



# Evaluating the importance of individual parameters in structural equation modeling: the need for type I error control

Robert A. Cribbie\*

*University of Manitoba, Winnipeg, Manitoba, Canada R3T 2N2*

Received 23 March 1999; received in revised form 5 August 1999; accepted 1 October 1999

---

## Abstract

The use of structural equation modeling in personality research has been increasing steadily over the past few decades. In evaluating the adequacy of a particular model researchers are often interested in evaluating not only the overall fit of the model, but also which of the proposed parameters are significant. Researchers who apply unrestricted post hoc model modifications, or who evaluate the significance of individual parameters without adopting some form of type I error control, risk capitalizing on chance. A Monte Carlo study was used to demonstrate the effectiveness of simple Bonferroni-type procedures for controlling the rate of type I errors when multiple parameters are evaluated in the structural portion of a theoretical model. © 2000 Elsevier Science Ltd. All rights reserved.

---

## 1. Introduction

The use of structural equation modeling (SEM) in personality research is rapidly increasing, particularly with the increased dissemination of information regarding SEM and the availability of user-friendly SEM software packages. The primary purpose of SEM rests in assessing the adequacy of a predetermined theoretical model to explain relationships among observed and latent variables. In fact, the necessity for researchers to explicitly state their a

---

\* Tel.: +1-204-474-9338; fax: +1-204-474-7599.

E-mail address: umcribbi@cc.umanitoba.ca (R.A. Cribbie).

priori causal theory represents one of the strongest arguments in support of SEM (Martin, 1987).

In evaluating the adequacy of a particular model researchers are often interested not only in the overall fit of the model, but also which of the proposed relationships (parameters) in the model are (or are not) important. A diverse literature has emerged describing available options for determining the ‘importance’ or ‘significance’ of individual parameters, however two alternatives have dominated the empirical literature. The first option is to evaluate the significance of each parameter at a specified significance level ( $\alpha$ ), declaring statistically significant parameters important and nonsignificant parameters unimportant to the model. The second option is to use sequential model modification indices (e.g. Wald Test, Lagrange Multiplier) to identify which parameters to remove from or add to the model in order to improve the fit of the model to the sample data. Individual parameters are considered significant if they contribute significantly to the overall fit of the model. A common problem with both of the aforementioned methods is that they risk capitalizing on chance when evaluating the significance of multiple parameters in a structural model. This paper discusses how these two methods risk capitalizing on chance, as well as proposing an alternative method for evaluating the significance of structural model parameters.

### *1.1. Model modification*

Model modification involves sequential testing procedures for adding (deleting) parameters to (from) a model in order to improve the overall fit of the model to the sample data. This process has been labeled a ‘specification search’ (Kaplan, 1988; Long, 1983; MacCallum, 1986) and the ‘specification errors’ are the misspecified paths in the model that if modified may improve the fit of the model to the data. Several procedures for conducting specification searches have been discussed, including the modification index (Joreskog & Sorbom, 1984; MacCallum, 1986; Sorbom, 1989), the Lagrange multiplier test (Bentler, 1989; MacCallum, 1986) and the Wald test (Lee, 1985; Wald, 1943). For researchers interested in evaluating the significance of individual parameters, the criteria for significance would be that the parameter contribute significantly to the overall fit of the model. In fact, the precise use of modification indices for identifying which hypothesized relationships in a confirmatory model are significant is appealing, however this specific use of modification indices in confirmatory analyses has consistently been replaced by an unrestricted exploration for the best fitting model.

Although model modification strategies were initially proposed as exploratory techniques for examining possible model improvements, some researchers also recommend investigating modifications that may provide a good fitting model with an even better fit (see MacCallum, 1986). In the latter case, several issues must be considered. First, researchers adopting a specification search must acknowledge that the research is inductive and that no hypothesis is being tested, no theory is assessed and no causal model is confirmed (Biddle & Marlin, 1987). Regardless of the type of model modification, changes to the model are data-driven and thus do not provide any substantive evidence about the validity of the model. As Hancock (in press, p. 3) describes, model modifications “tempt the researcher to trade the hat of hypothesis confirmer for that of the hypothesis explorer”. The probability of making decision errors is related to the number of model modifications (MacCallum, 1986; MacCallum, Roznowski &

Necowitz, 1992; also see Hancock (in press) for a Scheffé procedure for controlling decision errors with multiple model modifications), the criteria adopted for declaring model modifications statistically significant (Green & Babyak, 1997; Hancock, in press), as well as the reliability of the model modification tests to locate the specification error (discussed below).

