

Running Head: RMSEA with Non-normal Data

Estimating the Maximum Likelihood Root Mean Square Error of Approximation (RMSEA) with
Non-normal Data: A Monte-Carlo Study

Chuanji Gao^{1,2}

Dexin Shi¹

Alberto Maydeu-Olivares^{1,3}

1. Department of Psychology, University of South Carolina
2. Institute for Mind and Brain, University of South Carolina
3. Department of Psychology, University of Barcelona

Structural Equation Modeling (2019)

Correspondence should be addressed to Chuanji Gao, Department of Psychology, and Institute for Mind and Brain, University of South Carolina, Barnwell College. 1512 Pendleton St., Columbia, SC 29208. E-mail: chuanji@email.sc.edu. This research was supported by the National Science Foundation under Grant No. SES-1659936.

Abstract

Recent research has provided formulae for estimating the maximum likelihood (ML) RMSEA when mean, or mean and variance, corrections for non-normality are applied to the likelihood ratio test statistic. We investigate by simulation which choice of corrections provides most accurate point RMSEA estimates, confidence intervals, and p -values for a test of close fit under normality, and in the presence of non-normality. We found that, overall, any robust corrections (choices MLM, MLMV, and MLR) provide better results than ML, which assumes normality. When they err, all choices tend to suggest that the model fits more poorly than it really does. Choice MLMV (mean and variance corrections) provided the most accurate RMSEA estimates and p -values for tests of close fit results but its performance decreases as the number of variables being modeled increases.

Keywords: Structural equation modeling (SEM), RMSEA, nonnormal data, fit index

Estimating the Maximum Likelihood Root Mean Square Error of Approximation (RMSEA) with Non-normal Data: A Monte-Carlo Study

Structural equation modeling (SEM) is a comprehensive approach to fit theoretical models involving systems of equations with observed as well as latent variables (i.e., common factors). A critical step when fitting a model is to determine whether it fits the data at hand (Bearden, Sharma, & Teel, 1982; Bollen, 1989; Lomax & Schumacker, 2004). When no structure is imposed on the intercepts of the model (i.e., in covariance structure analysis), the null and alternative hypotheses of model fit can be written as $H_0 : \Sigma = \Sigma_0$ and $H_1 : \Sigma \neq \Sigma_0$, where Σ denotes the population covariance matrix, and $\Sigma_0 = \Sigma(\theta)$ denotes the covariance matrix implied by the theoretical model under consideration, expressed as a function of the model parameters θ .

In this paper, we focus on models in which all dependent variables are treated as continuous¹. In such instances, presently, maximum likelihood (ML) estimation is almost invariably used (Maydeu-Olivares, 2017b). Under normality assumptions, the null hypothesis of exact model fit can be assessed in large samples using $X_{ML}^2 = N\hat{F}_{ML}$, where N denotes sample size and \hat{F}_{ML} denotes the minimum of the ML fitting function. X_{ML}^2 is the likelihood ratio (LR) test statistic. In this setting, p -values for X_{ML}^2 are obtained using a reference chi-square. However, when sample size is small, normality assumptions are violated, or a large number of variables is being modeled, p -values obtained in this fashion need not be accurate (Bentler & Bonett, 1980; Chou, Bentler, & Satorra, 1991; Curran, West, & Finch, 1996; Hoelter, 1983; Hu, Bentler, & Kano, 1992; Moshagen, 2012; Shi, Lee, & Terry, 2018). In particular, when the data are not normally distributed, X_{ML}^2 does not follow an asymptotic chi-square distribution.

For non-normal data, the most widely used test statistic when ML estimation is employed involves using a mean correction, or a mean and variance correction, to X_{ML}^2 and using a reference chi-square distribution to obtain p -values (Satorra & Bentler, 1994). In this article, we focus on the mean or mean and variance corrections implemented in the widely used SEM software Mplus (Muthén & Muthén, 2017). Also, we refer to them using the same nomenclature used in Mplus: 1) X_{MLM}^2 : mean corrected LR statistic proposed by Satorra and Bentler (1994); 2) X_{MLMV}^2 : mean and variance corrected LR statistic described in Asparouhov and Muthén (2010)²; and 3) X_{MLR}^2 : mean corrected LR statistic as described in Asparouhov and Muthén (2005).

These test statistics are related to the LR statistic as follows: $X_{MLM}^2 = X_{ML}^2 / c_{MLM}$,

$X_{MLMV}^2 = a + X_{ML}^2 / c_{MLMV}$, and $X_{MLR}^2 = X_{ML}^2 / c_{MLR}$. Somewhat confusingly, these test statistics are obtained in Mplus by invoking different “ESTIMATORS” (e.g., ML, MLM, MLMV, and MLR).

In this paper we refer to them as ‘choices’ of ML estimation because all of them result in the same parameter estimates, ML estimates, but different standard errors for the parameter estimates, and different test statistics (i.e., different ‘chi-square tests’). In particular, choices MLM and MLMV result in the same standard errors as the same formula is used to compute them, but yield different chi-square and RMSEA statistics. Choice MLM adjusts the LR statistic by its asymptotic mean, whereas choice MLMV adjusts it by its asymptotic mean and variance. In both cases, the expected information matrix is used to obtain the standard errors of the parameter estimates. In choice MLR, the observed information matrix is used to obtain standard errors of the parameter estimates. As a result, standard errors for choices MLR and MLM/MLMV are different in finite samples. Like choice MLM, choice MLR adjusts the LR statistic by its asymptotic mean. As a result, asymptotically, choices MLR and MLM result in the

same chi-square statistic. However, in finite samples the chi-square statistics for choices MLR and MLM differ because the mean adjustment involves the information matrix, which is computed differently in choices MLM and MLR. See Maydeu-Olivares (2017b) for a detailed account of the differences between the various implementations of ML estimation for continuous outcomes.

