STRUCTURAL EQUATION MODELING, 7(4), 534–556

Type I Error and Power of Latent Mean Methods and MANOVA in Factorially Invariant and Noninvariant Latent Variable Systems

Gregory R. Hancock
Department of Measurement, Statistics and Evaluation
University of Maryland, College Park

Frank R. Lawrence
Center for Educational Accountability
University of Alabama at Birmingham

Jonathan Nevitt
Department of Measurement, Statistics and Evaluation
University of Maryland, College Park

This article addresses the issue of Type I error rates and relative power of structured means, multiple-indicator multiple-cause, and multivariate analysis of variance approaches for testing construct mean differences within a one-factor, two-group design. This study crosses 3 dimensions factorially: group sample size (n = 100, 200, 400, 800), factor loading pattern (8 variations), and factor mean difference ($\Delta\kappa$ = .2, .5, .8). Cases addressed include those where all factor loadings in both groups are equal, where loadings differ within a factor but are the same across groups, and where factor loadings differ across groups. Type I error rates are investigated using Monte Carlo simulation methods, while a population analysis approach is used to assess power of modeling and multivariate analysis of variance methods.

Researchers wishing to investigate group differences on a construct, rather than on a single measured variable, have a number of methods at their disposal. Traditionally, the method of choice has been multivariate analysis of variance

---

Requests for reprints should be sent to Gregory R. Hancock, 1230 Benjamin Building, University of Maryland, College Park, MD 20742–1115. E-mail: ghancock@wam.umd.edu

(MANOVA). MANOVA evaluates group differences on a linear composite of observed variables constructed so as to maximally differentiate the groups in multivariate space. Two more recent alternatives to MANOVA are available using structural equation modeling (SEM), Sörbom's (1974) structured means modeling (SMM), and a derivative of multiple-indicator multiple-cause (MIMIC) models (Jöreskog & Goldberger, 1975; Muthén, 1989). SMM seeks to model variables' mean structure along with the covariance structure in a fashion allowing researchers to make inferences regarding the groups' underlying construct means. The second SEM method, a MIMIC approach, employs group code (dummy) predictors within a structural equation model, which in turn is applied to a single set of data from all groups of interest combined. This study will first compare the three methods in terms of Type I error rate control, followed by an assessment of their relative power. Each of these methods is next described briefly in the context of a two-group comparison, as used in this investigation.

SMM

In SMM a construct η is explicitly represented as "latent"; that is, the underlying construct theoretically exerts a causal influence on the observed Y variables, thereby necessitating the existence of covariance among those variables (when the variance of η is nonzero). For a set of p observed Y indicators of construct η, Y values in a single group may be expressed in a $p \times 1$ vector $\mathbf{y}$ as follows: $\mathbf{y} = \nu + \Lambda\eta + \varepsilon$, where ν is a $p \times 1$ vector of intercept values, Λ is a $p \times 1$ vector of λ loadings, and ε is a $p \times 1$ vector of normal errors. Thus, the first-order moment vector is $E[\mathbf{y}] = \mu = \nu + \Lambda\kappa$, where κ is the mean of factor η; this reduces to $E[\mathbf{y}] = \Lambda\kappa$ in this study as variables will be assumed to contain no measurement bias (i.e., $\nu = \mathbf{0}$). The second-order moment matrix, assuming η and the errors to be independent, is $E[(\mathbf{y} - \mu)(\mathbf{y} - \mu)'] = \Sigma = \Lambda\phi\Lambda' + \Theta$, where ϕ is the variance of η and Θ is the $p \times p$ covariance matrix of the errors in ε.

With two groups, SMM offers a test of construct mean equivalence with the pertinent null hypothesis being H_0: $\kappa_1 = \kappa_2$. This test is optimally conducted under the assumption that observed variables have the same degree of bias (or none) across groups (i.e., $\nu_1 = \nu_2$), as well as the assumption that both groups' measurement models are tau-equivalent (i.e., $\Lambda_1 = \Lambda_2$). Note that this second assumption does not necessitate that the implied covariance matrices are identical across groups; only if factor variances $\phi_1 = \phi_2$ and error covariance matrices $\Theta_1 = \Theta_2$ would $\Sigma_1 = \Sigma_2$. These latter conditions are generally considered overly restrictive for establishing factorial invariance (see, e.g., Byrne, 1994, pp. 160–161). In fact, although SMM often proceeds by constraining corresponding intercept and loading parameters to be identical across groups, Byrne, Shavelson, and Muthén (1989) suggested that SMM's loading invariance assumption may be relaxed in some cases and still yield a meaningful comparison of latent means. This study will include both invariant and noninvariant loading scenarios.

MIMIC MODELS[1]

Like SMM, MIMIC approaches for testing hypotheses about latent means also start by positing a factor measurement model $\mathbf{y} = \nu + \Lambda\eta + \varepsilon$. However, whereas SMM fits the model to the two groups' data separately (but simultaneously), the MIMIC approach combines the data from both groups and incorporates a dichotomous predictor X (reflecting group membership) into the structural model. Specifically, $\eta = \gamma X + \zeta$, where γ is the causal impact of the group code variable X on the construct η and ζ is the construct residual unexplained by X. Given that X utilizes codes of 0 for Group 1 and 1 for Group 2, the γ parameter can be shown to represent the difference between groups' construct means. For Group 1 $\mathrm{E}[\eta] = \kappa_1 = \gamma\mathrm{E}[X] = 0$, and for Group 2 $\mathrm{E}[\eta] = \kappa_2 = \gamma\mathrm{E}[X] = \gamma$; thus, $\gamma = \kappa_2 - \kappa_1 = \Delta\kappa$, where $\Delta\kappa$ represents the difference in κ between groups. Therefore, in MIMIC modeling the hypothesis tested is H_0: $\kappa_1 = \kappa_2$, or more simply H_0: $\Delta\kappa = \gamma = 0$.

In the MIMIC approach for two groups, data from $n_1 + n_2 = N$ cases exist for p + 1 measured variables (p indicators and one dummy). The data from both groups are treated as a single sample and the p + 1 variables are combined in a partitioned $(p + 1) \times 1$ vector $[\mathbf{y}' \mid X]'$. As shown in Appendix A, the first-order moment of this partitioned vector (again assuming X utilizes codes of 0 for Group 1 and 1 for Group 2) can be expressed as $\mathrm{E}[\mathbf{y}' \mid X]' = [\mu' \mid (n_2 / N)]' = [(n_2 / N)\Delta\kappa\Lambda_2' \mid (n_2 / N)]'$ if neither group contains measurement bias (i.e., $\nu_1 = \nu_2 = \mathbf{0}$). The second-order moment matrix of the partitioned vector, assuming X, ζ, and elements of ε to be independent (as well as assuming the earlier coding scheme), is shown in Appendix A to be the following partitioned matrix when no measurement bias exists:

$$\Sigma_{[\mathbf{y}'|X]'} = \left[\begin{array}{c|c} \begin{array}{l} (n_1/N)\Lambda_1\phi_1\Lambda_1' + (n_2/N)\Lambda_2\phi_2\Lambda_2' + \\ (n_1 n_2/N^2)(\Delta\kappa)^2\Lambda_2\Lambda_2' + \\ (n_1/N)\Theta_1 + (n_2/N)\Theta_2 \end{array} & (n_1 n_2/N^2)\Lambda_2(\Delta\kappa) \\ \hline (n_1 n_2/N^2)(\Delta\kappa)\Lambda_2' & (n_1 n_2/N^2) \end{array}\right]$$

Because a MIMIC approach results in only one model for the combined data from both groups, it is implicitly assumed that the same measurement model holds in both groups. This includes loadings, construct variance, and error variances. In effect, all sources of covariation among observed variables are treated as equal across groups,

[1]For the case of two groups discussed here, only a single group code (dummy) variable is necessary. As this is the sole causal variable impinging on η, the name multiple-indicator *multiple*-cause (i.e., MIMIC) is a bit of a misnomer. Nonetheless, the name MIMIC will be used throughout the article as it represents a general approach accommodating two or more groups.

making the assumption of identical measurement models tantamount to an assumption of equal Σ matrices. This assumption is clearly more rigorous than that involved in the SMM approach; in fact, the SMM approach has the additional flexibility of being able to accommodate (i.e., to model) sources of measurement heterogeneity. On the other hand, the assumption of homogeneity of Σ matrices allows the MIMIC approach to require estimation of fewer parameters, which generally translates into a smaller required sample size. The extent to which MIMIC models are robust to heterogeneous covariance matrices remains to be investigated.

