THE JOURNAL OF EXPERIMENTAL EDUCATION, *80*(1), 26–44, 2012
Copyright © Taylor & Francis Group, LLC
ISSN: 0022-0973 print/1940-0683 online
DOI: 10.1080/00220973.2010.531299

Using the Modification Index and Standardized Expected Parameter Change for Model Modification

Tiffany A. Whittaker
The University of Texas at Austin

Model modification is oftentimes conducted after discovering a badly fitting structural equation model. During the modification process, the modification index (MI) and the standardized expected parameter change (SEPC) are 2 statistics that may be used to aid in the selection of parameters to add to a model to improve the fit. The purpose of this study was to extend the literature by examining the performance of the MI and the SEPC used independently and in conjunction with one another in terms of arriving at the correct confirmatory factor model. The results indicated that, in general, the SEPC outperformed the MI when arriving at the correct confirmatory factor model. However, they performed more similarly as factor loading size, sample size, and misspecified parameter size increased. The author provides recommendations on when the MI and SEPC perform more optimally.

Keywords *model misspecification, model modification, modification index, specification searches, standardized expected parameter change*

WHEN TESTING STRUCTURAL EQUATION MODELS, various statistics (e.g., the chi-square test statistic, χ^2) and fit indices (e.g., the comparative fit index) are examined to determine whether a theoretical model fits the data adequately. When these statistics and/or fit indices suggest inadequate fit of a structural equation model, the model may be modified, or respecified, followed by retesting of the modified model (MacCallum, Roznowski, & Necowitz, 1992). Because structural equation models are a priori models and based on hypothesized theoretical relations among observed and latent variables, poor fit of a model may be seen by some as evidence that the theoretical model is not plausible and/or poorly conceived and, thus, model modification should not be conducted. In contrast, some may view poor fit of a model as evidence of specification errors in the model, which would indicate a discrepancy between the theoretically plausible model hypothesized and the true model in the population.

Specification errors may be the inclusion of irrelevant relations or the exclusion of relevant relations (MacCallum, 1986). Most often, poor fit of a model constitutes the exclusion of relevant relations and is considered by some to have more serious consequences than the inclusion of irrelevant relations (Saris, Satorra, & van der Veld, 2009). As a result, applied researchers may add parameters to the initial model to improve fit. Although true structural equation models are unknown in practice and hypothesized models only represent approximations of the truth (Cudeck

Address correspondence to Tiffany A. Whittaker, Educational Psychology Department, The University of Texas at Austin, 1 University Station, D5800, SZB 504, Austin, TX 78712, USA. E-mail: tiffany.whittaker@mail.utexas.edu

& Browne, 1983), some may believe that model modification should be conducted in order to find a model that acceptably represents the relations among observed and latent variables (Saris et al., 2009).

Regardless of the differing views taken about model modification, it is commonly agreed upon that modification of a structural equation model is no longer confirmatory or a priori in nature, but rather exploratory. Similar to subset selection in multiple regression, modifying and retesting structural equation models on the basis of model fit indices capitalizes on the chance occurrences within the sample in which the models are being tested. Thus, when models are modified, they should be done so on the basis of relevant theory. In addition, cross-validation is highly recommended to help ensure the predictive validity of modified structural equation models (MacCallum et al., 1992).

Modification Index and Model Modification

In the process of model respecification or modification, various types of information may be evaluated and relied upon to aid in the selection of parameters to add to a model in order to improve the fit. The most popular piece of information that is evaluated is the modification index (MI) or Lagrange multiplier, which provides an estimated value in which the model's chi-square (χ^2) test statistic would decrease if a fixed parameter were added to the model and freely estimated (for a more detailed explanation about the computation of the MI, see Satorra, 1989 and Sörbom, 1989). Fixed parameters associated with a large MI value (e.g., larger than a χ^2 critical value of 3.84, which corresponds with 1 degree of freedom at an alpha level of .05) would then be examined to decide whether they would be theoretically plausible to include in the model and be freely estimated. In general, a specification search using the MI is carried out by first examining whether adding any fixed parameters to the model would significantly reduce the model's χ^2 test statistic. If so, researchers would typically examine the set of statistically significant, potential respecifications to determine which would lead to the largest decrease in the model's χ^2. If the respecification leading to the largest reduction in χ^2 is theoretically plausible, it could be added in order to improve model fit. This process would be repeated until adding any fixed parameters would not significantly reduce the model's χ^2 or until none of the statistically significant, potential respecifications are theoretically plausible to include in the model (Bollen, 1989).

Research concerning the accuracy of the MI in terms of helping researchers arrive at the correct structural equation model has demonstrated that the MI is not as accurate as one would hope. For example, MacCallum (1986) conducted a simulation study in which the performance of the MI in terms of arriving at the correct structural model was examined under various conditions. Four different types of misspecified (incorrect) structural models were fitted to two different data sets that were created according to the correctly specified (true) structural model under small ($n = 100$) and moderately small ($n = 300$) sample sizes. When fitting each of the misspecified (incorrect) structural models to the data, the results of the MI were recorded and the model was retested after adding the parameter suggested by the largest, statistically significant MI using an alpha of .01. Restricted specification searches were conducted in which only theoretically reasonable parameters (i.e., parameters actually in the true model) suggested by the MI were added. Unrestricted specification searches were also conducted in which all of the parameters suggested by the MI were added. The initial models represented different levels of misspecification and contained either one or two specification errors or parameter departures from the true structural model. Depending on the initial structural model tested, the MI should have followed a certain pattern in terms of identifying correct paths to be added to the model to significantly improve the fit of the model and ultimately arrive at the true model. The findings indicated that specification searches based on the MI are unsuccessful when the initial model contains a higher number of specification errors. Restricted searches led to more successful outcomes than did unrestricted searches. In addition, there were no successful outcomes under the small sample size of 100. Overall, the results were discouraging and MacCallum (1986) cautioned researchers with respect to the confidence one should place on using the MI when modifying structural equation models. Studies conducted by Kaplan (1988) and Silvia and MacCallum (1988) supported MacCallum's (1986) findings.

In contrast, a simulation study conducted by Chou and Bentler (1990) found the performance of the MI to be more promising. That is, the MI correctly suggested freely estimating the set of three fixed parameters not freely estimated in the misspecified model in all 100 replications under different sample size conditions (100, 200, 400, and 800). Notwithstanding, the MI incorrectly suggested freely estimating a fixed parameter, in particular when sample size was moderate to large. The success of the MI in this study may be the result of severity of model misspecification when testing whether the set of fixed parameters should be freely estimated to significantly reduce the model's χ^2.