Second, recent evidence has shown that specification searches can be unreliable detectors of specification errors. For example, Kaplan (1988), MacCallum (1986) and Silvia and MacCallum (1988) have found the modification index to be an unreliable indicator of specification errors. In addition, Hutchinson (1993) found that the univariate and multivariate Lagrange Multiplier and Wald tests performed poorly for most types of model misspecifications. Her findings contradict those of Chou and Bentler (1990) who report that the multivariate Lagrange Multiplier is a reliable indicator of model misspecifications. In addition, MacCallum et al. (1992) found that model modifications carried out in sequential specification searches can be very unstable, especially for small or moderately-large sample sizes. That is, the ‘best fitting’ model found with one sample may not generalize to an independent sample from the same population. Given the complexity of covariance (correlation) matrices there is a high probability that in any model there will be modifications available that would improve the fit of the model to the data, however, such modifications may be specific to the particular sample and not generalize to other samples in the population.

A final issue concerning post hoc model modification is the meaningfulness of the suggested modifications. Researchers have been warned against making model modifications that are not justified theoretically (Green & Babyak, 1997; Long, 1983; MacCallum et al., 1992; Sorbom, 1989) in order to control against modifying parameters simply to improve the fit of the model. For example, Steiger (1990, p. 175) asks “what percentage of researchers would find themselves unable to think up a ‘theoretical justification’ for freeing a parameter?”. In a review of SEM studies conducted by Breckler (1990) and MacCallum et al. (1992) only one study (of the 37 that acknowledged making model modifications) declined to make a model modification because it would not make sense theoretically to do so. Thus, the validity of justifications made for data-driven model modifications may be very suspect.

Therefore, although model modification indices are based on improving the fit of the model and thus improving the probability of accurately determining the importance of individual parameters in the model, the risk of making decision errors with model modifications is too extreme to recommend their use in confirmatory SEM analyses.

### *1.2. Statistical multiplicity*

The multiplicity problem in statistical inference refers to selecting significant findings from a large set of findings that support particular hypotheses. In SEM, researchers are often exploring models with numerous parameters to be estimated, and the probability that any parameter will be significant by chance increases as a function of the number of parameters to be tested in the model. Thus, a researcher who is examining the significance of several parameters in a model, each at a specified significance level ( $\alpha$ ), may actually have a much greater probability than  $\alpha$  of falsely declaring relationships significant in the model.

One suggestion is that SEM researchers adopt a strategy that controls the probability of falsely declaring parameters significant (i.e. committing type I errors) when multiplicity exists,

although there are a few arguments for NOT applying this method. First, there is often a high degree of interrelatedness between parameters in a model (Kaplan & Wenger, 1993), and therefore methods for controlling the rate of type I errors become overly conservative. Second, as Kaplan and Wenger (1993) argue, it is more important in tests of parameters to maximize power (control type II errors) because failure to include important parameters can bias parameter estimates. Lastly, an issue that has plagued multiple comparison researchers for decades (see Games, 1971; Hancock & Klockars, 1996), is how to define a particular set (family) of hypotheses when controlling the rate of type I errors in SEM.

First, the fact that parameters in a model are often highly intercorrelated does affect the conservativeness of type I error controlling procedures. Specifically, the size and pattern of the input covariance matrix determines the extent to which the power of type I error controlling procedures are affected. However, since researchers are typically unaware of the degree to which parameters within their model are correlated, and the correlations between parameter estimates decrease when a model is well-identified, researchers must be conscious of the possible risk of falsely declaring parameters significant. Thus, besides assuming that every parameter in the model was correlated perfectly with every other parameter, any degree of independence between parameters will increase the probability of making type I errors above the nominal significance level. MacCallum (1995) explains that researchers should also be aware that models with several (often highly correlated) parameters, although tending to fit the data well, are not disconfirmable and thus make finding a good fit meaningless.