MANOVA

As explained earlier, in SMM and MIMIC modeling methods the construct is explicitly represented as latent. In MANOVA, on the other hand, the construct is implicitly treated as "emergent." Here the measured variables assume the role of causal agents, leaving the construct to emerge as a linear composite of those observed variables on which it is dependent (Bollen & Lennox, 1991; Cohen, Cohen, Teresi, Marchi, & Velez, 1990; Cole, Maxwell, Arvey, & Salas, 1993). The composite $W = \mathbf{y}'\mathbf{a}$, where weights in the $p \times 1$ vector $\mathbf{a}$ are derived to maximally differentiate the groups in p-variate space. The hypothesis formally tested in the two-group MANOVA (also known as Hotelling's T^2) is H_0: $\mu_{W1} = \mu_{W2}$, where μ_W is a group's population mean on the optimal linear composite W. Note that, although SMM and MIMIC methods are able to segregate variables' true score (i.e., construct-related) variance from their error variance, MANOVA creates a composite using observed scores (thereby including variables' error variance). Thus, tests in MANOVA are not performed on precisely the same "constructs" as in the SEM methods (see Cole et al., 1993, for a nice treatment of SMM versus MANOVA).

Despite differences in "constructs" involved in testing, one similarity between the MIMIC approach and MANOVA is their assumption regarding equality of groups' covariance matrices. Although no literature explicitly addresses robustness of the MIMIC approach under covariance heterogeneity, MANOVA models have been well investigated under such circumstances. Specifically, work on the two-group case (e.g., Algina & Oshima, 1990; Hakstian, Roed, & Linn, 1979; Holloway & Dunn, 1967; Hopkins & Clay, 1963) as well as the general case (e.g., Olson, 1974) may be summarized as follows. In the "positive" condition where relatively smaller samples are drawn from populations with relatively smaller generalized variance (i.e., determinant of Σ), Type I error rate decreases when H_0 is true and power decreases when H_0 is false. Alternatively, in the "negative" condition where relatively smaller samples are drawn from populations with relatively larger generalized variance, Type I error rate increases when H_0 is true and power increases when H_0 is false. As detailed in individual studies, the degree to which Type I error and power are affected by covariance heterogeneity depends on a number of factors: total sample size, sample size disparities among groups, true differences among population means, magnitude of generalized variance, and generalized variance disparities among groups.

THIS STUDY

Deciding whether an SMM, MIMIC, or MANOVA approach is most appropriate for a given situation can be a challenge for researchers. As discussed by Cole et al. (1993), the primary decision should involve an appeal to theory, determining whether a latent or emergent variable system appears to best define the construct of interest. If emergent, a MANOVA approach may be most prudent; if latent, the choice between SMM and MIMIC methods hinges on concerns such as sample size and degree of model noninvariance across groups. Unfortunately, the default behavior has historically been the use of MANOVA with little regard to the nature of the underlying variable system. This article investigates the potential cost of such behavior in the single-factor two-group situation, using a Monte Carlo simulation to investigate Type I error control and a population analysis to examine relative power of the three methods under a variety of conditions. Conditions investigated, which are detailed more fully later, are as follows: factorially invariant groups with equal loadings, factorially invariant groups with varied loadings, factorially noninvariant groups with approximately equivalent generalized variance, and factorially noninvariant groups with disparate generalized variance. In addition to loading magnitude and configuration, the magnitude of the true factor mean difference between groups as well as sample size conditions (with equal and unequal cases) are manipulated.

Finally, to clarify, a comparison of latent mean modeling approaches with MANOVA may seem like an unfair comparison, particularly within a latent variable system for which MANOVA is not designed. However, the point of this comparison stems from the common recommendation to use MANOVA when interested in mean differences among multiple dependent variables, without regard to the nature of the variable system at hand. MANOVA is specifically designed for emergent variable systems; thus, regardless of its relative performance, this study is not seeking to endorse MANOVA when faced with a latent variable system. Rather, this study will show the benefits that may be reaped when choosing from the more appropriate latent variable methods instead of the default multivariate approach.

METHOD

Model and Conditions

In this experiment a single-factor model with three measured indicators (i.e., latent rather than emergent variable system) was used in each of two groups. The choice of three indicators was made as it is a commonly recommended minimum (e.g., Bentler & Chou, 1987). For simplicity, variables were created so as to contain no bias in either population (i.e., $\mathbf{v}_1 = \mathbf{v}_2 = \mathbf{0}$). The value of the factor mean was set to zero for Group 1's population (i.e., $\kappa_1 = 0$), whereas for Group 2's population the factor mean was varied: $\kappa_2 = \Delta\kappa = 0, .2, .5$, and .8. The population factor variance ϕ was set to one.

The factor loadings in Λ were varied to create eight population scenarios, four factorially invariant across groups (i.e., $\Lambda_1 = \Lambda_2$) and four factorially noninvariant (i.e., $\Lambda_1 \neq \Lambda_2$). In three of the four factorially invariant scenarios $\Lambda_1' = \Lambda_2' = [\lambda\ \lambda\ \lambda]'$, where λ was varied to be .4, .6, and .8; in the fourth factorially invariant scenario $\Lambda_1' = \Lambda_2' = [.4\ .6\ .8]'$. To create factorially noninvariant scenarios, the first condition had $\Lambda_1' = [.6\ .6\ .6]'$ and $\Lambda_2' = [.4\ .6\ .8]'$; the second had $\Lambda_1' = [.4\ .6\ .8]'$ and $\Lambda_2' = [.6\ .6\ .6]'$. In the third condition $\Lambda_1' = [.4\ .4\ .4]'$ and $\Lambda_2' = [.8\ .8\ .8]'$, while in the fourth $\Lambda_1' = [.8\ .8\ .8]'$ and $\Lambda_2' = [.4\ .4\ .4]'$. As for residual errors, for each group Θ was set to $\mathbf{I}$-diag($\Lambda\Lambda'$), therefore yielding variables with unit variance. Thus, for measurement models with $\Lambda' = [.4\ .4\ .4]'$, $\Lambda' = [.6\ .6\ .6]'$, $\Lambda' = [.8\ .8\ .8]'$, and $\Lambda' = [.4\ .6\ .8]'$, the generalized population variance det(Σ) = 0.9314, 0.7045, 0.2955, and 0.6833, respectively. This means that populations with $\Lambda_1' = [.4\ .6\ .8]'$ and $\Lambda_2' = [.6\ .6\ .6]'$ (or vice versa) have approximately equivalent generalized variance, whereas groups with $\Lambda_1' = [.4\ .4\ .4]'$ and $\Lambda_2' = [.8\ .8\ .8]'$ (or vice versa) have greatly disparate generalized variance.