MacCallum et al. (1992) further demonstrated the instability of model modification. In two sampling studies, several sets of data were created by randomly selecting cases from an existing data set to form various sample size conditions (100, 150, 200, 250, 325, 400, 800, and 1200). Samples were used to fit an initial model to the data, modify the model using the MI, and refit the modified model to the data. This sequence (initial fit, modification, and refitting) was conducted 3–4 times. The results indicated that the particular model modifications in each sequence were inconsistent across replication samples, especially when sample size was small ($n = 100$).

In addition, Hutchinson's (1993) population study compared the performance of the univariate MI and the multivariate MI (which considers freely estimating fixed parameters simultaneously) with respect to arriving at the correct model. Four increasingly misspecified models at measurement and structural levels were fitted to the data to examine which fixed parameters were suggested as being freed by the univariate and multivariate MI tests using a more stringent alpha of .01. The results indicated that the univariate and multivariate MI tests correctly suggested freeing parameters in the less misspecified model. However, both of the tests became less reliable as model misspecification increased. Overall, the multivariate MI test did not outperform its univariate counterpart. It must be noted that incorrect parameters were also included in the misspecified models in order to examine the performance of the Wald statistic (which estimates whether a parameter may be dropped from the model without significantly decreasing the fit of the model). Although this is a more realistic situation, model misspecification may have been confounded in this respect and, thus, may not be directly comparable to the aforementioned studies.

Expected Parameter Change and Model Modification

Hutchinson (1993) noted a possible limitation in her study with respect to not using the expected parameter change in conjunction with the modification indices which could have improved the success of specification searches and suggested further research using this combination approach.

The unstandardized expected parameter change (EPC) was first proposed by Saris, Satorra, and Sörbom (1987) to aid in the selection of parameters to add to model in order to improve model fit and to detect model misspecifications. The EPC indicates the estimated value of a fixed parameter if it were added to a model and freely estimated (for a more detailed explanation about the EPC, see Saris et al., 1987). Thus, the EPC provides "a direct estimate of the size of the misspecification for the restricted parameters" (Saris et al., 1987, p. 120). Similar to model modification using the MI, fixed parameters associated with the largest EPC value, indicating the most model misspecification, would be evaluated in terms of theoretical plausibility and could be freely estimated in the model. Again, this process would be repeated until the fixed parameters are not associated with large EPC values relative to others in the set or until none of the fixed parameters associated with large EPC values relative to others are theoretically plausible.

Saris et al. (1987) outlined four situations for model respecification in which the EPC could play an important role. First, when a large, significant MI coincides with a large EPC value for a parameter, the likely conclusion would be to freely estimate that parameter. Second, when a large, significant MI coincides with a small EPC value, freely estimating that parameter would not be recommended as this may be attributable to the sample size sensitivity of the MI. Third, when a small, nonsignificant MI coincides with a large EPC value, the decision is unclear and a power analysis is recommended. Fourth, when a small, nonsignificant MI coincides with a small EPC value for a parameter, it would not be recommended that the parameter be freely estimated. It must be noted that cutoff criteria were not suggested by Saris et al. (1987) in terms of how large an EPC value should be when deciding whether to add a particular parameter.

Luijben and Boomsma (1988) conducted a simulation study that compared the performance of the MI to that of the EPC under different sample size conditions ($n = 200$ and $n = 400$), different factor loading sizes (.40, .60, and .80), different factor loading structures (where no observed variables cross-loaded on more than one factor and where some of the observed variables loaded on two factors), and different factor intercorrelations (.30, .40, and .50). The data were generated according to a correlated two-factor confirmatory model. A misspecified model in which the two factors were not allowed to correlate was fit to the generated data. The number of times the largest MI value and the largest EPC value was associated with the correlation between the two factors was tallied out of 300 replications. The findings indicated that the MI and the EPC both improved with respect to correctly suggesting estimation of the factor intercorrelation as sample size increased, as factor loading size increased, and as the size of the correlation between the two factors increased. In addition, the EPC outperformed the MI in terms of indicating that the factor intercorrelation should be estimated in the model.

Luijben and Boomsma (1988) also demonstrated that the EPC, in its unstandardized version as was used in their study, is dependent upon the scaling of the variables in the model. In particular, the factor variances were set to a value of 1.0 and the factor loadings freely estimated in their study when examining the performance of the EPC as presented above. They subsequently conducted a smaller scale simulation study in which the path of the first indicator variable for each of the two factors was set to a value of 1.0 to examine the scale dependency of the unstandardized EPC. Thus, the factor variances were freely estimated when fitting the misspecified, uncorrelated factor model to the data. In 300 replications of sample sizes of 200, the unstandardized EPC never indicated that the missing factor intercorrelation should be estimated. Because of the scale noninvariance of the unstandardized EPC, reasonable comparisons among a collection of suggested fixed parameters will prove to be difficult.

Kaplan (1989) proposed a standardized version of the EPC (referred to here as P-SEPC) and examined the use of the MI, EPC, and the proposed standardized EPC (P-SEPC) in conjunction with one another in specification searches for two models fitted to existing data sets. Using the combined criteria in which the MI and the EPC/P-SEPC are substantially large for a fixed parameter, Kaplan (1989) found significant improvements in model fit after these parameters were freely estimated. Kaplan (1989) also concluded that the MI suggested estimating unlikely parameters, whereas the EPC and P-SEPC suggested estimating reasonable parameters in a model. It is important to note that a cutoff value greater than .10 for the EPC was regarded as representing a substantially large parameter estimate value in this study.

Nonetheless, the proposed, partially standardized version of the EPC (P-SEPC; Kaplan, 1989) still depends on the metrics of variables in the model (see Chou & Bentler, 1993 and Luijben, 1989). Thus, Chou and Bentler (1993) suggested fully standardizing the EPC, now referred to as the SEPC, which is invariant to the rescaling of observed and latent variables (for a more detailed explanation, see Chou & Bentler, 1993). The major difference between the two standardizations is that SEPC values for residual covariances between latent variables and between observed variables are standardized using their respective residual variances instead of their variances as in the P-SEPC standardization. That is, everything is standardized to unit variance under the SEPC standardization, whereas everything but the residual variables is standardized to unit variance under the P-SEPC standardization.

The fully standardized EPC (SEPC) has been available in the EQS software package since its implementation and is labeled *standardized change*. The SEPC became available in version 5 of the Mplus software package (2007; referred to as "StdYX E.P.C."). The P-SEPC was calculated in previous versions of Mplus. The P-SEPC is calculated in the current version of LISREL (Version 8) and is called the *completely standardized expected change*. In addition, the unstandardized EPC is provided with MI tests in the current version of AMOS (Version 18), which is termed *par change*.

