidental, non-controllable] *magnitude of factor loadings*, and [incidental, controllable] *number of indicators per latent variable*.	**Misspec. in meas. comp.** - Acceptable (> .80) under 79% of conditions - Under many conditions, an increasing function of [incidental, non-controllable] *degree of misspecification in measurement component* - When *degree of structural misspecification* at highest level, acceptable power achieved irrespective of *degree of misspecification in measurement component*
Global Assessment: **No misspec. in meas. comp.**: Increasing function of degree of structural misspecification and, in the main, delivering of a not unreasonable Type I error rate and Power. **Misspec. in meas. comp.**: Inflated Type I error rate and sensitive to degree of misspecification in measurement component. **Overall**: Though, in broad stroke, it behaves in a manner consonant with its intended purpose, it suffers from the defect of being prone to rejecting structurally correct models that happen to feature misspecification in the measurement component.		
RMSEA$_S$, .05	**No misspec. in meas. comp.** - Under 100% of conditions, egregiously high, frequently approaching unity	**No misspec. in meas. comp.** - Under 100% of conditions, approaches unity
	Misspec. in meas. comp. - Egregiously high	**Misspec. in meas. comp.** - Under 100% of conditions, approaches unity

Continued on next page

Table 3 – *Continued from previous page*

<table>
<tr><th>GFICV Pairing</th><th>Type I Error Rate</th><th>Power</th></tr>
<tr><td colspan="3">Global Assessment:
No misspec. in meas. comp.: Model will always be rejected, even if containing no misspecification in structural component.
Misspec. in meas. comp.: Model will almost always be rejected, even if containing no misspecification in structural component.
Overall: Employment of this GFICV pairing is logically equivalent to rejecting the structural component of each and every candidate model. That is to say, it will reject both every structurally correct candidate model, and every structurally misspecified model.</td></tr>
<tr><td>SRMR_S, .08</td><td>No misspec. in meas. comp.
- Near zero under all conditions
Misspec. in meas. comp.
- For all but one condition, beneath .05</td><td>No misspec. in meas. comp.
- Acceptable (> .80) only when degree of structural misspecification at highest level (near-zero, otherwise), upon which, irrespective of condition, it was essentially unity
Misspec. in meas. comp.
- Acceptable (> .80) only when degree of structural misspecification at highest level, upon which, irrespective of condition, it was essentially unity</td></tr>
<tr><td colspan="3">Global Assessment:
No misspec. in meas. comp.: As long as there is a high degree of structural misspecification to be detected, delivers a very good Type I error/Power balance.
Misspec. in meas. comp.: As long as there is a high degree of structural misspecification to be detected, delivers a very good Type I error/Power balance.
Overall: A well-behaved detector of structural misspecification as long as the latter is present at a relatively high degree.</td></tr>
<tr><td>CFI_S, .95</td><td>No misspec. in meas. comp.
- Ranging over [0,.691], exceeded .05 under 58% of conditions
- Decreasing function of each of N and [incidental, non-controllable] magnitude of factor loadings</td><td>No misspec. in meas. comp.
- Under all conditions, approaches unity</td></tr>
</table>

Continued on next page

Table 3 – *Continued from previous page*

GFICV Pairing	Type I Error Rate	Power
	Misspec. in meas. comp. - Exceeded .05 under 81% of conditions (egregiously high under 42% of these) - Decreasing function of N and [incidental, non-controllable] magnitude of factor loadings - Entirely satisfactory at N = 1000 (save for condition under which magnitude of factor loadings small and number of indicators per latent variable equal to 5)	**Misspec. in meas. comp.** - Acceptable (> .80) under 87% of conditions - Increasing function of degree of structural misspecification

Global Assessment:

No misspec. in meas. comp.: Sensitive to [incidental, non-controllable] magnitude of factor loadings, but, in the main, delivers a not unreasonable Type I error rate/Power balance.

Misspec. in meas. comp.: High Type I error rate, good Power.

Overall: Matching its relatively poor Type I error rate, and delivering, perhaps, slightly better power, delivers a detection performance very similar to that of [RMSEA-P, .05].

Table 4

Chi-Square Test with alpha=.05: Type I error rate and Power (Means; Standard Deviations) under condition of no misspecification in measurement component

			Type of Misspecification in Structural and Measurement Components			
N	# indic.	Mag. factor loadings	0-0	1-0	2-0	3-0
			Type I error rate		**Power**	
150	5	Small	.331 (430.314; 31.093)[a]	.378 (434.388; 3.789)[b]	.400 (437.006; 3.613)[c]	.704 (462.255; 32.830)[d]
	5	Large	.355 (431.129; 31.278)[a]	.473 (442.033; 31.761)[b]	.576 (45.17; 32.132)[c]	.982 (516.198; 34.649)[d]
	10	Small	.999 (2027.386; 71.38)[e]	.999 (2034.388; 71.409)[f]	.999 (2091.314; 72.625)[h]	1 (2091.314; 72.625)[h]
	10	Large	.998 (2026.943; 71.077)[e]	1 (2046.817; 72.087)[f]	1 (2061.653; 73.034)[g]	1 (215.667; 73.034)[h]
250	5	Small	.194 (414.896; 3.023)[a]	.29 (42.563; 29.932)[b]	.267 (424.449; 3.285)[c]	.778 (469.218; 33.007)[d]
	5	Large	.194 (415.402; 29.825)[a]	.377 (432.539; 31.044)[b]	.509 (445.615; 32.041)[d]	1 (555.860; 37.283)[d]
	10	Small	.905 (1875.671; 63.029)[e]	.936 (1887.79; 63.045)[f]	.947 (1895.884; 63.388)[g]	.998 (1981.581; 65.961)[h]
	10	Large	.903 (1876.805; 63.19)[e]	.963 (1909.898; 64.033)[f]	.987 (1933.931; 65.104)[g]	1 (2076.913; 68.168)[h]
1000	5	Small	.078 (400.784; 28.508)[a]	.216 (419.164; 29.646)[b]	.365 (433.179; 3.661)[d]	1 (61.227; 41.457)[d]
	5	Large	.083 (401.06; 28.437)[a]	.771 (468.132; 32.912)[b]	.986 (519.477; 36.942)[c]	1 (953.811; 55.151)[d]
	10	Small	.173 (1740.999; 58.061)[e]	.415 (1784.095; 6.224)[f]	.594 (1814.526; 6.698)[g]	1 (2153.156; 71.288)[h]
	10	Large	.175 (1741.126; 58.648)[e]	.889 (1869.571; 62.825)[f]	.995 (1961.893; 66.786)[h]	1 (2547.044; 79.068)[h]

Note. [a] *df= 394.* [b] *df= 395.* [c] *df= 396.* [d] *df= 397.* [e] *df= 1700.* [f] *df= 1701.* [g] *df= 1702.,* [h] *df= 1702.*

Table 5

RMSEA-P < .05: Type I error rate and Power (Means; Standard Deviations) under condition of no misspecification in measurement component

			Type of Misspecification in Structural and Measurement Components			
N	# indic.	Mag. factor loadings	0-0	1-0	2-0	3-0
			Type I error rate		**Power**	
150	5	Small	.105 (.014;.028)[a]	.321 (.037;.041)[b]	.419 (.048;.043)[c]	.995 (.157;.033)[d]
	5	Large	.173 (.023;.033)[a]	.842 (.105;.046)[b]	.956 (.132;.041)[c]	1 (.278;.031)[d]
	10	Small	.146 (.020;.031)[e]	.689 (.077;.045)[f]	.829 (.096;.042)[g]	1 (.237;.032)[h]
	10	Large	.213 (.027;.036)[e]	.982 (.154;.041)[f]	.999 (.188;.038)[g]	1 (.339;.029)[h]
250	5	Small	.058 (.013;.022)[a]	.351 (.045;.034)[b]	.482 (.056;.032)[c]	1 (.165;.026)[d]
	5	Large	.104 (.019;.027)[a]	.931 (.110;.033)[b]	.991 (.135;.029)[c]	1 (.280;.023)[d]
	10	Small	.095 (.018;.027)[e]	.774 (.084;.034)[f]	.932 (.104;.030)[g]	1 (.241;.024)[h]
	10	Large	.115 (.019;.027)[e]	.999 (.157;.030)[f]	1 (.190;.029)[g]	1 (.336;.022)[h]
1000	5	Small	0 (.008;.013)[a]	.401 (.055;.017)[b]	.746 (.069;.014)[c]	1 (.171;.013)[d]
	5	Large	.001 (.010;.015)[a]	1 (.114;.014)[b]	1 (.139;.014)[c]	1 (.280;.012)[d]
	10	Small	0 (.009;.013)[e]	.984 (.091;.015)[f]	1 (.109;.014)[g]	1 (.242;.012)[h]
	10	Large	0 (.009;.013)[e]	1 (.159;.014)[f]	1 (.190;.014)[g]	1 (.339;.011)[h]

Note. [a]*df* = 394. [b]*df* = 395. [c]*df* = 396. [d]*df* = 397. [e]*df* = 1699. [f]*df* = 1700. [g]*df* = 1701. [h]*df* = 1702.