Finally, total sample sizes of N = 200, 400, 800, and 1,600 across both groups were investigated in n_1:n_2 ratios of 1:1, 2:3, and 1:3. That is, for N = 200 the three cases were 100:100, 80:120, and 50:150; for N = 400 the three cases were 200:200, 160:240, and 100:300; for N = 800 the three cases were 400:400, 320:480, and 200:600; and for N = 1,600 the three cases were 800:800, 640:960, and 400:1,200. Notice that for populations with $\Lambda_1' = [.4\ .4\ .4]'$ and $\Lambda_2' = [.8\ .8\ .8]'$, the sample size ratios 2:3 and 1:3 represent the negative pairing condition (described earlier), whereas for populations with $\Lambda_1' = [.8\ .8\ .8]'$ and $\Lambda_2' = [.4\ .4\ .4]'$, the sample size ratios 2:3 and 1:3 represent the positive pairing condition.

Simulation and Type I Error Rate Estimation

Raw data matrices were generated in GAUSS (Aptech Systems, 1996) to achieve the desired sample size and covariance structure. For each group (j = 1, 2) an $n_j \times 3$ matrix of random normal deviates was premultiplied against the Cholesky factorization of that group's population correlation matrix. The population correlation matrix for each group is $\mathbf{P}_j = \Lambda_j\Lambda_j' + [\mathbf{I}\text{- diag}(\Lambda_j\Lambda_j')]$, which is also the population covariance matrix because variables were created to have unit variance. For each of the sample size and loading conditions, 1,000 pairs of data matrices (i.e., one per group) were generated for analysis and Type I error rate evaluation.

Each set of simulated raw data was analyzed using MANOVA as well as the SEM approaches. All MANOVA analyses were conducted directly within GAUSS, tallying the number of null hypothesis rejections (assuming $\alpha = .05$). SEM analyses were conducted using EQS 5.7 (Bentler, 1998). Specifically, SMM was run twice on each data set (once with loading constraints and once without); for each the number of times was tallied in which the construct mean of Group 2 differed statistically significantly ($\alpha = .05$) from the zero-constrained latent mean of Group 1. Finally, for the MIMIC approach, the raw data for the two groups were first concatenated into a single data file with an additional dummy column vector inserted to designate group membership. The number of times the group code predictor made a statistically significant ($\alpha = .05$) contribution to the prediction of the latent construct was counted. Again, 1,000 replications were conducted per cell of the design; in the event that an SMM or MIMIC analysis yielded an error code or a convergence failure, that pair of data matrices was completely discounted and another was generated in its place for analysis using the MANOVA, SMM, and MIMIC approaches.

Population Analysis and Power Estimation

In this portion of this study no data were simulated; rather, the model-implied population variances and covariances were analyzed assuming the various sample size combinations listed earlier. Such a population analysis approach, which has been implemented in similar SEM investigations (e.g., Kaplan & George, 1995), can be used in place of Monte Carlo simulations when the sampling behavior of the relevant test statistics is not of direct interest. Thus, the population covariance matrices were generated for all conditions of interest, submitted to SMM, MIMIC, and MANOVA methods for estimation of the relevant distribution's noncentrality parameter, and power was estimated as detailed following.

SMM. A population mean vector μ and covariance matrix Σ were generated for both groups, where for a given group $\mu = \Lambda\kappa$ and $\Sigma = \Lambda\Lambda' + \Theta$ (because $\nu = 0$ and $\phi = 1$, as described previously). A single factor model was fit to both groups, first imposing cross-group equality constraints on loadings and intercepts and then with loading constraints freed. Mean and covariance structures were estimated using maximum likelihood estimation in EQS 5.7 (Bentler, 1998). Although the factor mean for Group 1 was set to zero, each model was run both with the Group 2 mean free to vary as well as fixed to zero. The difference between the two model χ^2 values represents a noncentrality parameter δ for the noncentral χ^2 distribution with one degree of freedom. From this δ value the power was determined for an $\alpha = .05$ level test of factor mean equivalence, following from the discussion of Saris and Satorra (1993) for power analysis without specific alternatives. Other approaches to estimating power, such as using a Lagrange Multiplier χ^2 or Wald test χ^2, are asymptotically equivalent (see Bollen, 1989; Buse, 1982; Satorra, 1989; Satorra & Saris, 1983, 1985).

MIMIC. Because data from both groups are combined in this approach, a single covariance matrix is needed that includes the group code (dummy) variable. For this study the partitioned matrix presented earlier (and derived in Appendix A) is appropriate, with $\phi_1 = \phi_2 = 1$. As described previously, a model was fit to the combined group in which a single dummy predictor has a causal influence on the factor η. Maximum likelihood estimation was used within EQS 5.7 (Bentler, 1998). As the path from the dummy X to η captures the expected factor mean difference between groups, the model was run both with this path free to vary as well as fixed to zero. The difference between the two model χ^2 values represents a noncentrality parameter δ for the noncentral χ^2 distribution with 1 *df*, from which power was estimated as mentioned previously.[2]

MANOVA. As derived in Appendix B, the pooled within sums-of-squares matrix is $\mathbf{W} = n_1(\Lambda_1\Lambda_1' + \Theta_1) + n_2(\Lambda_2\Lambda_2' + \Theta_2)$ and the total sums-of-squares matrix is $\mathbf{T} = \mathbf{W} + (n_1n_2 / N)(\Delta\kappa)^2\Lambda_2\,\Lambda_2'$. Power determination for an F test at the .05 level was made based on a noncentral F distribution with 3 and $N - 4$ *df*, where the noncentrality parameter δ for this scenario reduces to $\delta = N[(|\mathbf{T}| / |\mathbf{W}|) - 1]$ as follows from Cohen (1988; see chapter 10). Power calculations from this noncentral distribution were made using the power analysis program STAT-POWER 2.2 (Bavry, 1993).

[2]For both the structured means modeling and multiple-indicator multiple-cause approaches, power was originally estimated using the z value associated with the path representing the factor mean difference. The square of this z value represents an estimate of the same noncentrality parameter δ (for a noncentral χ^2 distribution with one degree of freedom) as the χ^2 value that was ultimately used in this study to estimate power. Unfortunately, in the course of our investigation the standard errors for the relevant path did not appear invariant to, for example, changes in which variable is designated the factor's scale indicator. Any change in the model yields a change in the information matrix, which in turn seems to affect standard errors and hence the z value used in power estimation. The χ^2 difference value, on the other hand, is not affected by such changes, a phenomenon explored recently by Lawrence and Hancock (1998). For this reason, the χ^2 difference values were used in power estimation. Methodologists conducting population analyses may wish to take note that z values (and most likely Wald and Lagrange χ^2 values for similar reasons) may not provide for as stable an estimate of power as the χ^2 difference. Certainly further exploration of this issue is needed.

RESULTS

Type I Error Rates

The four scenarios in which both groups had identical measurement models (and hence homogeneous covariance matrices) are contained in Table 1, where Type I error results for SMM, MIMIC, and MANOVA approaches are detailed. The four noninvariant scenarios follow in Table 2. In both tables, any error rate falling at or beyond Bradley's (1978) liberal criterion of $\alpha \pm 1/2\alpha$ (i.e., $\leq .025$ or $\geq .075$) is underlined. Note that, although SMM was run with and without loading constraints, only the constrained results are presented in the tables. Type I error rates from the two conditions never differed by more than .017 (almost always differing by only a few thousandths), and showed no systematic tendency to be higher in the constrained or unconstrained condition.

As would be expected in theory, in loading-invariant cases Type I error control is maintained at acceptable levels regardless of loading pattern or sample size ratio. Similarly, when loadings are noninvariant but have nearly equivalent generalized variances (as in the first two column blocks of Table 2), Type I error rates are completely acceptable. In fact, even when loading patterns precipitate disparate generalized variances but sample sizes are equal across groups, all three methods appear to control Type I error satisfactorily. However, when sample size differences accompany the disparity in generalized variance, acceptable control over Type I error is no longer assured.