The Use of the MI and EPC in Applied Research

To assess the frequency in which applied researchers use the EPC in conjunction with the MI, I conducted a cited reference search in the Social Sciences Citation Index from which I collected and briefly examined articles that were published in English within the past 10 years citing Kaplan (1989) or Saris et al. (1987) given that they had originally recommended the use of the MI with the EPC. The cited reference search discovered a total of 31 articles. Because the interest of this review was the use of the EPC in combination with the MI in applied research, I excluded discussion/review pieces and simulation/methodological studies, resulting in 17 applied articles.

The MI was used solely in four applied studies, and the MI was used in conjunction with the EPC or the P-SEPC in the remaining 13 applied studies. In the majority of the 13 articles in which the combination criteria (i.e., MI with the EPC/P-SEPC) was used, the authors simply alluded to examining high values of the MI associated with high values of the EPC/P-SEPC when modifying the initial structural equation model. Hence, cutoff values were not explicitly mentioned in these articles. Nonetheless, in 2 of these 13 articles, the authors indicated that the P-SEPC values were either too low (e.g., .153) or that the path recommended by the P-SEPC would be in the opposite direction than expected (i.e., a negative loading versus a positive loading) and, thus, theoretically unreasonable. In addition, the authors of one article indicated that a P-SEPC value of .17 associated with a correlation between item residuals was substantial enough to include in the respecified model, and the authors of yet another article indicated that EPC values of .172 and .118 associated with two direct path loadings were indicative of model misspecification and were thus estimated in the respecified model.

Purpose of the Study

To summarize, the preponderance of research concerning the MI has generally demonstrated less than acceptable performance when attempting to arrive at the correct structural equation model. In contrast, the research concerning the performance of the EPC has been more promising. Kaplan (1989) found that using the MI in conjunction with the EPC/P-SEPC, as proposed by Saris et al. (1987), led to significant model improvements and more meaningful respecifications. Nonetheless, Kaplan (1989) used this combination criteria (MI with EPC/P-SEPC) when fitting models to existing data sets in which the true model is unknown. Although we never know truth, it would be useful for applied researchers to understand under what conditions the EPC performs more optimally. In addition, the EPC and P-SEPC, as used in Kaplan's (1989) study, are both scale noninvariant, which may have resulted in unreasonable comparisons among the set of fixed parameters suggested by the ECP/P-SEPC. Luijben and Boomsma (1988) conducted a simulation study in which they found that the EPC outperformed the MI in terms of suggesting the correct fixed parameters to be freely estimated in a model. Again, the EPC used in their simulation study was unstandardized and depended on the scaling of latent variables in the model, as demonstrated in their study.

Saris et al. (2009) demonstrated how the SEPC can be sensitive to model parameters not related to the model misspecification. For example, using population data, they fit a model that excluded a correlation between two observed variables' residuals which was .20 in the population. They generated standardized data for a sample size of 400 in which different values of one other parameter in the model (a direct path) ranged from .1 to .9. The expected value of the misspecification (the excluded residual correlation) was .2 in all conditions and was deemed a substantial misspecification. The MI, the standard error of the EPC, the 95% confidence interval of the EPC, and the power of the MI test were calculated for each of the nine sets of population data. The findings of this illustration were that the power of the MI test increased, the standard error of the EPC became smaller, and, thus, the 95% confidence interval of the EPC became smaller as values of the irrelevant parameter became bigger. Nevertheless, the power of the MI test was relatively low for all the varied values of the irrelevant parameter in the model.

In the end, Saris et al. (2009) suggested examining the EPC in conjunction with the MI as well as the power associated with the MI test when trying to detect misspecification errors in a structural equation model. The cursory glance at the use of the MI and the EPC in applied research, however, indicated that the combination criteria are not widely used among applied researchers. To date, no comprehensive simulation study has examined the performance of the MI used in conjunction with the fully standardized version of the EPC (SEPC), which may lead to improved success when respecifying models (Hutchinson, 1993). Further, no comprehensive simulation study has examined the use of cutoff values when implementing the SEPC. Kaplan (1989) used an EPC cutoff value greater than .10 to represent a large parameter estimate and the authors of two applied articles indicated that certain EPC or P-SEPC values (ranging from .12 to .17) were suggestive of model misspecification. Thus, cutoff values of the SEPC may be implemented when respecifying models. When examining the SEPC of a fixed parameter, a standardized value of .2 or larger was suggested by Saris et al. (2009) as representing a large value and, thus, may begin to be implemented as a cutoff value of the SEPC in applied research. The dearth of research concerning the SEPC makes it difficult to properly evaluate its performance. Given the small number of studies in this area, the purpose of this study was to extend the literature by examining the performance of the MI and the standardized EPC (SEPC) used independently and in conjunction with one another as well as applying the use of the SEPC suggested cutoff value of .20 when identifying model misspecification.

METHOD

I conducted a simulation study to evaluate the performance of the MI and SEPC under varied sample size, factor loading size, and misspecified parameter size conditions. The EQS Structural Program Software (Version 6.1; Bentler, 1995) was used to generate normally distributed, raw data according to a correlated two-factor model. The correct factor model used in the present study was similar to the model used in Luijben and Boomsma (1988) with the exception of the number of indicator variables per factor. In particular, each factor was indicated by four observed variables in the present study as opposed to three indicator variables per factor as used in Luijben and Boomsma's (1988) study. Data for the true, correlated two-factor model (see Figure 1) were generated given specific sample sizes, factor loading sizes, and factor intercorrelation sizes as subsequently described.

Study Conditions

Sample Size (n)

Large sample sizes are necessary when testing structural equation models. Sample sizes of 200 are generally recommended to avoid convergence failure and improper solutions (Loehlin, 1998). However, some researchers use sample sizes as small as 100 (e.g., see MacCallum & Austin, 2000). Various sample size conditions were examined in the present study and were based on Breckler's (1990) review of 72 structural equation modeling studies published in personality and social psychology journals. The range of sample sizes used in these studies was between 40 and 8,650. Sample sizes greater than 500 were used in 25% of these studies, whereas sample sizes between 100 and 500 were used in approximately 50% of the studies. Thus, data were generated for sample sizes of 100, 200, 400, 600, and 800 to represent a range of realistic sample sizes.

