Multivariate Behavioral Research, 46:1–32, 2011

ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273171.2011.543027

Psychology Press
Taylor & Francis Group

Accuracy in Parameter Estimation for the Root Mean Square Error of Approximation: Sample Size Planning for Narrow Confidence Intervals

Ken Kelley
University of Notre Dame

Keke Lai
University of Notre Dame

The root mean square error of approximation (RMSEA) is one of the most widely reported measures of misfit/fit in applications of structural equation modeling. When the RMSEA is of interest, so too should be the accompanying confidence interval. A narrow confidence interval reveals that the plausible parameter values are confined to a relatively small range at the specified level of confidence. The accuracy in parameter estimation approach to sample size planning is developed for the RMSEA so that the confidence interval for the population RMSEA will have a width whose expectation is sufficiently narrow. Analytic developments are shown to work well with a Monte Carlo simulation study. Freely available computer software is developed so that the methods discussed can be implemented. The methods are demonstrated for a repeated measures design where the way in which social relationships and initial depression influence coping strategies and later depression are examined.

Structural equation modeling (SEM) is widely used in many disciplines where variables tend to be measured with error and/or latent constructs are hypothesized to exist. The behavioral, educational, and social sciences literature, among others, has seen tremendous growth in the use of SEM in the last decade. The general

Correspondence concerning this article should be addressed to Ken Kelley, Department of Management, Mendoza College of Business, University of Notre Dame, Notre Dame, IN 46545. E-mail: KKelley@ND.edu

goal of SEM is to recover the population covariance matrix, Σ, of k manifest (observed) variables by fitting a theoretical model that describes the relationships among the k measured variables and the specified latent variables.

The root mean square error of approximation (RMSEA; Browne & Cudeck, 1992; Steiger & Lind, 1980) has become one of the most, if not the most, widely used assessment of misfit/fit in the applications of SEM (e.g., Jackson, Gillaspy, & Purc-Stephenson, 2009; Taylor, 2008). Unlike many other fit indices, the RMSEA is used both descriptively (i.e., sample estimates) and inferentially (with confidence intervals and hypothesis tests). The two most important features of the RMSEA are (a) it is a standardized measure not wedded to the scales of the measured or latent variables and (b) its approximate distributional properties are known, which makes it possible to obtain parametric confidence intervals and perform hypothesis tests.

There are two popular ways to use the RMSEA to assess a model's fit. The first method is to examine the point estimate and compare it with a certain fixed cutoff value, say c. In this context, if $\hat{\varepsilon} < c$, the model is considered to have a certain degree of fit (e.g., close fit, mediocre fit, etc.) where ε refers to the population RMSEA and $\hat{\varepsilon}$ refers to the point estimate. The second method is to conduct hypothesis testing to infer if the null hypothesis $H_0 : \varepsilon \geq c$ can be rejected at a specified significance level (e.g., MacCallum, Browne, & Sugawara, 1996). If the null hypothesis is rejected, it is concluded that $\varepsilon < c$ and the model fit is better than the degree of fit corresponding to the cutoff value of the specified null hypothesis. For both of the methods, choosing a proper cutoff value (c) is critically important, and a widely used convention is that $\varepsilon \leq 0.05$ refers to close fit, $\varepsilon \leq 0.08$ mediocre fit, and $\varepsilon > 0.10$ poor fit (see, e.g., Browne & Cudeck, 1992; MacCallum, Browne, & Sugawara, 1996). Besides these two conventional methods, a third way to assess the model fit is to form a confidence interval for the population RMSEA. Instead of comparing $\hat{\varepsilon}$ to c or testing a null hypothesis, a confidence interval for ε is interested in the value of ε itself and thus is not wedded to any cutoff value c.

In applied research, sample estimates almost certainly differ from their corresponding population parameter. A confidence interval acknowledges such uncertainty and provides a range of plausible values for the population parameter at some specified confidence level (e.g., .90, .95, .99). Confidence intervals "quantify our knowledge, or lack thereof, about a parameter" (Hahn & Meeker, 1991, p. 29), and correspondingly we know more, holding everything else constant, about parameters that have narrow confidence intervals as compared with wider confidence intervals. From a scientific perspective, the accuracy of the estimate is of key importance, and in order to facilitate scientific gains by building cumulative knowledge, researchers should work to avoid "embarrassingly large" confidence intervals (Cohen, 1994, p. 1002) so that the accuracy of the parameter estimate is respectable and appropriate for the intended use.

A general approach to sample size planning termed *accuracy in parameter estimation* (AIPE; e.g., Kelley, 2007b, 2007c, 2008; Kelley & Maxwell, 2003; Kelley, Maxwell, & Rausch, 2003; Kelley & Rausch, 2006) permits researchers to obtain a sufficiently narrow confidence interval so that the parameter estimate will have a high degree of expected accuracy at a specified level of confidence.

The AIPE approach to sample size planning is an important alternative or supplement to the traditional power analytic approach (e.g., Cohen, 1988; see Maxwell, Kelley, & Rausch, 2008, for a review and comparison of AIPE and power analysis approaches to sample size planning). In structural equation modeling, planning sample size so that the RMSEA is estimated with a sufficiently narrow confidence interval will facilitate model evaluation and descriptions of the extent to which data is consistent with a specified model. In this article we first briefly review confidence interval formation for the population RMSEA and then develop a method to plan sample size so that the expected confidence interval width for the RMSEA is sufficiently narrow. Our sample size planning method is then evaluated with an extensive Monte Carlo simulation study so that its effectiveness is verified in situations commonly encountered in practice. We then show an example of how our methods can be used in an applied setting. Additionally, we have implemented the sample size planning procedure into R (R Development Core Team, 2010) so that the methods can be readily applied by researchers.

POINT ESTIMATE AND CONFIDENCE INTERVAL FOR THE RMSEA

In this section we briefly review the confidence interval formation for RMSEA and define our notation. Readers who want to implement the sample size planning methods may wish to only browse this section as it is not necessary to fully understand the theory behind confidence interval formation in order to implement the sample size planning methods. Nevertheless, this section is necessary to fully understand the rationale of the methodological developments we make.

Let $\mathbf{\Sigma}$ be the population covariance matrix of k manifest variables and $\mathbf{S}$ be the sample covariance matrix based on N individuals. Further, let $\boldsymbol{\theta}^*$ be a $q \times 1$ vector of potential parameter values for a postulated covariance structure where the q values are each identified. The $k \times k$ model-implied covariance matrix is denoted $\mathbf{M}(\boldsymbol{\theta}^*)$. The model's degrees of freedom, ν, is then $k(k+1)/2 - q$.

For a correctly specified model, there exists a particular $\boldsymbol{\theta}^*$, denoted $\boldsymbol{\theta}$, such that $\mathbf{M}(\boldsymbol{\theta}) = \mathbf{\Sigma}$, where $\boldsymbol{\theta}$ is the $q \times 1$ vector of the population parameters. Of course, $\boldsymbol{\theta}$ is unknown in practice and must be estimated. Estimation of $\boldsymbol{\theta}$ can be done in several ways (e.g., maximum likelihood, generalized least squares, asymptotic distribution-free methods) with the most widely used estimation procedure being normal theory maximum likelihood. We use normal theory maximum likelihood estimation throughout the article and use $\hat{\boldsymbol{\theta}}$ to denote the maximum likelihood estimate of $\boldsymbol{\theta}$. Values for $\hat{\boldsymbol{\theta}}$ can be obtained by minimizing the discrepancy function with respect to $\boldsymbol{\theta}^*$ (e.g., McDonald, 1989; Bollen, 1989),

$$F\left(\mathbf{S}, \mathbf{M}(\boldsymbol{\theta}^*)\right) = \log\left(\left|\mathbf{M}(\boldsymbol{\theta}^*)\right|\right) + tr\left(\mathbf{S}\mathbf{M}(\boldsymbol{\theta}^*)^{-1}\right) - \log|\mathbf{S}| - k, \tag{1}$$

where $tr(\cdot)$ refers to the trace of the matrix and an exponent of -1 denotes the inverse of the matrix. Because $\hat{\boldsymbol{\theta}}$ minimizes Equation 1,

$$\min F\left(\mathbf{S}, \mathbf{M}(\boldsymbol{\theta}^*)\right) = F\left(\mathbf{S}, \mathbf{M}\left(\hat{\boldsymbol{\theta}}\right)\right) \equiv \hat{F}, \tag{2}$$

where $\hat{F}$ is the value of the maximum likelihood discrepancy function evaluated at $\hat{\boldsymbol{\theta}}$. Based on Equation 1, $F\left(\mathbf{S}, \mathbf{M}(\boldsymbol{\theta}^*)\right)$ is zero only when $\mathbf{S}$ equals $\mathbf{M}(\boldsymbol{\theta}^*)$ and increases without bound as $\mathbf{S}$ and $\mathbf{M}(\boldsymbol{\theta}^*)$ become more discrepant.

For a correctly specified model, when the assumptions of independent observations and multivariate normality hold and sample size is not too small, Steiger, Shapiro, and Browne (1985, Theorem 1; see also Browne & Cudeck, 1992) showed that the quantity

$$T = \hat{F} \times (N - 1) \tag{3}$$

approximately follows a central χ^2 distribution with ν degrees of freedom. For an incorrectly specified model there exists no $\boldsymbol{\theta}^*$ such that $\mathbf{M}(\boldsymbol{\theta}^*) = \boldsymbol{\Sigma}$. The discrepancy between the population covariance matrix and the population model-implied covariance matrix can be measured as

$$\min F\left(\boldsymbol{\Sigma}, \mathbf{M}(\boldsymbol{\theta}^*)\right) = F\left(\boldsymbol{\Sigma}, \mathbf{M}(\boldsymbol{\theta}_0)\right) \equiv F_0, \tag{4}$$

where $\boldsymbol{\theta}_0$ is a vector of population model parameters and F_0 is always larger than zero for a misspecified model. For such a misspecified model, when the assumptions of independent observations and multivariate normality hold, N is not too small, and the discrepancy is not too large. Steiger et al. (1985, Theorem 1; see also Browne & Cudeck, 1992) showed that the quantity T from Equation 3 approximately follows a noncentral χ^2 distribution with ν degrees of freedom and noncentrality parameter[1]

$$\text{𐤋} = F_0 \times (N - 1). \tag{5}$$

[1]The symbol 𐤋 introduced in Equation 5 is the Phoenician letter lamd, which was a precursor to the Greek letter "Λ/λ" (lambda) and the Latin letter "L/l" (ell; Powell, 1991). Although λ and Λ are sometimes used to denote the noncentrality parameter of the χ^2 distribution, in general λ and Λ are more often associated with noncentrality parameters from t distributions and F distributions, respectively. Further, we use λ to denote path coefficients in a forthcoming section; the symbol 𐤋 is used for the χ^2 noncentrality parameter to avoid potential confusion.