Table 6

RMSEAs < .05: Type I error rate and Power (Means; Standard Deviations) under condition of no misspecification in measurement component

			Type of Misspecification in Structural and Measurement Components			
N	# indic.	Mag. factor loadings	0-0	1-0	2-0	3-0
			Type I error rate		**Power**	
150	5	Small	.982 (.380;.190)[a]	.996 (.439;.177)[b]	1 (.476;.167)[c]	1 (.552;.129)[d]
	5	Large	.942 (.239;.146)[a]	1 (.369;.129)[b]	1 (.418;.115)[c]	1 (.518;.090)[d]
	10	Small	.977 (.324;.185)[e]	1 (.425;.164)[f]	1 (.453;.145)[g]	1 (.546;.112)[h]
	10	Large	.807 (.134;.088)[e]	1 (.315;.078)[f]	1 (.372;.071)[g]	1 (.483;.052)[h]
250	5	Small	.986 (.358;.197)[a]	1 (.428;.162)[b]	1 (.474;.157)[c]	1 (.558;.124)[d]
	5	Large	.911 (.173;.109)[a]	1 (.337;.105)[b]	1 (.391;.095)[c]	1 (.497;.068)[d]
	10	Small	.967 (.269;.172)[e]	1 (.386;.140)[f]	1 (.434;.128)[g]	1 (.533;.098)[h]
	10	Large	.716 (.095;.059)[e]	1 (.298;.055)[f]	1 (.357;.050)[g]	1 (.474;.038)[h]
1000	5	Small	.962 (.203;.125)[a]	1 (.356;.121)[b]	1 (.413;.113)[c]	1 (.512;.078)[d]
	5	Large	.636 (.077;.042)[a]	1 (.291;.042)[b]	1 (.351;.037)[c]	1 (.466;.026)[d]
	10	Small	.839 (.110;.06)[e]	1 (.307;.062)[f]	1 (.365;.058)[g]	1 (.477;.038)[h]
	10	Large	.268 (.044;.026)[e]	1 (.284;.025)[f]	1 (.344;.025)[g]	1 (.461;.017)[h]

Note. [a] *df* = 394. [b] *df* = 395. [c] *df* = 396. [d] *df* = 397. [e] *df* = 1700. [f] *df* = 1701. [g] *df* = 1702.

Table 7

$SRMR_S$ < .08: Type I error rate and Power (Means; Standard Deviations) under condition of no misspecification in measurement component

			Type of Misspecification in Structural and Measurement Components			
N	# indic.	Mag. factor loadings	0-0	1-0	2-0	3-0
			Type I error rate		**Power**	
150	5	Small	.011 (.035;.016)[a]	.100 (.053;.021)[b]	.210 (.062;.023)[c]	.998 (.174;.035)[d]
	5	Large	0 (.020;.008)[a]	.004 (.042;.013)[b]	.026 (.051;.014)[c]	1 (.172;.022)[d]
	10	Small	0 (.025;.010)[e]	.020 (.046;.016)[f]	.070 (.054;.016)[g]	1 (.175;.027)[h]
	10	Large	0 (.015;.007)[e]	.001 (.040;.011)[f]	.005 (.049;.011)[g]	1 (.172;.019)[h]
250	5	Small	.002 (.029;.012)[a]	.042 (.049;.017)[b]	.116 (.057;.018)[c]	1 (.176;.029)[d]
	5	Large	0 (.015;.006)[a]	.001 (.040;.011)[b]	.003 (.049;.011)[c]	1 (.172;.018)[d]
	10	Small	0 (.020;.008)[e]	.003 (.043;.013)[f]	.021 (.051;.013)[g]	1 (.174;.020)[h]
	10	Large	0 (.011;.005)[e]	0 (.039;.008)[f]	0 (.048;.009)[g]	1 (.172;.015)[h]
1000	5	Small	0 (.015;.006)[a]	0 (.041;.010)[b]	.004 (.050;.010)[c]	1 (.173;.014)[d]
	5	Large	0 (.008;.003)[a]	0 (.038;.005)[b]	0 (.048;.005)[c]	1 (.171;.009)[d]
	10	Small	0 (.010;.004)[e]	0 (.039;.006)[f]	0 (.048;.007)[g]	1 (.172;.010)[h]
	10	Large	0 (.006;.002)[e]	0 (.038;.004)[f]	0 (.047;.004)[g]	1 (.171;.007)[h]

Note. [a] *df* = 394. [b] *df* = 395. [c] *df* = 396. [d] *df* = 397. [e] *df* = 1700. [f] *df* = 1701. [g] *df* = 1702.

Table 8

$CFI_S < .95$: Type I error rate and Power (Means; Standard Deviations) under condition of no misspecification in measurement component

			Type of Misspecification in Structural and Measurement Components			
N	# indic.	Mag. factor loadings	0-0	1-0	2-0	3-0
			Type I error rate		**Power**	
150	5	Small	.691 (.903;.080)[a]	.874 (.850;.096)[b]	.972 (.793;.109)[c]	1 (.683;.102)[d]
	5	Large	.259 (.960;.044)[a]	.805 (.898;.060)[b]	.974 (.847;.069)[c]	1 (.734;.064)[d]
	10	Small	.485 (.932;.065)[e]	.872 (.869;.079)[f]	.976 (.823;.088)[g]	1 (.710;.080)[h]
	10	Large	.034 (.986;.017)[e]	.778 (.924;.034)[f]	.982 (.875;.043)[g]	1 (.759;.041)[h]
250	5	Small	.548 (.918;.076)[a]	.883 (.863;.084)[b]	.976 (.805;.097)[c]	1 (.694;.091)[d]
	5	Large	.087 (.978;.029)[a]	.801 (.914;.047)[b]	.980 (.865;.054)[c]	1 (.750;.048)[d]
	10	Small	.303 (.951;.058)[e]	.862 (.890;.067)[f]	.980 (.839;.075)[g]	1 (.723;.069)[h]
	10	Large	.001 (.993;.008)[e]	.775 (.932;.023)[f]	.998 (.884;.030)[g]	1 (.765;.031)[h]
1000	5	Small	.150 (.971;.037)[a]	.816 (.905;.056)[b]	.988 (.851;.065)[c]	1 (.737;.054)[d]
	5	Large	0 (.995;.005)[a]	.806 (.935;.017)[b]	1 (.888;.021)[c]	1 (.770;.021)[d]
	10	Small	.012 (.991;.011)[e]	.811 (.929;.026)[f]	.998 (.880;.033)[g]	1 (.764;.028)[h]
	10	Large	0 (.998;.002)[e]	.865 (.939;.011)[f]	1 (.892;.015)[g]	1 (.774;.014)[h]

Note. [a] *df* = 394. [b] *df* = 395. [c] *df* = 396. [d] *df* = 397. [e] *df* = 1700. [f] *df* = 1701. [g] *df* = 1702.