Regardless of the outcome of the model fit test, it is necessary to assess the magnitude of model misfit, the discrepancy between the unknown data generating mechanism and the fitted model (Browne & Cudeck, 1993; Maydeu-Olivares, 2017a). The most popular (Jackson, Gillaspay, & Purc-Stephenson, 2009; McDonald & Ho, 2002) effect size of model misfit is the Root Mean Square Error of Approximation (RMSEA: Browne & Cudeck, 1993; Steiger, 1990; Steiger & Lind, 1980). To a large extent, the popularity of the RMSEA comes from Browne and Cudeck's (1993) proposal of replacing the null hypothesis of model fit, which can be stated in terms of the RMSEA as $H_0 : RMSEA = 0$, by a more lenient null hypothesis of close fit, $H_0^* : RMSEA \leq k$, where k is an arbitrary value of the RMSEA, typically .05. Testing for close fit instead of exact fit makes particular sense when sample size is very large (as any trivial model fit discrepancy will be detected by the chi-square test of exact fit), or when many variables are being modeled (as finding a well-fitting model is increasingly difficult as the number of variables to be modeled increases) (Maydeu-Olivares, 2017a).

Following Browne and Cudeck (1993; Cudeck & Henly, 1991), let $\hat{\Sigma}$ be the best fit of the model to the population covariance matrix Σ in terms of the fit function used to estimate the model, $F(\Sigma, \Sigma_0)$. We note that both Σ , $\hat{\Sigma}$ are unknown fixed matrices. Then, the RMSEA is defined in the population as (Browne & Cudeck, 1993)

$$RMSEA = \sqrt{\frac{F(\Sigma, \tilde{\Gamma})}{df}}, \quad (1)$$

where $F(\Sigma, \tilde{\Gamma})$ denotes the minimum of the discrepancy between the population covariance matrix and the covariance matrix implied by the model. It is clear from (1) that the population RMSEA depends on the estimation method employed. When ML estimation is used the population RMSEA (i.e., $RMSEA_{ML}$) is

$$RMSEA_{ML} = \sqrt{\frac{\tilde{I}}{df}}, \quad (2)$$

and \tilde{I} is obtained by minimizing

$$F_{ML} = \ln|\Sigma_0| - \ln|\Sigma| + \text{tr}(\Sigma\Sigma_0^{-1}) - p, \quad (3)$$

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

$$RMSEA_{ML} = \sqrt{\max\left(\frac{X_{ML}^2 - df}{N \times df}, 0\right)} \quad (4)$$

where N denotes sample size. A 90% confidence interval (CI) for $RMSEA_{ML}$ can be obtained as

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

where L and U are the solutions to

$$F_{\chi^2}(X_{ML}^2; df, L) = .95, \text{ and } F_{\chi^2}(X_{ML}^2; df, U) = .05, \quad (6)$$

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

$$1 - F_{\chi^2} \left(X_{ML}^2; df, N \times df \times k^2 \right). \quad (7)$$

How to estimate the $RMSEA_{ML}$ when data are non-normal? The most popular approach is to simply replace X_{ML}^2 by X_{MLM}^2 or X_{MLMV}^2 , respectively, in equations (4), (6) and (7). We refer to this approach as the *naïve approach*, and it is the approach implemented in most popular software programs (e.g., Mplus). However, the use of this *naïve approach* results in estimating a parameter value that is not the $RMSEA_{ML}$ (2) leading to difficulty of interpretation (and comparison) of RMSEA values (Brosseau-Liard, Savalei, & Li, 2012; Savalei, 2018).

How can we obtain consistent³ estimates of the population parameter $RMSEA_{ML}$, regardless of whether data are normal or non-normal, and regardless of the choice of mean or mean and variance correction? By using

$$RMSEA_{ML} = \sqrt{\max \left(\frac{c(X^2 - df)}{N \times df}, 0 \right)} \quad (8)$$

$$\sqrt{\max \left(0, \frac{cL}{N \times df} \right)}, \sqrt{\max \left(0, \frac{cU}{N \times df} \right)} \quad (9)$$

$$1 - F_{\chi^2} \left(X^2; df, N \times df \times k^2 / c \right), \quad (10)$$

instead of equations (4), (5), and (7) (Brosseau-Liard et al., 2012; Li & Bentler, 2006; Savalei, 2018). We refer to the use of formula (8), (9), and (10) as the *correct approach*. In these formula, $X^2 = X_{ML}^2$, X_{MLM}^2 , X_{MLR}^2 , or X_{MLMV}^2 , and $c = c_{ML}$, c_{MLM} , c_{MLR} or c_{MLMV} denotes its associated correction used. We provide in Table 1 formula for computing the constant c for different choices of corrections⁴ to the likelihood ratio test statistic. In particular, for X_{ML}^2 , suitable for normal outcomes, $c = 1$, and the formula reduce to (4), (5), and (7). Furthermore, when MLM and MLMV are used, different chi-square statistics are obtained but under the correct approach

described above, the same RMSEA ML point estimates are obtained. However, different confidence intervals and p -values for tests of close fit are obtained for MLM and MLMV when the correct approach is used. For details see Savalei (2018).

 Insert Table 1 about here

What choice of uncorrected, mean, or mean and variance corrected chi-square statistics (i.e., ML, MLM, MLMV, or MLR) leads to more accurate point estimates of the $RMSEA_{ML}$ parameter, confidence intervals, and tests of close fit? This has not been investigated. To fill this gap in the literature, in this article we compare the performance of different procedures to estimate the $RMSEA_{ML}$ parameter using a Monte Carlo simulation study. In our simulation design, the performance of point estimates, CIs and tests of close fit were examined under various degrees of non-normality, model size, sample size, and degree of model misspecifications. We close our presentation by offering some practical guidance of estimating the RMSEA when data are non-normal.