The population RMSEA is defined as

$$\varepsilon = \sqrt{\frac{F_0}{\nu}}. \tag{6}$$

Although a seemingly reasonable estimate of ε would be based on substituting $\hat{F}$ from Equation 3 for F_0 in Equation 6, $\hat{F}$ is an appreciably biased estimator of F_0 (McDonald, 1989; Wald, 1943). Because $\hat{F}$ approximately follows a noncentral χ^2 distribution, the expected value of $\hat{F}$ is approximately[2]

$$\mathrm{E}[\hat{F}] = F_0 + \frac{\nu}{N-1}. \tag{7}$$

Thus, a better estimate of F_0 is

$$\hat{F}_0 = \hat{F} - \frac{\nu}{N-1}. \tag{8}$$

However, it is possible for $\frac{\nu}{N-1}$ to be larger than $\hat{F}$, implying a negative value for $\hat{F}_0$, which should always be a nonnegative value (e.g., Browne, 1984). Recognizing that $\hat{F}_0$ may be negative, the estimated RMSEA is defined as

$$\hat{\varepsilon} = \sqrt{\max\left\{0, \frac{\hat{F}_0}{\nu}\right\}}. \tag{9}$$

Technically, $\hat{\varepsilon}$ as defined in Equation 9 cannot be an unbiased estimate due to possible truncation correcting for the possibility of negative values. Although not unbiased, Equation 9 is an optimal estimator of ε (see Steiger, 2000, for a review of estimation and inference regarding the RMSEA). However, point estimates in and of themselves do not convey the uncertainty associated with their use as estimates of population quantities. Explicitly acknowledging the uncertainty of the estimate is the reason it is widely recommended in the methodological literature to report confidence intervals for ε in addition to $\hat{\varepsilon}$.

Steiger and Lind (1980) presented the idea of forming confidence intervals for ε, which was later formalized in Steiger (1989) and Browne and Cudeck (1992). Confidence intervals for ε can be obtained by using the *inversion confidence interval principle* and the *confidence interval transformation principle*, both of which have been discussed extensively elsewhere (Cumming & Finch, 2001; Fleishman, 1980; Kelley, 2007a; Smithson, 2001; Steiger, 1989, 2000; Steiger

[2]The expectation of a $\chi^2(\nu, \lambda)$ variate is $\nu + \lambda$, where ν is the degrees of freedom and λ is the noncentrality parameter. In the present case $(N-1)\hat{F} \sim \chi^2(\nu, \lambda)$, and thus $\mathrm{E}\left[(N-1)\hat{F}\right] = \nu + (N-1)F_0$ and Equation 7 follows.

& Fouladi, 1997). In order to form a confidence interval for ε, one can first form a confidence interval for Λ and then transform the confidence limits onto the scale of ε because there is a one-to-one monotonic relationship between Λ and ε. The monotonic relationship can be shown by rewriting ε in terms of Λ:

$$\varepsilon = \sqrt{\frac{\Lambda}{\nu(N-1)}}. \tag{10}$$

Also note that confidence interval formation for ε only depends on summary statistics, and thus it does not require raw data. A supplement for this article is available at https://repository.library.nd.edu/view/1/AIPE_RMSEA_MBR_Supplement.pdf and it discusses how confidence intervals for ε can be formed, as well as the other methods we discuss, with the freely available MBESS (Kelley, 2007a, 2007b; Kelley & Lai, 2010) R (R Development Core Team, 2010) package.

Unfortunately, the confidence interval for ε does not always work well, in the sense that the nominal confidence interval coverage can differ from the empirical confidence interval coverage in certain situations. Curran, Bollen, Chen, Paxton, and Kirby (2003) found that for some misspecified models the nominal confidence interval coverage rate performed poorly. For example, they showed that when $\varepsilon = .25$ and $N = 800$, the empirical confidence interval coverage rate was 76% even though the nominal confidence interval coverage was 90%. Although the situation noted was based on a rather extreme situation (i.e., where the ε value was large for a very simple model), concern of the possible failure of the confidence interval procedure was part of the reason it is necessary to perform a Monte Carlo simulation study to evaluate the effectiveness of the sample size planning method we develop, which we do in a forthcoming section.

AIPE FOR THE RMSEA

Accuracy is a function of bias and precision so that obtaining a more precise estimate without increasing bias implies that the estimate is more accurate. We use the term *accuracy in parameter estimation* as a way to describe this approach to sample size planning because precision is being improved without increasing bias.[3] Let ω be the desired width of the confidence interval for ε and w be the confidence interval width obtained in a particular situation. Thus ω is

[3]The RMSEA is not unbiased, but its bias decreases as sample size is increased. Correspondingly, not only is the precision with which ε is estimated improved when sample size is increased but also the bias is reduced.

a constant, specified a priori, whereas w is a random variable, varying across different samples. We seek to develop a method where the sample size can be planned so that the expected value of the observed confidence interval width is sufficiently narrow. In doing so, ν, the confidence level (i.e., $1-\alpha$), and a value for ε must be specified, which we denote ε^*. Given a specific model of interest, the value of ν is known a priori and fixed, and $1-\alpha$ is a constant set by the researcher. However, ε is an unknown parameter that is of interest. Our sample size approach is based on the assumption that $\varepsilon^* = \varepsilon$. However, multiple reasonable estimates of ε can be used so that the researcher can evaluate the necessary sample sizes based on different ε^* values. Because the knowledge of some (in general unknown) population parameters is necessary to calculate N, as is generally the case of any formalized sample size planning methods, researchers need to note that the sample size returned should be viewed as approximations, conditional on the appropriateness of input information. Some ideas for choosing reasonable estimates of ε are (a) basing it on prior research conducted with a similar population, (b) meta-analyses, or (c) norms for what is considered "close" fit (e.g., $\varepsilon^* = .05$) or "mediocre" fit (e.g., $\varepsilon^* = .08$; e.g., MacCallum et al., 1996). Furthermore, we believe using several values of ε^* and ω, and in some cases multiple degrees of freedom values, is helpful to understand the impact of each component on the necessary sample size.

In order to plan sample size so that the expected confidence confidence interval width is sufficiently narrow, we first seek to find the expected value of $\hat{\varepsilon}$, denoted $E[\hat{\varepsilon}]$, and compute the confidence interval width for ε given $E[\hat{\varepsilon}]$, ν, N, and $1-\alpha$. Given ν, N, and $1-\alpha$, all of which are design factors, the value of w is solely determined by $\hat{\varepsilon}$. We regard the confidence interval width at $E[\hat{\varepsilon}]$ with the particular combination of ν, N, $1-\alpha$ as the expected confidence interval width for ε. That is, the expected confidence interval width is $E[w|(E[\hat{\varepsilon}], \nu, N, 1-\alpha)]$, which for simplicity we write as $E[w]$, with the implicit understanding that $E[w]$ necessarily depends on $E[\hat{\varepsilon}]$, ν, N, and the confidence level.[4]

Given ε, ν, and N, from Equation 9, the expectation of $\hat{\varepsilon}$ can be written as

$$E[\hat{\varepsilon}] = E\left[\sqrt{\max\left\{0, \frac{\hat{F}_0}{\nu}\right\}}\right]. \tag{11}$$

[4]Given ν, N, and $1-\alpha$, the value of w is solely determined by $\hat{\varepsilon}$, and we use $w = h(\hat{\varepsilon})$ to denote such relationship, where $h(\cdot)$ refers to a nonlinear confidence interval formation function (i.e., the confidence interval formation function discussed previously). Using the Taylor expansion to expand $h(\hat{\varepsilon})$ at $E[\hat{\varepsilon}]$, one would obtain that $h(\hat{\varepsilon}) = h(E[\hat{\varepsilon}]) +$ remainder (e.g., Casella & Berger, 2002, p. 241). Taking the expectation on both sides of the equation, it leads to $E[h(\hat{\varepsilon})] \approx E[h(E[\hat{\varepsilon}])]$. Because $E[h(\hat{\varepsilon})] = E[w]$ according to the definition of w and $E[\hat{\varepsilon}]$ is a constant, it follows that $E[w] \approx h(E[\hat{\varepsilon}])$, meaning the expectation of w is approximately equal to the confidence interval width obtained based on $E[\hat{\varepsilon}]$.

If $\hat{F}_0$ is negative, the expectation in Equation 11 is simply zero. Thus, the focus of the derivation is the expectation of $\sqrt{\frac{\hat{F}_0}{\nu}}$. In situations where $\hat{F}_0$ is larger than zero, which is usually the case, Equation 11 becomes

$$\mathrm{E}[\hat{\varepsilon}] = \mathrm{E}\left[\sqrt{\frac{\hat{F}_0}{\nu}}\right]. \tag{12}$$

Because ν is a constant given a specific model, we can move the expectation calculation to the numerator of the right-hand side of the equation:

$$\mathrm{E}[\hat{\varepsilon}] = \frac{1}{\sqrt{\nu}}\mathrm{E}\left[\sqrt{\hat{F}_0}\right]. \tag{13}$$

In general, the expectation of a variable's function does not equal the function of the variable's expectation. Nevertheless, applying the Taylor expansion, one can show that $\mathrm{E}[g(\hat{F}_0)] \approx g\left(\mathrm{E}[\hat{F}_0]\right)$, where $g(\cdot)$ refers to a differentiable function of $\hat{F}_0$ under some fairly general conditions (e.g., Casella & Berger, 2002, p. 241). In the present context, $g\left(\hat{F}_0\right)$ is $\sqrt{\hat{F}_0}$. Because the expectation of $\hat{F}_0$ is F_0 (by combining Equations 7 and 8), Equation 13 becomes

$$\mathrm{E}[\hat{\varepsilon}] \approx \sqrt{\frac{F_0}{\nu}}, \tag{14}$$

with the right-hand side being equal to the definition of ε from Equation 6. Thus,

$$\mathrm{E}[\hat{\varepsilon}] \approx \varepsilon, \tag{15}$$

with the bias decreasing as ν decreases and/or N increases.

As previously discussed, the confidence interval width is a function of (a) the $\hat{\varepsilon}$ value, (b) degrees of freedom, (c) confidence level, and (d) sample size. If a, b, and c were known, the confidence interval width would be determined solely by the sample size. More specifically, the sample size planning procedure aims to find the smallest value for N such that, for a specified value of ε (i.e., ε^*), ν, and α,

$$\mathrm{E}_N[w] \leq \omega, \tag{16}$$

where ω is the desired confidence interval width specified by the researcher, and the subscript N on the expectation emphasizes that the expectation of w is determined by N, the only factor that varies in the sample size planning procedure.

Algorithm to Obtain the Necessary Sample Size

Given the previous discussion, we now discuss the algorithm for obtaining the necessary sample size. The way in which the sample size planning procedure begins is at an initial sample size value, say N_0, such that $\mathrm{E}_{N_0}[w] > \omega$ (i.e., at a sample size where the expected confidence interval width is wider than desired). The next step is to increase the sample size by 1 and then evaluate the expected confidence interval width at a sample size of $N_{(0+1)}$. If $\mathrm{E}_{N_{(0+1)}}[w] \leq \omega$, the procedure stops and $N_{(0+1)}$ is the necessary sample size. If, however, $\mathrm{E}_{N_{(0+1)}}[w] > \omega$, the sample size is increased by 1 and a check is performed, as before, to determine if the sample size leads to an expected confidence interval width that is sufficiently narrow. This process continues until the sample size yields an expected confidence interval width that is sufficiently narrow, that is, an iterative process of evaluating $\mathrm{E}_{N_{(0+i)}}[w]$ to determine at the minimum sample size where the expected confidence interval width is less than or equal to ω, where i in the subscript denotes the particular iteration and where $N_{(0+i)}$ is the sample size for the particular iteration. A supplement that supports this article is available at https://repository.library.nd.edu/view/1/AIPE_RMSEA_MBR_Supplement.pdf and provides information on how the necessary sample size can be easily planned using the MBESS R package.