To elaborate, consider the performance of each of the three methods under the most disparate loading patterns in the last two column blocks of Table 2. SMM appears to maintain Type I error control within the limits of Bradley's liberal criterion under all conditions, even with the loadings constrained. The MIMIC approach, on the other hand, tends to become conservative when the most extreme sample size ratio is negatively paired with generalized variance, and tends to become liberal in similar positive pairing conditions. This behavior is somewhat opposed to that typically observed for MANOVA, where extreme negative conditions are seen to result in the inflated error rates as expected and extreme positive pairings result in conservative error rates (though not outside Bradley's interval in this study).

Power in Invariant Scenarios

The four scenarios in which both groups had identical measurement models are contained in Table 3, where power analysis results for SMM, MIMIC, and MANOVA approaches are presented. Note that, although SMM was run with and without loading constraints, only the constrained results are presented in Table 3.

TABLE 1
Type I Error Rates of SMM, MIMIC, and MANOVA Under Factorial Invariance

	$\Lambda_1' = [.4\ .4\ .4]$ $\Lambda_2' = [.4\ .4\ .4]$			$\Lambda_1' = [.6\ .6\ .6]$ $\Lambda_2' = [.6\ .6\ .6]$			$\Lambda_1' = [.8\ .8\ .8]$ $\Lambda_2' = [.8\ .8\ .8]$			$\Lambda_1' = [.4\ .6\ .8]$ $\Lambda_2' = [.4\ .6\ .8]$		
n_1:n_2	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*
100:100	.032	.038	.032	.041	.043	.037	.064	.064	.055	.037	.039	.044
80:120	.037	.046	.031	.042	.047	.050	.026	.028	.047	.045	.045	.048
50:150	.046	.037	.039	.060	.056	.049	.057	.057	.041	.039	.047	.050
200:200	.042	.047	.046	.044	.043	.038	.051	.051	.042	.040	.043	.044
160:240	.042	.040	.047	.051	.052	.056	.048	.047	.054	.039	.041	.046
100:300	.037	.037	.034	.045	.051	.050	.055	.050	.059	.037	.037	.042
400:400	.043	.044	.038	.044	.044	.040	.056	.056	.046	.038	.038	.043
320:480	.037	.038	.037	.049	.048	.045	.045	.046	.051	.037	.037	.045
200:600	.055	.052	.052	.045	.046	.044	.034	.036	.053	.058	.057	.043
800:800	.044	.044	.058	.047	.047	.049	.051	.051	.057	.051	.051	.045
640:960	.046	.045	.047	.046	.046	.047	.061	.060	.047	.040	.039	.043
400:1200	.041	.042	.047	.053	.053	.053	.050	.049	.050	.039	.036	.043

Note. SM = structured means modeling; MI = multiple-indicator multiple-cause; MA = multivariate analysis of variance.

TABLE 2
Type I Error Rates of SMM, MIMIC, and MANOVA Under Factorial Noninvariance

	$\Lambda_1' = [.6\ .6\ .6]$ $\Lambda_2' = [.4\ .6\ .8]$			$\Lambda_1' = [.4\ .6\ .8]$ $\Lambda_2' = [.6\ .6\ .6]$			$\Lambda_1' = [.4\ .4\ .4]$ $\Lambda_2' = [.8\ .8\ .8]$			$\Lambda_1' = [.8\ .8\ .8]$ $\Lambda_2' = [.4\ .4\ .4]$		
$n_1:n_2$	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*
100:100	.033	.037	.040	.046	.043	.054	.048	.045	.048	.068	.056	.050
80:120	.051	.048	.045	.041	.043	.044	–.049	.039	.067	+.058	.081	.051
50:150	.038	.039	.050	.051	.052	.051	–.052	.023	.109	+.070	.089	.039
200:200	.044	.045	.052	.039	.053	.047	.049	.054	.053	.051	.056	.039
160:240	.055	.052	.048	.059	.061	.063	–.044	.034	.062	+.054	.065	.048
100:300	.039	.038	.060	.053	.050	.049	–.037	.017	.080	+.050	.079	.038
400:400	.044	.046	.044	.057	.068	.056	.052	.054	.035	.045	.047	.047
320:480	.045	.046	.043	.060	.054	.060	–.047	.041	.072	+.049	.059	.045
200:600	.055	.051	.058	.067	.069	.070	–.052	.034	.104	+.047	.074	.034
800:800	.058	.058	.052	.056	.048	.039	.040	.041	.047	.058	.039	.060
640:960	.048	.047	.052	.041	.042	.047	–.040	.029	.062	+.048	.059	.046
400:1200	.049	.048	.051	.045	.048	.046	–.052	.025	.112	+.040	.080	.034

Note. A minus (–) sign indicates a negative pairing of sample size with generalized variance pertaining to all three methods; a plus (+) sign indicates a positive pairing. Any error rate falling at or beyond Bradley's (1978) liberal criterion of $\alpha \pm 1/2\alpha$ (i.e., $\leq .025$ or ≥ 075) is underlined. SM = structured means modeling; MI = multiple-indicator multiple-cause; MA = multivariate analysis of variance.

TABLE 3
Power of SMM, MIMIC, and MANOVA Under Factorial Invariance

	$\Lambda_1' = [.4\ .4\ .4]$ $\Lambda_2' = [.4\ .4\ .4]$			$\Lambda_1' = [.6\ .6\ .6]$ $\Lambda_2' = [.6\ .6\ .6]$			$\Lambda_1' = [.8\ .8\ .8]$ $\Lambda_2' = [.8\ .8\ .8]$			$\Lambda_1' = [.4\ .6\ .8]$ $\Lambda_2' = [.4\ .6\ .8]$		
n_1:n_2	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*
$\Delta\kappa = .2$												
100:100	.136	.136	.096	.200	.201	.133	.252	.253	.165	.221	.222	.146
80:120	.132	.132	.094	.193	.194	.129	.243	.244	.160	.214	.215	.141
50:150	.113	.114	.083	.160	.162	.110	.199	.201	.133	.176	.178	.120
200:200	.225	.225	.148	.352	.352	.232	.447	.448	.303	.392	.393	.261
160:240	.218	.218	.144	.340	.341	.224	.432	.434	.292	.379	.380	.252
100:300	.180	.181	.122	.276	.278	.182	.351	.354	.233	.307	.310	.203
400:400	.398	.399	.266	.608	.609	.440	.735	.735	.569	.665	.666	.496
320:480	.385	.386	.256	.591	.592	.424	.717	.718	.550	.647	.648	.478
200:600	.313	.314	.206	.489	.491	.338	.608	.611	.443	.541	.544	.382
800:800	.673	.673	.503	.886	.886	.762	.956	.956	.883	.922	.922	.821
640:960	.655	.656	.486	.873	.873	.743	.948	.948	.868	.911	.912	.803
400:1200	.549	.551	.388	.781	.783	.624	.886	.887	.765	.832	.834	.688
$\Delta\kappa = .5$												
100:100	.559	.562	.396	.788	.790	.635	.890	.892	.775	.838	.840	.700
80:120	.542	.545	.382	.771	.774	.615	.876	.879	.757	.822	.825	.680
50:150	.442	.450	.303	.657	.671	.501	.777	.793	.638	.713	.727	.561
200:200	.849	.850	.711	.975	.975	.927	.995	.995	.980	.987	.987	.956
160:240	.833	.835	.692	.969	.970	.916	.993	.993	.975	.983	.984	.948
100:300	.731	.738	.572	.919	.925	.829	.972	.976	.928	.947	.953	.879
400:400	.989	.989	.960	1.000	1.000	.999	1.000	1.000	1.000	1.000	1.000	1.000
320:480	.986	.986	.952	1.000	1.000	.998	1.000	1.000	1.000	1.000	1.000	1.000

(continued)

TABLE 3 (Continued)