Simulation Study

Method

Given the popularity of confirmatory factor models in empirical studies, we used confirmatory factor models as data generating and fitted models. More specifically, the fitted model was a one-factor model in all cases; the data generating model was a two-factor model with the following characteristics. First, model size includes 16 or 32 items, which represented a small and a large population model, respectively. For a given population model, each item loaded on a single factor. Item loading values ranged from .5 to .8; true population loading values were repeated in sets of four (i.e., values of [.5 .6 .7 .8] repeated four times in the small

population model and eight times in the large population model). Uniquenesses values were specified so that the population covariance matrix was a correlation matrix. Second, three different skewness and kurtosis values were examined. The values were selected to parallel previous studies (e.g., West, Finch, & Curran, 1995): skewness = 0, kurtosis = 0 (i.e., normal); skewness = 1, kurtosis = 7; and skewness = 3, kurtosis = 21. Non-normal data were obtained by following the procedure described in Vale and Maurelli (1983). Third, sample sizes of 200, 500, and 1000 were included to examine sample size conditions encountered in applied settings (Wolf, Harrington, Clark, & Miller, 2013). Fourth, one correctly specified population model and two misspecified models were used. The different levels of misspecification were obtained by setting the inter-factor correlation, ρ to 1 (correctly specified), .8 (small misspecification), and .4 (large misspecification). In all cases, the mean of the data were zero and the variances of the factors were set to one.

Taken together, the simulation study consisted of a crossed design of 54 conditions: 2 model sizes (16, 32) \times 3 distributional conditions \times 3 sample sizes (200, 500, 1000) \times 3 levels of model misspecification.

Data Analysis

For each of the 54 simulated conditions, 5,000 replications were obtained. Model estimation was performed using *Mplus* (Muthén & Muthén, 2017). Replications that showed convergence problems or improper solutions were noted and eliminated from further analysis. First, the sample RMSEA was estimated as described in Table 1 for ML, MLM, MLMV, and MLR, and averaged across replications. Second, the 90% CIs for the population $RMSEA_{ML}$ were estimated as described in Table 1 and coverage rates (percentage of intervals that contain the true parameter) were obtained for each choice of test statistic. Third, empirical rejection rates at a 5%

significance level of a test that the $RMSEA_{ML}$ is less than or equal to its population value were obtained. Note that we only offer results for the “correct” approach.

Results

Table 2 shows the resulting population values in our simulation and the average of sample RMSEA across replications. We note that the results for MLM and MLMV are the same in this table because the formula for MLMV simplifies to that of MLM for RMSEA point estimates. We begin by briefly summarizing the results of the population values of RMSEA. First, the population values of RMSEA were equal to 0 for all properly specified models ($\rho = 1$) regardless of model size ($p = 16$ or 32). Second, consistent with prior studies (Chen, Curran, Bollen, Kirby, & Paxton, 2008; Fan, Thompson, & Wang, 1999; Hu & Bentler, 1999), population RMSEA increased as the degree of model misspecification increased. Third, in accordance with previous research (Breivik & Olsson, 2001; Kenny & McCoach, 2003; Savalei, 2012; Shi, Lee, & Maydeu-Olivares, 2019), we found that the population values of the RMSEA tend to decrease as model size increases holding model misspecification constant.

Next, we focus on the sample estimates of the RMSEA and their behavior under different conditions. When data were normally distributed, estimates close to population values were obtained regardless of the choice of test statistic. Though the results produced by ML (i.e., normality assumptions) were accurate overall when data were normal, we observed that the sample RMSEA tended to overestimate the population RMSEA when data were non-normal, and the bias increased with decreasing sample size, decreasing model misspecification, or increasing model size. For example, the average sample ML RMSEA is .084 using X^2_{ML} when the population ML RMSEA = 0 if *kurtosis* = 3, *skewness* = 21, $N = 200$, $\rho = 1$, and $p = 32$.

Robust methods (MLM/MLMV and MLR) provide similar results than choice ML when data are normal, but more accurate estimates when data are non-normal. For the condition above, MLM and MLMV yielded an average of .064 and MLR an average RMSEA of .075. In general, and across conditions, MLM/MLMV provided more accurate RMSEA point estimates than MLR. Finally, we note that all test statistics appear to converge to the population RMSEA values, i.e., bias decreases as sample size increases.

 Insert Table 2 about here

Table 3 shows the coverage rates (i.e., percentage of confidence intervals that include the true parameter) for 90% CIs around the population RMSEA. We have highlighted in Table 3 conditions where acceptable coverage rates (between 85% and 95%) were obtained. As Table 3 reveals, the coverage of the 90% CIs obtained under normality (i.e., choice ML) was generally poor. As was the case for point estimates, the coverage rates for ML CIs were strongly influenced by the degree of non-normality. Choice ML produced some acceptable CIs only when data were normal. In general, negligible differences were observed in terms of coverage of CIs between choices ML, MLM, MLMV, and MLR when data were normal. However, when data were not normal, coverage of CIs of robust methods (MLM, MLR and MLMV) was better than for choice ML. For example, when fitting models with non-normal data (*kurtosis* = 3, *skewness* = 21) with $p = 16$ and $\rho = .8$, population RMSEA = .067 and $N = 1000$, coverage rate for 90% CIs under ML was .215, for MLM and MLR was .784 and .721, respectively, and for MLMV was .925. Among robust methods, choice MLMV negligibly outperformed choices that involved only a mean correction (MLM and MLR).

Furthermore, the coverage rates of the RMSEAs robust to non-normality (i.e., using choices MLM, MLR, MLMV) need not improve with increasing sample size. This is because the mean (or mean and variance) adjusted statistics do not follow asymptotically a chi-square distribution. They only agree in mean (or in mean and variance) with a reference a chi-square distribution. As a result, their empirical distribution need not converge to that of the reference distribution as sample increases. For instance, coverage for 90% CIs around the population RMSEA using MLMV when $skewness = 1$, $kurtosis = 7$, $\rho = .8$, $p = 16$, improves from $N = 200$ to 500, but worsens from $N = 500$ to 1000: coverage rates are .925, .908, and .879, respectively.

 Insert Table 3 about here

Table 4 shows the results for tests of close fit, more specifically, empirical rejection rates at a 5% significance level of tests of the RMSEA being less than or equal to its population value. We considered rejection rates between 0.03 and 0.08 as reasonably accurate and they are highlighted in Table 4. Empirical rejection rates for choice ML (i.e., under normality) were strongly influenced by the non-normality of the data. When data were not normal, no acceptable empirical rejection rates were obtained using choice ML regardless of model size, model misspecification or sample size. Empirical rejection rates were also strongly influenced by model size. In general, robust choices (MLMV, MLM and MLR) yield more accurate results than choice ML across conditions. However, the main finding that can be extracted from this table is that assessing close fit using the RMSEA in large models ($p = 32$) often fails.