MONTE CARLO SIMULATION STUDY

Recall that Equation 15, which the sample size planning procedure depends on, is an approximation. However, empirical support for Equation 15 exists. Curran et al. (2003) were able to show in a variety of situations that when $N \geq 200$, $\mathrm{E}[\hat{\varepsilon}] \approx \varepsilon$. Thus, Curran et al.'s finding is an empirical demonstration of the effectiveness of Equation 15 as an approximation, at least in certain situations. Additionally, the confidence interval procedure itself has been shown at times not to work well (Curran et al., 2003). The impact of the approximation in Equation 15, the lack of an exact confidence interval procedure in all cases, and the possible truncation of the lower confidence interval limit at zero was not clear. Correspondingly, a Monte Carlo simulation study was needed to verify the effectiveness of our proposed procedures in realistic situations.

The simulation study was conducted in the context of four models representative of applied research: (a) a confirmatory factor analysis (CFA) model (Model 1) based on Holzinger & Swineford (1939), (b) an autoregressive model (Model 2) based on Curran et al. (2003), (c) a complex SEM model (Model 3) based on Maruyama and McGarvey (1980), and (d) a more complex SEM model (Model 4). Path diagrams and model parameters are provided in Figures 1 to 4. Model 4 extends Model 3 in that (a) reciprocal relationship between endogenous variables

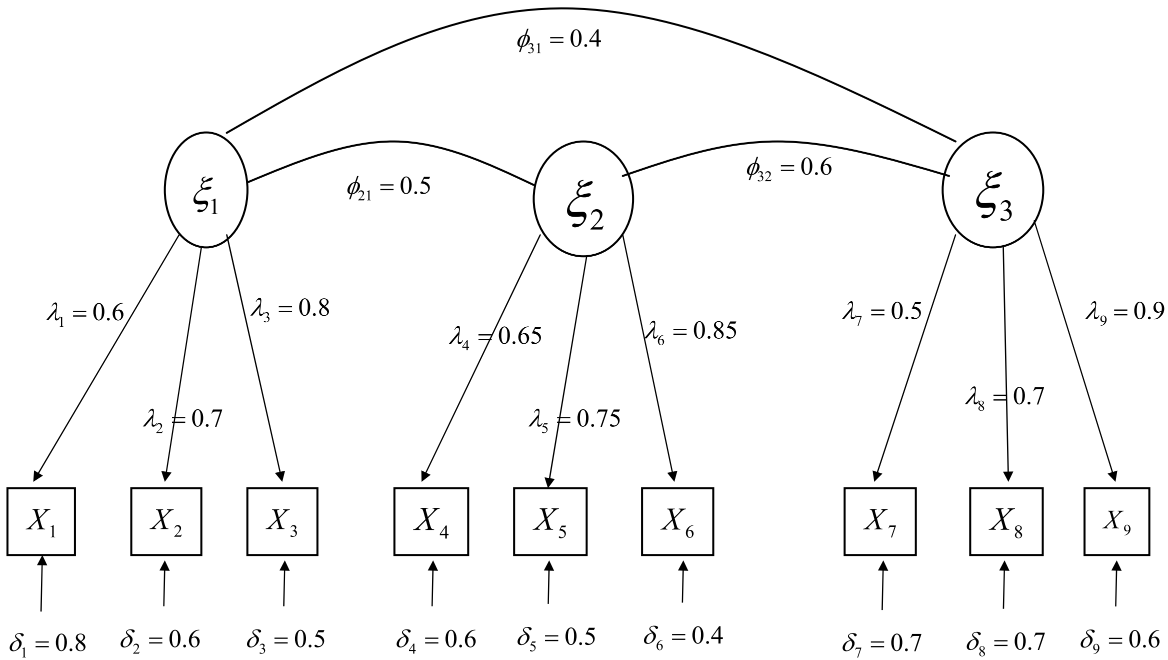

FIGURE 1 Population Model 1 used in the Monte Carlo simulation study. This is a confirmatory factor analysis model, where each of the three latent variables is measured by three manifest variables. The variances of the latent variables are fixed at 1. Numbers shown are population unstandardized model parameters. *Note*. ξ represents a factor, ϕ a covariance, λ a factor loading, X a measured variable, and δ the error variance of a measured variable.

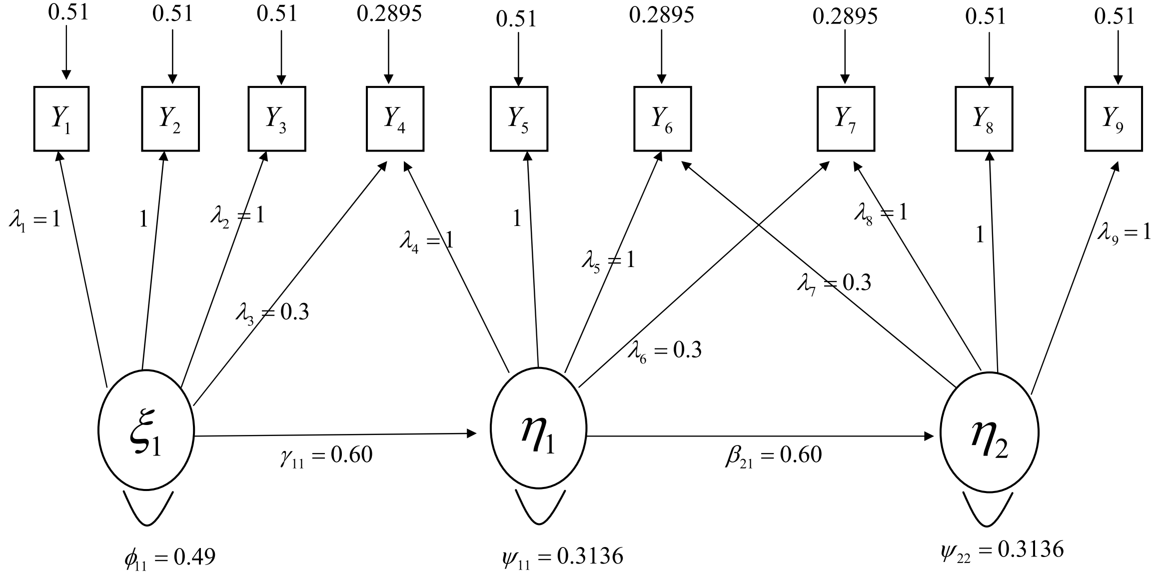

FIGURE 2 Population Model 2 used in the Monte Carlo simulation study. This structural equation model contains three latent variables and nine manifest variables. Some manifest variables serve as indicators for more than one latent variable. Numbers shown are population unstandardized model parameters. *Note*. ξ and η represent latent variables, ϕ and ψ the variances of latent variables, γ and β structural coefficients, λ a factor loading, and Y a measured variable.

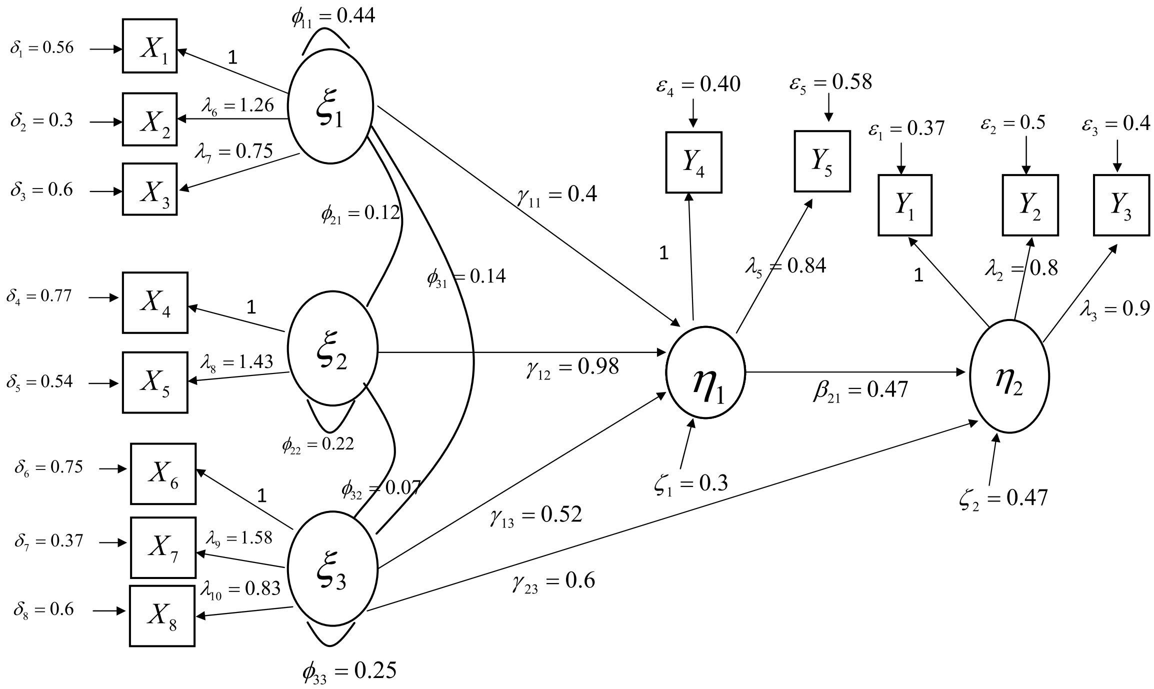

FIGURE 3 Population Model 3 used in the Monte Carlo simulation study. This structural equation model contains three exogenous latent variables and two endogenous latent variables. There is a mediation effect among three latent variables. Numbers shown are population unstandardized model parameters. *Note.* ξ and η represent latent variables, ϕ a variance or covariance, λ a factor loading, X and Y measured variables, and δ and ε the error variances of measured variables.