Second, although power is important in all statistical tests, no increase in power is meaningful if the rate of type I errors is not controlled. Therefore, although maximizing power ensures that important parameters are retained in a model, an increase in power is not justified if there is a large risk of falsely retaining nonsignificant parameters.

Lastly, the goal of specifying a family of hypotheses is to select a set that is not so large that it is impossible to ever reject any hypothesis, yet is not so small that it does not provide adequate control of type I errors. Kirk (1995) states that a family of tests should consist of those tests that are related in terms of their content and intended use. There has been much discussion in the literature regarding separate analyses of the measurement model (MM) and the structural model (SM) in SEM (e.g. Anderson & Gerbing, 1988; Hunter & Gerbing, 1982; Lance, Cornwell & Mulaik, 1988). Lance et al. (1988, p. 185) summarize that “relations in both submodels (SM and MM) represent causal hypotheses, but relations among variables in the structural portion of the model are usually of greater substantive interest”. This paper recommends that the SM and MM be treated as separate families and propose a method for controlling the rate of type I errors in the SM. This is not to say that controlling type I errors in the MM is not important, yet only beyond the scope of this investigation.

### *1.3. Methods for controlling type I errors in SEM*

Selecting the best method for controlling type I errors in SEM is not unlike selecting the best multiple comparison procedure within an analysis of variance framework, the goal is to find the most powerful method that is also able to maintain the rate of type I errors at a specified level. As in multiple comparison research though, the controversy is how to specify a level at which to control the rate of type I errors. The typical approach in SEM has been to set the

error rate per parameter (ERPP). With this method, the significance of individual parameters in a model are evaluated at a specified significance level (e.g.  $\alpha = 0.05$ ). This method ignores the issue of statistical multiplicity and thus allows the overall rate of type I errors to inflate with the number of parameters tested. Another possibility is to control the familywise error rate (FWE), where the error rate is set for a family of conceptually related tests (e.g. Bonferroni). As stated above, the method adopted in this paper is to control the error rate for the SM family of hypotheses. This family includes all hypotheses to be tested in the SM. A third possibility is to control the false discovery rate (FDR). The FDR was presented by Benjamini and Hochberg (1995) and is defined as the expected proportion of the number of erroneous rejections to the total number of rejections. Research on the FDR has shown that it represents a compromise between the often liberal ERPP and the often conservative FWE methods of controlling for type I errors when multiplicity exists (Benjamini & Hochberg, 1995; Benjamini, Hochberg & Kling, 1994; Keselman, Cribbie & Holland, 1999). The FDR should be especially appealing to SEM researchers who are not willing to accept the large losses in per-parameter power typically associated with FWE procedures, but also do not want to completely ignore the multiplicity issue.

Therefore, the goal of this paper is to compare the power and FWE control of multiple testing strategies for evaluating the significance of multiple parameters in the structural portion of a latent variable model.

## 2. Method

A Monte Carlo study was used to compare the FWE control and per-parameter power of the ERPP method typically adopted by SEM researchers with that of two FWE controlling procedures and the FDR controlling procedure due to Benjamini and Hochberg (1995). The FWE was controlled over the number of hypotheses tested in the SM of two theoretical models<sup>1</sup>, and the per-parameter power is defined as the probability of correctly declaring individual parameters in the SM significant.

### 2.1. Theoretical models

Two theoretical models were investigated, each proposing a saturated SM (i.e. tested hypotheses concerning the significance of all covariances between latent variables). Misspecifications were established by generating data with at least one of the structural model covariances equal to zero. Model A (Fig. 1) contained three hypothesized covariances (three latent variables) in the SM, one which was misspecified. Model B (Fig. 2) contained 10 hypothesized covariances (five latent variables) in the SM, three of which were misspecified. Two observed variables were generated for each latent variable in each model using

---

<sup>1</sup> This paper demonstrates a method for controlling the FWE over the number of hypotheses tested in the structural portion of a latent variable model. This implies that the rate of type I errors was controlled only for parameters which contain testable hypotheses. For example, the error term for a latent variable may be estimated, although there is typically no direct hypothesis tested and therefore no type I error control is required or imposed.