 Insert Table 4 about here

A numerical example: Fitting a confirmatory factor analysis model to the short version of the SPSI-R

The Social Problem Solving Inventory-Revised (SPSI-R: D’Zurilla, Nezu, & Maydeu-Olivares, 2002) is a 52 item questionnaire designed to measure the five dimensions of the social problem solving model of D’Zurilla and colleagues (D’Zurilla, Nezu, & Maydeu-Olivares, 2004; Maydeu-Olivares & D’Zurilla, 1996). Here, we report the results of fitting a five-factor confirmatory model to the 25 items of the short version of the questionnaire. The US normative young adult sample was used. Sample size is $N = 1,020$. Each item consists of a rating using five response alternatives from ‘very untrue of me’ to ‘very true of me’. Data appears quite normally distributed: skewness ranges from $-.78$ to $.81$ with an average of $.18$; excess kurtosis ranges from -1.03 to $.04$ with an average of $-.59$. We fitted the model with maximum likelihood choices ML, MLM, MLR, and MLMV. The first is suitable for normal outcomes; the others yield standard errors and X^2 statistics robust to non-normality. Goodness of fit results are presented in Table 5. We include in this table the X^2 statistics, and constant c computed using the formula in Table 1, naïve (incorrect) RMSEA results, and the correct RMSEA results (computed using the formula in Table 1). We see in Table 5 that regardless of the X^2 statistic used, the hypothesis of exact model fit is to be rejected. This is not surprising since we are fitting a quite restrictive model (265 df) to 25 observed variables. We note, however, that the MLMV X^2 is smallest and that the ML X^2 is largest. This is consistent with simulation results by Maydeu-Olivares (2017b) who found that the ML X^2 rejects too often in models of this size, even with normally distributed data; in contrast, the MLMV X^2 maintained adequate rejection rates.

Now, the naïve RMSEA results are obtained by using formula (4) and (5) for all choices of test statistic implemented in the current version of Mplus; that is, if MLMV is used, by

replacing χ^2_{ML} by χ^2_{MLMV} in these equations. The correct RMSEA results are obtained using the formula in Table 1. We see in Table 5 that when the naïve approach is used, much lower RMSEA estimates are obtained when using robust methods (MLM, MLMV, or MLR) than when normality is assumed (choice ML). For instance, under ML the estimated RMSEA is .050, whereas when MLMV is used the estimate is .040. In fact, robust CIs for the RMSEA do not include the RMSEA estimate obtained using ML.

The correct robust RMSEA results (obtained using Lavaan, Rosseel, 2012) reveal a different picture: they are much more in line with those obtained under normality. For instance, the MLMV results suggest that the point estimate is .048. Given our simulation results, we conclude that the MLMV RMSEA results are likely to be the most accurate. We also note that in this example the effect of using the naïve approach to estimate RMSEA results is to underestimate the degree of model misfit, giving the impression that the model fits better than it does.

Discussion

Our findings are in line with previous research (Brosseau-Liard et al., 2012; Curran, Bollen, Chen, Paxton, & Kirby, 2003; Fan et al., 1999; Hu & Bentler, 1999). Sample size and model size influence the bias of RMSEA estimation because the RMSEA is a function of the χ^2 statistic, whose performance is influenced, in turn, by both sample size and model size. Shi et al. (2019) found that the bias of the RMSEA increased as model size increased. This is in contrast to results by Kenny and McCoach (2003) who reported the opposite effect. Shi et al. (2019) speculated that this might be due to the size of the models considered in each study (4 to 25 variables in Kenny and McCoach vs. 10 to 120 in Shi et al.). The present results support Shi et al.'s conclusions. Brosseau-Liard et al. (2012) have also suggested that the bias of the sample

RMSEA may increase with decreasing model size because it involves population RMSEA values close to zero.

As reported in the literature, we found that RMSEA point estimates, CIs and tests of close fit obtained under normality (i.e., choice ML) were adversely affected by non-normality. ML estimation assumes multivariate normality, which is frequently violated in practice (Micceri, 1989). Although the ML estimates are consistent when data are not normal, test statistics and standard errors are inflated when data are non-normal (Chou et al., 1991; Curran et al., 1996; Muthén & Kaplan, 1985). Repeatedly, research has also found inflated sample estimates of RMSEA (Brosseau-Liard et al., 2012; Curran, Bollen, Chen, Paxton, & Kirby, 2003; Fan et al., 1999), poor coverage of CIs (Brosseau-Liard et al., 2012; Maydeu-Olivares, 2017b; Maydeu-Olivares, Shi, & Rosseel, 2018), and inflated false positive error rate of tests of close fit (Maydeu-Olivares, 2017b; Maydeu-Olivares et al., 2018) due to non-normality. In other words, when data are not normal, inflated RMSEA estimates may be obtained, and models may be rejected by tests of close fit because of non-normality, not because of lack of fit proper.

Brosseau-Liard et al. (2012) distinguished two approaches to estimate the $RMSEA_{ML}$ in the presence of non-normal data. One is the *naïve approach* we have described which directly substitutes X^2 with relevant robust statistic (which they refer to as the *population-corrected robust RMSEA*); the other is the *correct approach* examined in the current study (which they refer to as the *sample-corrected robust RMSEA*). They report that the correct formula used here yielded consistent estimates of the population RMSEA and did a good job at reducing bias due to non-normality. In contrast, the *naïve approach* results in a statistic that decreases with increasing non-normality. In fact, the RMSEA computed based on the *naïve approach* have population values change with the degree of non-normality (Brosseau-Liard et al., 2012; Curran et al., 1996;

Nevitt & Hancock, 2000; Savalei, 2018). Therefore, the authors recommended using the formulae presented in Table 1. However, MLM was the focus of Brosseau-Liard et al. (2012); few studies have addressed the estimation of the ML RMSEA based on MLMV and MLR. Recently, research provided insights on how to correctly compute the ML RMSEA and CI with MLM, MLMV, and MLR (Maydeu-Olivares et al., 2018; Savalei, 2018).