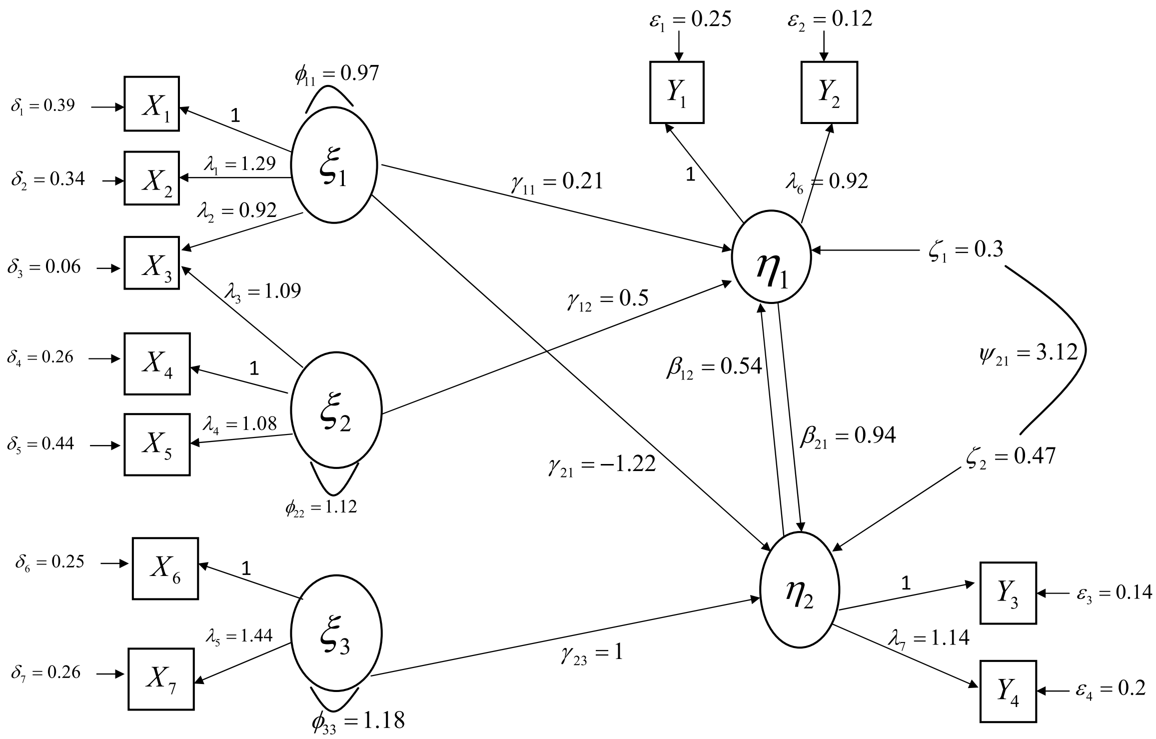

FIGURE 4 Population Model 4 used in the Monte Carlo simulation study. This structural equation model contains three exogenous latent variables and two endogenous latent variables. There is a mediation effect and a reciprocal effect among the latent variables. Numbers shown are population unstandardized model parameters. *Note.* ξ and η represent latent variables, ϕ and ψ variances of latent variables, ζ the error variance of an endogenous variable, ψ the covariance between the errors for endogenous variables, γ and β structural coefficients, λ a factor loading, X and Y measured variables, and δ and ε the error variances of measured variables.

(i.e., η_1 and η_2) is included, (b) errors for the endogenous variables are allowed to covary, and (c) a manifest variable is included in more than one measurement model for the latent variables. Some modifications of both the original model and/or parameters were made so that these models are more generally applicable. All simulation and analysis was conducted in R. The MASS package (Venables & Ripley, 2002, 2010) was used to generate the multivariate normal data, the models were fitted with the sem (Fox, 2006) package, and the MBESS (Kelley, 2007a, 2007b; Kelley & Lai, 2010) package was used to obtain a confidence interval for the population RMSEA.

After specifying the (true) model and (true) model parameters, the model implied covariance matrix, $\mathbf{M}(\boldsymbol{\theta})$, can be obtained, and it is used as the population covariance matrix of the manifest variables (i.e., $\boldsymbol{\Sigma} = \mathbf{M}(\boldsymbol{\theta})$; e.g., Davey & Savla, 2010, Chapter 4) to generate random data. Then the models are intentionally misspecified to different extents.[5] Each model is misspecified in three ways to create three models with different values of ε. In particular, Model 1 is misspecified in the manner that (a) $\lambda_4 = \lambda_1 = \lambda_7$, $\lambda_5 = \lambda_2 = \lambda_8$, $\lambda_6 = \lambda_3 = \lambda_9$; (b) the specifications in (a) plus $\delta_1 = \delta_4 = \delta_7$, $\delta_2 = \delta_5 = \delta_8$, $\delta_3 = \delta_6 = \delta_9$; and (c) the specification in (a) plus all δ's being equal and $\phi_{21} = 0$, so that the population RMSEA values are 0.0242, 0.0526, and 0.123, respectively. Model 2 is misspecified in the manner that (a) $\lambda_6 = 0$; (b) $\lambda_3 = \lambda_6 = \lambda_7 = 0$; and (c) $\lambda_3 = \lambda_6 = \lambda_7 = 0$, $\delta_1 = \delta_2 = \delta_3$, $\delta_4 = \delta_5 = \delta_6$, $\delta_7 = \delta_8 = \delta_9$, so that the population RMSEA values are 0.0268, 0.0614, and 0.0865, respectively. Model 3 is misspecified in the manner that (a) $\beta_{21} = 0$ and ψ_{21} added (i.e., the path from η_1 to η_2 is replaced by the covariance between the two); (b) $\beta_{21} = \phi_{21} = \phi_{32} = 0$ and ψ_{21} and γ_{21} added; and (c) $\beta_{21} = \gamma_{11} = \gamma_{12} = \gamma_{23} = 0$, $\phi_{21} = \phi_{31} = \phi_{32} = 0$ and β_{12}, γ_{21}, and γ_{22} added in the model, so that the population RMSEA values are 0.028, 0.0435, and 0.09, respectively. Model 4 is misspecified in the manner that (a) $\beta_{21} = \beta_{12} = 0$; (b) $\psi_{21} = 0$; and (c) errors for X's all being equal, $\varepsilon_1 = \varepsilon_2$, $\varepsilon_3 = \varepsilon_4$, and $\psi_{21} = 0$, so that the population RMSEA values are 0.0386, 0.0525, and 0.0833, respectively. The desired confidence interval width for ε is set from .01 to .05 with an increment of .01. Given ε, ω, and the model degrees of freedom, the necessary sample size N is returned by the sample size planning procedure as discussed in the previous section and as illustrated in the supplement available at https://repository.library.nd.edu/view/1/AIPE_RMSEA_MBR_Supplement.pdf. A random sample of size N is generated from a particular multivariate normal distribution with population

[5] Another way to create misspecified models is to use the Cudeck-Browne procedure (Cudeck & Browne, 1992). This procedure is implemented in the function `Sigma.2.SigmaStar()` in the MBESS package. See the supplement available at https://repository.library.nd.edu/view/1/AIPE_RMSEA_MBR_Supplement.pdf for detailed documentation of this function and its possible use when performing Monte Carlo simulations.

covariance matrix $\mathbf{\Sigma}$, and then it is used to fit the (misspecified) model and obtain a 95% confidence interval for ε. The results from each condition are based on 5,000 replications. All programming scripts and code are available from the authors upon request.

Simulation results are reported in detail in Tables 1 to 4 where (a) "M_w" and "Mdn_w" refer to the mean and median of the 5,000 random confidence interval widths, respectively; (b) "P_{99}," "P_{97}," "P_{95}," "P_{90}," "P_{80}," and "P_{70}" refer to the respective percentiles of the random widths; and (c) "α_{up}," "α_{low}," and "α" refer to the empirical Type I error rate on the upper tail, lower tail, and both tails, respectively.

Results show that the method is effective at producing confidence intervals whose mean (and median) value is ω, which is the goal of the sample size planning procedure. In the worst cases, the mean of the random w's exceeds ω by only a trivial amount. For example, M_w is 0.0556 in Table 2 when $\varepsilon = .0526$, $\omega = .05$, and $N = 317$, exceeding the desired confidence interval width by about only .006; $M_w = 0.0562$ in Table 3 when $\varepsilon = .0903$, $\omega = .05$, and $N = 146$, again exceeding the desired value by about only 0.006. For another, M_w is 0.0534 at $N = 191$, given $\varepsilon = .0834$ and $\nu = 45$ (Table 4), exceeding the desired value $\omega = .05$ by just 0.0034. As sample size becomes larger, the discrepancy between empirical mean and ω further reduces. In practice, obtaining a confidence interval that exceeds the desired width by such a small amount is very unlikely to produce any substantive impact. Thus, the procedure is effective at accomplishing its stated goal of planning sample size so that the expected confidence interval width is sufficiently narrow.

In addition, in some cases, given the same ε and ν, the sample size planning procedure returns the same N for different ω values. In particular, $N = 1{,}263$ for both $\omega = .03$ and .04 (Table 2; $\varepsilon = .0268$, $\nu = 23$); $N = 1{,}107$ for both $\omega = .02$ and .03 (Table 3; $\varepsilon = .028, \nu = 58$); $N = 458$ for both $\omega = .04$ and .05 (Table 3; $\varepsilon = .0436, \nu = 58$); $N = 586$ for both $\omega = .04$ and .05 (Table 4; $\varepsilon = .0386, \nu = 38$). This somewhat paradoxical phenomenon can be explained by the truncation of the lower confidence limit when it is theoretically below zero. To better understand this, it is helpful to plot the width of the 95% confidence interval for ε as a function of $\hat{\varepsilon}$, N, and ν. Figure 5 shows the 95% confidence interval width for ε given N (for 250 and 1,000) and ν (for 10 and 25) at different values of $\hat{\varepsilon}$ (for values between 0.00 and 0.15). What is immediately obvious from Figure 5 is the nonmonotonic relationship between $\hat{\varepsilon}$ and w. Theoretically, there is a monotonic decreasing relationship between w and $\hat{\varepsilon}$, but due to truncation of the lower confidence limit when it would otherwise extend below zero, the nonmonotonic and discontinuous relationship develops. Nevertheless, after the break in the functional relationship between w and $\hat{\varepsilon}$, it is clear that larger values of $\hat{\varepsilon}$ lead to narrower confidence intervals, holding constant N, ν, and the confidence level.