	$\Lambda_1' = [.4\ .4\ .4]$ $\Lambda_2' = [.4\ .4\ .4]$			$\Lambda_1' = [.6\ .6\ .6]$ $\Lambda_2' = [.6\ .6\ .6]$			$\Lambda_1' = [.8\ .8\ .8]$ $\Lambda_2' = [.8\ .8\ .8]$			$\Lambda_1' = [.4\ .6\ .8]$ $\Lambda_2' = [.4\ .6\ .8]$		
$n_1{:}n_2$	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*
200:600	.954	.957	.886	.997	.998	.990	1.000	1.000	.999	.999	.999	.996
800:800	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
640:960	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400:1200	.999	.999	.996	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\Delta\kappa = .8$												
100:100	.917	.918	.819	.992	.992	.973	.999	.999	.995	.996	.996	.987
80:120	.905	.908	.802	.989	.990	.967	.998	.999	.994	.995	.995	.983
50:150	.815	.830	.687	.958	.967	.913	.988	.992	.974	.975	.982	.946
200:200	.997	.997	.989	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
160:240	.996	.997	.986	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
100:300	.982	.985	.951	1.000	1.000	.998	1.000	1.000	1.000	1.000	1.000	1.000
400:400	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
320:480	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
200:600	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
800:800	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
640:960	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
400:1200	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Note. SM = structured means modeling; MI = multiple-indicator multiple-cause; MA = multivariate analysis of variance.

This is because all loading constraints were true in the population for these invariant scenarios, thus yielding identical power estimates with and without loading constraints.

As expected, the power of all methods is greater for larger values of N, $\Delta\kappa$, and λ. Also, an increased disparity in the n_1:n_2 ratio leads to decreases in power, again as expected. As for differences between the two SEM-based methods, the MIMIC approach is unilaterally equivalent or superior (though never by much); the magnitude of this superiority appears to vary as a function of the other conditions. For a small latent effect size ($\Delta\kappa = .2$), SMM and MIMIC approaches are almost identical in power for all sample size and loading configurations; MIMIC's superiority never exceeds .003 (i.e., 0.3%) for the case of all-equal loadings as well as for the varied loading case. For a medium latent effect size ($\Delta\kappa = .5$), the MIMIC method is better in the all-equal loading cases by a maximum of 1.6%, whereas for the varied loading case that difference is 1.4%. For a large latent effect size ($\Delta\kappa = .8$), the MIMIC method's maximum superiority in the all-equal loading cases is 1.5%, whereas for the varied loading case it is a mere 0.7%.

If a researcher had chosen to use the traditional MANOVA in these invariant scenarios, which is actually inconsistent with the latent variable system used in this investigation, a loss in power would result in all but those maximum power situations (i.e., where both power estimates are 1.000). Consider a comparison of MANOVA with the MIMIC approach (the more powerful of the two SEM methods, as presented earlier). For a small latent effect size, MANOVA is always less powerful than the MIMIC approach; the power difference is as high as 16.8% in the all-equal loading cases and 17.0% in the varied loading case. For a medium latent effect size, MANOVA is still never more powerful; the MIMIC method is better in the all-equal loading cases by as much as 17.0%, whereas for the varied loading case the maximum difference is 16.6%. Finally, for a large latent effect size, the MIMIC method's superiority over MANOVA reaches 14.3% in the all-equal loading cases, whereas for the varied loading case it is 3.6%.

Power in Noninvariant Scenarios

The four scenarios in which groups had different measurement models are contained in Table 4, where power analysis results for SMM, MIMIC, and MANOVA approaches are detailed. The two scenarios with approximately equivalent generalized variance will be addressed first (in the first two column blocks in Table 4), followed by those two scenarios with disparate generalized variance (in the last two column blocks in Table 4). As done previously, only SMM results with constrained loadings are presented; the very slight power improvement that occurred on releasing loading constraints is mentioned later.

TABLE 4
Power of SMM, MIMIC, and MANOVA Under Factorial Noninvariance

	$\Lambda_1' = [.6\ .6\ .6]$ $\Lambda_2' = [.4\ .6\ .8]$			$\Lambda_1' = [.4\ .6\ .8]$ $\Lambda_2' = [.6\ .6\ .6]$			$\Lambda_1' = [.4\ .4\ .4]$ $\Lambda_2' = [.8\ .8\ .8]$			$\Lambda_1' = [.8\ .8\ .8]$ $\Lambda_2' = [.4\ .4\ .4]$		
$n_1{:}n_2$	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*	*SM*	*MI*	*MA*
$\Delta\kappa = .2$												
100:100	.217	.218	.148	.193	.195	.134	.305	.306	.200	.112	.113	.083
80:120	.212	.213	.143	.188	.191	.130	–.309	.284	.185	+.107	.113	.083
50:150	.177	.178	.120	.158	.161	.111	–.269	.220	.144	+.091	.104	.078
200:200	.384	.386	.266	.339	.342	.234	.538	.539	.377	.178	.178	.120
160:240	.375	.376	.255	.330	.334	.225	–.543	.502	.346	+.166	.180	.121
100:300	.308	.309	.204	.270	.276	.182	–.479	.388	.258	+.133	.160	.110
400:400	.655	.656	.505	.590	.594	.444	.829	.829	.683	.308	.308	.202
320:480	.641	.643	.485	.575	.582	.427	–.834	.794	.638	+.285	.312	.205
200:600	.542	.543	.385	.479	.488	.339	–.771	.659	.489	+.220	.274	.179
800:800	.916	.917	.829	.872	.875	.767	.985	.985	.948	.541	.541	.380
640:960	.908	.909	.810	.861	.866	.747	–.986	.976	.926	+.504	.547	.385
400:1200	.832	.833	.693	.770	.780	.626	–.969	.918	.814	+.389	.484	.332
$\Delta\kappa = .5$												
100:100	.831	.834	.709	.773	.778	.640	.946	.948	.873	.440	.442	.297
80:120	.817	.822	.687	.757	.766	.619	–.949	.929	.837	+.407	.447	.301
50:150	.713	.727	.565	.647	.668	.503	–.907	.834	.692	+.309	.393	.260

200:200	.985	.986	.960	.970	.971	.929	.999	.999	.995	.727	.728	.562
160:240	.982	.983	.952	.965	.968	.918	–.999	.998	.991	+.686	.734	.568
100:300	.948	.953	.882	.913	.924	.830	–.997	.986	.954	+.546	.667	.496
400:400	1.00	1.000	1.000	1.000	1.000	.999	1.00	1.000	1.000	.953	.953	.878
320:480	1.00	1.000	1.000	1.000	1.000	.998	–1.000	1.000	1.000	+.934	.955	.883
200:600	.999	.999	.996	.997	.998	.990	–1.000	1.000	1.000	+.837	.923	.822
800:800	1.00	1.000	1.000	1.000	1.000	1.000	1.00	1.000	1.000	.999	.999	.996
640:960	1.00	1.000	1.000	1.000	1.000	1.000	–1.000	1.000	1.000	+.998	.999	.996
400:1200	1.00	1.000	1.000	1.000	1.000	1.000	–1.000	1.000	1.000	+.986	.998	.989
$\Delta\kappa = .8$												
100:100	.996	.996	.988	.990	.991	.974	1.00	1.000	.999	.820	.822	.676
80:120	.995	.995	.985	.987	.989	.968	–1.000	1.000	.999	+.781	.827	.683
50:150	.975	.982	.948	.954	.967	.914	–.999	.996	.986	+.636	.766	.606
200:200	1.000	1.000	1.000	1.000	1.000	1.000	1.00	1.000	1.000	.983	.983	.947
160:240	1.000	1.000	1.000	1.000	1.000	1.000	–1.000	1.000	1.000	+.973	.984	.950
100:300	1.000	1.000	1.000	.999	1.000	.998	–1.000	1.000	1.000	+.907	.967	.910
400:400	1.000	1.000	1.000	1.000	1.000	1.000	1.00	1.000	1.000	1.00	1.000	1.000
320:480	1.000	1.000	1.000	1.000	1.000	1.000	–1.000	1.000	1.000	+1.000	1.000	1.000
200:600	1.000	1.000	1.000	1.000	1.000	1.000	–1.000	1.000	1.000	+.996	1.000	.998
800:800	1.000	1.000	1.000	1.000	1.000	1.000	1.00	1.000	1.000	1.00	1.000	1.000
640:960	1.000	1.000	1.000	1.000	1.000	1.000	–1.000	1.000	1.000	+1.000	1.000	1.000
400:1200	1.000	1.000	1.000	1.000	1.000	1.000	–1.000	1.000	1.000	+1.000	1.000	1.000

Note. A minus (–) sign indicates a negative pairing of sample size with generalized variance pertaining to all three methods; a plus (+) sign indicates a positive pairing. SM = structured means modeling; MI = multiple-indicator multiple-cause; MA = multivariate analysis of variance.