In accordance with prior research (Brosseau-Liard et al., 2012), under non-normal data, the robust $RMSEA_{ML}$ computed using the correct approach generally reduced the biases of uncorrected $RMSEA_{ML}$ point estimates, CIs, and tests of close fit. However, when the data were normally distributed, comparison of the point estimates of different corrections revealed that the robust RMSEA did not change much of the biases. This finding indicated that the robust methodology corrects for non-normality but not biases from other causes, which is also in line with Brosseau-Liard et al. (2012).

Our study revealed that MLMV consistently had a better performance than MLM, and MLR. For the point estimates of RMSEA, the numeric values of MLM are equal to MLMV (see Savalei, 2018 for a detailed explanation). The MLM could lead to a poor test statistic compared to MLMV because the MLM performs only a mean adjustment instead of a mean and variance adjustment in MLMV. Though MLM and MLR yield similar CIs and tests of close fit because both of them perform a mean adjustment, MLM and MLR are based on different test statistics because of different information function estimates (Asparouhov & Muthén, 2005; Yuan & Bentler, 2000). Studies have suggested that the test statistics based on Asparouhov and Muthén (2005) such as MLR could lead to under-rejection or over-rejection of models (Maydeu-Olivares, 2017b; Savalei, 2010).

There are limitations with this study. As with many simulation studies, we used a single hypothesized model; also, factor loadings were kept constant across conditions. In addition, alternative types of model misspecifications could be included. Our study used omitted factor correlation; future studies could include omitted cross-loadings to provide a comprehensive picture of the behavior of the robust RMSEA. In addition, different data generation method might also influence the simulation results. One might wonder whether the results would differ for ordinal data modeled as if it were continuous. To test this possibility, we repeated our study using Muthén and Kaplan's (1985) approach: Non-normal data were obtained by generating multivariate normal data and discretizing it into five categories according to a set of thresholds to obtain data with the desired skewness and kurtosis. This method has been used to generate non-normal data in previous simulation studies (e.g, Forero & Maydeu-Olivares, 2009; Maydeu-Olivares, Coffman, & Hartmann, 2007). Tables summarizing the results obtained are provided as supplementary materials. As can be seen in the supplementary materials, similar results were found: MLMV outperformed other choices of test statistics in tests of close fit. Interestingly, when using the Muthén and Kaplan (1985) approach to generate the data, MLMV confidence intervals and test of close fit results are much more accurate than using the Vale and Maurelli (1983) approach to generate data using similar values of skewness and kurtosis⁵. In other words, confidence intervals and tests of close fit appear to be more precise when fitting models to ordinal data (treating it as continuous, e.g., Rhemtulla, Brosseau-Liard, & Savalei, 2012), than when data are truly continuous. Maydeu-Olivares, Fairchild and Hall (2017) investigated the effect of the number of response alternatives (including truly continuous data) on the average sample RMSEA_{ML} estimates. They used the naïve approach to estimate the population parameter from for the mean and variance corrected statistic (i.e., MLMV)⁶. Further research is clearly

needed to investigate the effect of the number of response alternatives (including truly continuous data) on coverage rates for the $RMSEA_{ML}$.

In closing, the main goal of the current study was to compare the performances of different choices for estimating the $RMSEA_{ML}$ with non-normal data. Results from our simulation study support that when the correct approach (formula) is applied, MLMV is the best choice by providing the most accurate point estimates, CIs and tests of close fit (type I error rates). However, its performance is unsatisfactorily when data are truly continuous in large models; its performance is better when data are ordinal treated as continuous. Though more studies are needed to replicate and extend the results to other situations, we hope this study could provide some insights for applied researchers to evaluate the extent of misfit in SEM models under non-normal data.

References

- Asparouhov, T., & Muthén, B. (2005). Multivariate statistical modeling with survey data. In *Proceedings of the Federal Committee on Statistical Methodology (FCSM) Research Conference*.
- Asparouhov, T., & Muthén, B. (2010). *Simple second order chi-square correction scaled chi-square statistics* (Technical appendix). Los Angeles, CA.
- Bearden, W. O., Sharma, S., & Teel, J. E. (1982). Sample Size Effects on Chi Square and Other Statistics Used in Evaluating Causal Models. *Journal of Marketing Research*, 19(4), 425.
<http://doi.org/10.2307/3151716>
- Bentler, P. M., & Bonett, D. D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88(3), 588–606.
<http://doi.org/10.1037//0033-2909.88.3.588>
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Breivik, E., & Olsson, U. H. (2001). Adding variables to improve fit: The effect of model size on fit assessment in LISREL. In R. Cudeck, K. G. Jöreskog, & S. du Toit (Eds.), *Structural Equation Modeling: Present and Future: A Festschrift in honor of Karl Jöreskog* (pp. 169–194). Chicago, IL: SSI International.
- Brosseau-Liard, P. E., Savalei, V., & Li, L. (2012). An Investigation of the Sample Performance of Two Nonnormality Corrections for RMSEA. *Multivariate Behavioral Research*, 47(6), 904–930. <http://doi.org/10.1080/00273171.2012.715252>
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. s. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

- 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.
<http://doi.org/10.1177/0049124108314720>
- Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, 44(2), 347–357. <http://doi.org/10.1111/j.2044-8317.1991.tb00966.x>
- Cudeck, R., & Henly, S. J. (1991). Model selection in covariance structures analysis and the “problem” of sample size: A clarification. *Psychological Bulletin*, 109(3), 512–519.
<http://doi.org/10.1037/0033-2909.109.3.512>
- 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*, 32(2), 208–252. <http://doi.org/10.1177/0049124103256130>
- Curran, P., West, S., & Finch, J. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. *Psychological Methods*, 1(1), 16–29.
<http://doi.org/10.1037/1082-989X.1.1.16>
- D’Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2002). Manual of the Social Problem-Solving Inventory-Revised. North Tonawanda, NY: Multi-Health Systems, Inc.
- D’Zurilla, T. J., Nezu, A. M., & Maydeu-Olivares, A. (2004). Social problem solving: Theory and assessment. In E. C. Chang, T. J. D’Zurilla, & L. J. Sanna (Eds.), *Social problem solving: Theory, research, and training* (pp. 11–27). Washington, DC, US: American Psychological Association. <http://doi.org/10.1037/10805-000>