TABLE 1
Empirical Distributions of Confidence Interval for RMSEA in the Context of Misspecified Model 1

	$\varepsilon = 0.0242 \quad \nu = 30$				
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	5,858	1,901	1,504	894	421
M_w	0.0100	0.0224	0.0268	0.0360	0.0497
Mdn_w	0.0100	0.0201	0.0237	0.0362	0.0488
P_{99}	0.0105	0.0315	0.0354	0.0459	0.0661
P_{95}	0.0103	0.0309	0.0349	0.0449	0.0632
P_{90}	0.0102	0.0302	0.0342	0.0439	0.0606
P_{80}	0.0101	0.0282	0.0327	0.0419	0.0571
P_{70}	0.0101	0.0242	0.0311	0.0400	0.0541
P_{60}	0.0100	0.0206	0.0287	0.0382	0.0516
α_{up}	0.021	0.024	0.027	0.00	0.00
α_{low}	0.025	0.024	0.026	0.028	0.026
α	0.046	0.049	0.053	0.028	0.026
	$\varepsilon = 0.0526 \quad \nu = 36$				
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	4,439	1,224	617	398	317
M_w	0.0100	0.0201	0.0310	0.0460	0.0556
Mdn_w	0.0100	0.0200	0.0300	0.0401	0.0497
P_{99}	0.0101	0.0210	0.0527	0.0669	0.0749
P_{95}	0.0100	0.0206	0.0332	0.0657	0.0738
P_{90}	0.0100	0.0204	0.0322	0.0640	0.0724
P_{80}	0.0100	0.0203	0.0312	0.0602	0.0696
P_{70}	0.0100	0.0202	0.0307	0.0507	0.0663
P_{60}	0.0100	0.0201	0.0303	0.0411	0.0615
α_{up}	0.030	0.027	0.026	0.025	0.023
α_{low}	0.028	0.024	0.028	0.026	0.031
α	0.059	0.052	0.054	0.052	0.055
	$\varepsilon = 0.1232 \quad \nu = 39$				
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	3,973	1,018	470	277	187
M_w	0.0100	0.0200	0.0300	0.0400	0.0501
Mdn_w	0.0100	0.0200	0.0300	0.0400	0.0499
P_{99}	0.0100	0.0201	0.0304	0.0412	0.0531
P_{95}	0.0100	0.0200	0.0302	0.0407	0.0518
P_{90}	0.0100	0.0200	0.0302	0.0405	0.0513
P_{80}	0.0100	0.0200	0.0301	0.0403	0.0507
P_{70}	0.0100	0.0200	0.0300	0.0402	0.0504
P_{60}	0.0100	0.0200	0.0300	0.0401	0.0502
α_{up}	0.024	0.025	0.032	0.023	0.029
α_{low}	0.025	0.026	0.026	0.027	0.029
α	0.049	0.051	0.058	0.050	0.058

Note. ε refers to the population RMSEA, ν the model degrees of freedom, ω the desired expected confidence interval width, N the sample size calculated with our sample size planning method. M_w and Mdn_w refer to the mean and median of the 5,000 random confidence interval widths, respectively. P_{99}, P_{97}, P_{95}, P_{90}, P_{80}, and P_{70} refer to the respective percentiles of the random widths. α_{up}, α_{low}, and α refer to the empirical Type I error rate on the upper tail, lower tail, and both tails, respectively.

TABLE 2
Empirical Distributions of Confidence Interval for RMSEA in the Context of Misspecified Model 2

			$\varepsilon = 0.0268 \quad \nu = 23$		
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	7,301	2,216	1,263	1,263	572
M_w	0.0100	0.0211	0.0316		0.0472
Mdn_w	0.0100	0.0201	0.0293		0.0466
P_{99}	0.0103	0.0306	0.0407		0.0603
P_{97}	0.0102	0.0301	0.0403		0.0594
P_{95}	0.0102	0.0294	0.0399		0.0587
P_{90}	0.0101	0.0265	0.0390		0.0570
P_{80}	0.0101	0.0214	0.0374		0.0543
P_{70}	0.0101	0.0208	0.0355		0.0517
α_{up}	0.026	0.024	0.028		0.0
α_{low}	0.032	0.030	0.033		0.024
α	0.057	0.054	0.060		0.024
			$\varepsilon = 0.0614 \quad \nu = 25$		
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	6,273	1,657	797	491	347
M_w	0.0100	0.0200	0.0302	0.0414	0.0542
Mdn_w	0.0100	0.0200	0.0300	0.0401	0.0502
P_{99}	0.0100	0.0204	0.0326	0.0633	0.0762
P_{97}	0.0100	0.0203	0.0318	0.0603	0.0755
P_{95}	0.0100	0.0203	0.0315	0.0527	0.0746
P_{90}	0.0100	0.0202	0.0310	0.0436	0.0722
P_{80}	0.0100	0.0201	0.0306	0.0419	0.0636
P_{70}	0.0100	0.0201	0.0304	0.0411	0.0527
α_{up}	0.025	0.027	0.030	0.029	0.027
α_{low}	0.028	0.026	0.029	0.035	0.030
α	0.053	0.053	0.059	0.063	0.057
			$\varepsilon = 0.0865 \quad \nu = 31$		
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	5,022	1,303	611	367	253
M_w	0.0100	0.0200	0.0301	0.0404	0.0509
Mdn_w	0.0100	0.0200	0.0300	0.0401	0.0501
P_{99}	0.0100	0.0202	0.0335	0.0432	0.0802
P_{97}	0.0100	0.0201	0.0335	0.0432	0.0552
P_{95}	0.0100	0.0201	0.0307	0.0432	0.0540
P_{90}	0.0100	0.0201	0.0305	0.0418	0.0527
P_{80}	0.0100	0.0200	0.0303	0.0410	0.0521
P_{70}	0.0100	0.0200	0.0302	0.0406	0.0514
α_{up}	0.021	0.021	0.047	0.074	0.098
α_{low}	0.019	0.019	0.017	0.020	0.024
α	0.040	0.040	0.065	0.094	0.122

Note. ε refers to the population RMSEA, ν the model degrees of freedom, ω the desired expected confidence interval width, N the sample size calculated with our sample size planning method. M_w and Mdn_w refer to the mean and median of the 5,000 random confidence interval widths, respectively. P_{99}, P_{97}, P_{95}, P_{90}, P_{80}, and P_{70} refer to the respective percentiles of the random widths. α_{up}, α_{low}, and α refer to the empirical Type I error rate on the upper tail, lower tail, and both tails, respectively. Additionally, simulations were not conducted for the situations where $\varepsilon = 0.0268$, $\nu = 23$, and $\omega = .04$ because the required sample size was the same as the preceding condition when ω was .01 units smaller.

TABLE 3
Empirical Distributions of Confidence Interval for RMSEA in the Context of Misspecified Model 3

			$\varepsilon = 0.0280$ $\nu = 58$		
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	3,181	1,107	1,107	779	316
M_w	0.0100	0.0265		0.0351	0.0500
Mdn_w	0.0100	0.0268		0.0372	0.0498
P_{99}	0.0106	0.0374		0.0446	0.0679
P_{97}	0.0104	0.0372		0.0443	0.0655
P_{95}	0.0104	0.0370		0.0439	0.0640
P_{90}	0.0103	0.0364		0.0431	0.0615
P_{80}	0.0102	0.0351		0.0415	0.0574
P_{70}	0.0101	0.0337		0.0402	0.0546
α_{up}	0.028	0.025		0.024	0.0
α_{low}	0.029	0.029		0.026	0.011
α	0.057	0.054		0.050	0.011
			$\varepsilon = 0.0436$ $\nu = 58$		
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	2,889	867	471	458	458
M_w	0.0100	0.0203	0.0395	0.0408	
Mdn_w	0.0100	0.0200	0.0300	0.0306	
P_{99}	0.0101	0.0391	0.0574	0.0582	
P_{97}	0.0101	0.0216	0.0570	0.0578	
P_{95}	0.0101	0.0213	0.0566	0.0575	
P_{90}	0.0101	0.0210	0.0556	0.0566	
P_{80}	0.0101	0.0206	0.0537	0.0546	
P_{70}	0.0100	0.0203	0.0512	0.0521	
α_{up}	0.019	0.019	0.018	0.023	
α_{low}	0.025	0.028	0.024	0.026	
α	0.044	0.047	0.042	0.048	
			$\varepsilon = 0.0903$ $\nu = 61$		
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	2,580	687	333	207	146
M_w	0.0100	0.0200	0.0300	0.0404	0.0562
Mdn_w	0.0100	0.0200	0.0300	0.0398	0.0497
P_{99}	0.0100	0.0203	0.0313	0.0444	0.1025
P_{97}	0.0100	0.0202	0.0310	0.0431	0.1012
P_{95}	0.0100	0.0202	0.0308	0.0424	0.0997
P_{90}	0.0100	0.0201	0.0306	0.0416	0.0950
P_{80}	0.0100	0.0201	0.0304	0.0409	0.0521
P_{70}	0.0100	0.0200	0.0302	0.0405	0.0511
α_{up}	0.022	0.026	0.021	0.025	0.021
α_{low}	0.022	0.026	0.024	0.029	0.037
α	0.044	0.052	0.045	0.054	0.058

Note. ε refers to the population RMSEA, ν the model degrees of freedom, ω the desired expected confidence interval width, N the sample size calculated with our sample size planning method. M_w and Mdn_w refer to the mean and median of the 5,000 random confidence interval widths, respectively. P_{99}, P_{97}, P_{95}, P_{90}, P_{80}, and P_{70} refer to the respective percentiles of the random widths. α_{up}, α_{low}, and α refer to the empirical Type I error rate on the upper tail, lower tail, and both tails, respectively. Additionally, simulations were not conducted for the situations where $\varepsilon = 0.0280$, $\nu = 58$, and $\omega = .03$ and $\varepsilon = 0.0436$, $\nu = 58$, and $\omega = .05$ because the required sample size was the same as the preceding conditions when ω was .01 units smaller.

TABLE 4
Empirical Distributions of Confidence Interval for RMSEA
in the Context of Misspecified Model 4

	$\varepsilon = 0.0386$ $\nu = 38$				
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	4,351	1,280	687	586	586
M_w	0.0100	0.0203	0.0351	0.0403	
Mdn_w	0.0100	0.0200	0.0301	0.0348	
P_{99}	0.0101	0.0227	0.0503	0.0546	
P_{97}	0.0101	0.0219	0.0500	0.0542	
P_{95}	0.0101	0.0216	0.0496	0.0539	
P_{90}	0.0101	0.0211	0.0485	0.0530	
P_{80}	0.0100	0.0206	0.0459	0.0508	
P_{70}	0.0100	0.0204	0.0420	0.0483	
α_{up}	0.022	0.024	0.021	0.023	
α_{low}	0.023	0.024	0.024	0.030	
α	0.045	0.048	0.046	0.053	
	$\varepsilon = 0.0526$ $\nu = 37$				
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	4,324	1,195	604	390	317
M_w	0.0100	0.0201	0.0310	0.0460	0.0546
Mdn_w	0.0100	0.0200	0.0300	0.0400	0.0461
P_{99}	0.0101	0.0209	0.0532	0.0673	0.0746
P_{97}	0.0100	0.0207	0.0498	0.0668	0.0742
P_{95}	0.0100	0.0206	0.0334	0.0663	0.0737
P_{90}	0.0100	0.0204	0.0322	0.0650	0.0722
P_{80}	0.0100	0.0203	0.0312	0.0606	0.0693
P_{70}	0.0100	0.0202	0.0307	0.0503	0.0656
α_{up}	0.022	0.021	0.031	0.031	0.030
α_{low}	0.024	0.026	0.025	0.024	0.027
α	0.046	0.047	0.056	0.055	0.058
	$\varepsilon = 0.0834$ $\nu = 45$				
ω	*0.01*	*0.02*	*0.03*	*0.04*	*0.05*
N	3,485	921	442	272	191
M_w	0.0100	0.0201	0.0301	0.0404	0.0534
Mdn_w	0.0100	0.0200	0.0300	0.0400	0.0499
P_{99}	0.0100	0.0243	0.0350	0.0447	0.0925
P_{97}	0.0100	0.0202	0.0311	0.0440	0.0906
P_{95}	0.0100	0.0202	0.0309	0.0430	0.0891
P_{90}	0.0100	0.0201	0.0306	0.0420	0.0538
P_{80}	0.0100	0.0201	0.0304	0.0410	0.0524
P_{70}	0.0100	0.0200	0.0302	0.0406	0.0514
α_{up}	0.024	0.032	0.030	0.038	0.037
α_{low}	0.019	0.020	0.023	0.027	0.024
α	0.043	0.052	0.053	0.066	0.060

Note. ε refers to the population RMSEA, ν the model degrees of freedom, ω the desired expected confidence interval width, N the sample size calculated with our sample size planning method. M_w and Mdn_w refer to the mean and median of the 5,000 random confidence interval widths, respectively. P_{99}, P_{97}, P_{95}, P_{90}, P_{80}, and P_{70} refer to the respective percentiles of the random widths. α_{up}, α_{low}, and α refer to the empirical Type I error rate on the upper tail, lower tail, and both tails, respectively. Additionally, simulations were not conducted for the situation where $\varepsilon = 0.0386$, $\nu = 38$, and $\omega = .05$ because the required sample size was the same as the preceding condition when ω was .01 units smaller.