Table 9

Chi-Square Test with α=.05: Type I error rate and Power (Means; Standard Deviations) under condition of misspecification in measurement component

			Type of Misspecification in Structural and Measurement Components											
N	# indic.	Mag. factor loadings	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-1
				Type I error rate						Power				
150	5	Small	1 (626.434; 41.261)	1 (1376.308; 56.688)	1 (2100.889; 73.014)	1 (630.07; 40.88)	1 (1377.715; 58.872)	1 (2101.238; 69.158)	1 (632.012; 40.853)	1 (1382.898; 57.36)	1 (2107.692; 70.842)	1 (659.21; 42.174)	1 (1407.572; 60.56)	1 (2134.852; 70.911)
150	5	Large	1 (616.177; 40.694)	1 (1299.231; 55.768)	1 (1945.89; 70.135)	1 (643.606; 42.121)	1 (1304.621; 56.132)	1 (1949.817; 69.752)	1 (633.003; 41.621)	1 (1314.782; 57.813)	1 (1962.647; 70.633)	1 (714.084; 44.689)	1 (1377.7; 59.288)	1 (2025.456; 72.006)
150	10	Small	1 (3101.012; 91.466)	1 (5741.721; 115.506)	1 (7336.88; 132.482)	1 (3105.078; 92.019)	1 (5745.742; 114.897)	1 (7344.04; 135.767)	1 (3114.66; 91.061)	1 (5754.578; 118.182)	1 (7352.072; 135.192)	1 (3167.06; 90.842)	1 (5806.056; 118.498)	1 (7405.765; 132.787)
150	10	Large	1 (3045.862; 88.735)	1 (5569.368; 114.057)	1 (7126.82; 129.608)	1 (3062.643; 89.15)	1 (5585.107; 114.692)	1 (7144.526; 128.953)	1 (3080.139; 89.68)	1 (5602.279; 115.311)	1 (7161.737; 130.213)	1 (3168.023; 90.288)	1 (5691.155; 113.86)	1 (7249.382; 130.312)
250	5	Small	1 (744.253; 45.592)	1 (1998.158; 71.193)	1 (3201.132; 91.206)	1 (745.471; 45.225)	1 (1999.411; 71.006)	1 (3204.912; 88.565)	1 (752.168; 46.278)	1 (2007.311; 69.901)	1 (3212.552; 89.337)	1 (794.3; 47.779)	1 (2046.057; 72.162)	1 (3259.877; 90.271)
250	5	Large	1 (721.446; 45.183)	1 (1861.31; 69.312)	1 (2938.99; 86.909)	1 (733.385; 46.379)	1 (1869.444; 70.355)	1 (2946.254; 86.661)	1 (748.093; 46.653)	1 (1886.345; 71.113)	1 (2966.679; 87.221)	1 (854.371; 51.619)	1 (1990.555; 72.711)	1 (3069.491; 89.658)
250	10	Small	1 (3670.652; 97.808)	1 (8070.76; 136.42)	1 (10731.60; 160.311)	1 (3698.877; 98.049)	1 (8074.304; 134.591)	1 (10738.22; 159.313)	1 (3686.892; 97.264)	1 (8090.854; 135.732)	1 (10752.261; 159.657)	1 (3773.633; 98.557)	1 (8175.731; 137.605)	1 (10834.85; 157.185)
250	10	Large	1 (3574.259; 94.311)	1 (7778.906; 133.823)	1 (10373.89; 155.727)	1 (3601.542; 94.935)	1 (7806.215; 134.299)	1 (10402.95; 156.937)	1 (3630.178; 94.96)	1 (7833.469; 134.526)	1 (10429.602; 155.645)	1 (3777.105; 97.524)	1 (7980.875; 135.019)	1 (10576.045; 157.043)
1,000	5	Small	1 (1721.843; 73.925)	1 (6741.821; 135.384)	1 (11562.18; 172.205)	1 (1725.671; 74.269)	1 (6738.672; 134.073)	1 (11572.745; 174.588)	1 (1745.905; 73.887)	1 (6770.145; 133.48)	1 (11608.003; 172.753)	1 (1912.96; 81.497)	1 (6925.838; 135.215)	1 (11775.474; 175.423)
1,000	5	Large	1 (1622.316; 72.727)	1 (6183.067; 127.485)	1 (10498.165; 166.345)	1 (1668.549; 73.333)	1 (6215.168; 127.676)	1 (10527.295; 167.279)	1 (1724.868; 75.94)	1 (6278.698; 130.107)	1 (10602.50; 164.523)	1 (2147.829; 86.529)	1 (6691.402; 134.448)	1 (11012.08; 168.449)
1,000	10	Small	1 (8908.935; 156.135)	1 (26530.76; 248.236)	1 (37168.66; 288.301)	1 (8935.182; 155.587)	1 (26551.58; 249.71)	1 (37191.801; 287.653)	1 (8980.032; 158.584)	1 (26602.14; 248.704)	1 (35853.69; 288.139)	1 (9322.338; 161.794)	1 (26939.91; 247.396)	1 (37578.06; 288.668)
1,000	10	Large	1 (8535.16; 156.434)	1 (25356.753; 240.517)	1 (35731.732; 284.392)	1 (8642.215; 157.953)	1 (25460.959; 240.401)	1 (35844.567; 280.147)	1 (8747.393; 159.587)	1 (25566.476; 240.352)	1 (35942.108; 282.703)	1 (9339.37; 163.168)	1 (26152.821; 241.18)	1 (36527.353; 283.775)

Table 10

RMSEA-P < .05: Type I error rate and Power (Means; Standard Deviations) under condition of misspecification in measurement component.

N	# indic.	Mag. factor loadings	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-1
			Type I error rate			**Power**								
150	5	Small	.111 (.016;.028)	.196 (.025;.035)	.369 (.047;.045)	.173 (.024;.033)	.153 (.022;.032)	.352 (.045;.046)	.359 (.042;.041)	.462 (.053;.042)	.705 (.080;.046)	.991 (.152;.035)	.992 (.149;.035)	.993 (.166;.037)
150	5	Large	.212 (.026;.037)	.236 (.030;.038)	.353 (.042;.043)	.878 (.112;.046)	.634 (.074;.047)	.641 (.075;.047)	.948 (.125;.041)	.947 (.124;.041)	.953 (.130;.040)	1 (.270;.030)	1 (.269;.031)	1 (.270;.031)
150	10	Small	.142 (.020;.031)	.167 (.021;.035)	.197 (.024;.036)	.517 (.058;.042)	.456 (.052;.047)	.486 (.060;.045)	.794 (.090;.042)	.823 (.093;.041)	.887 (.102;.039)	1 (.235;.031)	1 (.235;.033)	1 (.241;.032)
150	10	Large	.255 (.032;.038)	.361 (.044;.044)	.459 (.053;.045)	.958 (.142;.044)	.984 (.147;.039)	.986 (.149;.039)	1 (.184;.037)	1 (.187;.036)	1 (.191;.034)	1 (.338;.029)	1 (.339;.029)	1 (.340;.029)
250	5	Small	.077 (.015;.024)	.171 (.026;.031)	.463 (.053;.038)	.124 (.022;.028)	.156 (.027;.030)	.447 (.053;.035)	.362 (.047;.033)	.571 (.063;.032)	.83 (.089;.034)	1 (.156;.027)	1 (.158;.026)	.997 (.172;.026)
250	5	Large	.132 (.021;.028)	.175 (.026;.031)	.349 (.044;.037)	.835 (.091;.034)	.716 (.077;.036)	.739 (.079;.037)	.980 (.127;.031)	.991 (.126;.030)	.996 (.133;.029)	1 (.273;.024)	1 (.270;.023)	1 (.271;.024)
250	10	Small	.102 (.02;.027)	.294 (.036;.035)	.265 (.037;.034)	.810 (.088;.036)	.616 (.068;.032)	.660 (.073;.036)	.920 (.099;.03)	.928 (.103;.028)	.977 (.105;.027)	1 (.240;.025)	1 (.239;.027)	1 (.244;.027)
250	10	Large	.175 (.027;.032)	.376 (.045;.036)	.474 (.054;.039)	.995 (.146;.031)	.996 (.149;.030)	.999 (.153;.029)	1 (.187;.028)	1 (.189;.027)	1 (.192;.027)	1 (.339;.022)	1 (.339;.022)	1 (.340;.022)
1,000	5	Small	.002 (.012;.016)	.090 (.034;.021)	.746 (.072;.019)	.016 (.024;.018)	.051 (.031;.019)	.688 (.067;.017)	.607 (.063;.014)	.893 (.076;.014)	1 (.099;.014)	1 (.164;.013)	1 (.165;.013)	1 (.179;.014)
1,000	5	Large	.002 (.012;.015)	.030 (.022;.019)	.235 (.046;.019)	.988 (.095;.016)	.909 (.082;.016)	.939 (.084;.015)	1 (.129;.014)	1 (.128;.014)	1 (.135;.014)	1 (.273;.011)	1 (.270;.011)	1 (.272;.011)
1,000	10	Small	.049 (.035;.018)	.167 (.051;.011)	.333 (.057;.016)	.909 (.079;.015)	.905 (.081;.013)	1 (.103;.028)	1 (.111;.013)	1 (.116;.009)	1 (.103;.015)	1 (.243;.012)	1 (.243;.014)	1 (.236;.012)
1,000	10	Large	.031 (.025;.019)	.262 (.049;.019)	.483 (.059;.019)	1 (.149;.014)	1 (.151;.014)	1 (.156;.014)	1 (.189;.013)	1 (.190;.013)	1 (.193;.013)	1 (.339;.010)	1 (.339;.011)	1 (.339;.011)