Approximately equivalent generalized variance. Again, power of all methods is greater for larger values of N, $\Delta\kappa$, and λ, whereas an increased disparity in the n_1:n_2 ratio leads to decreases in power. As for differences between the two SEM-based methods, the MIMIC approach is still seen to be equivalent or superior when loadings in SMM are constrained, although the magnitude of this superiority remains quite small. For a small latent effect size ($\Delta\kappa = .2$), SMM and MIMIC approaches are almost identical in power for all sample size and loading configurations; MIMIC's superiority never exceeds 1.0%. For a medium latent effect size ($\Delta\kappa = .5$), the MIMIC method is never more than 2.1% higher (almost always much less), whereas for a large latent effect size ($\Delta\kappa = .8$) the MIMIC method's maximum superiority is 1.3%. If the SMM loading constraints are freed in these scenarios, the method's power shows a slight increase; the power gained is enough to make SMM generally superior to the MIMIC approach, but never by more than 1.9%.

If MANOVA had been used in these first two noninvariant scenarios, a loss in power would result in all cases but those with maximum power. A comparison of MANOVA with the MIMIC approach shows that, for a small latent effect size, MANOVA is always less powerful than the MIMIC approach and by as much as 15.8%. For a medium latent effect size, MANOVA is still never more powerful than the MIMIC approach; using MANOVA can result in as much as a 16.5% loss in power. When a large latent effect size is present, the superiority of the MIMIC approach reaches a maximum of 5.3%.

Disparate generalized variance. As before, results follow the expected patterns with respect to N, $\Delta\kappa$, and λ. However, with respect to the n_1:n_2 ratio, some curious but consistent phenomena appear in the disparate generalized variance cases. Specifically, in the $\Lambda_1' = [.4\ .4\ .4]'$ and $\Lambda_2' = [.8\ .8\ .8]'$ scenarios, SMM generally experiences a very slight power increase moving from n_1:n_2 = 1:1 to the negative n_1:n_2 = 2:3; then, as expected, power for the negative n_1:n_2 = 1:3 case drops to below that in either preceeding sample size ratio. Both MANOVA and the MIMIC approach behave as expected, with consistent decreases in power with increased n_1:n_2 disparity. Conversely, in the $\Lambda_1' = [.8\ .8\ .8]'$ and $\Lambda_2' = [.4\ .4\ .4]'$ scenarios, MANOVA and MIMIC results exhibit a very slight power increase moving from n_1:n_2 = 1:1 to the positive n_1:n_2 = 2:3; then, as expected, power for the positive n_1:n_2 = 1:3 case drops to below that in either preceeding sample size ratio. Meanwhile, SMM exhibits consistent decreases in power with increased n_1:n_2 disparity as expected. To explain these phenomena fully would require a substantial amount of space for what is, for current purposes, a tangential curiosity. Suffice it to say, the reader can replicate this type of phenomenon with an independent-groups pooled-variance t test where population variances (and hence sample variances) are unequal. The important point is that, for large sample size disparities, power tends to degrade as expected.

In general, the noninvariant scenarios with disparate generalized variance show less power for all methods in the positive conditions than in the negative conditions as expected. As for the methods' relative power, first consider the comparison of the two SEM-based methods. When sample sizes are equal, the two methods are virtually identical in power; the MIMIC approach tends to have a slight edge, but never by more than 0.2%. When sample sizes are unequal, the negative conditions consistently favor SMM (by as much as 9.1%). The positive conditions, on the other hand, consistently favor the MIMIC approach (by as much as 13.0%). However, all noninvariant power results must also be considered in light of the previous Type I error results. Specifically, in negative conditions the inferiority of the MIMIC approach is likely to be the result of a general conservatism as evidenced by low Type I error rates, whereas in positive conditions the apparent power advantage of the MIMIC method may simply be an artifact of the unacceptably liberal Type I error control. Finally, the release of loading constraints in SMM did not alter its power relative to that of the MIMIC approach in these final scenarios; this is because error variances were not constrained across groups, thereby allowing for equivalent nonstandardized loadings without inflating the model χ^2 values.

If MANOVA had been used in these last noninvariant scenarios, a fairly substantial loss in power would often result. In the negative pairings with $\Lambda_1' = [.4\ .4\ .4]'$ and $\Lambda_2' = [.8\ .8\ .8]'$, where SMM is the better SEM-based method and MANOVA is inferior to SMM in power by as much as 28.2% even with MANOVA's failure to control Type I error rates to acceptable levels. In the positive scenarios with $\Lambda_1' = [.8\ .8\ .8]'$ and $\Lambda_2' = [.4\ .4\ .4]'$, where the MIMIC approach is the better SEM method because of a failure to control Type I error, MANOVA is inferior to that MIMIC approach in power by as much as 17.1%, and is inferior to SMM by as much as 16.5%.

DISCUSSION AND CONCLUSIONS

When faced with latent (as opposed to emergent) construct differences, MANOVA constitutes the wrong model; this study offers evidence of the potential price of choosing that wrong model. Although isolated high-power conditions exist where this traditional approach is competitive, under most scenarios examined in this investigation SEM-based methods were superior and often substantially so. This may not be terribly surprising given that the population data in this study were created assuming a latent variable system (for which the SMM and MIMIC approaches are tailored), thus stacking the deck against MANOVA from the start. But the point of such a comparison draws from the common recommendation to use MANOVA when interested in mean differences among multiple dependent variables, without regard to the nature of the variable system at hand. And seen here, even in negative pairing scenarios when MANOVA has a known power boost (at the expense of Type I error control), this traditional multivariate method still falls well short of the more powerful SEM approaches.

Much more interesting is the comparison between the two SEM-based methods. First, SMM appears to control Type I error acceptably in all scenarios addressed in this study, whether or not loading constraints are in place. A MIMIC approach, on the other hand, controls the error rate when approximately equal generalized variances and/or equal sample sizes are present, but not with both sample size and generalized variance disparities. To elaborate, in the bias-free MIMIC models considered in this study (i.e., where $\nu_1 = \nu_2 = \mathbf{0}$) a group code variable is explaining variance in observed variables, but only indirectly as mediated by the latent construct (explained variance is systematic in the population power analyses, and random in the Monte Carlo Type I error simulations). In general, when there are bigger loadings the path from the group code variable to the construct need not be as large to explain a given amount of observed variable variance, and when there are smaller loadings the relevant path must become larger to accommodate. Now when two samples with different loadings are combined and analyzed with a single model, the magnitude of the single set of loadings that results will reflect contributions of the original samples' loadings relative to their sample sizes. This means that if the sample with the smaller loadings is given more weight due to a larger sample size, this will yield relatively smaller combined-group loadings and hence an inflated path from the group code variable must result; this inflated path will reject the null hypothesis more often, leading to more power but also to liberal Type I error control. Conversely, if the sample with the smaller loadings is given less weight due to a smaller sample size, this will yield relatively larger combined-group loadings and hence a smaller required path from the group code variable; this attenuated path will reject the null hypothesis less often, leading to less power as well as conservative Type I error control. With disparate generalized variances it is only when two (reasonably large) samples are equal in size that the combined result will mirror what would be expected with the combination of two theoretically infinite populations, thereby controlling the Type I error rate.