- Fan, X., Thompson, B., & Wang, L. (1999). Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 56–83.
<http://doi.org/10.1080/10705519909540119>
- Foldnes, N., & Olsson, U. H. (2015). Correcting Too Much or Too Little? The Performance of Three Chi-Square Corrections. *Multivariate Behavioral Research*, 50(5), 533–543.
<http://doi.org/10.1080/00273171.2015.1036964>
- Forero, C. G., & Maydeu-Olivares, A. (2009). Estimation of IRT graded response models: limited versus full information methods. *Psychological Methods*, 14(3), 275–99.
<http://doi.org/10.1037/a0015825>
- Hoelter, J. W. (1983). The Analysis of Covariance Structures: Goodness-of-Fit Indices. *Sociological Methods & Research*, 11(3), 325–344.
<http://doi.org/10.1177/0049124183011003003>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55.
<http://doi.org/10.1080/10705519909540118>
- Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, 112(2), 351–362. <http://doi.org/10.1037/0033-2909.112.2.351>
- Jackson, D. L., Gillaspay, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods*, 14(1), 6–23. <http://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: A Multidisciplinary Journal*, 10(3), 333–351. http://doi.org/10.1207/S15328007SEM1003_1
- Li, L., & Bentler, P. M. (2006). *Robust statistical tests for evaluating the hypothesis of close fit of misspecified mean and covariance structural models* (UCLA Statistics Preprint No. 506). Los Angeles, CA.
- Lomax, R. G., & Schumacker, R. E. (2004). *A beginner's guide to structural equation modeling* (3rd ed., Vol. 200). Mahwah, NJ: Lawrence Erlbaum.
- <http://doi.org/10.4135/9781412973380.n17>
- Maydeu-Olivares, A. (2017a). Assessing the Size of Model Misfit in Structural Equation Models. *Psychometrika*, 82(3), 533–558. <http://doi.org/10.1007/s11336-016-9552-7>
- Maydeu-Olivares, A. (2017b). Maximum likelihood estimation of structural equation models for continuous data: Standard errors and goodness of fit. *Structural Equation Modeling*, 24(3), 383–394. <http://doi.org/10.1080/10705511.2016.1269606>
- Maydeu-Olivares, A., Coffman, D. L., & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods*, 12(2), 157–176. <http://doi.org/10.1037/1082-989X.12.2.157>
- Maydeu-Olivares, A., & D’Zurilla, T. J. (1996). A factor-analytic study of the Social Problem-Solving Inventory: An integration of theory and data. *Cognitive Therapy and Research*, 20(2), 115–133. <http://doi.org/10.1007/BF02228030>
- Maydeu-Olivares, A., Fairchild, A. J., & Hall, A. G. (2017). Goodness of Fit in Item Factor Analysis: Effect of the Number of Response Alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(4), 495–505. <http://doi.org/10.1080/10705511.2017.1289816>
- Maydeu-Olivares, A., Shi, D., & Rosseel, Y. (2018). Assessing Fit in Structural Equation

- Models: A Monte-Carlo Evaluation of RMSEA Versus SRMR Confidence Intervals and Tests of Close Fit. *Structural Equation Modeling: A Multidisciplinary Journal*, 25(3), 389–402. <http://doi.org/10.1080/10705511.2017.1389611>
- McDonald, R. P., & Ho, M.-H. R. (2002). Principles and practice in reporting structural equation analyses. *Psychological Methods*, 7(1), 64–82. <http://doi.org/10.1037//1082-989X.7.1.64>
- 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. <http://doi.org/10.1080/10705511.2012.634724>
- Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, 38(2), 171–189. <http://doi.org/10.1111/j.2044-8317.1985.tb00832.x>
- Muthén, L. K., & Muthén, B. (2017). MPLUS 8 [Computer program]. Los Angeles, CA: Muthén & Muthén.
- Nevitt, J., & Hancock, G. R. (2000). Improving the Root Mean Square Error of Approximation for Nonnormal Conditions in Structural Equation Modeling. *The Journal of Experimental Education*, 68(3), 251–268. <http://doi.org/10.1080/00220970009600095>
- Rhemtulla, M., Brosseau-Liard, P. É. E., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, 17(3), 354–73. <http://doi.org/10.1037/a0029315>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36. <http://doi.org/10.18637/jss.v048.i02>
- Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in

- covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variable analysis. Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
- Savalei, V. (2010). Expected versus observed information in SEM with incomplete normal and nonnormal data. *Psychological Methods*, 15(4), 352–67. <http://doi.org/10.1037/a0020143>
- 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. <http://doi.org/10.1177/0013164412452564>
- Savalei, V. (2018). On the Computation of the RMSEA and CFI from the Mean-And-Variance Corrected Test Statistic with Nonnormal Data in SEM. *Multivariate Behavioral Research*, 0(0), 1–11. <http://doi.org/10.1080/00273171.2018.1455142>
- Shi, D., Lee, T., & Maydeu-Olivares, A. (2019). Understanding the Model Size Effect on SEM Fit Indices. *Educational and Psychological Measurement*, 79(2), 310–334. <http://doi.org/10.1177/0013164418783530>
- Steiger, J. H. (1990). Structural Model Evaluation and Modification: An Interval Estimation Approach. *Multivariate Behavioral Research*, 25(2), 173–180. http://doi.org/10.1207/s15327906mbr2502_4
- Steiger, J. H., & Lind, J. C. (1980). Statistically-based tests for the number of common factors. Iowa: Paper presented at the Annual Meeting of the Psychometric Society.
- West, S. G., Finch, J. F., & Curran, P. J. (1995). Structural equation models with nonnormal variables: Problems and remedies. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 56–75). Thousand Oaks, CA: Sage.
- Wolf, E. J., Harrington, K. M., Clark, S. L., & Miller, M. W. (2013). Sample Size Requirements

for Structural Equation Models. *Educational and Psychological Measurement*, 73(6), 913–934. <http://doi.org/10.1177/0013164413495237>

Yuan, K.-H., & Bentler, P. M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociological Methodology*, 30(1), 165–200. <http://doi.org/10.1111/0081-1750.00078>

Footnotes

¹ Some exogenous variables can be dummy variables, as in regression.