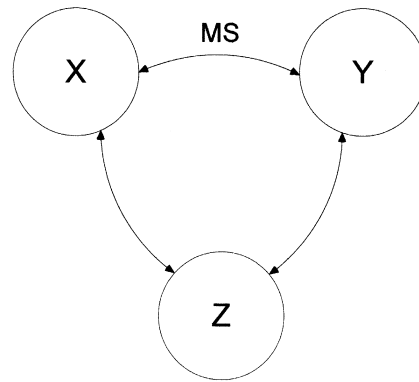


Fig. 1. Structural model A.  $X$ ,  $Y$  and  $Z$  represent latent variables, MS represents a misspecified parameter (i.e.  $\text{COV}_{xy} = 0$ ).

multivariate data generation procedures outlined in Kaiser and Dickman (1962). Structural model factor correlations were set at 0.15, minimizing floor and ceiling effects. There were 200 cases per replication. For identification and scaling of the models one parameter from each latent variable in the MM was set at unity and all other parameters (including error terms) were estimated. This resulted in 9 degrees of freedom for model A and 35 degrees of freedom for model B. All SEM analyses were performed using SAS PROC CALIS. In addition, all analyses utilized the sample covariance matrix and maximum likelihood estimation.

## 2.2. Multiple testing procedures

### 2.2.1. Bonferroni method

In this well known procedure, the  $p$ -values corresponding to the  $k$  parameter estimates are compared to an  $\alpha_{pp} = \alpha_{fw}/k$ , where  $\alpha_{pp}$  is the per-parameter type I error rate and  $\alpha_{fw}$  is a predetermined, acceptable, FWE. The Bonferroni method controls the actual FWE at  $\alpha_{fw}$  by declaring any parameter with  $p \leq \alpha_{pp}$  significant.

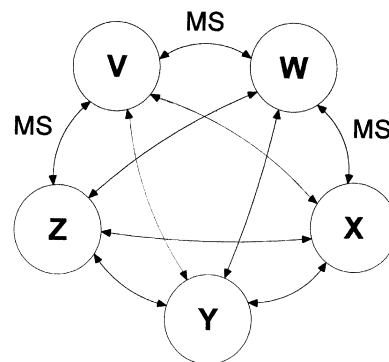


Fig. 2. Structural model B.  $V$ ,  $W$ ,  $X$ ,  $Y$  and  $Z$  represent latent variables, MS represents a misspecified parameter (i.e.  $\text{COV}_{vw} = 0$ ,  $\text{COV}_{vz} = 0$ ,  $\text{COV}_{xw} = 0$ ).

### 2.2.2. Hochberg's (1988) sequentially acceptive step-up Bonferroni procedure

In this procedure, the  $p$ -values corresponding to the  $k$  parameter estimates are ordered from smallest to largest. Then, for any  $I = k, k - 1, \dots, 1$ , if  $p_I \leq \alpha/(k - I + 1)$ , the Hochberg procedure rejects all parameters with  $p_{I'}$  ( $I' \leq I$ ). Therefore, one begins by testing the largest  $p$ -value,  $p_k$ , and declares all parameters significant if  $p_k \leq \alpha$ . If  $p_k > \alpha$  then the parameter associated with  $p_k$  is declared nonsignificant and one proceeds to compare  $p_{k-1}$  to  $\alpha/2$ . If  $p_{k-1} \leq \alpha/2$  then all parameters associated with  $p_I$  ( $I = k - 1, \dots, 1$ ) are declared significant, but if  $p_{k-1} > \alpha/2$  then the parameter associated with  $p_{k-1}$  is declared nonsignificant and one proceeds to compare  $p_{k-2}$  to  $\alpha/3$ , and so on.