Regarding power, in the invariant scenarios the MIMIC approach was only slightly more powerful than SMM, with a maximum superiority of 1.6%. In noninvariant scenarios with approximately equivalent generalized variances the MIMIC approach was only slightly more powerful when SMM employed constrained loadings; SMM became generally but only slightly more powerful when the constraints were released (not tabled). As for disparate generalized variance cases, the two SEM methods had virtually identical power with equal sample sizes, but not with unequal sample sizes. As mentioned earlier, the MIMIC method's apparent power superiority when large loadings are paired with small samples comes at the expense of Type I error control, whereas SMM's superiority when large loadings are paired with large samples does not sacrifice Type I error control.

Thus, given a latent variable system, the choice between MIMIC and SMM approaches could be viewed as dependent on sample size. That is, when sample sizes are equal, there seems to be little practical difference between methods in terms of power or Type I error control. When sample sizes become increasingly disparate, however, SMM would seem to be favored. Although slightly less powerful than the MIMIC approach in some scenarios, researchers are generally without a priori knowledge of potential loading disparities across populations. Thus, whatever minute reduction in power that may result by choosing SMM over a MIMIC strategy would seem to be a small sacrifice to gain flexibility in accommodating loading invariance and to avoid the MIMIC approach's potential to lose control over Type I error.

Certainly further comparison of the two SEM methods is warranted. This study dealt with a limited number of loading and sample size conditions, and only with the case of three indicator variables. Also interesting would be the extension of this work beyond the two-group case, both in terms of the SEM methods' relation to each other as well as to MANOVA. A full understanding of such scenarios is necessary for the further development of experimental and nonexperimental design at the latent variable level.

REFERENCES

Algina, J., & Oshima, T. C. (1990). Robustness of the independent samples Hotelling's T^2 to variance-covariance heteroscedasticity when sample sizes are unequal and in small ratios. *Psychological Bulletin, 108*, 308–313.

Aptech Systems. (1996). *GAUSS system and graphics manual*. Maple Valley, WA: Author.

Bavry, J. L. (1993). *STAT–POWER 2.2*. Chicago: Scientific Software International.

Bentler, P. M. (1998). *EQS structural equations program*. Encino, CA: Multivariate Software.

Bentler, P. M., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*, 78–117.

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

Bollen, K. A., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin, 110*, 305–314.

Bradley, J. V. (1978). Robustness? *British Journal of Mathematical and Statistical Psychology, 31*, 144–152.

Buse, A. (1982). The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *The American Statistician, 36*, 153–157.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

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

Cohen, P., Cohen, J., Teresi, M., Marchi, M., & Velez, C. N. (1990). Problems in the measurement of latent variables in structural equations causal models. *Applied Psychological Measurement, 14*, 183–196.

Cole, D. A., Maxwell, S. E., Arvey, R., & Salas, E. (1993). Multivariate group comparisons of variable systems: MANOVA and structural equation modeling. *Psychological Bulletin, 114*, 174–184.
Hakstian, A. R., Roed, J. C., & Linn, J. C. (1979). Two-sample T^2 procedure and the assumption of homogeneous covariance matrices. *Psychological Bulletin, 86*, 1255–1263.
Holloway, L. N., & Dunn, O. J. (1967). The robustness of Hotelling's T^2. *Journal of the American Statistical Association, 62*, 124–136.
Hopkins, J. W., & Clay, P. P. F. (1963). Some empirical distributions of bivariate T^2 and homoscedasticity criterion M under unequal variance and leptokurtosis. *Journal of the American Statistical Association, 58*, 1048–1053.
Jöreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association, 70*, 631–639.
Kaplan, D., & George, R. (1995). A study of the power associated with testing factor mean differences under violations of factorial invariance. *Structural Equation Modeling, 2*, 101–118.
Lawrence, F. R., & Hancock, G. R. (1998, April). *Finite sample behavior of the likelihood ratio, Wald, and Lagrange Multiplier tests: Bias and variability in univariate noncentrality parameter estimation.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.
Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika, 54*, 557–585.
Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association, 69*, 894–908.
Saris, W. E., & Satorra, A. (1993). Power evaluations in structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 181–204). Newbury Park, CA: Sage.
Satorra, A. (1989). Alternative test criteria in covariance structure analysis: A unified approach. *Psychometrika, 54*, 131–151.
Satorra, A., & Saris, W. E. (1983). The accuracy of a procedure for calculating the power of the likelihood ratio test as used within the LISREL framework. In C. P. Middendorp, B. Niemoller, & W. E. Saris (Eds.), *Sociometric research 1982* (pp. 127–190). Amsterdam: Sociometric Research Foundation.
Satorra, A., & Saris, W. E. (1985). Power of the likelihood ratio test in covariance structure analysis. *Psychometrika, 50*, 83–90.
Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology, 27*, 229–239.

APPENDIX A

First-order and Second-order Moments for MIMIC Approach

In general for Group 1, $\mathbf{y}_1 = \nu_1 + \Lambda_1\eta + \varepsilon_1$. In the current study $\nu_1 = \nu_2 = \mathbf{0}$; therefore, $E[\mathbf{y}_1] = \mu_1 = \Lambda_1\kappa_1$. Further, because κ_1 was set to 0, $E[\mathbf{y}_1] = \mu_1 = \mathbf{0}$.

In general for Group 2, $\mathbf{y}_2 = \nu_2 + \Lambda_2\eta + \varepsilon_2$. In the current study $\nu_1 = \nu_2 = \mathbf{0}$; therefore, $E[\mathbf{y}_2] = \mu_2 = \Lambda_2\kappa_2$. Further, as $\kappa_2 = \Delta\kappa$ (because $\kappa_1 = 0$), $E[\mathbf{y}_2] = \Lambda_2\Delta\kappa$.

When groups are combined, the first moment vector (in partitioned format to include the group code dummy X with $X_1 = 0$ and $X_2 = 1$) is as follows:

$$\begin{aligned}
&(n_1/N)\mathrm{E}[\mathbf{y}_1' \mid X_1]' + (n_2/N)\mathrm{E}[\mathbf{y}_2' \mid X_2]' \\
&= [(n_1/N)\mathrm{E}[\mathbf{y}_1'] + (n_2/N)\mathrm{E}[\mathbf{y}_2'] \mid (n_1/N)\mathrm{E}[X_1] + (n_2/N)\mathrm{E}[X_2]]' \\
&= [(n_1/N)\mathrm{E}[\mathbf{y}_1'] + (n_2/N)\mathrm{E}[\mathbf{y}_2'] \mid (n_2/N)]' \\
&= [(n_2/N)\mu_2' \mid (n_2/N)]' \\
&= [(n_2/N)\Delta\kappa\Lambda_2' \mid (n_2/N)]'.
\end{aligned}$$

The second moment matrix for the combined groups (in partitioned format) may be derived as follows:

$$\begin{aligned}
&\mathrm{E}[[[\mathbf{y}' \mid X]' - [\,\mu' \mid (n_2/N)]'][[\mathbf{y}' \mid X]' - [\,\mu' \mid (n_2/N)]']'] \\
&= \mathrm{E}[[[\mathbf{y}' \mid X]' - [\,\mu' \mid (n_2/N)]'][[\mathbf{y}' \mid X] - [\,\mu' \mid (n_2/N)]]] \\
&= \mathrm{E}[[\mathbf{y}' - \mu' \mid X - (n_2/N)]'[\mathbf{y}' - \mu' \mid X - (n_2/N)]]
\end{aligned}$$

$$= \left[\begin{array}{c|c} \mathrm{E}[[\mathbf{y}-\mu][\mathbf{y}-\mu]'] & \mathrm{E}[[\mathbf{y}-\mu][X-(n_2/N)]] \\ \hline \mathrm{E}[[X-(n_2/N)][\mathbf{y}-\mu]'] & \mathrm{E}[X-(n_2/N)]^2 \end{array}\right] \qquad \text{Matrix 1}$$

At this point, it is easier to focus on individual quadrants of Matrix 1.