² There are two versions of the mean and variance corrected statistic, one proposed by Satorra and Bentler (1994) and another proposed by Asparouhov and Muthén (2010). Here we use the latter, as it is simpler to describe. In practice, differences between both statistics are very small (Foldnes & Olsson, 2015).

³ The sample RMSEA (4) is not an unbiased estimator of the population RMSEA (2) when data are normally distributed. It is a consistent estimator, in other words, it will converge to the population value as sample size increases. With non-normal data, the “naïve” approach results in a quantity that does not converge to the population RMSEA (2) as sample size increases; it converges to a different population quantity.

⁴ For mean corrected statistics, the constant c is generally printed by software programs. For mean and variance corrections it is not currently printed by software programs.

⁵ The results of this additional set of simulations is available from the authors upon request.

⁶ The correct formula was not available at the time.

Table 1. Formula for estimating the scaling correction c for different choices of corrections for goodness-of-fit test statistics

Choice	Suitable for	χ^2 statistic	Scaling correction, c
ML	Normal outcomes	likelihood ratio (LR) statistic	1
MLM	Non-normal outcomes	Satorra and Bentler (1994) mean-adjusted LR	$\frac{\chi^2_{ML}}{\chi^2_{MLM}}$
MLMV	Non-normal outcomes	Asparouhov and Muthén (2010) mean- and variance-adjusted LR	$\frac{\chi^2_{ML}}{\chi^2_{MLM}} - \frac{\chi^2_{MLM}}{\chi^2_{MLMV}}$
MLR	Non-normal outcomes	Asparouhov and Muthén (2005) mean-adjusted LR	$\frac{\chi^2_{ML}}{\chi^2_{MLR}}$

Table 2. Population RMSEA and Average of RMSEA estimates across Replications

ρ	N	Skew..	Kurt..	$p = 16$				$p = 32$			
				Pop.	MLM			Pop.	MLM		
					ML	& MLMV	MLR		ML	& MLMV	MLR
.4	200	3	21	.138	.151	.141	.147	.105	.125	.116	.122
	500	3	21	.138	.144	.138	.140	.105	.114	.107	.110
	1000	3	21	.138	.141	.137	.138	.105	.110	.105	.106
.8	200	3	21	.067	.104	.084	.094	.054	.096	.081	.090
	500	3	21	.067	.086	.071	.075	.054	.076	.063	.066
	1000	3	21	.067	.077	.068	.070	.054	.067	.057	.059
1	200	3	21	0	.085	.053	.069	0	.084	.064	.075
	500	3	21	0	.058	.026	.035	0	.057	.034	.041
	1000	3	21	0	.042	.014	.020	0	.042	.019	.024
.4	200	1	7	.138	.140	.138	.140	.105	.110	.108	.110
	500	1	7	.138	.139	.137	.138	.105	.107	.105	.106
	1000	1	7	.138	.138	.137	.138	.105	.106	.105	.105
.8	200	1	7	.067	.078	.072	.075	.054	.068	.064	.067
	500	1	7	.067	.072	.068	.069	.054	.060	.056	.057
	1000	1	7	.067	.069	.067	.067	.054	.058	.055	.055
1	200	1	7	0	.041	.026	.034	0	.044	.034	.041
	500	1	7	0	.027	.012	.016	0	.028	.015	.019
	1000	1	7	0	.020	.007	.009	0	.020	.008	.010
.4	200	0	0	.138	.136	.136	.138	.105	.105	.105	.106
	500	0	0	.138	.137	.137	.138	.105	.105	.105	.105
	1000	0	0	.138	.137	.137	.138	.105	.105	.105	.105
.8	200	0	0	.067	.068	.068	.068	.054	.057	.058	.058
	500	0	0	.067	.067	.067	.067	.054	.055	.055	.055
	1000	0	0	.067	.067	.067	.067	.054	.054	.054	.054
1	200	0	0	0	.014	.015	.015	0	.017	.018	.019
	500	0	0	0	.008	.008	.008	0	.007	.007	.007
	1000	0	0	0	.005	.005	.005	0	.004	.004	.004

Notes: pop = population value; ML = likelihood ratio (LR) test statistic; MLM = Satorra-Bentler (1994) mean adjusted LR; MLMV = Asparouhov and Muthén (2010) mean and variance adjusted LR, MLR = Asparouhov and Muthén (2005) mean adjusted LR.