### 2.2.3. Benjamini and Hochberg's (1995) FDR procedure

Like Hochberg's step-up procedure, the  $p$ -values corresponding to the  $k$  parameter estimates are ordered from smallest to largest. Then, for any  $I = k, k - 1, \dots, 1$ , if  $p_I \leq \alpha(I/k)$ , the FDR procedure rejects all parameters associated with  $p_{I'}$  ( $I' \leq I$ ). Therefore, one begins by testing the largest  $p$ -value,  $p_k$ , and declares all parameters significant if  $p_k \leq \alpha$ . If  $p_k > \alpha$  then the parameter associated with  $p_k$  is declared nonsignificant and one proceeds to compare  $p_{k-1}$  to  $\alpha(k - 1/k)$ . If  $p_{k-1} \leq \alpha(k - 1/k)$  then all parameters associated with  $p_I$  ( $I = k - 1, \dots, 1$ ) are declared significant, but if  $p_{k-1} > \alpha(k - 1/k)$  then the parameter associated with  $p_{k-1}$  is declared nonsignificant and one proceeds to compare  $p_{k-2}$  to  $\alpha/(k - 2/k)$ , and so on.

## 3. Results

All results were averaged over 2500 simulations using a nominal significance level of 0.05. Preliminary analyses were conducted to determine the fit of the models to the data. The mean goodness of fit index (GFI), adjusted goodness of fit index (AGFI) and root mean square error of approximation (RMSEA) for models A and B are presented in Table 1. Model A and B both fit the data well (average GFIs  $> 0.97$ , AGFIs  $> 0.94$ , RMSEAs  $< 0.02$ ) even though both models contained misspecified parameters.

### 3.1. FWE control

For model A, each of the multiple testing methods maintained the FWE rate for the SM at or below 0.05 (ERPP = 0.030, Bonferroni = 0.008, Hochberg = 0.025, FDR = 0.026, see Fig. 3).

Table 1  
Average fit indices (and range) for models A and B<sup>a</sup>

	GFI	AGFI	RMSEA
Model A	0.990 (0.962–1.000)	0.966 (0.868–0.999)	0.019 (0.000–0.123)
Model B	0.976 (0.951–0.992)	0.948 (0.893–0.982)	0.014 (0.000–0.073)

<sup>a</sup> GFI = goodness of fit index; AGFI = goodness of fit index adjusted for degrees of freedom; RMSEA = root mean square error of approximation.

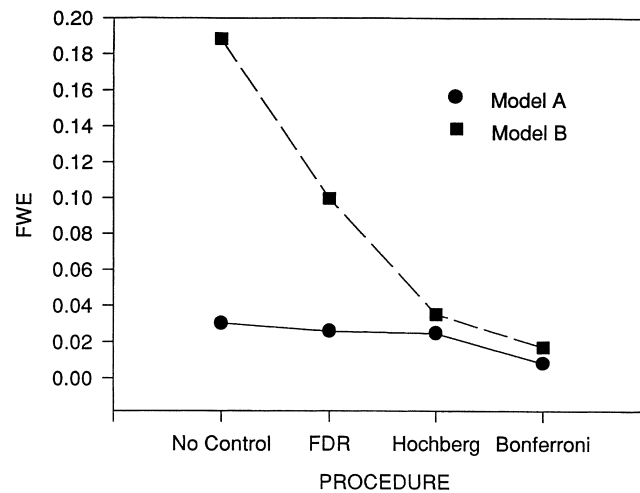


Fig. 3. The FWE for each multiple testing procedure for models A and B.

Only one null effect parameter was tested in model A and therefore it was expected that all methods would control the FWE rate at the nominal level. The FWE rate increased dramatically in model B (which contained three null effect parameters) for the ERPP method (0.188). In contrast, the Bonferroni and Hochberg procedures maintained the FWE at approximately 0.05 (0.017 and 0.035, respectively). The FDR procedure, as expected, represents a compromise between no type I error control and strict FWE control (FWE = 0.099).

