

# Hierarchically Nested Covariance Structure Models for Multitrait-Multimethod Data

Keith F. Widaman  
University of California at Riverside

A taxonomy of covariance structure models for representing multitrait-multimethod data is presented. Using this taxonomy, it is possible to formulate alternate series of hierarchically ordered, or nested, models for such data. By specifying hierarchically nested models, significance tests of differences between competing models are available. Within the proposed framework, specific model comparisons may be formulated to test the significance of the convergent and the discriminant validity shown by a set of measures as well as the extent of method variance. Application of the proposed framework to three multitrait-multimethod matrices allowed resolution of contradictory conclusions drawn in previously published work, demonstrating the utility of the present approach.

Investigating the construct validity of psychological measures is an involved process, requiring the collation of evidence from a variety of types of studies. In the article that standardized the use of the term "construct validity," Cronbach and Meehl (1955) enumerated types of research that would contribute to the construct validity of a measure. One of these types of study involves verifying that a measure correlates highly with other variables that purportedly measure the same construct. To the extent that such correlations are found, mea-

asures display *convergent validity*. Perhaps the most elegant form of such study utilizes the multitrait-multimethod matrix proposed by Campbell and Fiske (1959), in which correlations are computed among measures of two or more traits gathered using two or more methods of measurement. Use of the multitrait-multimethod matrix leads to an important criterion, beyond those stated by Cronbach and Meehl (1955), for establishing the construct validity of a measure: In addition to correlating highly with other measures of the same construct, a measure should fail to correlate as highly with measures of different, distinct constructs. To the extent the latter pattern of correlations is found, the measures display *discriminant validity*.

Campbell and Fiske (1959) suggested several types of comparison of correlations in a multitrait-multimethod matrix in order to demonstrate the degree of convergent and discriminant validity of the measured variables. These comparisons essentially involve determining the proportion of times certain convergent and discriminant validity criteria are satisfied; the higher the proportion of times a given criterion is satisfied, the clearer and stronger the pattern of convergent or discriminant validity.

The comparison procedures proposed by Campbell and Fiske (1959), though rather straightforward to follow, do have a number of shortcomings, as noted by several authors (e.g., Althausen & Heberlein, 1970; Bagozzi, 1978; Kenny, 1976). First, the four Campbell and Fiske criteria involve simply

---

APPLIED PSYCHOLOGICAL MEASUREMENT  
Vol. 9, No. 1, March 1985, pp. 1-26  
© Copyright 1985 Applied Psychological Measurement Inc.  
0146-6216/85/010001-26\$2.55

comparing the magnitudes of specified correlations in a situation in which testing the statistical significance of the overall pattern is of questionable value and appropriateness, due to the lack of independence of correlations used in the various comparisons. Although there have been recent attempts to develop significance tests for hypotheses regarding differences in level of correlations that reflect the Campbell and Fiske criteria (e.g., Hubert & Baker, 1978, 1979), developments to date have been rather limited. For example, the statistical tests derived by Hubert and Baker (1978) are not sensitive to the particular pattern of correlations among traits, though Hubert and Baker mentioned that unspecified extensions to their procedures could be made that might be sensitive to differential patterns of trait relationships.

A second problem with the Campbell and Fiske (1959) criteria is that precise estimates of the amounts of trait-related and method-related variance for each measure are not obtainable, though such estimates would be very useful for indicating which measures should be refined for use in further research. A third problem is that the Campbell and Fiske criteria are evaluated on the observed correlations among measures, and differences among variables in their level of reliability will distort both the correlations among measures and any summary measures derived from the correlations.

A number of researchers have noted the shortcomings of and ambiguities associated with the Campbell and Fiske (1959) criteria, and have proposed alternative approaches to evaluating multitrait-multimethod data. Some developed one-step (Jackson, 1969, 1971; but see Conger, 1971) or two-step (Golding & Seidman, 1974; Jackson, 1975; but see Golding, 1977, Jackson, 1977) component analytic approaches to identifying trait and method factors. Others have applied structural equation modeling procedures to multitrait-multimethod data. Of the latter, some proposed procedures have been path analytic in nature or conception (Althauser, 1974; Althauser & Heberlein, 1970; Althauser, Heberlein, & Scott, 1971; Alwin, 1974; Werts & Linn, 1970), others have been based on confirmatory factor analysis (Bagozzi, 1978, 1980; Boruch & Wol-

ins, 1970; Jöreskog, 1971, 1974; Kalleberg & Kluegel, 1975; Kenny, 1976; Lee, 1980; Marsh & Hocevar, 1983; Schmitt, 1978; Schmitt, Coyle, & Saari, 1977; Schmitt & Saari, 1978; Schwarzer, 1982), while still others employed direct product models (Bentler & Lee, 1979; Browne, 1984).