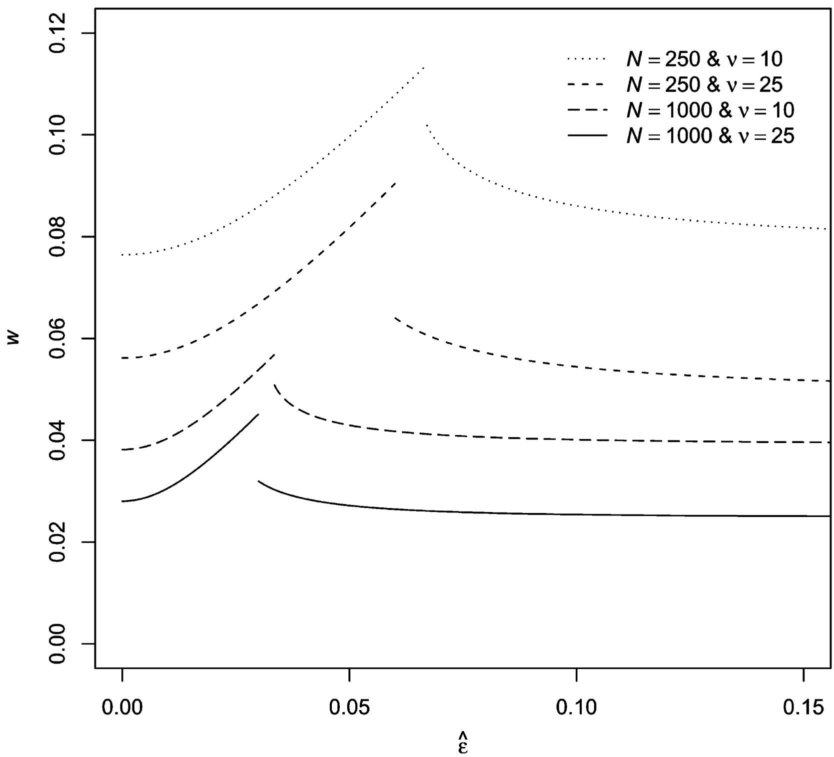

FIGURE 5 Width of the 95% confidence interval for ε as a function of $\hat{\varepsilon}$, N, and ν. *Note.* w is the observed confidence interval width, $\hat{\varepsilon}$ is the estimate of the population value of the root mean square error of approximation (RMSEA), N is the sample size, and ν is the value of the degrees of freedom. The discontinuity in the confidence interval width is due to truncation of the lower confidence interval limit at zero, due to the fact that the population RMSEA is necessarily nonnegative.

Understanding what the figure conveys is important because it shows that as $\hat{\varepsilon}$ gets close to zero, holding everything else constant, the confidence interval is artificially truncated, which has the impact of reducing the confidence interval width. Now returning to the tables, as a function of ε and ν, holding these two terms constant at a place near the functional break points in Figure 5, the necessary sample size at a particular ω tends to change more slowly compared with the change in ω (i.e., a change in N requires larger change in ω). Put another way, in the four cases where different ω values yield the same value for N, the difference of 0.01 in ω is too small to cause a change in N because ε changes continuously but N changes as an integer. This is analogous to the case where one inputs, say $\omega = 0.030$ and $\omega = 0.031$, other things being the same, and the sample size planning procedure returns (unsurprisingly) the same N. Similar issues arise in power analysis where power changes as a continuous function theoretically, but sample size necessarily changes as a step-function (i.e., whole numbers).

Further note in the Monte Carlo simulation study results that in some situations the confidence interval formation method itself did not work well. For example, although the nominal Type I error rate was set to .05, the empirical α values are (a) .026 when $\varepsilon = 0.0242$, $\nu = 30$, and $N = 421$ (Table 1); (b) .122 when $\varepsilon = 0.0865$, $\nu = 31$, and $N = 253$ (Table 2); (c) .011 when $\varepsilon = 0.0280$, $\nu = 58$, and $N = 316$ (Table 3); and (d) .066 when $\varepsilon = 0.0834$, $\nu = 45$, and $N = 272$ (Table 4). In these four conditions the empirical confidence coverage is quite different from the nominal value. Recall that we used the typical method to form confidence intervals for ε, namely, the one based on a noncentral chi-square distribution. The failure of the confidence interval method is thus not due to our sample size planning method but rather is a problem with the confidence interval procedure itself. Such a finding is consistent with the work of Curran et al. (2003), who report confidence interval empirical coverage rates as low as 76% when the nominal value was specified as 90%. The conditions we included in our Monte Carlo simulation study show that the discrepancy found by Curran et al. was not unique to their specific situations.[6] Nevertheless, the confidence interval procedure worked well in the vast majority of situations.

As stated previously, the method we have proposed is necessarily limited by the existing methods of maximum likelihood estimation and RMSEA confidence interval formation in the SEM literature. Therefore, instead of our sample size planning procedure having a problem, the reason the empirical α does not equal the nominal α in a few cases lies in some combination of maximum likelihood estimation, formation methods for ε, and the fact that $\hat{\varepsilon}$ only approximately follows a noncentral chi-square distribution. Even so, in the conditions noted earlier, the mean and median of those confidence interval widths were all very close to the desired ω value.

PLANNING SAMPLE SIZE: AN EMPIRICAL APPLICATION

The sample size planning method we developed requires the following information to be specified in order to obtain a confidence interval for population RMSEA that is sufficiently narrow: (a) model degrees of freedom (i.e., ν), (b) a

[6]During the peer review of this article the effectiveness of the sem R package and optimization routine was called into question. To show that this was not specific to the sem package, we performed a Monte Carlo simulation study to obtain $\hat{\varepsilon}$ for Model 2c using Mplus (Muthén & Muthén, 2010). The results were essentially identical with 88.10% of the 10,000 Mplus replications correctly bracketing the population value when confidence intervals for ε were formed (recall that the reported value here was 87.80% with 5,000 replications).

presumed population RMSEA value (i.e., ε^*), (c) desired confidence interval width (i.e., ω), and (d) confidence level (i.e., $1-\alpha$). In SEM the model or models of interest are typically known before the data are collected: the model of interest represents the theoretically hypothesized relationships that are the driving force of the research. Correspondingly, the model degrees of freedom (a from previous text) will be known once a model is chosen. Determining the ε^* value that needs to be specified poses a more difficult problem than determining the degrees of freedom. The value(s) chosen for ε^* are the values that the researcher presumes to be true in order to implement the sample size planning procedure. When $\varepsilon^* = \varepsilon$, the sample size planning procedure we developed is optimal with the procedure tending to work less well when ε^* and ε become more discrepant. When a single reasonable estimate of ε^* is not available, the effects of several reasonable values on necessary sample size can be explored in a sensitivity analysis. The desired confidence interval width is based on the goals of the researcher and is context specific—the more accurately the RMSEA is to be estimated, the smaller ω needs to be. Finally, the confidence level needs to be specified. In general, a 90% or 95% confidence interval is typically used, but certainly any reasonable value could be used (e.g., 99%). Note that higher confidence levels (e.g., .99 as compared with .90) imply a larger necessary sample size for a particular expected confidence interval width, all other things being equal.

Our empirical example is based on Holahan, Moos, Holahan, and Brennan (1997), who studied the relations between social context and depressive symptoms at Time 1, coping strategies at Time 2, and depressive symptoms at Time 3. Figure 6 presents the path diagram for the model of interest. Based on Holahan et al.'s results, $\chi^2(30) = 35.66$, $N = 183$, $p = .44$, RMSEA $= .0322$. Given these statistics, we can form a 95% confidence interval for the population RMSEA: $CI_{.95} = [0 \leq \varepsilon \leq .0729]$. Although $\hat{\varepsilon}$ is small, supporting a close-fitting model, the confidence interval for ε is wide. The implication of the wide confidence interval is that plausibly the population RMSEA can be as small as 0 (a perfect fit) or as large as .073 (a reasonable model fit). Thus, there is a fairly large degree of uncertainty about the population value of ε and thus about to what extent the model explains the relationships among the variables.

Suppose a researcher plans to replicate the analysis on a sample from another well-defined population (e.g., patients undergoing treatment for cancer). The study is initially designed to follow the procedures, measures, and model (which has 30 degrees of freedom) used by Holahan et al. (1997). The researcher believes that demonstrating a small value of the RMSEA on the new sample with a narrow 95% confidence interval is an important goal. Suppose the researcher believes that 0.035 is the ideal width for a confidence interval for the population RMSEA, with a confidence interval width larger than 0.05 too wide to be informative and a confidence interval width smaller than 0.02 unnecessarily narrow. Thus, the researcher investigates the necessary sample size for confi-

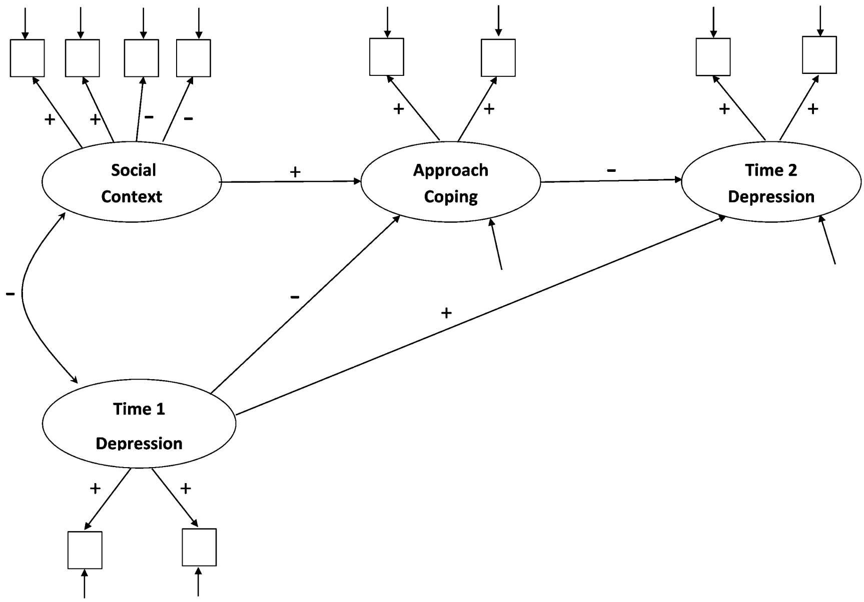

FIGURE 6 A simplified path diagram for the model in Holahan et al. (1997). Squares represent manifest variables. The positive/negative signs next to the paths refer to the signs of the model parameter estimates obtained in Holahan et al.

dence interval widths of 0.02 (minimum), 0.035 (ideal), and 0.05 (maximum) as the input for ω.