Upper left

$$\begin{aligned}
&\mathrm{E}[[\mathbf{y}-\mu][\mathbf{y}-\mu]'] \\
&= (n_1/N)\mathrm{E}[[\mathbf{y}_1-\mu][\mathbf{y}_1-\mu]'] + (n_2/N)\mathrm{E}[[\mathbf{y}_2-\mu][\mathbf{y}_2-\mu]'] \\
&= (n_1/N)\mathrm{E}[[\mathbf{y}_1-\mu_1+\mu_1-\mu][\mathbf{y}_1-\mu_1+\mu_1-\mu]'] + (n_2/N)\mathrm{E}[[\mathbf{y}_2-\mu_2+\mu_2-\mu][\mathbf{y}_2-\mu_2+\mu_2- \\
&\qquad \mu]'] \\
&= (n_1/N)[\mathrm{E}[[\mathbf{y}_1-\mu_1][\mathbf{y}_1-\mu_1]'] + \mathrm{E}[[\mu_1-\mu][\mu_1-\mu]']] + (n_2/N)[\mathrm{E}[[\mathbf{y}_2-\mu_2][\mathbf{y}_2-\mu_2]'] + \\
&\qquad \mathrm{E}[[\mu_2-\mu][\mu_2-\mu]']] \\
&= (n_1/N)[\Sigma_1 + [\mathbf{0} - (n_2/N)\Lambda_2\Delta\kappa][\mathbf{0}' - (n_2/N)\Delta\kappa\Lambda_2']] \\
&\qquad + (n_2/N)[\Sigma_2 + [\Lambda_2\Delta\kappa - (n_2/N)\Lambda_2\Delta\kappa][\Delta\kappa\Lambda_2' - (n_2/N)\Delta\kappa\Lambda_2']] \\
&= (n_1/N)[\Sigma_1 + (n_2/N)^2(\Delta\kappa)^2\Lambda_2\Lambda_2'] + (n_2/N)[\Sigma_2 + [(n_1/N)\Lambda_2\Delta\kappa][(n_1/N)\Delta\kappa\Lambda_2']] \\
&= (n_1/N)[\Sigma_1 + (n_2/N)^2(\Delta\kappa)^2\Lambda_2\Lambda_2'] + (n_2/N)[\Sigma_2 + (n_1/N)^2(\Delta\kappa)^2\Lambda_2\Lambda_2'] \\
&= (n_1/N)\Sigma_1 + (n_2/N)\Sigma_2 + (n_1/N)(n_2/N)^2(\Delta\kappa)^2\Lambda_2\Lambda_2' + (n_2/N)(n_1/N)^2(\Delta\kappa)^2\Lambda_2\Lambda_2' \\
&= (n_1/N)\Sigma_1 + (n_2/N)\Sigma_2 + (n_1n_2/N^2)(\Delta\kappa)^2\Lambda_2\Lambda_2' \\
&= (n_1/N)\Lambda_1\phi_1\Lambda_1' + (n_2/N)\Lambda_2\phi_2\Lambda_2' + (n_1n_2/N^2)(\Delta\kappa)^2\Lambda_2\Lambda_2' + (n_1/N)\Theta_1 + (n_2/N)\Theta_2
\end{aligned}$$

Lower left (which equals the transpose of the upper right)

$$\begin{aligned}
&\mathrm{E}[[X-(n_2/N)][\mathbf{y}-\mu]'] \\
&= (n_1/N)\mathrm{E}[[X_1-(n_2/N)][\mathbf{y}_1'-\mu']] + (n_2/N)\mathrm{E}[[X_2-(n_2/N)][\mathbf{y}_2'-\mu']] \\
&= (n_1/N)\mathrm{E}[[0-(n_2/N)][\mathbf{y}_1'-\mu']] + (n_2/N)\mathrm{E}[[1-(n_2/N)][\mathbf{y}_2'-\mu']] \\
&= (n_1/N)(-n_2/N)\mathrm{E}[\mathbf{y}_1'-\mu'] + (n_2/N)(n_1/N)\mathrm{E}[\mathbf{y}_2'-\mu'] \\
&= (-n_1n_2/N^2)\mathrm{E}[\mathbf{y}_1'-\mu'] + (n_1n_2/N^2)\mathrm{E}[\mathbf{y}_2'-\mu']
\end{aligned}$$

$= (-n_1n_2/N^2)[\mathbf{0}' - (n_2/N)\Delta\kappa\Lambda_2'] + (n_1n_2/N^2)[\Delta\kappa\Lambda_2' - (n_2/N)\Delta\kappa\Lambda_2']$
$= (n_1n_2/N^2)(n_2/N)\Delta\kappa\Lambda_2' + (n_1n_2/N^2)(n_1/N)\Delta\kappa\Lambda_2'$
$= (n_1n_2/N^2)\Delta\kappa\Lambda_2'$

Lower right
$\mathrm{E}[X - (n_2/N)]^2$
$= (n_1/N)\mathrm{E}[X_1 - (n_2/N)]^2 + (n_2/N)\mathrm{E}[X_2 - (n_2/N)]^2$
$= (n_1/N)[0 - (n_2/N)]^2 + (n_2/N)\mathrm{E}[1 - (n_2/N)]^2$
$= (n_1/N)(n_2/N)^2 + (n_2/N)(n_1/N)^2$
$= n_1n_2/N^2$

Q.E.D.

APPENDIX B

Within and Total Sums-of-Squares Matrices for MANOVA

The Within Sums-of-Squares matrix **W** is as follows:

$$\begin{aligned}\mathbf{W} &= n_1\Sigma_1 + n_2\Sigma_2\\ &= n_1(\Lambda_1\Lambda_1' + \Theta_1) + n_2(\Lambda_2\Lambda_2' + \Theta_2)\,.\end{aligned}$$

The Between Sums-of-Squares matrix **B** is as follows:

$$\begin{aligned}\mathbf{B} &= n_1[\mu_1 - \mu][\mu_1 - \mu]' + n_2[\mu_2 - \mu][\mu_2 - \mu]'\\ &= n_1[\mathbf{0} - (n_2/N)\Lambda_2\Delta\kappa][\mathbf{0}' - (n_2/N)\Delta\kappa\Lambda_2'] + n_2[\Lambda_2\Delta\kappa - (n_2/N)\Lambda_2\Delta\kappa][\Delta\kappa\Lambda_2' - (n_2/N)\Delta\kappa\Lambda_2']\\ &= n_1(n_2/N)^2(\Delta\kappa)^2\Lambda_2\Lambda_2' + n_2(n_1/N)^2(\Delta\kappa)^2\Lambda_2\Lambda_2'\\ &= (n_1n_2/N)(\Delta\kappa)^2\Lambda_2\Lambda_2'\,.\end{aligned}$$

The Total Sums-of-Squares matrix **T** is as follows:

$$\begin{aligned}\mathbf{T} &= \mathbf{W} + \mathbf{B}\\ &= \mathbf{W} + (n_1n_2/N)(\Delta\kappa)^2\Lambda_2\Lambda_2'\,.\end{aligned}$$

Q.E.D.