Table 3. Coverage Rates for 90% Confidence Intervals around the Population RMSEA

ρ	N	Skew.	Kurt.	$p = 16$					$p = 32$				
				Pop.	ML	MLM	MLMV	MLR	Pop.	ML	MLM	MLMV	MLR
.4	200	3	21	.138	.425	.744	.924	.562	.105	.050	.341	.720	.143
	500	3	21	.138	.509	.750	.903	.665	.105	.157	.569	.872	.436
	1000	3	21	.138	.522	.763	.882	.678	.105	.233	.614	.861	.536
.8	200	3	21	.067	.056	.564	.868	.287	.054	.000	.025	.338	.002
	500	3	21	.067	.114	.752	.927	.608	.054	.000	.324	.792	.133
	1000	3	21	.067	.215	.784	.925	.721	.054	.002	.568	.891	.407
1	200	3	21	0	.001	.343	.833	.085	0	.000	.000	.135	.000
	500	3	21	0	.000	.661	.955	.342	0	.000	.041	.719	.003
	1000	3	21	0	.000	.805	.978	.586	0	.000	.256	.945	.065
.4	200	1	7	.138	.723	.780	.891	.705	.105	.489	.612	.878	.478
	500	1	7	.138	.716	.808	.889	.725	.105	.540	.643	.837	.581
	1000	1	7	.138	.702	.785	.849	.714	.105	.568	.641	.788	.602
.8	200	1	7	.067	.581	.778	.925	.666	.054	.123	.380	.824	.198
	500	1	7	.067	.650	.809	.908	.779	.054	.256	.656	.901	.563
	1000	1	7	.067	.688	.807	.879	.793	.054	.346	.707	.877	.672
1	200	1	7	0	.281	.734	.954	.501	0	.006	.122	.790	.026
	500	1	7	0	.235	.827	.973	.748	0	.005	.530	.965	.280
	1000	1	7	0	.204	.912	.974	.856	0	.001	.755	.985	.595
.4	200	0	0	.138	.764	.765	.828	.741	.105	.659	.656	.843	.622
	500	0	0	.138	.755	.778	.805	.723	.105	.641	.642	.732	.615
	1000	0	0	.138	.757	.777	.794	.714	.105	.634	.634	.686	.611
.8	200	0	0	.067	.824	.819	.888	.817	.054	.714	.690	.903	.684
	500	0	0	.067	.825	.827	.853	.825	.054	.738	.734	.842	.734
	1000	0	0	.067	.817	.816	.830	.815	.054	.739	.740	.794	.738
1	200	0	0	0	.914	.901	.956	.899	0	.708	.660	.960	.658
	500	0	0	0	.936	.933	.954	.932	0	.884	.875	.963	.874
	1000	0	0	0	.942	.941	.950	.941	0	.925	.920	.961	.920

Notes: skew = skewness, kurt = excess kurtosis, pop = population value; ML = likelihood ratio (LR) test statistic; MLM = Satorra-Bentler (1994) mean adjusted LR; MLMV = Asparouhov and Muthén (2010) mean and variance adjusted LR, MLR = Asparouhov and Muthén (2005) mean adjusted LR. Shaded results indicate acceptable coverage (between .85 and .95).

Table 4. Test of close fit results. Empirical rejection rates at a 5% significance level of a test that the RMSEA is less than or equal to its population value

ρ	N	Skew.	Kurt.	$p = 16$					$p = 32$				
				Pop.	ML	MLM	MLMV	MLR	Pop.	ML	MLM	MLMV	MLR
.4	200	3	21	.138	.550	.186	.060	.393	.105	.950	.659	.289	.861
	500	3	21	.138	.432	.122	.047	.228	.105	.841	.331	.113	.508
	1000	3	21	.138	.374	.094	.043	.187	.105	.737	.228	.083	.337
.8	200	3	21	.067	.945	.431	.134	.717	.054	1.00	.977	.671	.998
	500	3	21	.067	.887	.208	.067	.371	.054	1.00	.676	.212	.869
	1000	3	21	.067	.780	.141	.054	.220	.054	.998	.405	.109	.582
1	200	3	21	0	.999	.658	.171	.915	0	1.00	1.00	.870	1.00
	500	3	21	0	1.00	.340	.045	.658	0	1.00	.959	.284	.997
	1000	3	21	0	1.00	.196	.022	.414	0	1.00	.744	.055	.935
.4	200	1	7	.138	.194	.113	.052	.210	.105	.480	.329	.116	.485
	500	1	7	.138	.175	.065	.033	.152	.105	.366	.193	.087	.289
	1000	1	7	.138	.174	.081	.053	.158	.105	.314	.170	.101	.230
.8	200	1	7	.067	.413	.192	.072	.316	.054	.881	.625	.184	.807
	500	1	7	.067	.324	.123	.061	.163	.054	.745	.306	.098	.418
	1000	1	7	.067	.267	.102	.067	.118	.054	.646	.225	.100	.270
1	200	1	7	0	.720	.268	.049	.501	0	.994	.879	.215	.974
	500	1	7	0	.767	.129	.027	.253	0	.995	.471	.035	.722
	1000	1	7	0	.798	.088	.027	.145	0	.999	.245	.015	.406
.4	200	0	0	.138	.096	.099	.072	.138	.105	.198	.211	.100	.261
	500	0	0	.138	.097	.079	.068	.140	.105	.158	.161	.119	.207
	1000	0	0	.138	.100	.080	.074	.147	.105	.162	.163	.139	.202
.8	200	0	0	.067	.115	.122	.084	.126	.054	.261	.293	.102	.298
	500	0	0	.067	.092	.094	.080	.096	.054	.183	.189	.117	.192
	1000	0	0	.067	.088	.090	.083	.091	.054	.172	.175	.139	.176
1	200	0	0	0	.086	.101	.045	.103	0	.295	.343	.041	.345
	500	0	0	0	.065	.067	.046	.068	0	.116	.127	.037	.126
	1000	0	0	0	.058	.060	.050	.059	0	.075	.080	.039	.080

Notes: pop = population value; ML = likelihood ratio (LR) test statistic; MLM = Satorra-Bentler (1994) mean adjusted LR; MLMV = Asparouhov and Muthén (2010) mean and variance adjusted LR, MLR = Asparouhov and Muthén (2005) mean adjusted LR. Shaded results indicate acceptable rejection rates (between .03 and .08).

Table 5. A five factor confirmatory factor analysis model fitted to the short version of the SPSSI-R: RMSEA parameter estimates, 90% confidence intervals, and p-values for a test that RMSEA \leq .05 for different choices of overall test statistics with ML estimates

Estimator	χ^2	<i>c</i>	Naïve			Correct		
			Est	90% CI	<i>p</i>	Est	90% CI	<i>p</i>
ML	933.527	1	<i>na</i>	<i>na</i>	<i>na</i>	.050	(.046, .053)	.544
MLM	798.599	1.1690	.044	(.041, .048)	.995	.048	(.044, .052)	.796
MLMV	692.512	1.4590	.040	(.036, .043)	1	.048	(.043, .052)	.765
MLR	799.497	1.1676	.044	(.041, .048)	.995	.048	(.044, .052)	.795

Notes: $N = 1,020$; $df = 265$; *na* = not applicable. The naïve approach is currently implemented in most software packages, it uses Browne and Cudeck's formula (4), (6) and (7) regardless of choice of test statistic. The correct results are obtained using formula (8), (9), and (10) –see also Table 1.