The researcher also needs to chose a value for ε^* to implement the sample size planning procedure. Not knowing the population value of the ε for the new population leads the researcher to consider several reasonable values of the ε^* so that the researcher can better understand the effects of various ε^* values on the sample size calculated. Based on the Holahan et al. (1997) study, the values of ε^* for the model of interest are .02 to .06 by .01. Table 5 shows how values of the ε^* and desired confidence interval width interact to produce the necessary sample size for 95% confidence intervals with 30 degrees of freedom. Based on the assumptions of how well the model fits and how narrow a confidence interval is desired, the necessary sample size can be planned. For example, if the researcher believes that .04 is the population value of the ε and desires a confidence interval that has an expected width of 0.035, a sample size of 643 would be necessary (Table 5).

Suppose that a third indicator were available for each of the constructs with only two indicators: (a) Time 1 depression, (b) Approach coping, and (c) Time 2 depression. This new model now with four indicators for Social context and three

TABLE 5
Necessary Sample Size for the RMSEA (ε) to Have a 95% Confidence Interval With an Expected Width That Is Sufficiently Narrow for Models With 30 Degrees of Freedom for 5 Selected RMSEA Values and 3 Desired Confidence Interval Widths (ω)

	ω		
ε	*.02*	*.035*	*.05*
.02	2,208	1,053	364
.03	1,711	982	562
.04	1,541	643	553
.05	1,455	572	355
.06	1,406	531	307

Note. Tabled values were calculated with the `ss.aipe.rmsea()` function from the MBESS R package. Details on how to implement the function to plan sample size in other situations is given in the supplement.

each for Time 1 depression, Approach coping, and Time 2 depression would have 60 degrees of freedom. In general, increasing the number of indicators improves the quality of estimating latent variables. Because the interest is in the relations among the latent variables and the extended model has the same structural paths as the original model, the extended model can be used to study the same phenomena.

As can be seen in Table 6, the extended model with 60 degrees of freedom requires smaller sample sizes than does the original model with 30 degrees of

TABLE 6
Necessary Sample Size for the RMSEA (ε) to Have a 95% Confidence Interval With an Expected Width That Is Sufficiently Narrow for Models with 60 Degrees of Freedom for 5 Selected RMSEA Values and 3 Desired Confidence Interval Widths (ω)

	ω		
ε	*.02*	*.035*	*.05*
.02	2,171	647	232
.03	1,014	966	343
.04	875	544	544
.05	802	349	349
.06	759	309	243

Note. Tabled values were calculated with the `ss.aipe.rmsea()` function from the MBESS R package. Details on how to implement the function to plan sample size in other situations is given in the supplement.

freedom. For the example noted earlier, if the researcher believes that .04 is the population value of ε and desires a confidence interval that has an expected width of 0.035, a sample size of 544 would be necessary (recall that it was 643 for the original model). A seemingly odd outcome stands out in Table 6 where for ε^* values of .04 and .05 the necessary sample sizes for widths of 0.035 and 0.05 require the same value of sample size. The reason for this is the truncation of confidence intervals that would theoretically extend below 0 as discussed previously. Nevertheless, the tables illustrate with a specific example how the sample size can be planned based on the specifics of the model and the goals of the researcher.

DISCUSSION

Our sample size planning procedure requires an input value for what is believed to be the population value of ε and the model degrees of freedom. Thus, the necessary sample size returned is conditional on this input information as is the case with any sample size planning method. Any sample size planning method that requires information about population parameter(s) should be viewed as an approximation as the presumed parameter value used for the sample size planning procedure and the actual parameter value will almost certainly differ. Nevertheless, planning the appropriate sample size for the specified goal can be very useful in planning and evaluating a research design provided the input is reasonable (e.g., Cohen, 1988, pp. 12–13; MacCallum, Browne, & Cai, 2006, p. 34).

The ideal sample size being based on ε is not unique to our work; in fact, it is a strategy that has already been implemented in the current literature of SEM sample size planning. For example, MacCallum et al. (1996) propose a method to plan sample size so that there is sufficient power to reject the null hypothesis that $\varepsilon \geq .05$ (or $\varepsilon \geq .10$). MacCallum et al. (2006) provide guidelines for specifying reasonable values of what we call ε^* to plan sample size. Although our sample size planning method is from the AIPE perspective instead of the power analytic perspective, those guidelines are equally applicable to our method. Of course, our method and the method of MacCallum et al. (2006) differ at a fundamental level: their method seeks to reject a null hypothesis, whereas our method seeks to bracket the population value with a narrow confidence interval. The choice of the appropriate method depends on the researcher's goals for how inference will be made based on the RMSEA.

The cutoff values for the RMSEA are arbitrary conventions yet they are widely used in practice. There is increasing doubt and criticism in the recent literature on the appropriateness of using fixed cutoff values across a wide range of models and situations. For example, some research indicates that the RMSEA value for a model to be considered close fit/poor fit can vary among different substantive areas or model characteristics (e.g., Beauducel & Wittmann, 2005; Marsh, Hau, & Wen, 2004; Raykov, 1998). In addition, Chen, Curran, Bollen, Kirby, and Paxton (2008) argue that there is little empirical support for using .05, or any other fixed value, as a universal cutoff value for RMSEA to determine adequate model fit (p. 462). Instead of being wedded to a certain fixed cutoff value of the RMSEA (i.e., .05, .08, .10), the AIPE approach strives to obtain a narrow confidence interval for ε so that more accurate information about ε is available in order for the researcher and reader to evaluate the model fit on a case-by-case basis. Although MacCallum et al. (1996) discuss sample size planning in the power analysis context, they argue that the confidence interval for RMSEA is more informative than a significance test and "strongly urge" (p. 130) the use of confidence intervals. They also warn that conventional cutoff values are aids for interpretation rather than absolute thresholds (p. 134). Thus, there is an identified need and an existing framework for the AIPE method we develop to be used in applications of SEM.

Rather than there being a requirement that a single point estimate of ε be available in order to plan sample size, a reasonable range of estimates of ε can be used to plan sample size to ensure that the sample size produces an expected width that is at least as narrow as desired. The way in which the size of $\hat{\varepsilon}$ relates to confidence interval width (recall Figure 5) implies that when $\hat{\varepsilon}$ is larger than ε^*, the confidence interval width will tend to be narrower than ω, holding everything else constant. Conversely, when $\hat{\varepsilon}$ is smaller than ε^*, the confidence interval width will tend to be wider than ω, holding everything else constant. In practice, researchers may consider several alternative models representing competing theoretical hypotheses and have some uncertainty about ε and what value(s) ε^* to use to plan sample size. Figures 7 to 9 depict the necessary sample size as a function of ν, ε, and ω. From these figures, there are four observations. First, fixing ε and ω, the necessary sample size decreases or remains the same as ν increases, holding everything else constant. Second, smaller ε values imply larger necessary sample size, holding everything else constant. Third, the sample size changes more slowly after ν exceeds a certain point, as can be seen in the curves for smaller RMSEA values (e.g., .04, .05), with this "certain point" being dependent on several factors. Fourth, within each plot, the curves for the larger values of the specified ε (e.g., .10, .08, and .06) are more tightly clustered than the smaller values of the specified ε (e.g., .05 and .04), which indicates that the change in N is not proportional to the change in ε, holding everything else constant. These observations illustrate the usefulness of the MBESS (Kelley, 2007a, 2007b; Kelley & Lai, 2010) R (R Development Core Team, 2010) package that contains the `ss.aipe.rmsea()` function for planning sample size for the RMSEA from the AIPE framework.

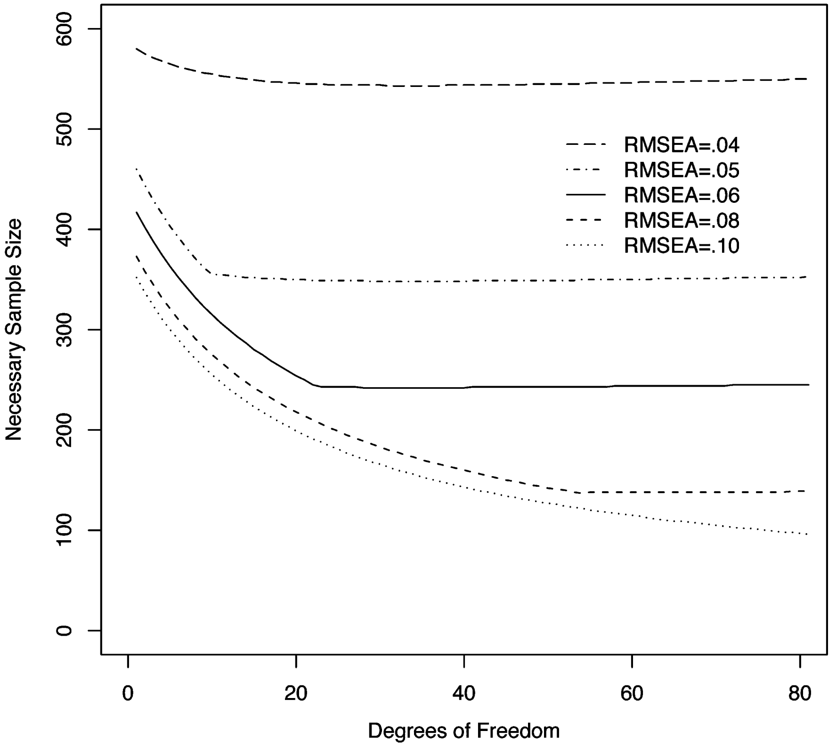

FIGURE 7 The necessary sample sizes for various population RMSEA values and model degrees of freedom, holding the desired expected 95% confidence interval width at 0.05. *Note*. The figure shows the relationship between the necessary sample size and degrees of freedom for five values of the population RMSEA value for a desired 95% confidence interval width of .05.

As can be seen from Tables 5 and 6 and Figures 7, 8, and 9, in some situations (e.g., small degrees of freedom) where the desired confidence interval width is rather small (e.g., $\omega = .02$), the necessary sample size is very large. These large sample sizes are larger than some researchers would reasonably be able to collect given limited time and resources. An implication of this is that not all studies will be able to have confidence intervals for ε that are very narrow due to the large sample size required and lack of resources to obtain such large sample size. Kelley and Rausch (2006) discuss the implicit issue of trading “embarrassingly large” confidence intervals for “distressingly large" sample sizes (p. 369). They argue that such knowledge is still beneficial because it will be known a priori that the confidence interval will likely be wider than desired, which would alleviate any unrealistic expectations about the width of the confidence interval a priori.