### 3.2. Per-parameter power

The per-parameter power for the SM of models A and B, for each of the multiple testing

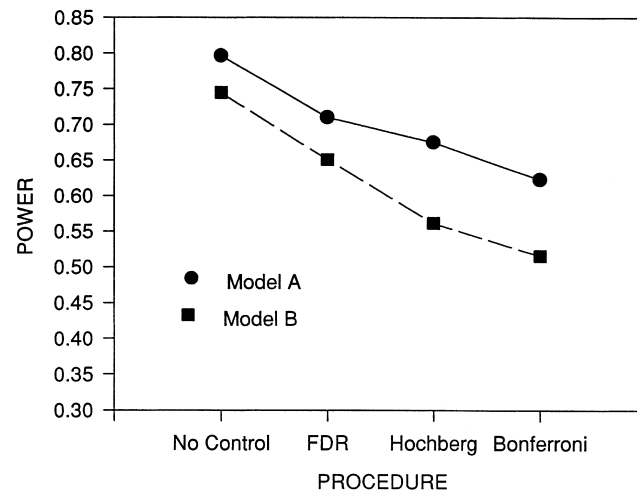


Fig. 4. The per-parameter power for each multiple testing procedure for models A and B.



procedures, is presented in Fig. 4. The per-parameter power of models A and B decreased, as expected, when type I error control was imposed. In model A the per-parameter power of the Bonferroni (0.623) and Hochberg (0.675) procedures was moderately lower than that of the ERPP method (0.797), with the power of the FDR procedure (0.710) falling between the no error control and FWE control procedures. In model B, the per-parameter power of the FDR procedure (0.651) again represented a compromise between the ERPP method (0.745) and FWE controlling procedures (0.515 and 0.561 for the Bonferroni and Hochberg procedures, respectively).

#### 4. Discussion

This paper explored different methods for determining the importance of individual parameters in SEM. The results of this investigation illustrate the dramatic increase in the SM FWE rate when the significance of multiple parameters were evaluated without type I error control, confirming the need for SEM researchers to incorporate type I error control into their confirmatory analyses. Previous recommendations for determining the significance of individual parameters have focused on the use of model modification indices, however, the uninhibited use of model modification indices removes the researcher from a position of hypothesis testing, and instead results in an inflated risk of making ‘sample-specific’ modifications. Therefore, a more direct approach is needed for evaluating the significance of individual parameters.

This paper highlights the effectiveness of simple Bonferroni-type procedures for reducing the risk of falsely declaring parameters in the SM significant. Even when the overall fit of the model is good, there is no guarantee that all of the proposed relationships in the model are significant (i.e. that the model is correctly specified; see Shevlin & Miles, 1998). In fact, as the number of parameters to be evaluated increases there is an expected increase in the probability of falsely declaring individual parameters significant. Therefore, as the size and complexity of studies (and hence models) in the social and behavioral sciences increases, the need for proper error control measures becomes increasingly important.

The results of this investigation also extend the findings of Benjamini and Hochberg (1995) and Keselman et al. (1999) that researchers not willing to adopt strict FWE control because of the large losses in power can instead adopt more liberal FDR control, a compromise between no type I error control and strict FWE control. Researchers adopting FDR control when testing multiple hypotheses gain considerable power over FWE controlling procedures, while still reducing the probability of making type I errors.

Although this paper outlines possible methods for controlling the rate of type I errors in the structural portion of a latent variable model, future research needs to be conducted exploring the issues and implications of using Bonferroni-type procedures to control the rate of type I errors in SEM. For example, this investigation explored the need to control the rate of type I errors in the SM, although further investigation needs to be conducted into controlling the rate of type I errors in the MM as well. In addition, this paper explored hypothesized relationships concerning latent variable covariances with fixed effect sizes, however, further research is also needed into controlling the rate of type I errors in the SM with different forms of hypothesized

relationships, different patterns of misspecification, as well as varying effect sizes across the relationships.

To summarize, as the number of researchers in the social and behavioral sciences using SEM increases, the need for clear guidelines for conducting analyses increases as well. Researchers who do not control for statistical multiplicity in SEM, or who utilize unrestricted specification searches to determine the significance of parameters in their models, risk ‘capitalizing on chance’. Therefore, in order to improve the reliability and validity of the results of SEM analyses researchers are recommended to apply simple Bonferroni-type procedures that control for statistical multiplicity when determining the importance of individual parameters in SEM analyses.