Table 11

$RMSEA_S$ < .05: Type I error rate and Power (Means; Standard Deviations) under condition of misspecification in measurement component.

N	# indic.	Mag. factor loadings	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-1
150	5	Small	.989 (.403;.197)	.989 (.383;.193)	.995 (.456;.210)	.998 (.426;.172)	.997 (.382;.161)	.997 (.425;.173)	.998 (.454;.156)	1 (.451;.159)	1 (.510;.170)	1 (.560;.133)	1 (.507;.130)	1 (.537;.134)
150	5	Large	.924 (.247;.167)	.940 (.235;.143)	.946 (.253;.152)	.948 (.757;.809)	.997 (.313;.131)	.996 (.319;.134)	1 (.406;.130)	1 (.381;.115)	1 (.394;.113)	1 (.527;.106)	1 (.487;.083)	1 (.482;.085)
150	10	Small	.978 (.376;.203)	1 (.423;.225)	.972 (.474;.238)	1 (.444;.184)	1 (.468;.207)	1 (.506;.208)	1 (.495;.157)	1 (.506;.152)	1 (.550;.161)	1 (.590;.128)	1 (.603;.118)	1 (.623;.126)
150	10	Large	.876 (.170;.111)	.927 (.228;.150)	.940 (.253;.155)	.999 (.318;.103)	1 (.344;.106)	1 (.360;.106)	1 (.384;.080)	1 (.407;.092)	1 (.426;.099)	1 (.496;.063)	1 (.510;.070)	1 (.519;.073)
250	5	Small	.990 (.354;.188)	.998 (.379;.184)	.996 (.451;.195)	.995 (.365;.165)	.993 (.381;.186)	1 (.436;.179)	1 (.424;.149)	1 (.434;.149)	1 (.503;.150)	1 (.525;.124)	1 (.519;.131)	1 (.545;.120)
250	5	Large	.896 (.180;.124)	.919 (.175;.105)	.949 (.204;.112)	.998 (.305;.111)	.997 (.276;.105)	.994 (.282;.110)	1 (.363;.087)	1 (.350;.087)	1 (.363;.093)	1 (.479;.070)	1 (.464;.064)	1 (.463;.063)
250	10	Small	.981 (.361;.195)	.991 (.479;.201)	1 (.497;.184)	.982 (.587;.649)	1 (.472;.166)	1 (.463;.155)	1 (.499;.156)	1 (.539;.142)	1 (.552;.196)	1 (.579;.106)	1 (.619;.121)	1 (.632;.114)
250	10	Large	.813 (.130;.083)	.929 (.188;.103)	.959 (.213;.113)	1 (.296;.063)	1 (.320;.074)	1 (.341;.086)	1 (.368;.056)	1 (.385;.066)	1 (.403;.078)	1 (.482;.040)	1 (.494;.050)	1 (.502;.054)
1,000	5	Small	.961 (.222;.139)	.998 (.297;.159)	1 (.476;.182)	.998 (.269;.132)	.998 (.288;.135)	1 (.434;.161)	1 (.344;.100)	1 (.373;.120)	1 (.497;.131)	1 (.471;.080)	1 (.458;.078)	1 (.540;.101)
1,000	5	Large	.661 (.080;.042)	.793 (.095;.043)	.977 (.142;.045)	1 (.254;.044)	1 (.225;.043)	1 (.238;.042)	1 (.322;.037)	1 (.309;.035)	1 (.324;.034)	1 (.447;.025)	1 (.433;.023)	1 (.436;.023)
1,000	10	Small	1 (.401;.186)	1 (.593;.18)	1 (.742;.176)	1 (.437;.158)	1 (.653;.197)	1 (.917;.082)	1 (.519;.128)	1 (.634;.125)	1 (.314;.143)	1 (.584;.095)	1 (.667;.103)	1 (.701;.160)
1,000	10	Large	.752 (.087;.039)	.975 (.143;.047)	.996 (.171;.052)	1 (.275;.028)	1 (.290;.032)	1 (.311;.034)	1 (.350;.025)	1 (.361;.028)	1 (.374;.029)	1 (.468;.017)	1 (.475;.020)	1 (.482;.021)

Table 12

$SRMR_S$ < .08: Type I error rate and Power (Means; Standard Deviations) under condition of misspecification in measurement component

N	# indic.	Mag. factor loadings	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-1
150	5	Small	.014 (.036;.016)	.026 (.040;.018)	.081 (.048;.020)	.040 (.045;.018)	.049 (.044;.019)	.160 (.053;.025)	.111 (.056;.020)	.135 (.058;.019)	.287 (.070;.023)	.993 (.161;.034)	.979 (.152;.035)	.986 (.160;.039)
150	5	Large	0 (.020;.009)	0 (.021;.009)	0 (.022;.009)	.071 (.055;.037)	.003 (.034;.012)	0 (.033;.013)	.018 (.048;.013)	.014 (.047;.013)	.022 (.048;.014)	1 (.173;.023)	1 (.164;.023)	1 (.164;.023)
150	10	Small	0 (.025;.011)	.008 (.026;.011)	0 (.028;.012)	.008 (.040;.014)	.007 (.039;.016)	.014 (.041;.015)	.024 (.050;.015)	.031 (.052;.015)	.070 (.054;.015)	1 (.172;.027)	1 (.168;.029)	1 (.169;.026)
150	10	Large	0 (.015;.006)	0 (.016;.007)	0 (.017;.007)	0 (.038;.011)	0 (.038;.010)	0 (.037;.010)	.004 (.048;.010)	.004 (.048;.010)	.006 (.048;.010)	1 (.171;.019)	1 (.170;.019)	1 (.169;.020)
250	5	Small	.002 (.029;.013)	.005 (.034;.014)	.021 (.042;.016)	.003 (.036;.015)	.014 (.037;.016)	.044 (.046;.016)	.037 (.048;.015)	.055 (.053;.016)	.190 (.064;.019)	.998 (.158;.028)	.998 (.154;.028)	.992 (.158;.030)
250	5	Large	0 (.015;.007)	0 (.016;.007)	0 (.019;.007)	0 (.035;.010)	0 (.031;.010)	0 (.030;.010)	.003 (.045;.010)	.001 (.044;.010)	.004 (.045;.010)	1 (.167;.018)	1 (.164;.017)	1 (.164;.018)
250	10	Small	0 (.020;.009)	0 (.024;.009)	0 (.025;.008)	.024 (.047;.024)	0 (.036;.011)	0 (.037;.011)	.007 (.047;.012)	.010 (.048;.011)	0 (.048;.010)	1 (.172;.022)	1 (.168;.022)	1 (.168;.026)
250	10	Large	0 (.012;.005)	0 (.013;.005)	0 (.014;.005)	0 (.037;.008)	0 (.036;.008)	0 (.036;.008)	0 (.047;.008)	0 (.047;.008)	.001 (.047;.008)	1 (.171;.015)	1 (.169;.015)	1 (.168;.015)
1,000	5	Small	0 (.016;.006)	0 (.023;.008)	0 (.037;.009)	0 (.024;.009)	0 (.025;.009)	0 (.037;.009)	0 (.040;.008)	.002 (.044;.008)	.013 (.057;.009)	1 (.158;.014)	1 (.152;.015)	1 (.157;.016)
1,000	5	Large	0 (.008;.003)	0 (.009;.004)	0 (.013;.004)	0 (.032;.005)	0 (.028;.005)	0 (.027;.005)	0 (.043;.005)	0 (.041;.005)	0 (.042;.005)	1 (.166;.009)	1 (.163;.009)	1 (.163;.009)
1,000	10	Small	0 (.014;.004)	0 (.017;.003)	0 (.018;.004)	0 (.034;.006)	0 (.033;.006)	0 (.036;.009)	0 (.044;.006)	0 (.044;.005)	0 (.044;.007)	1 (.170;.010)	1 (.164;.011)	1 (.154;.011)
1,000	10	Large	0 (.007;.002)	0 (.009;.002)	0 (.010;.003)	0 (.036;.004)	0 (.035;.004)	0 (.035;.004)	0 (.046;.004)	0 (.045;.004)	0 (.046;.004)	1 (.170;.007)	1 (.168;.007)	1 (.167;.007)