Although many contributions have been made to the modeling of multitrait-multimethod data, a unified and comprehensive strategy for testing structural models for such data has never been developed. The current state of the literature is perhaps best exemplified by Schmitt (1978) and Bagozzi (1978), the most extensive presentations in the literature of structural equation modeling of multitrait-multimethod data. Schmitt and Bagozzi each analyzed the same two multitrait-multimethod matrices, which came from studies published by Ostrom (1969) and Kothandapani (1971). Although they analyzed the same data and used very similar confirmatory factor analytic approaches, Schmitt and Bagozzi arrived at contradictory conclusions regarding the two multitrait-multimethod matrices: Schmitt stated that the Kothandapani matrix evidenced a greater degree of convergent and discriminant validity than did the Ostrom matrix, whereas Bagozzi claimed that the Ostrom matrix was the one showing the clearer, stronger pattern of convergent and discriminant validity. These contradictory conclusions, coupled with the presence of a number of calculational errors in Schmitt (1978) and several invalid interpretations of model comparisons by Bagozzi (1978), prompted the present attempt to standardize the fitting of structural models to multitrait-multimethod data.

The aim of the present paper is to describe a general procedure for specifying and testing a hierarchically nested set of models for multitrait-multimethod data. Within the proposed approach, it is a simple matter to formulate and interpret tests of the degree of convergent and of discriminant validity shown by a set of measures, as well as to estimate the amount of method variance in the measures. Applying the proposed approach to the Ostrom (1969) and Kothandapani (1971) multitrait-multimethod matrices will demonstrate both how the present approach is implemented and the utility

of the proposed approach over those previously presented. In so doing, a variety of misinterpretations by Bagozzi (1978) are resolved.

**Method**

**Confirmatory Factor Model  
For Multitrait-Multimethod Data<sup>1</sup>**

Assume that each person in a sample of  $N$  subjects has a score on each of  $mt$  measures that represent  $t$  trait constructs assessed under each of  $m$  methods. The correlation matrix  $\Sigma$  of the  $mt$  observed variables may be expressed, in the standard factor analytic decomposition into  $r$  common factors, as

$$\Sigma = \Lambda \Phi \Lambda' + \Psi, \quad (1)$$

where  $\Lambda$  is an  $mt \times r$  matrix of factor loadings,  
 $\Phi$  is an  $r \times r$  matrix of correlations among factors, and  
 $\Psi$  is an  $mt \times mt$  diagonal matrix of unique factor variances.

If the  $t$  trait measures are arranged within methods, the factor analytic decomposition of  $\Sigma$  represented by Equation 1 may be more easily adapted to fitting multitrait-multimethod data if  $\Lambda$  is partitioned in the following way:

$$\Lambda = [\Lambda_T | \Lambda_M]$$

$$= \left[ \begin{array}{c|ccc} T_1 & M_1 & 0 & \dots & 0 \\ T_2 & 0 & M_2 & \dots & 0 \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & 0 \\ T_m & 0 & 0 & \dots & M_m \end{array} \right], \quad (2)$$

<sup>1</sup>In the present paper, for ease of presentation and with no loss of generality, models are proposed under the assumption that observed variables have been standardized to zero mean and unit variance, and thus that  $\Sigma$  is a correlation matrix. It should be noted, however, that the generalization of the procedures outlined to analyses of covariance matrices is straightforward. The decision to employ the metric of correlations was made in order to conform to the standard way of presenting multitrait-multimethod data, which is in the form of a correlation, as opposed to a covariance, matrix.

where  $\Lambda_T$  is an  $mt \times t$  submatrix of  $\Lambda$  that contains loadings of observed variables on the  $t$  trait factors,

$\Lambda_M$  is an  $mt \times m$  submatrix of  $\Lambda$  that contains method factor loadings,

the  $T_i$  are  $t \times t$  diagonal submatrices of  $\Lambda_T$  containing trait factor loadings for the  $t$  measures gathered using method  $i$ , and

the  $M_i$  are column vectors within  $\Lambda_M$  containing  $t$  loadings on method factor  $i$ .

It is also useful to partition  $\Phi$  as

$$\Phi = \begin{bmatrix} \Phi_{TT} & \Phi_{TM} \\ \Phi_{MT} & \Phi_{MM} \end{bmatrix}, \quad (3)$$

where  $\Phi_{TT}$  is a  $t \times t$  symmetric submatrix of  $\Phi$  that contains trait factor intercorrelations,

$\Phi_{MM}$  is an  $m \times m$  symmetric submatrix of  $\Phi$  that contains method factor intercorrelations, and

$\Phi_{MT}$  ( $= \Phi'_{TM}$ ) is an  $m \times t$  rectangular submatrix of  $\Phi$  that contains correlations of the  $m$  method factors with the  $t$  trait factors.