Factor Loading Size (λ)

Standardized factor loading sizes of .40, .50, .60, and .70 were examined to represent moderate to high factor loading sizes. Factor loadings equal to or greater than .30 or .40 are commonly deemed to be meaningful to interpret (Brown, 2006). However, it has been suggested that factor loadings lower than .40 are typically disregarded (Saris et al., 2009) and too low with respect to factor reliability. Thus, the factor loadings in the present study were chosen to represent fair to

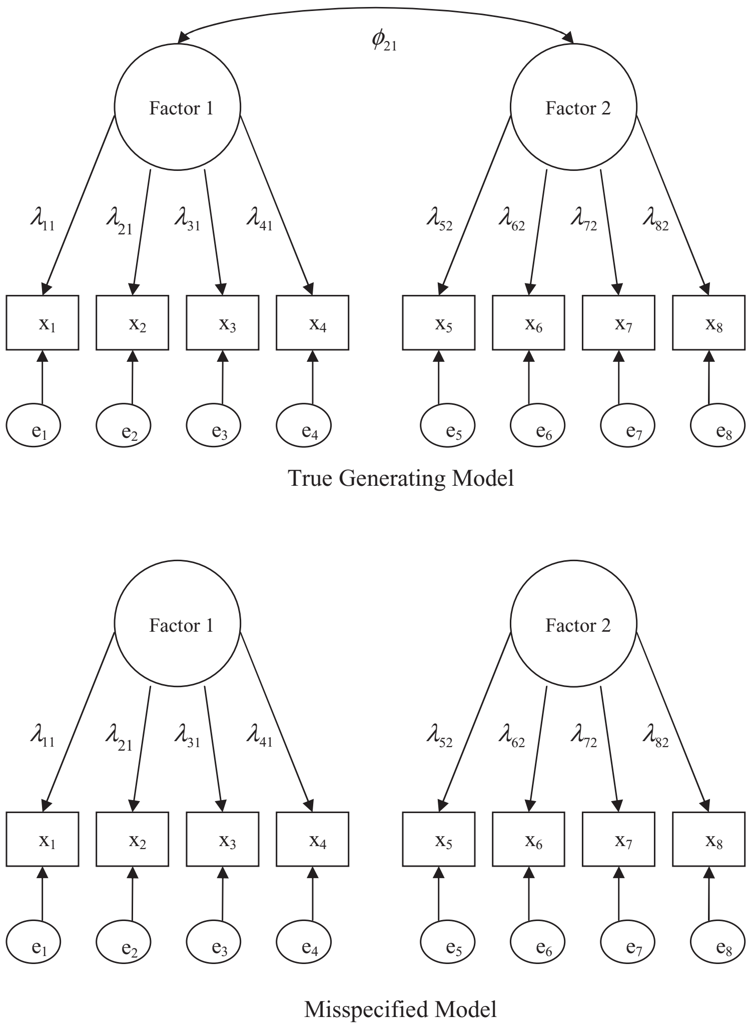

FIGURE 1 Correlated two-factor generating model and uncorrelated two-factor misspecified model.

acceptable factor reliability (.16 to .49). All observed variables loaded equally on each factor in a given condition. Error paths and factor variances were set to equal 1.0 to provide the scale of measurement.

Factor Intercorrelation Size (ϕ)

The correlation between factors was also varied at .20, .30, .40, .50, .60, and .70. These values were selected in order to examine a full range of plausible and practically meaningful factor intercorrelations. In terms of statistics, a factor intercorrelation of .20 may be considered too low to include in a structural equation model, rendering any results under this condition immaterial. In practice, however, a factor intercorrelation of .20 may provide enough support for a theoretically plausible relation in a structural equation model and an SEPC value of .20 or larger was deemed as substantial by Saris et al. (2009).

Procedure

For each combination of sample size and factor loading size, 2,000 sets of raw data were generated according the correct factor model. The misspecified model, which set the factor intercorrelation to zero, was fitted to the data in each replication using maximum likelihood estimation (see Figure 1). The focus on the factor intercorrelation was done to compare the findings in the present study with the findings of previous research and to serve as a baseline for future comparisons. If improper solutions were encountered, indicating that the correlations estimated may be greater than 1.0 or that estimated residual variances may be negative, or if the estimates failed to converge, additional data were generated to achieve 2,000 total replications. The number of iterations that were allowed for estimates to converge was set to 100.

Several outcome measures were of interest and were documented. These outcome measures included the following: (a) the number of times the largest and statistically significant ($p < .05$) MI was associated with the fixed intercorrelation between factors; (b) the number of times the MI associated with the fixed intercorrelation between factors was statistically significant at $p < .05$, though not necessarily the largest MI value relative to others; (c) the number of times the SEPC value was the largest for the fixed intercorrelation between factors; (d) the number of times the SEPC value for the fixed intercorrelation between factors was greater than the recommended cutoff of .20 suggested by Saris et al. (2009); and (e) the number of times the MI associated with the factor intercorrelation was statistically significant ($p < .05$) and the SEPC value was the largest for the fixed intercorrelation between factors. It must be noted that the default Lagrange multiplier test in EQS was specified to search for possible fixed parameters to be freed. Thus, residual covariances were not examined as possible fixed parameters to be released in the present study. The P-SEPC was not included for comparison in the present study because it is considered to be a partially standardized value which may lead to different conclusions depending upon the scaling of variables in the model (Chou & Bentler, 1993; Luijben, 1989).

RESULTS

The results are displayed in Tables 1 through 4 as a function of factor intercorrelation size and sample size at each factor loading size condition. Columns 1 through 4 in the Tables 1 through 4 illustrate the selection rates of the factor intercorrelation parameter when using the MI and SEPC independently, corresponding to outcome measures *a* through *d* described earlier in the Procedure section. Column 5 in Tables 1 through 4 illustrate the selection rates of the factor intercorrelation parameter in each condition under the joint criteria when using the MI and SEPC in conjunction with one another, corresponding to outcome measure *e* described above. When examining the results in Tables 1 through 4, the accuracy of the MI, the SEPC, and the combination of the MI and the SEPC was deemed to be adequate if they were associated with the factor intercorrelation in at least 70% of the 2,000 replications in a given condition.

TABLE 1
Percentage of Times (Out of 2,000 Replications) Each Criterion Indicated the Factor Intercorrelation be Freely Estimated at Factor Loading Size of .40 as a Function of Factor Intercorrelation and Sample Size