Table 13

CFIs < .95: Type I error rate and Power (Means; Standard Deviations) under condition of misspecification in measurement component

N	# indic.	Mag. factor loadings	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-1
150	5	Small	.714 (.896;.083)	.695 (.896;.083)	.815 (.858;.099)	.847 (.859;.090)	.826 (.876;.083)	.893 (.845;.094)	.963 (.810;.100)	.951 (.803;.103)	.981 (.749;.118)	.998 (.687;.101)	1 (.714;.099)	1 (.675;.104)
150	5	Large	.276 (.956;.052)	.260 (.960;.043)	.313 (.954;.050)	.836 (.763;.277)	.650 (.921;.058)	.646 (.916;.061)	.959 (.856;.074)	.952 (.868;.067)	.951 (.857;.067)	1 (.731;.073)	1 (.756;.060)	1 (.756;.062)
150	10	Small	.561 (.917;.074)	.667 (.901;.088)	.761 (.882;.098)	.839 (.870;.084)	.846 (.858;.099)	.889 (.836;.109)	.976 (.813;.091)	.969 (.802;.092)	1 (.775;.100)	1 (.699;.086)	1 (.692;.085)	1 (.666;.089)
150	10	Large	.097 (.979;.028)	.215 (.964;.044)	.285 (.957;.047)	.693 (.925;.044)	.774 (.916;.047)	.833 (.909;.047)	.988 (.874;.046)	.988 (.863;.052)	.993 (.852;.056)	1 (.759;.045)	1 (.753;.048)	1 (.747;.050)
250	5	Small	.582 (.918;.074)	.683 (.905;.075)	.815 (.868;.090)	.735 (.894;.079)	.764 (.882;.092)	.899 (.848;.096)	.949 (.841;.086)	.966 (.824;.092)	.990 (.766;.102)	1 (.725;.089)	1 (.719;.093)	1 (.684;.094)
250	5	Large	.113 (.975;.034)	.096 (.977;.027)	.166 (.969;.033)	.647 (.928;.049)	.536 (.938;.044)	.565 (.934;.047)	.960 (.882;.048)	.952 (.887;.049)	.976 (.876;.054)	1 (.766;.049)	1 (.774;.046)	1 (.770;.046)
250	10	Small	.531 (.926;.067)	.789 (.884;.078)	.898 (.881;.068)	.861 (.817;.226)	.920 (.863;.076)	.915 (.862;.074)	.983 (.815;.087)	1 (.790;.085)	1 (.785;.110)	1 (.713;.071)	1 (.688;.079)	1 (.676;.079)
250	10	Large	.032 (.987;.015)	.112 (.976;.025)	.171 (.970;.030)	.685 (.936;.026)	.784 (.927;.031)	.831 (.918;.038)	.996 (.883;.031)	.997 (.876;.037)	.998 (.865;.044)	1 (.769;.030)	1 (.764;.035)	1 (.758;.037)
1,000	5	Small	.207 (.965;.042)	.422 (.940;.058)	.912 (.863;.084)	.444 (.941;.055)	.537 (.931;.057)	.928 (.857;.083)	.916 (.894;.053)	.964 (.873;.066)	1 (.784;.081)	1 (.777;.052)	1 (.779;.052)	1 (.706;.068)
1,000	5	Large	0 (.995;.005)	.001 (.993;.006)	.005 (.986;.009)	.459 (.950;.017)	.237 (.959;.015)	.354 (.954;.016)	.995 (.905;.020)	.989 (.910;.019)	.998 (.899;.020)	1 (.787;.019)	1 (.794;.018)	1 (.789;.019)
1,000	10	Small	.618 (.916;.067)	1 (.843;.081)	1 (.779;.086)	.881 (.883;.072)	1 (.783;.095)	1 (.642;.049)	1 (.810;.072)	1 (.744;.072)	.978 (.904;.057)	1 (.719;.060)	1 (.666;.067)	1 (.642;.104)
1,000	10	Large	0 (.995;.004)	.003 (.987;.008)	.008 (.982;.011)	.683 (.945;.010)	.787 (.904;.012)	.920 (.932;.014)	1 (.894;.014)	1 (.889;.015)	1 (.882;.016)	1 (.779;.013)	1 (.777;.015)	1 (.773;.015)

Author Contact

Moritz Heene, Department of Psychology, Ludwig-Maximilians-Universität München, Germany. ORCID iD: https://orcid.org/0009-0007-4956-2417. Correspondence regarding this article should be addressed to Moritz Heene, Email: heene@psy.lmu.de

Michael Maraun, Department of Psychology, Simon Fraser University, Burnaby, B.C., Canada.

Nadine Juliana Glushko, Independent scholar.

Sunthud Pornprasertmanit, Faculty of Psychology, Chulalongkorn University, Bangkok, Thailand.

Conflict of Interest and Funding

The authors declare that there are no conflicts of interest with respect to the research, authorship, and/or publication of this article. The authors received no specific funding for this work.

Author Contributions

Conceptualization: Moritz Heene, Michael D. Maraun, Nadine J. Glushko and Sunthud Pornprasertmanit.
Data Curation: Moritz Heene and Sunthud Pornprasertmanit.
Formal Analysis: Moritz Heene, Michael D. Maraun, Nadine J. Glushko and Sunthud Pornprasertmanit.
Investigation: Moritz Heene, Michael D. Maraun, Nadine J. Glushko and Sunthud Pornprasertmanit.
Methodology: Moritz Heene, Michael D. Maraun, Nadine J. Glushko and Sunthud Pornprasertmanit.
Project Administration: Moritz Heene, Nadine J. Glushko and Sunthud Pornprasertmanit.
Resources: Moritz Heene, Nadine J. Glushko and Sunthud Pornprasertmanit.
Software: Moritz Heene and Sunthud Pornprasertmanit.
Supervision: Moritz Heene.
Validation: Moritz Heene, Nadine J. Glushko and Sunthud Pornprasertmanit.
Visualization: Moritz Heene and Nadine J. Glushko.
Writing - Original Draft Preparation: Moritz Heene, Michael D. Maraun and Nadine J. Glushko.
Writing - Review & Editing: Moritz Heene, Michael D. Maraun, Nadine J. Glushko, and Sunthud Pornprasertmanit

Open Science Practices

This article earned the Open Materials and Open Code badge for making materials and code openly available. It has been verified that the analysis reproduced the results presented in the article. The entire editorial process, including the open reviews, is published in the online supplement.

References

Bentler, P. M., & Yuan, K.-H. (1999). Structural equation modeling with small samples: Test statistics. *Multivariate Behavioral Research*, *34*(2), 181–197. https://doi.org/10.1207/S15327906Mb340203

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons.