## References

- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin*, *103*, 411–423.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B*, *57*, 289–300.
- Benjamini, Y., Hochberg, Y. & Kling, Y. (1994). False discovery rate controlling procedures for pairwise comparisons. Unpublished manuscript.
- Bentler, P. M. (1989). *EQS structural equations manual*. Los Angeles: BMDP Statistical Software.
- Biddle, B. J., & Marlin, M. M. (1987). Causality, confirmation, credulity and structural equation modeling. *Child Development*, *58*, 4–17.
- Breckler, S. J. (1990). Application of covariance structure modeling in psychology: cause for concern? *Psychological Bulletin*, *107*, 260–273.
- 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*, 115–136.
- Games, P. A. (1971). Multiple comparisons of means. *American Educational Research Journal*, *8*, 531–565.
- Green, S. B., & Babyak, M. A. (1997). Control of type I errors with multiple tests of constraints in structural equation modeling. *Multivariate Behavioral Research*, *32*, 39–51.
- Hancock, G. R. (in press). A sequential Scheffe-type respecification procedure for controlling type I error in exploratory structural equation model modification. *Structural Equation Modeling: a Multidisciplinary Journal*.
- Hancock, G. R., & Klockars, A. J. (1996). The quest for  $\alpha$ : developments in multiple comparison procedures in the quarter century since Games. *Review of Educational Research*, *66*, 269–306.
- Hochberg, Y. (1988). A sharper Bonferroni procedure for multiple tests of significance. *Biometrika*, *75*, 800–802.
- Hunter, J. E., & Gerbing, D. W. (1982). Unidimensional measurement, second order factor analysis and causal models. *Research in Organizational Behavior*, *4*, 267–320.
- Hutchinson, S. R. (1993). Univariate and multivariate specification search indices in covariance structure modeling. *Journal of Experimental Education*, *61*, 171–181.
- Joreskog, K. G., & Sorbom, D. (1984). *Lisrel VI users guide* (3rd ed.). Mooreville, IN: Scientific Software.
- Kaiser, H. F., & Dickman, F. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary correlation matrix. *Psychometrika*, *27*, 179–182.
- 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., & Wenger, R. N. (1993). Asymptotic independence and separability in covariance structure models: Implications for specification error, power and model modification. *Multivariate Behavioral Research*, *28*, 467–482.
- Keselman, H. J., Cribbie, R. A., & Holland, B. (1999). The pairwise multiple comparison multiplicity problem: an alternative approach to familywise and comparisonwise type I error control. *Psychological Methods*, *4*, 58–69.

- Kirk, R. E. (1995). *Experimental design: procedures for the behavioral sciences*. Toronto, Canada: Brooks/Cole.
- Lance, C. E., Cornwell, J. M., & Mulaik, S. A. (1988). Limited information parameter estimates for latent or mixed manifest and latent variable models. *Multivariate Behavioral Research*, 23, 171–187.
- Lee, S.-Y. (1985). On testing functional constraints in structural equation models. *Biometrika*, 57, 239–251.
- Long, J. S. (1983). *Covariance structure models: an introduction to LISREL*. Beverly Hills: Sage.
- MacCallum, R. C. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, 100, 107–120.
- MacCallum, R. C. (1995). Model specification: procedure, strategies, and related issues. In R. H. Hoyle, *Structural equation modeling: concepts, issues and applications* (pp. 16–36). Newbury Park, CA: Sage.
- 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.
- Martin, J. A. (1987). Structural equation modeling: a guide for the perplexed. *Child Development*, 58, 33–37.
- Shevlin, M., & Miles, J. N. V. (1998). Effects of sample size, model specification and factor loadings on the GFI in confirmatory factor analysis. *Personality and Individual Differences*, 25, 85–90.
- Silvia, E. S. M., & MacCallum, R. C. (1988). Some factors affecting the success of specification searches in covariance structure modeling. *Multivariate Behavioral Research*, 23, 297–326.
- Sorbom, D. (1989). Model modification. *Psychometrika*, 54, 371–384.
- Steiger, J. H. (1990). Structural model evaluation and modification: an interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, 54, 426–482.