Condition	*Largest, sig. MI*	*Sig. MI*	*Largest SEPC*	*SEPC ≥.20 cutoff value*	*Sig. MI and largest SEPC*
$\lambda = .40/\phi_{21} = .20$					
$n = 100$	7.30	11.95	32.30	40.40	10.90
$n = 200$	13.65	20.20	45.15	40.75	19.50
$n = 400$	25.90	39.40	64.75	45.00	38.25
$n = 600$	37.20	51.80	77.25	44.80	50.95
$n = 800$	49.90	67.65	87.00	47.25	67.05
$\lambda = .40/\phi_{21} = .30$					
$n = 100$	13.45	20.60	41.85	52.90	20.15
$n = 200$	24.20	38.25	63.65	63.25	37.50
$n = 400$	52.50	71.20	86.50	76.15	70.25
$n = 600$	70.90	87.15	95.15	81.95	86.65
$n = 800$	82.80	94.85	98.10	86.55	94.65
$\lambda = .40/\phi_{21} = .40$					
$n = 100$	20.45	31.55	53.75	67.10	30.75
$n = 200$	44.85	62.85	80.30	81.65	61.40
$n = 400$	77.95	91.75	96.70	94.00	91.25
$n = 600$	90.65	98.60	99.15	96.85	98.20
$n = 800$	96.50	99.80	99.85	99.15	99.70
$\lambda = .40/\phi_{21} = .50$					
$n = 100$	32.95	47.30	66.15	79.10	45.60
$n = 200$	63.90	81.15	89.60	92.30	79.85
$n = 400$	90.95	98.75	99.65	99.45	98.65
$n = 600$	97.80	99.90	100.00	99.80	99.90
$n = 800$	99.65	100.00	100.00	100.00	100.00
$\lambda = .40/\phi_{21} = .60$					
$n = 100$	43.30	60.80	76.00	87.60	58.75
$n = 200$	77.05	91.40	95.05	97.40	90.35
$n = 400$	96.95	99.75	99.80	99.80	99.65
$n = 600$	99.40	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .40/\phi_{21} = .70$					
$n = 100$	58.20	76.40	84.40	94.10	73.85
$n = 200$	85.30	96.90	97.05	99.15	95.55
$n = 400$	99.55	100.00	100.00	100.00	100.00
$n = 600$	99.95	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00

Note. Largest, sig. MI = largest and statistically significant modification index; MI = modification index; SEPC = standardized expected parameter change.

Performance of the Modification Index

In the first column of Tables 1 through 4, the number of times out of 2,000 replications in which the MI value associated with the correct parameter (the factor intercorrelation) was the largest and was statistically significant ($p < .05$) was documented. This is similar to the outcome

TABLE 2
Percentage of Times (Out of 2,000 Replications) Each Criterion Indicated the Factor Intercorrelation be Freely Estimated at Factor Loading Size of .50 as a Function of Factor Intercorrelation and Sample Size

Condition	*Largest, sig. MI*	*Sig. MI*	*Largest SEPC*	*SEPC ≥.20 cutoff value*	*Sig. MI and largest SEPC*
$\lambda = .50/\phi_{21} = .20$					
$n = 100$	12.20	17.95	39.15	44.05	16.95
$n = 200$	25.20	34.60	57.30	45.30	33.40
$n = 400$	47.75	61.75	78.25	47.05	60.25
$n = 600$	64.30	79.40	90.35	46.95	78.10
$n = 800$	77.65	89.10	94.80	49.95	88.25
$\lambda = .50/\phi_{21} = .30$					
$n = 100$	25.55	36.20	58.90	66.70	35.00
$n = 200$	51.05	65.95	83.00	76.20	64.65
$n = 400$	80.75	91.25	96.45	85.30	90.60
$n = 600$	93.40	98.80	99.40	91.10	98.60
$n = 800$	97.75	99.85	99.80	95.00	99.70
$\lambda = .50/\phi_{21} = .40$					
$n = 100$	43.00	56.35	73.70	82.35	54.60
$n = 200$	76.50	87.55	93.60	93.40	86.55
$n = 400$	96.10	99.70	99.70	99.10	99.50
$n = 600$	99.30	100.00	99.95	99.85	99.95
$n = 800$	100.00	100.00	100.00	99.90	100.00
$\lambda = .50/\phi_{21} = .50$					
$n = 100$	62.90	77.40	85.70	93.25	75.10
$n = 200$	90.30	97.45	98.55	99.05	97.05
$n = 400$	99.30	100.00	100.00	99.95	100.00
$n = 600$	99.90	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .50/\phi_{21} = .60$					
$n = 100$	77.60	90.20	93.40	97.65	88.25
$n = 200$	96.85	99.50	99.70	99.85	99.40
$n = 400$	99.95	100.00	100.00	100.00	100.00
$n = 600$	99.95	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .50/\phi_{21} = .70$					
$n = 100$	88.15	96.35	96.45	99.40	94.85
$n = 200$	99.35	100.00	100.00	100.00	100.00
$n = 400$	100.00	100.00	100.00	100.00	100.00
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00

Note. MI = modification index; SEPC = standardized expected parameter change.

measured in MacCallum's (1986) study with the exception that an alpha of .05 was used in the present study. The largest, significant MI criterion performed the worst in the smallest factor loading ($\lambda = .40$), smallest factor intercorrelation ($\phi = .20$), and smallest sample size ($n = 100$) condition (see Table 1) wherein it was associated with the factor intercorrelation in only 7.30% of the replications. Further, the largest, significant MI criterion did not work well when sample size was 100. In particular, it reached 70% accuracy (identifying that the factor intercorrelation

TABLE 3
Percentage of Times (Out of 2,000 Replications) Each Criterion Indicated the Factor Intercorrelation be Freely Estimated at Factor Loading Size of .60 as a Function of Factor Intercorrelation and Sample Size

Condition	*Largest, sig. MI*	*Sig. MI*	*Largest SEPC*	*SEPC ≥.20 cutoff value*	*Sig. MI and largest SEPC*
$\lambda = .60/\phi_{21} = .20$					
$n = 100$	20.75	28.05	50.15	48.75	26.95
$n = 200$	38.50	49.30	69.30	49.25	48.05
$n = 400$	68.00	79.50	90.40	50.20	78.45
$n = 600$	83.95	93.30	96.60	48.95	92.55
$n = 800$	93.15	97.90	99.15	48.80	97.55
$\lambda = .60/\phi_{21} = .30$					
$n = 100$	41.10	52.70	72.55	73.45	51.15
$n = 200$	71.40	82.50	92.10	82.10	81.40
$n = 400$	94.40	98.55	99.45	91.15	98.20
$n = 600$	98.75	99.95	99.85	95.95	99.85
$n = 800$	99.75	100.00	100.00	98.20	100.00
$\lambda = .60/\phi_{21} = .40$					
$n = 100$	66.45	77.35	87.85	90.00	75.95
$n = 200$	92.35	97.95	99.10	98.25	97.70
$n = 400$	99.65	99.95	99.95	99.80	99.95
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .60/\phi_{21} = .50$					
$n = 100$	85.80	94.50	96.95	98.70	93.50
$n = 200$	98.55	99.90	99.95	99.90	99.85
$n = 400$	100.00	100.00	100.00	100.00	100.00
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .60/\phi_{21} = .60$					
$n = 100$	94.55	98.95	99.00	99.80	98.35
$n = 200$	99.85	100.00	100.00	100.00	100.00
$n = 400$	100.00	100.00	100.00	100.00	100.00
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .60/\phi_{21} = .70$					
$n = 100$	98.60	99.95	99.80	100.00	99.80
$n = 200$	99.95	100.00	100.00	100.00	100.00
$n = 400$	100.00	100.00	100.00	100.00	100.00
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00

Note. MI = modification index; SEPC = standardized expected parameter change.

should be estimated in 70% of the 2,000 replications) in less than half of the smallest sample size conditions. It can be seen, however, that as factor loading, factor intercorrelation, and sample size increased, so did the accuracy of the largest, significant MI criterion. The selection rates of the largest, significant MI criterion reached 70% accuracy in 90 (75%) of the 120 total conditions.