Breivik, E., & Olsson, U. H. (2001). Adding variables to improve fit: The effect of model size on fit assessment in lisrel. In R. Cudeck, S. Du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future: A festschrift in honour of karl jöreskog* (pp. 169–194). Scientific Software International.

Browne, M. W., MacCallum, R. C., Kim, C. T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, *7*(4), 403–421. https://doi.org/10.1037//1082-989x.7.4.403

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

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

Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*(4), 381–398. https://doi.org/10.1037/1082-989X.12.4.381

Curran, P. J., Bollen, K. A., Paxton, P., Kirby, J., & Chen, F. (2002). The noncentral chi-square distribution in misspecified structural equation models: Finite sample results from a monte carlo simulation. *Multivariate Behavioral Research*, *37*(1), 1–36. https://doi.org/10.1207/S15327906MBR3701_01

Driel, O. P. (1978). On various causes of improper solutions in maximum likelihood factor analysis.

Psychometrika, *43*(2), 225–243. https://doi.org/10.1007/BF02293865

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

Grice, J. W., Cohn, A., Ramsey, R. R., & Chaney, J. M. (2015). On muddled reasoning and mediation modeling. *Basic and Applied Social Psychology*, *37*(4), 214–225. https://doi.org/10.1080/01973533.2015.1049350

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

Hayduk, L. A. (2014). Seeing perfectly fitting factor models that are causally misspecified: Understanding that close-fitting models can be worse. *Educational and Psychological Measurement*, *74*(6), 905–926. https://doi.org/10.1177/0013164414527449

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods*, *16*(3), 319–336. https://doi.org/10.1037/a0024917

Heene, M., Hilbert, S., Freudenthaler, H. H., & Bühner, M. (2012). Sensitivity of sem fit indexes with respect to violations of uncorrelated errors. *Structural Equation Modeling*, *19*(1), 36–50. https://doi.org/10.1080/10705511.2012.634710

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

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, *14*(1), 6–23. https://doi.org/10.1037/a0014694

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

Kline, R. B. (2015). The mediation myth. *Basic and Applied Social Psychology*, *37*(4), 202–213. https://doi.org/10.1080/01973533.2015.1049349

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

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indexes in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin*, *103*(3), 391–410. https://doi.org/10.1037/0033-2909.103.3.391

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

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

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

Peterson, R. A. (2000). A meta analysis of variance accounted for and factor loadings in exploratory factor analysis. *Marketing Letters*, *11*(3), 261–275. https://doi.org/10.1023/A:1008191211004

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology*, *90*(1), 175–181. https://doi.org/10.1037/0021-9010.90.1.175

Pornprasertmanit, S., Miller, P., & Schoemann, A. (2015). *Simsem: Simulated structural equation modeling*. http://CRAN.R-project.org/package=simsem

R Core Team. (2015). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Saris, W. E., & Satorra, A. (1988). Characteristics of structural equation models which affect the power of the likelihood ratio test. *Sociometric Research*, *2*, 220–236.

Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Sage.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or detection of misspecifications. *Structural Equation Modeling*, *16*, 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.

Schonemann, P. H. (1981). Power as a function of communality in factor analysis. *Bulletin of the Psychonomic Society*, *17*(1), 57–60.

Shah, R., & Goldstein, S. M. (2006). Use of structural equation modeling in operations management research: Looking back and forward. *Journal of Operations Management*, *24*(2), 148–169. https://doi.org/10.1016/j.jom.2005.05.001

Strobl, C., Boulesteix, A.-L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, *9*(1), 307. https://doi.org/10.1186/1471-2105-9-307

Tate, C. U. (2015). On the overuse and misuse of mediation analysis: It may be a matter of timing. *Basic and Applied Social Psychology*, *37*(4), 235–246. https://doi.org/10.1080/01973533.2015.1062380

Westland, J. C. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications*, *9*(6), 476–487. https://doi.org/10.1016/j.elerap.2010.07.003

Williams, L. J., & O'Boyle, E. H. J. (2011). The myth of global fit indices and alternatives for assessing latent variable relations. *Organizational Research Methods*, *14*(2), 350–369. https://doi.org/10.1177/1094428110391472

Xia, Y., & Yang, Y. (2018). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1055-2

Yuan, K.-H. (2005). Fit indices versus test statistics. *Multivariate Behavioral Research*, *40*(1), 115–148. https://doi.org/10.1207/s15327906mbr4001_5

Yuan, K.-H., Yang, M., & Jiang, G. (2017). Empirically corrected rescaled statistics for sem with small n and large p. *Multivariate Behavioral Research*, *52*(6), 673–698.

Appendix A

Table A1

RMSEA < .05: Type I error rate and Power (Means; Standard Deviations) under condition of no misspecification in measurement component

N	# indic.	Mag. factor loadings	0-0	1-0	2-0	3-0
			Type I error rate		Power	
150	5	Small	0 (.022;.012)	0 (.024;.011)	0 (.024;.011)	0 (.032;.01)
	5	Large	0 (.023;.011)	0 (.026;.011)	0 (.029;.010)	.003 (.044;.007)
	10	Small	0 (.036;.004)	0 (.036;.004)	0 (.036;.004)	0 (.039;.004)
	10	Large	0 (.036;.004)	0 (.037;.004)	0 (.037;.004)	0 (.042;.003)
250	5	Small	0 (.013;.009)	0 (.014;.009)	0 (.015;.009)	0 (.026;.007)
	5	Large	0 (.013;.009)	0 (.018;.009)	0 (.021;.009)	0 (.04;.005)
	10	Small	0 (.020;.004)	0 (.021;.004)	0 (.021;.004)	0 (.025;.003)
	10	Large	0 (.020;.004)	0 (.022;.003)	0 (.023;.003)	0 (.03;.003)
1000	5	Small	0 (.004;.004)	0 (.007;.005)	0 (.023;.002)	
	5	Large	0 (.004;.004)	0 (.013;.003)	0 (.017;.003)	0 (.037;.002)
	10	Small	0 (.004;.003)	0 (.007;.003)	0 (.008;.002)	0 (.016;.001)
	10	Large	0 (.004;.003)	0 (.010;.002)	0 (.012;.002)	0 (.022;.001)

Appendix B

Table B1

SRMR < .08: Type I error rate and Power (Means; Standard Deviations) under condition of no misspecification in measurement component

N	# indic.	Mag. factor loading	0-0 Type I error rate	1-0	2-0 Power	3-0
150	5	Small	0 (.060;.003)	0 (.061;.003)	0 (.061;.003)	.020 (.068;.005)
	5	Large	0 (.048;.003)	0 (.050;.004)	0 (.052;.004)	.853 (.092;.011)
	10	Small	0 (.065;.002)	0 (.066;.002)	0 (.066;.002)	.155 (.075;.004)
	10	Large	0 (.053;.003)	0 (.056;.003)	0 (.057;.003)	.986 (.102;.010)
250	5	Small	0 (.046;.002)	0 (.047;.002)	0 (.047;.002)	.020 (.056;.005)
	5	Large	0 (.037;.002)	0 (.040;.003)	0 (.041;.003)	.746 (.087;.01)
	10	Small	0 (.050;.001)	0 (.051;.001)	0 (.051;.001)	.063 (.063;.004)
	10	Large	0 (.041;.002)	0 (.044;.002)	0 (.045;.002)	.963 (.094;.008)
1000	5	Small	0 (.023;.001)	0 (.024;.001)	0 (.025;.001)	.000 (.038;.003)
	5	Large	0 (.019;.001)	0 (.024;.002)	0 (.026;.002)	.511 (.080;.005)
	10	Small	0 (.025;.001)	0 (.027;.001)	0 (.027;.001)	.000 (.045;.003)
	10	Large	0 (.020;.001)	0 (.026;.001)	0 (.028;.001)	.982 (.089;.004)

Appendix C

Table C1

CFI < .95: Type I error rate and Power (Means; Standard Deviations) under condition of no misspecification in measurement component