Substituting Equations 2 and 3 into Equation 1 results in the following representation:

$$\Sigma = [\Lambda_T | \Lambda_M] \begin{bmatrix} \Phi_{TT} & \Phi_{TM} \\ \Phi_{MT} & \Phi_{MM} \end{bmatrix} \begin{bmatrix} \Lambda'_T \\ \Lambda'_M \end{bmatrix} + \Psi. \quad (4)$$

Assuming that the factor analytic model of  $\Sigma$  represented by Equation 1 is valid and that scores on the  $mt$  observed measures follow a multivariate normal distribution, Jöreskog (1969, 1971) developed procedures for obtaining maximum likelihood estimates of all model parameters in  $\Lambda$ ,  $\Phi$ , and  $\Psi$ . The significance of each of the parameters in the model may be tested by forming a  $z$ -ratio of the parameter estimate divided by its asymptotic standard error; parameter estimates with  $z$ -ratios greater than  $|2.00|$  are typically considered significant beyond the .05 level. In addition to the test of each individual parameter estimate, maximum likelihood estimation yields an overall  $\chi^2$  goodness-of-fit test. The  $\chi^2$  test is a test of the difference in fit

between a given model and a completely saturated model that perfectly represents the data. Thus, if the  $\chi^2$  value associated with a model is statistically significant, there is a statistical basis for rejecting the given model in favor of one with fewer restrictions, so that the patterns in the data may be modeled more accurately.

In addition to the statistical fit of a model, the level of practical fit of a covariance structure model must also be considered. Because the  $\chi^2$  test is directly related to sample size, a model that represents well a set of data, that is, that leaves small residual covariances, may still have a significant  $\chi^2$  if the sample is rather large. On the other hand, a model that fails to represent a data set well may have a nonsignificant  $\chi^2$ , which would support acceptance of the model, if the sample size was rather small. To remedy this situation, Bentler and Bonett (1980) proposed two measures of practical, as opposed to statistical, fit of covariance structure models. These two measures are termed rho and delta and are calculated in the following manner:

$$\text{rho} = \frac{(\chi_n^2/df_n) - (\chi_s^2/df_s)}{(\chi_n^2/df_n) - 1}, \quad (5)$$

and

$$\text{delta} = \frac{(\chi_n^2 - \chi_s^2)}{\chi_n^2}, \quad (6)$$

where  $\chi_n^2$  is the chi-square associated with the null model,

$df_n$  is the degrees of freedom for the null model,

$\chi_s^2$  is the chi-square associated with a substantive model under consideration, and

$df_s$  is the degrees of freedom for the substantive model.

The first measure, rho, is a generalization to restricted covariance structure models of the Tucker and Lewis (1973) reliability coefficient for unrestricted factor analysis models. Rho is a relative measure of off-diagonal covariation among observed variables explained by a model, a relative measure because the fit of each model is evaluated with regard to the degrees of freedom for the model.

The second measure, delta, is an absolute measure of fit, because delta generally represents the proportion of off-diagonal covariation explained by a model, regardless of degrees of freedom. Bentler and Bonett (1980) stated that rho and/or delta should attain values of .90 or above for a model to be accepted, because models with fit indices below .90 can usually be simply and substantially improved. Although standards for evaluating differences in measures of practical fit have not been developed, in the present paper differences between models in either rho or delta of less than .01 were considered unimportant on practical grounds.

A third general type of consideration, which may be used in conjunction with statistical significance and practical importance when evaluating covariance structure models, is the stability of parameter estimates after respecification of a model. If Model B is a respecification of Model A obtained by adding one or more theoretically meaningful parameter estimates to Model A, the stability of estimates of parameters that are common to Models A and B is of interest. If the common parameter estimates show a high level of stability, Model B introduced additional parameters whose effects were relatively independent of those of parameters in Model A. If, on the other hand, the estimates of common parameters are rather different in Model B than in Model A, then it may be concluded that, regardless of differences between Models A and B in levels of statistical and practical significance, Model B should be accepted. All other considerations (e.g., magnitude of standard errors) being equal, Model B may have allowed less biased estimation of parameters common to the two models. This stability consideration is related to the well-known "third variable," or omitted variable, problem in path analysis (Duncan, 1975; Kenny, 1979). Although highly restricted models are preferred on theoretical and practical grounds, specification of overly restricted models that omit variables important to the system will result in bias in estimation of parameters. By examining the stability of parameter estimates across competing covariance structure models estimated from the same data, it is possible to observe whether bias is apparent in particular models.