The second column in Tables 1 through 4 indicates the percentage of times the MI value associated with the factor intercorrelation was statistically significant at an alpha of .05, which is

TABLE 4
Percentage of Times (Out of 2,000 Replications) Each Criterion Indicated the Factor Intercorrelation be Freely Estimated at Factor Loading Size of .70 as a Function of Factor Intercorrelation and Sample Size

Condition	*Largest, sig. MI*	*Sig. MI*	*Largest SEPC*	*SEPC ≥.20 cutoff value*	*Sig. MI and largest SEPC*
$\lambda = .70/\phi_{21} = .20$					
$n = 100$	26.90	33.55	60.30	48.95	33.00
$n = 200$	51.25	61.35	81.75	49.55	60.60
$n = 400$	80.90	89.15	96.00	51.55	88.60
$n = 600$	93.75	97.25	99.00	50.95	97.10
$n = 800$	97.50	99.45	99.70	49.70	99.35
$\lambda = .70/\phi_{21} = .30$					
$n = 100$	57.15	66.45	85.55	79.30	65.90
$n = 200$	86.30	93.20	97.20	88.15	92.75
$n = 400$	98.60	99.65	99.80	94.30	99.50
$n = 600$	99.85	100.00	100.00	98.10	100.00
$n = 800$	100.00	100.00	100.00	99.20	100.00
$\lambda = .70/\phi_{21} = .40$					
$n = 100$	81.20	88.65	95.20	94.90	88.10
$n = 200$	97.60	99.40	99.80	98.80	99.35
$n = 400$	99.95	100.00	100.00	99.95	100.00
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .70/\phi_{21} = .50$					
$n = 100$	95.75	98.00	99.05	99.30	97.85
$n = 200$	99.85	100.00	100.00	100.00	100.00
$n = 400$	100.00	100.00	100.00	100.00	100.00
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .70/\phi_{21} = .60$					
$n = 100$	99.25	99.85	99.85	99.95	99.80
$n = 200$	100.00	100.00	100.00	100.00	100.00
$n = 400$	100.00	100.00	100.00	100.00	100.00
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00
$\lambda = .70/\phi_{21} = .70$					
$n = 100$	99.95	100.00	100.00	100.00	100.00
$n = 200$	100.00	100.00	100.00	100.00	100.00
$n = 400$	100.00	100.00	100.00	100.00	100.00
$n = 600$	100.00	100.00	100.00	100.00	100.00
$n = 800$	100.00	100.00	100.00	100.00	100.00

Note. MI = modification index; SEPC = standardized expected parameter change.

also a realistic consideration of applied researchers when determining whether to freely estimate a fixed parameter. As one would expect, the significant MI criterion outperformed the largest, significant MI criterion, correctly selecting the factor intercorrelation more often in each of the conditions, with the exception of conditions in which the largest, significant MI criterion and the significant MI criterion both reached 100% accuracy. Similar to the largest, significant MI criterion, the significant MI criterion did not perform well in the smallest factor loading, smallest factor intercorrelation, and smallest sample size condition (see Table 1) wherein the MI associated with the factor intercorrelation was statistically significant in only 11.95% of the replications. Likewise, the significant MI criterion did not tend to work well when sample size was 100. That is, it reached 70% accuracy (identifying that the factor intercorrelation should be estimated in 70% of the 2,000 replications) in only half of the smallest sample size conditions. The accuracy of the significant MI criterion also improved as factor loading, factor intercorrelation, and sample size increased, reaching 70% accuracy in 97 (80.8%) of the 120 total conditions.

Performance of the SEPC

The third column of Tables 1 through 4 indicates the percentage of times (out of 2,000 replications) the largest *positive* value of the SEPC ("Largest SEPC") was associated with the factor intercorrelation, indicating a large misspecification in the model. Only positive values were considered in order to mimic legitimate modifications to the model. This would correspond best with the more realistic situation in which a positive relation between factors would be expected and hypothesized given that the generating model included a positive correlation between factors. Similar to the MI, the SEPC did not perform well in the smallest factor loading ($\lambda = .40$), smallest factor intercorrelation ($\phi = .20$), and smallest sample size ($n = 100$) condition (see Table 1), but did improve as factor loading, factor intercorrelation, and sample size increased. For example, the largest SEPC was correctly associated with the factor intercorrelation in more than 70% of the 2,000 replications in 107 (89%) of the 120 conditions. In addition, the results illustrate that the SEPC did tend to outperform the MI in several conditions with respect to correctly selecting the factor intercorrelation.

SEPC Value Cutoff

Column 4 of Tables 1 through 4 includes the percentage of times the estimated SEPC value associated with the factor intercorrelation was equal to or greater than a positive value of .20. This was examined to represent the application of cutoff values when deciding whether to freely estimate a fixed parameter as suggested by Saris et al. (2009). Again, only positive values were considered to represent a valid modification to the model. The cutoff criterion performed better than the MI or largest SEPC with respect to correctly selecting the factor intercorrelation in the smallest sample size conditions until factor loadings reached a value of .70 (see Table 4). As would be expected, the SEPC cutoff criterion of .20 became more accurate as the factor intercorrelation increased and performed more similarly to the largest SEPC criterion as factor loading, factor intercorrelation, and sample size increased. Further, it reached 70% accuracy (identifying that the factor intercorrelation should be estimated in 70% of the 2,000 replications) in 80% of the 120 conditions.

Joint Performance of the MI and SEPC

Column 5 in Tables 1 through 4 contains the percentage of times that both the MI associated with the factor intercorrelation was statistically significant at an alpha of .05 and the largest positive value of the SEPC was associated with the factor intercorrelation. Again, the joint criteria did not perform well in the smallest factor loading, smallest factor intercorrelation, and smallest sample size condition (see Table 1). As one would expect given the previously discussed results, the more stringent joint criteria performed similarly to, but slightly worse than, the significant MI criterion in several conditions. The significant MI criterion and the largest positive SEPC criterion correctly agreed more with one another as factor loading, factor intercorrelation, and sample size increased, identifying that the factor intercorrelation should be estimated in 70% of the 2,000 replications in 96 (80%) of the 120 conditions.