N	# indic.	Mag. factor loading	0-0 Type I error rate	1-0	2-0 Power	3-0
150	5	Small	.415 (.954;.034)	.466 (.951;.034)	.496 (.949;.034)	.786 (.920;.039)
	5	Large	.011 (.983;.013)	.021 (.979;.013)	.039 (.976;.014)	.585 (.947;.015)
	10	Small	.999 (.845;.033)	.999 (.841;.033)	.999 (.840;.033)	1 (.815;.033)
	10	Large	.834 (.936;.014)	.886 (.933;.014)	.925 (.930;.014)	.996 (.913;.014)
250	5	Small	.074 (.981;.018)	.094 (.979;.019)	.106 (.977;.019)	.568 (.945;.024)
	5	Large	0 (.993;.007)	0 (.989;.008)	0 (.986;.008)	.244 (.957;.010)
	10	Small	.609 (.944;.020)	.677 (.940;.020)	.729 (.938;.020)	.981 (.911;.020)
	10	Large	.002 (.978;.008)	.003 (.974;.008)	.005 (.972;.008)	.347 (.953;.008)
1000	5	Small	0 (.997;.004)	0 (.995;.005)	0 (.993;.005)	.128 (.959;.008)
	5	Large	0 (.999;.001)	0 (.995;.002)	0 (.992;.003)	0 (.962;.004)
	10	Small	0 (.996;.004)	0 (.993;.005)	0 (.990;.005)	.017 (.962;.006)
	10	Large	0 (.998;.001)	0 (.995;.002)	0 (.992;.002)	0 (.974;.002)

Appendix D

Technical Elucidation of Departure of Mean and Standard Deviation of RMSEAs from Theoretical Counterparts

Because RMSEA$_S$ is simply the standard RMSEA computed on the structural component of a candidate model, and, under particular assumptions, the sampling distribution of the RMSEA is known (e.g., Rigdon, 1996; Steiger, 1990), the parameters of the sampling distribution of RMSEA$_S$ can be analytically derived. If the particular assumptions are satisfied, and H_0 is true, it can be shown that $\hat{\chi}^2_{misspec}$ of (4) has a central χ^2-distribution. Accordingly, setting aside those instances of a negative value of $\hat{\chi}^2_{misspec} - \hat{\chi}^2_{model}$, the squared counterpart of RMSEA$_S$, $\mathrm{RM\hat{S}EA}^2_S$, will also have approximately central χ^2. It follows, then, that:

$$E(\mathrm{RMSEA}^2_{\mathrm{S}}) = \frac{E(\hat{\chi}^2_{\mathrm{misspec}}) - E(\hat{\chi}^2_{\mathrm{model}})}{df_{\mathrm{model}}(N-1)} = 0;$$

and

$$\mathrm{Var}(\mathrm{RMSEA}^2_{\mathrm{S}}) = \frac{\mathrm{Var}(\hat{\chi}^2_{\mathrm{misspec}})}{df_{\mathrm{model}}(N-1)} = \frac{2\mathrm{df}}{[df(N-1)]^2} = \frac{2}{df(N-1)}.$$

As shown in Table 6, however, under H_0 true [i.e., a correctly specified structural component], the empirical means of RMSEAs are seldom close to zero. Now, as for the case of our models, with six indicators and 14 parameters to be estimated (three correlations, eight paths, and three error variances), df is equal to $(6(6+1)/2) - 14 = 7$. Finally, with sample sizes of $N = [150, 250, 1000]$, the theoretical standard deviation of RMSEA$_S$ is computed to be [.004, .002, .001]. It will be confirmed by inspection that the empirical standard deviations presented in Table 6 are not even close to these values. All told, it appears that the distribution of RMSEA$_S$ is not well approximated by the theoretical distributions, heretofore, by methodologists, derived. It would have to be concluded that the actual distribution of RMSEA$_S$ remains unknown; and there exists no logical basis for selecting cutoff values such as .05, by means of reference to the quantiles of distributions, the relevance of which, to the actual distribution of RMSEA$_S$, is unknown.

Appendix E

Table E1

RMSEA < .05: Type I error rate and Power (Means; Standard Deviations) under condition of misspecification in measurement component

N	# indic.	Mag. factor loadings	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-1
				Type I error rate						**Power**				
150	5	Small	.677 (.062;.006)	1 (.129;.004)	1 (.170;.004)	.705 (.063;.006)	1 (.129;.004)	1 (.170;.004)	.707 (.063;.006)	1 (.129;.004)	1 (.170;.004)	.869 (.066;.005)	1 (.130;.004)	1 (.171;.003)
150	5	Large	.566 (.061;.006)	1 (.124;.004)	1 (.162;.004)	.789 (.065;.006)	1 (.124;.004)	1 (.162;.004)	.993 (.063;.006)	1 (.124;.004)	1 (.162;.004)	.585 (.073;.005)	1 (.128;.004)	1 (.165;.004)
150	10	Small	1 (.074;.002)	1 (.126;.002)	1 (.149;.002)	.999 (.074;.002)	1 (.126;.002)	1 (.149;.002)	1 (.074;.002)	1 (.126;.002)	1 (.149;.002)	.999 (.076;.002)	1 (.127;.002)	1 (.149;.002)
150	10	Large	1 (.073;.002)	1 (.123;.002)	1 (.146;.002)	1 (.073;.002)	1 (.123;.002)	1 (.146;.002)	1 (.073;.002)	1 (.123;.002)	1 (.146;.002)	1 (.073;.002)	1 (.125;.002)	1 (.147;.002)
250	5	Small	.443 (.060;.004)	1 (.128;.003)	1 (.169;.003)	.444 (.059;.004)	1 (.127;.003)	1 (.169;.003)	.444 (.06;.004)	1 (.128;.003)	1 (.169;.003)	.487 (.063;.004)	1 (.129;.003)	1 (.170;.003)
250	5	Large	.269 (.058;.004)	1 (.122;.003)	1 (.161;.003)	.350 (.058;.004)	1 (.122;.003)	1 (.161;.003)	.350 (.058;.004)	1 (.122;.003)	1 (.161;.003)	.983 (.063;.004)	1 (.128;.003)	1 (.164;.003)
250	10	Small	1 (.068;.002)	1 (.122;.001)	1 (.146;.001)	1 (.069;.002)	1 (.122;.001)	1 (.146;.001)	1 (.069;.002)	1 (.122;.001)	1 (.146;.001)	1 (.068;.004)	1 (.123;.001)	1 (.146;.001)
250	10	Large	1 (.066;.002)	1 (.12;.001)	1 (.143;.001)	1 (.067;.002)	1 (.12;.001)	1 (.143;.001)	1 (.067;.002)	1 (.12;.001)	1 (.143;.001)	1 (.067;.002)	1 (.123;.001)	1 (.146;.001)
1000	5	Small	.105 (.058;.002)	1 (.127;.001)	1 (.168;.001)	.112 (.058;.002)	1 (.127;.001)	1 (.168;.001)	.112 (.058;.002)	1 (.127;.001)	1 (.168;.001)	.854 (.062;.002)	1 (.128;.001)	1 (.169;.001)
1000	5	Large	.006 (.056;.002)	1 (.121;.001)	1 (.16;.001)	.025 (.057;.002)	1 (.121;.001)	1 (.16;.001)	.025 (.057;.002)	1 (.121;.001)	1 (.16;.001)	.105 (.057;.002)	1 (.126;.001)	1 (.161;.001)
1000	10	Small	1 (.065;.001)	1 (.121;.001)	1 (.144;.001)	1 (.065;.001)	1 (.121;.001)	1 (.144;.001)	1 (.065;.001)	1 (.121;.001)	1 (.144;.001)	1 (.065;.001)	1 (.122;.001)	1 (.145;.001)
1000	10	Large	1 (.063;.001)	1 (.118;.001)	1 (.142;.001)	1 (.064;.001)	1 (.118;.001)	1 (.142;.001)	1 (.064;.001)	1 (.118;.001)	1 (.142;.001)	1 (.064;.001)	1 (.118;.001)	1 (.143;.001)

Appendix F

Table F1

SRMR < .08: Type I error rate and Power (Means; Standard Deviations) under condition of misspecification in measurement component