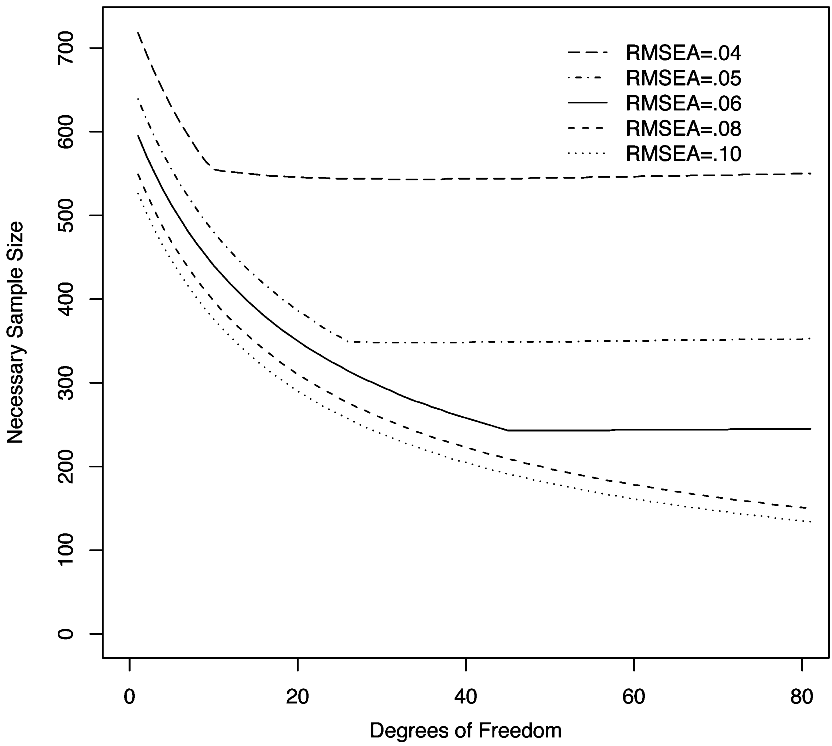

FIGURE 8 The necessary sample sizes for various population RMSEA values and model degrees of freedom, holding the desired expected 95% confidence interval width at 0.04. *Note.* The figure shows the relationship between the necessary sample size and degrees of freedom for five values of the population RMSEA value for a desired 95% confidence interval width of .04.

Researchers who are only able to obtain smaller sample sizes could use the methods to show that it would be difficult or impossible in some situations to obtain the required sample size so that the confidence interval for ε had an expected width that was sufficiently narrow, even if the sample size provides sufficient statistical power (e.g., for a large effect size).

In concluding, the RMSEA is widely used in applications of SEM and is often the only or the main overall assessment of model fit. The RMSEA is important because it provides an overall assessment of the extent to which a theoretical model is supported by the data. However, when a confidence interval for ε is wide, the uncertainty with which $\hat{\varepsilon}$ represents ε becomes apparent. What might initially appear to be a "good fitting" model might at best (i.e., the lower confidence interval limit) be an exceptional fitting model or at worst

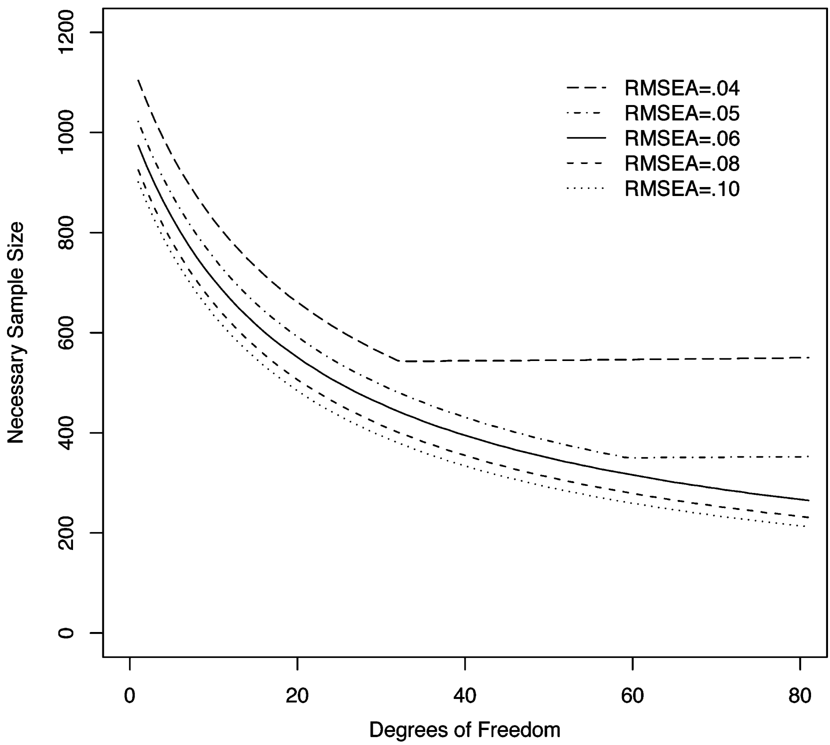

FIGURE 9 The necessary sample sizes for various population RMSEA values and model degrees of freedom, holding the desired expected 95% confidence interval width at 0.03. *Note.* The figure shows the relationship between the necessary sample size and degrees of freedom for five values of the population RMSEA value for a desired 95% confidence interval width of .03.

(i.e., the upper confidence interval limit) be a very poorly fitting model, at some specified level of confidence. The AIPE approach to sample size planning strives to obtain an accurate estimate of ε so that the researcher can evaluate the model fit on a case-by-case basis. Because of sampling error, $\hat{\varepsilon}$ (a point estimate) will differ from ε. A confidence interval provides a range of plausible values for the population parameter, and the narrower the confidence interval is, the more information is available about the population parameter, holding everything else constant. Therefore, in an effort to obtain an accurate estimate of ε, the AIPE approach to sample size planning should be considered when planning a research study that involves the evaluation of a theoretically interesting model.

ACKNOWLEDGMENT

This research was supported in part by the University of Notre Dame Center for Research Computing through the use of the high-performance computing cluster. This article contains a supplement available at https://repository.library.nd.edu/view/1/AIPE_RMSEA_MBR_Supplement.pdf that demonstrates how to implement the methods discussed with freely available software.

REFERENCES

Beauducel, A., & Wittmann, W. (2005). Simulation study on fit indices in confirmatory factor analysis based on data with slightly distorted simple structure. *Structural Equation Modeling, 12,* 41–75.

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

Browne, M. W. (1984). Asymptotic distribution free methods in the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology, 24,* 445–455.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods & Research, 21,* 230–258.

Casella, G., & Berger, R. L. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury Press.

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,* 462–494.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist, 49,* 997–1003.

Cudeck, R., & Browne, M. W. (1992). Constructing a covariance matrix that yields a specified minimizer and a specified minimum discrepancy function value. *Psychometrika, 57,* 357–369.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement, 61,* 532–574.

Curran, P. J., Bollen, K. A., Chen, F., Paxton, P., & Kirby, J. B. (2003). Finite sampling properties of the point estimates and confidence intervals of the RMSEA. *Sociological Methods & Research Methods, 32,* 208–252.

Davey, A., & Savla, J. (2010). *Statistical power analysis with missing data: A structural equation modeling approach.* New York, NY: Routledge.

Fleishman, A. I. (1980). Confidence intervals for correlation ratios. *Educational and Psychological Measurement, 40,* 659–670.

Fox, J. (2006). Structural equation modeling with the SEM package in R. *Structural Equation Modeling*, 465–486.

Hahn, G., & Meeker, W. (1991). *Statistical intervals: A guide for practitioners*. New York, NY: Wiley.

Holahan, C. J., Moos, R. H., Holahan, C. K., & Brennan, P. L. (1997). Social context, coping strategies, and depressive symptoms: An expanded model with cardiac patients. *Journal of Personality and Social Psychology, 72,* 918–928.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution.* Chicago, IL: The University of Chicago.

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

Kelley, K. (2007a). Confidence intervals for standardized effect sizes: Theory, application, and implementation. *Journal of Statistical Software, 20*(8), 1–24.

Kelley, K. (2007b). Methods for the behavioral, educational, and educational sciences: An R package. *Behavior Research Methods, 39,* 979–984.

Kelley, K. (2007c). Sample size planning for the coefficient of variation from the accuracy in parameter estimation approach. *Behavior Research Methods, 39,* 755–766.

Kelley, K. (2008). Sample size planning for the squared multiple correlation coefficient: Accuracy in parameter estimation via narrow confidence intervals. *Multivariate Behavioral Research, 43,* 524–555.

Kelley, K., & Lai, K. (2010). MBESS 3.0.4 (or greater) [Computer software and manual]. Retrieved from http://www.cran.r-project.org/

Kelley, K., & Maxwell, S. E. (2003). Sample size for multiple regression: Obtaining regression coefficients that are accurate, not simply significant. *Psychological Methods, 8,* 305–321.

Kelley, K., Maxwell, S. E., & Rausch, J. R. (2003). Obtaining power or obtaining precision: Delineating methods of sample size planning. *Evaluation and the Health Professions, 26,* 258–287.

Kelley, K., & Rausch, J. R. (2006). Sample size planning for the standardized mean difference: Accuracy in parameter estimation via narrow confidence intervals. *Psychological Methods, 11,* 363–385.

MacCallum, R. C., Browne, M. W., & Cai, L. (2006). Testing differences between nested covariance structure models: Power analysis and null hypotheses. *Psychological Methods, 11,* 19–35.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130–149.

Marsh, H. W., Hau, K., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizeing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11,* 320–341.

Maruyama, G., & McGarvey, B. (1980). Evaluating causal models: An application of maximum-likelihood analysis of structural equations. *Psychological Bulletin, 87,* 502–512.

Maxwell, S. E., Kelley, K., & Rausch, J. R. (2008). Sample size planning for statistical power and accuracy in parameter estimation. *Annual Review of Psychology, 59,* 537–563.

McDonald, R. P. (1989). An index of goodnesss of fit based on noncentrality. *Journal of Classification, 6,* 97–103.

Muthén, L. K., & Muthén, B. O. (2010). Mplus user's guide (Version 6) [Computer software and manual]. Los Angeles, CA: Author.

Powell, B. B. (1991). *Homer and the origin of the Greek alphabet* (2nd ed.). Cambridge, United Kingdom: Cambridge University Press.

R Development Core Team. (2010). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria: R Foundation for Statistical Computing.

Raykov, T. (1998). On the use of confirmatory factor analysis in personality research. *Personality and Individual Differences, 24,* 291–293.

Smithson, M. (2001). Correct confidence intervals for various regression effect sizes and parameters: The importance of noncentral distributions in computing intervals. *Educational and Psychological Measurement, 61,* 605–632.

Steiger, J. H. (1989). *Ezpath: A supplementary module for systat and sygraph*. Evanston, IL: SYSTAT.

Steiger, J. H. (2000). Point estimation, hypothesis testing, and interval estimation: Some comments and a reply to Hayduk and Glaser. *Structural Equation Modeling, 7,* 164–182.

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical methods. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there where no significance tests?* (pp. 221–257). Mahwah, NJ: Erlbaum.

Steiger, J. H., & Lind, J. C. (1980, May). *Statistically based tests for the number of common factors.* Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.

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

Taylor, A. B. (2008). *Two new methods of studying the performance of SEM fit indexes* (Doctoral dissertation, Arizona State University).

Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S* (4th ed.). New York, New York: Springer.

Venables, W. N., & Ripley, B. D. (2010). MASS [Computer software and manual]. Retrieved from http://www.cran.r-project.org/

Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society, 54,* 426–482.