Estimation Problems and Negative SEPC Values

As noted previously, if improper solutions or convergence problems were encountered in a replication, additional data were generated to achieve 2,000 total replications. Improper solutions would indicate that a standardized estimate may be greater than 1.0 and that the variances estimated may be negative (i.e., Heywood cases). Convergence problems would indicate that the estimates failed to converge on the model parameters before reaching 100 iterations. Convergence problems did not occur in the present study, however, improper solutions did occur under certain conditions. Although these problems were generally encountered infrequently (never resulting in fewer than 81% usable replications), they occurred more often when sample size was equal to 100 in low factor loading conditions (.40 and .50) as well as when sample size was equal to 200 in the lowest factor loading condition (.40), regardless of factor intercorrelation. Improper solutions decreased as factor loading and sample size increased.

When examining the performance of the SEPC, only positive values were considered in order to emulate the behavior of an applied researcher when considering the addition of valid parameters to a model. It must be noted that some of the estimated SEPC values were negative in certain conditions, though infrequently. Negative SEPC values were estimated more often when sample size was 100 in the smaller factor intercorrelation conditions ($\phi = .20$, .30, and .40).

Test Conditions

The results presented above are based on the misspecification (omission) of a factor intercorrelation in a simple, correlated two-factor confirmatory factor analysis (CFA) model. Models in the applied literature are oftentimes more complex and may contain more than one model misspecification. As a result, applied researchers may question whether the MI and/or SEPC might perform differently if the model is more complex and/or additional parameters are excluded from the model. To examine the effect of the form of the model on the performance of the MI and/or SEPC, eight different test conditions were analyzed using population data generated from a slightly more complex CFA model.

In particular, I used a correlated, three-factor model with three cross-loadings to generate data under the following conditions: (a) primary factor loadings of .50 or .70; (b) secondary factor loadings of .50 under the low factor loading (.50) condition and secondary factor loadings of .40 under the high factor loading (.70) condition; (c) factor intercorrelations of .40 under the low factor loading (.50) condition and factor intercorrelations of .30 under the high factor loading (.70) condition; and (d) sample sizes of 200 or 400. To examine the effect of model misspecification, the models were misspecified by either excluding the three cross-loadings or by excluding the three cross-loadings and the three-factor intercorrelations. The MI and the SEPC were then used to determine the sequence of correct parameters to add to the misspecified model. The findings from these test conditions generally mimicked the results from the more comprehensive set of simulations as reported and the use of the MI and the SEPC commonly led to the correct factor model. The MI and the SEPC did at times suggest the inclusion of incorrect model parameters when examined in isolation, in particular when the model was more misspecified.

DISCUSSION

Performance of MI and SEPC

The purpose of this study was to evaluate the performance of the MI and the SEPC when used independently and in conjunction with one another during a simulated model modification process under various conditions. To summarize the findings, all of the criteria examined tended to become more accurate as factor loading size, model misspecification size, and sample size increased. Overall, the significant MI criterion outperformed the largest, significant MI criterion with respect to accuracy. At the same time, the MI criteria (largest, statistically significant MI as well as statistically significant MI) were, in general, less accurate than were the largest SEPC criterion in a majority of the conditions. This finding supports results from the study conducted by Luijben and Boomsma (1988) in which the SEPC tended to outperform the MI when arriving at the correct confirmatory model. Given the recent recommendation that an SEPC cutoff of .20 would indicate a substantial misspecification (Saris et al., 2009), the performance of the SEPC cutoff criterion was of particular interest in the present study. The SEPC cutoff criterion only performed more accurately than the MI and the SEPC under certain conditions (small sample and factor loading sizes). The accuracy of the joint criteria (significant MI and largest SEPC) was of most interest in the present study, given that this has been recommended by Saris et al. (1987) and Kaplan (1988, 1989) and little studied. The joint criteria proved to be slightly less accurate than the significant MI.

Implications

Given the present findings, the SEPC generally performs more accurately with respect to arriving at the correct confirmatory factor model than the MI, the SEPC cutoff criterion, or the joint criteria. The SEPC and MI do perform more similarly, however, as factor loading and sample size increase but are also dependent upon the misspecified parameter value size. The accuracy of the significant MI and of the joint criteria was slightly augmented in the present study given that among the set of fixed parameters associated with significant MI values and the largest SEPC value, the correct parameter was known. Also, it must be noted that while the number of significant MI values for the factor intercorrelation increased as factor loading, factor intercorrelation, and sample size increased, so did the number of significant MI values for parameters other than the factor intercorrelation. As a consequence, applied researchers would have a larger list of parameters, correct and incorrect, from which to choose to add to a model. Nonetheless, it was also seen that the MI value associated with the factor intercorrelation, when statistically significant, tended to be the largest or second largest MI value.

The difficulty is to decide which parameters, among those that are statistically significant and substantial, would be most theoretically meaningful to estimate freely. Thus, when the MI and SEPC agree to free a fixed parameter, researchers will be more assured when making the suggested model respecification if theoretically plausible. It was rarely seen that the MI associated with the factor intercorrelation was significant but the SEPC was not the largest for the factor intercorrelation. It was seen, however, that a nonsignificant MI and a large SEPC could occur simultaneously. Although these results are not presented in Tables 1 through 4, this situation tended to occur more often when sample size was small and when the misspecified parameter value was small. In this case, as recommended by Saris et al. (1987) and Kaplan (1988, 1989), a power analysis may help substantiate the model respecification.

Other than agreeing to freely estimate a parameter, the MI and SEPC can agree with one another in terms of not freely estimating a parameter (nonsignificant MI and a low SEPC). It would be recommended then, in this case, that the fixed parameter not be freely estimated. This could lead to possible model misspecifications in certain conditions, such as when sample size and missing parameter values are small. Nonetheless, this recommendation would be the most suitable in this situation. This reiterates the importance of having reasonably sized data sets (>100) to help ensure proper estimates and help instill more confidence in the findings. Given the plethora of recommended sample size requirements in structural equation modeling, it would seem that applied researchers are aware that reasonably large data sets are necessary when fitting structural equation models. Nonetheless, some models have been fit to data sets containing 100 or fewer observations (see Brecker, 1990; MacCallum & Austin, 2000).

Applied researchers will not know the true model, which makes the model respecification process challenging. As a result, applied researchers will examine the MI and SEPC to aid in the model respecification process. The sole reliance upon the MI and SEPC, however, is not recommended. That is, these statistics should be considered in light of plausible and meaningful additions to a model. Another point of consideration is that the sequence in which parameters are freely estimated in a respecified model may affect model estimates, including the values of the MI and SEPC. Thus, after modifying a model, it is highly recommended that they be cross-validated to provide more support for the respecified structure of the model.

Limitations and Suggestions for Future Research

The findings presented may not generalize to all structural equation modeling situations and should be considered in light of the conditions manipulated and examined in the present study. The generating model in the comprehensive simulation study was a simple, correlated two-factor CFA model and a single parameter was excluded from the misspecified model, as was done in previous studies. This was done because no comprehensive simulation study had examined the performance of the MI and SEPC in conjunction with another or the use of cutoff criterion and thus serves as a baseline study to which future research may be compared.