N	# indic.	Mag. factor loadings	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-1
			Type I error rate			**Power**								
150	5	Small	.003 (.071;.003)	1 (.093;.004)	1 (.110;.005)	.001 (.071;.003)	1 (.093;.004)	1 (.110;.005)	.002 (.071;.003)	1 (.094;.004)	1 (.110;.004)	.283 (.078;.005)	1 (.099;.005)	1 (.117;.006)
150	5	Large	0 (.054;.004)	.004 (.068;.004)	.407 (.079;.005)	0 (.059;.004)	.012 (.069;.004)	.434 (.081;.005)	0 (.057;.004)	.019 (.074;.004)	.536 (.081;.005)	.966 (.098;.01)	.993 (.104;.011)	1 (.113;.01)
150	10	Small	.099 (.077;.002)	1 (.094;.003)	1 (.104;.003)	.116 (.077;.002)	1 (.095;.003)	1 (.104;.003)	.163 (.078;.002)	1 (.095;.003)	1 (.104;.003)	.952 (.086;.004)	1 (.102;.004)	1 (.111;.004)
150	10	Large	0 (.073;.003)	.003 (.071;.003)	.161 (.076;.004)	0 (.062;.003)	.01 (.072;.003)	.29 (.078;.004)	0 (.063;.003)	.023 (.073;.003)	.374 (.086;.004)	.997 (.105;.01)	1 (.112;.009)	1 (.116;.009)
250	5	Small	0 (.060;.002)	.97 (.085;.003)	1 (.104;.003)	0 (.06;.002)	.963 (.085;.003)	1 (.104;.003)	0 (.06;.002)	.977 (.086;.003)	1 (.104;.003)	.003 (.067;.004)	1 (.092;.004)	1 (.111;.005)
250	5	Large	0 (.045;.003)	.023 (.061;.003)	.023 (.073;.004)	0 (.047;.003)	.003 (.062;.003)	.035 (.073;.004)	0 (.048;.003)	.065 (.063;.003)	.848 (.089;.009)	.995 (.099;.009)	1 (.109;.008)	1 (.117;.008)
250	10	Small	0 (.065;.002)	.995 (.085;.002)	1 (.096;.002)	0 (.066;.002)	.996 (.085;.002)	1 (.096;.002)	0 (.066;.002)	.998 (.086;.002)	1 (.096;.002)	.101 (.076;.003)	1 (.094;.003)	1 (.103;.003)
250	10	Large	0 (.050;.002)	1 (.062;.002)	1 (.069;.003)	0 (.052;.002)	1 (.064;.003)	1 (.07;.003)	0 (.053;.002)	1 (.065;.003)	1 (.071;.003)	.88 (.1;.008)	1 (.107;.007)	1 (.111;.007)
1000	5	Small	0 (.045;.001)	.001 (.076;.001)	1 (.096;.002)	0 (.045;.001)	.001 (.076;.001)	1 (.096;.002)	0 (.045;.001)	.001 (.076;.001)	1 (.096;.002)	.729 (.054;.002)	.999 (.083;.002)	1 (.104;.004)
1000	5	Large	0 (.032;.001)	0 (.052;.001)	1 (.066;.002)	0 (.034;.001)	0 (.053;.001)	1 (.067;.002)	0 (.036;.001)	0 (.055;.001)	1 (.067;.002)	.003 (.062;.002)	.999 (.083;.002)	1 (.104;.004)
1000	10	Small	0 (.049;.001)	0 (.073;.001)	1 (.085;.001)	0 (.049;.001)	0 (.074;.001)	1 (.085;.001)	0 (.049;.001)	0 (.074;.001)	1 (.085;.001)	.999 (.074;.001)	1 (.083;.002)	.999 (.094;.001)
1000	10	Large	0 (.035;.001)	0 (.051;.001)	1 (.059;.001)	0 (.038;.001)	0 (.053;.001)	1 (.061;.001)	0 (.04;.001)	0 (.054;.001)	1 (.062;.001)	.093 (.061;.001)	1 (.1;.004)	1 (.105;.004)

Appendix G

Table G1

CFI < .95: Type I error rate and Power (Means; Standard Deviations) under condition of misspecification in measurement component

N	# indic.	Mag. factor loadings	0-1	0-2	0-3	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-1
			Type I error rate			**Power**								
150	5	Small	1 (.775;.04)	1 (.456;.036)	1 (.334;.03)	1 (.772;.04)	1 (.457;.037)	1 (.335;.03)	1 (.772;.039)	1 (.454;.037)	1 (.332;.031)	1 (.747;.041)	1 (.442;.038)	1 (.323;.031)
150	5	Large	.994 (.909;.017)	1 (.712;.024)	1 (.594;.026)	.999 (.89;.019)	1 (.711;.024)	1 (.593;.026)	1 (.903;.017)	1 (.708;.024)	1 (.59;.026)	.999 (.86;.02)	1 (.688;.024)	1 (.573;.026)
150	10	Small	1 (.559;.033)	1 (.308;.026)	1 (.243;.021)	1 (.559;.033)	1 (.307;.026)	1 (.241;.022)	1 (.555;.033)	1 (.306;.026)	1 (.24;.022)	1 (.539;.033)	1 (.297;.025)	1 (.234;.022)
150	10	Large	1 (.781;.019)	1 (.554;.023)	1 (.47;.022)	1 (.778;.019)	1 (.552;.023)	1 (.468;.022)	1 (.776;.019)	1 (.55;.023)	1 (.467;.022)	1 (.761;.019)	1 (.54;.022)	1 (.458;.022)
250	5	Small	1 (.791;.028)	1 (.461;.028)	1 (.337;.023)	1 (.791;.028)	1 (.461;.028)	1 (.336;.024)	1 (.787;.029)	1 (.459;.028)	1 (.335;.023)	1 (.763;.03)	1 (.445;.028)	1 (.324;.024)
250	5	Large	.998 (.919;.012)	1 (.718;.018)	1 (.597;.019)	.998 (.916;.012)	1 (.716;.018)	1 (.596;.019)	.999 (.912;.017)	1 (.713;.018)	1 (.593;.019)	.999 (.886;.013)	1 (.694;.018)	1 (.577;.019)
250	10	Small	1 (.600;.025)	1 (.319;.02)	1 (.25;.017)	1 (.592;.025)	1 (.319;.02)	1 (.249;.017)	1 (.598;.025)	1 (.317;.02)	1 (.248;.017)	1 (.58;.025)	1 (.309;.02)	1 (.241;.017)
250	10	Large	1 (.810;.013)	1 (.569;.017)	1 (.481;.017)	1 (.808;.013)	1 (.567;.017)	1 (.479;.017)	1 (.805;.013)	1 (.565;.017)	1 (.477;.017)	1 (.792;.017)	1 (.555;.017)	1 (.469;.017)
1000	5	Small	1 (.800;.012)	1 (.464;.014)	1 (.337;.012)	1 (.8;.012)	1 (.464;.014)	1 (.337;.012)	1 (.797;.012)	1 (.461;.014)	1 (.335;.012)	1 (.772;.013)	1 (.448;.014)	1 (.325;.011)
1000	5	Large	1 (.923;.005)	1 (.721;.008)	1 (.599;.009)	1 (.919;.005)	1 (.719;.009)	1 (.598;.009)	1 (.917;.005)	1 (.716;.009)	1 (.595;.009)	.999 (.891;.006)	.999 (.697;.009)	.999 (.579;.009)
1000	10	Small	1 (.623;.012)	1 (.325;.01)	1 (.253;.009)	1 (.622;.012)	1 (.325;.01)	1 (.253;.009)	1 (.619;.012)	1 (.323;.01)	1 (.251;.009)	1 (.601;.012)	1 (.314;.01)	1 (.26;.009)
1000	10	Large	1 (.824;.006)	1 (.576;.008)	1 (.486;.008)	1 (.822;.006)	1 (.574;.008)	1 (.484;.008)	1 (.819;.006)	1 (.572;.008)	1 (.483;.008)	1 (.804;.006)	1 (.562;.008)	1 (.474;.008)