Although the test conditions that used a more complex model with varying levels of model misspecification demonstrated similar findings to those found in the more comprehensive simulation study, additional research should examine the performance of the MI and SEPC when various parameters of different magnitudes are excluded from more complex models. In addition, these missing parameters should be examined at the measurement and/or structural levels. Another limitation is that all the observed variables in the generating model loaded equally on each factor in a given condition. This would not necessarily be a realistic outcome in all confirmatory factor models. Thus, future research should examine models with varying values of the factor loadings.

The factor intercorrelations examined in the present simulation study widely ranged from .20 to .70 in order to examine the performance of the SEPC under varying conditions possibly encountered. A factor intercorrelation of .20 may appear to be statistically low and possibly unrealistic to suggest the inclusion of correlated factors in a model. Thus, conditions in which the factor intercorrelation was .20 may not be of interest to many applied researchers. Nevertheless, it may be theoretically important to have correlated factors in a CFA model and the recent suggestion of a SEPC value of .20 or greater as being substantial warranted the inclusion of this factor intercorrelation value. In addition, it was discovered in the applied research that some considered P-SEPC values associated with a parameter that were greater than .10 to be substantial enough to include in the respecified model. In contrast, high intercorrelations among factors (e.g., .70 or greater) may begin to indicate that a single factor should be modeled and/or tested. Thus, if researchers detect high intercorrelations among factors, a chi-square difference test could be conducted to determine whether modeling a single factor model will result in a significant loss of fit over a correlated factor model.

The search of applied articles in which the combination of the MI/Lagrange multiplier and the EPC was used in model modification/respecification may be limited because only articles in which Kaplan (1989) or Saris et al. (1987) were cited were reviewed. While Kaplan (1989) and Saris et al. (1987) originally suggested using the MI in combination with the EPC, some applied researchers may have used the combination criteria without including these citations in their studies. Although the application of the combination criteria (MI with EPC) was not the primary focus of this study, it is important to examine what is applied in practice by researchers in the field to understand the implications of the findings from simulation studies.

It must again be noted that the default Lagrange multiplier test in EQS was specified in the present study, meaning that the residual covariances were not examined as possible fixed parameters to be released during the comprehensive set of simulations or during the eight test condition simulations. Residual covariances were not of interest given that the generating model assumed that the residual variances did not covary. Nevertheless, residual covariances are examined and modeled by applied researchers. Moreover, this could have affected the performance of the MI and SEPC in that it narrowed the list of possible parameters to free. Future research should examine the possibility of freeing residual covariances in models assumed to have and to not have covarying residuals. This study can also be extended to incorporate the examination of these statistics under other conditions commonly encountered in applications of structural equation modeling, such as nonnormality among the variables and missing data conditions.

General Conclusion

In addition to the reliance on theoretical modifications to a model, applied researchers will rely on statistics such as the MI and SEPC during the model modification process. This study indicates that the MI and SEPC do not perform perfectly in all situations. Nonetheless, this study suggests that the MI and SEPC perform optimally under certain conditions. It is hoped that this study provides further knowledge concerning the use of these statistics when respecifying structural equation models.

AUTHOR NOTE

Tiffany A. Whittaker is an assistant professor in the Educational Psychology Department at the University of Texas at Austin. Her research interests include the evaluation of statistical procedures under various simulated conditions, including multiple linear regression analysis, structural equation modeling, and multilevel modeling. She has also conducted research investigating issues in testing and measurement using item response theory.

REFERENCES

Bentler, P. M. (1995). *EQS structural equations program manual*. Los Angeles, CA: BMDP Statistical Software.

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

Breckler, S. J. (1990). Applications of covariance structure modeling in psychology: Cause for concern? *Psychological Bulletin*, *107*, 260–273.

Brown, T. A. (2006). *Confirmatory factor analysis for applied researchers*. New York, NY: Guilford Press.

Chou, C.-P., & Bentler, P. M. (1990). Model modification in covariance structure modeling: A comparison among likelihood ratio, lagrange multiplier, and Wald tests. *Multivariate Behavioral Research*, *25*(1), 115–136.

Chou, C.-P., & Bentler, P. M. (1993). Invariant standardized estimated parameter change for model modification in covariance structure analysis. *Multivariate Behavioral Research*, *28*(1), 97–110.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, *18*, 147–167.

Hutchinson, S. R. (1993). Univariate and multivariate specification search indices in covariance structure modeling. *Journal of Experimental Education*, *61*, 171–181.

Kaplan, D. (1988). The impact of specification error on the estimation, testing, and improvement of structural equation models. *Multivariate Behavioral Research*, *23*, 69–86.

Kaplan, D. (1989). Model modification in covariance structure analysis: Application of the expected parameter change statistic. *Multivariate Behavioral Research*, *24*, 285–305.

Loehlin, J. C. (1998). *Latent variable models* (3rd ed.). Mahwah, NJ: Erlbaum.

Luijben, T. C. (1989). *Statistical guidance for model modification in covariance structure analysis*. Amsterdam, The Netherlands: Sociometric Research Foundation.

Luijben, T. C., & Boomsma, A. (1988). Statistical guidance for model modification in covariance structure analysis. *Compstat*, *1988*, 335–340.

MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, *100*(1), 107–120.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504.

Saris, W. E., Satorra, A., & Sörbom, D. (1987). The detection and correction of specification errors in structural equation models. In C. C. Clogg (Ed.), *Sociological methodology* (pp. 105–129). San Francisco, CA: Jossey-Bass.

Saris, W. E., Satorra, A., & van der Veld, W. M. (2009). Testing structural equation models or d detection of misspecifications? *Structural Equation Modeling*, *16*, 561–582.

Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika*, *54*(1), 131–151.

Silvia, E. M., & MacCallum, R. C. (1988). Some factors affecting the success of specification searches in covariance structure modeling. *Multivariate Behavioral Research*, *23*, 297–326.

Sörbom, D. (1989). Model modification. *Psychometrika*, *54*, 371–384.

Routledge
Taylor & Francis Group

The Journal of Experimental Education

ISSN: 0022-0973 (Print) 1940-0683 (Online) Journal homepage: https://www.tandfonline.com/loi/vjxe20

Using the Modification Index and Standardized Expected Parameter Change for Model Modification

Tiffany A. Whittaker

To cite this article: Tiffany A. Whittaker (2012) Using the Modification Index and Standardized Expected Parameter Change for Model Modification, The Journal of Experimental Education, 80:1, 26-44, DOI: 10.1080/00220973.2010.531299

To link to this article: https://doi.org/10.1080/00220973.2010.531299

Published online: 13 Dec 2011.

Submit your article to this journal

Article views: 2424

View related articles

Citing articles: 53 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=vjxe20