### Hierarchically Nested Models For Multitrait-Multimethod Data

A taxonomy of structural models for multitrait-multimethod data. Given the general model for multitrait-multimethod data represented by Equations 2 through 4 above, it is possible to specify a variety of theoretically interesting models that are special cases of the general model. One way to generate the array of structural models in a systematic fashion is to consider separately the structures that may be specified for the trait and method factor spaces. Consider first the trait factor space. There are three types of models for the trait space, models with the following specifications: (1) no trait factors, (2)  $t$  trait factors with fixed intercorrelations, and (3)  $t$  trait factors with freely estimated intercorrelations. These three types of models were translated into the following trait structures considered in the present article:

- 1 = No trait factors
- 2 =  $t$  trait factors, fixed unit intercorrelations
- 2' =  $t$  trait factors, fixed zero intercorrelations
- 3 =  $t$  trait factors, freely estimated intercorrelations

The two structures with fixed trait factor intercorrelations are given identical Arabic numeral labels, 2 and 2' (with the superscript to distinguish the structures), because the two structures entail the same number of parameter estimates. An identical number of estimates is made under Structures 2 and 2', since in neither case are correlations among trait factors estimated: under Structure 2, the correlations among trait factors are fixed at unity, resulting in a model with a single, general trait factor, whereas under Structure 2' the trait factors are forced to be orthogonal. Because Structures 2 and 2' have the same number of estimates, it is not possible to obtain a statistical test of the difference in fit of the two structures, since neither structure is nested within the other. However, given the identical number of parameter estimates, the structure associated with the smaller  $\chi^2$  value would be preferred.

Finally, note that the magnitude of the Arabic numerals assigned to structures is a key to the nest-

ing of the structures. When comparing two structures, the structure labeled with the larger number is the more inclusive structure and has the larger number of estimates; the structure with the smaller number is therefore nested within the more inclusive structure. That is, Structure 3 is the most inclusive structure, Structures 2 and 2' are nested within Structure 3, and Structure 1 is nested within Structures 2 and 2' and therefore within Structure 3 as well.

A set of structures for the method factor space, parallel to those for the trait factor space, may also be specified. This results in the following short list of method factor structures:

- A = No method factors
- B =  $m$  method factors, fixed unit intercorrelations
- B' =  $m$  method factors, fixed zero intercorrelations
- C =  $m$  method factors, freely estimated intercorrelations

Distinctions like those made for the trait space structures may also be drawn for the method space structures. For example, Structures B and B' are alternative structures with the same number of estimates and thus cannot be compared statistically. In addition, the letters assigned are a key to the nesting of the method space structures. Of two method structures, the structure labeled with the higher letter (i.e., the letter occurring later in the alphabet) is the more inclusive structure having the greater number of estimates, and the structure with the smaller (i.e., earlier) letter is nested within the former structure.

A taxonomy of models for multitrait-multimethod data may then be generated by cross-classifying the four trait structures and the four method structures, as shown in Table 1. Each of the models in Table 1 may then be designated by one number and one letter, which would denote the trait and method structures, respectively, embodied in the model. For example, Model 3B' has  $t$  freely correlated trait factors and  $m$  orthogonal method factors.

Given the number and letter labels of the models in Table 1, it is a simple matter to determine whether one model is nested within another model. A given

Table 1  
 Taxonomy of Structural Models for Multitrait-Multimethod Data

Trait Structure	Method Structure			
	A	B	B'	C
1	Null model	1 General method	$m$ methods only (orthogonal)	$m$ methods only (oblique)
2	1 General trait	2 General factors	1 General + $m$ methods (orthogonal)	1 General + $m$ methods (oblique)
2'	$t$ traits only (orthogonal)	1 General + $t$ traits (orthogonal)	$t$ traits (orthogonal) + $m$ methods (orthogonal)	$t$ traits (orthogonal) + $m$ methods (oblique)
3	$t$ traits only (oblique)	1 General + $t$ traits (oblique)	$t$ traits (oblique) + $m$ methods (orthogonal)	$t$ traits (oblique) + $m$ methods (oblique)

model is nested within a second model if the second model has a number label that is equal to or larger than the number for the first, and the second model has a letter label that is equal to or larger than (i.e., later alphabetically) the letter for the first. Thus, Model 2A is nested within, and is therefore a restricted version of, Models 3A, 2C, and 3C, among others. But, Model 2A is not nested within Model 1C, because the former model has a higher number whereas the latter model has a higher letter. The above rule for determining nesting of models is generally applicable, given the stipulation that Structures B and B' cannot be compared, nor can Structures 2 and 2'.

The spatial layout of models in Table 1 can also be used to determine nesting of models. The two horizontal and two vertical lines in the body of Table 1 divide the table body into nine areas. Given these lines of demarcation, a particular model is nested within all other models separated by at least one line of demarcation that lie to the right and/or below the model in question. Thus, Model 2B' is nested within Models 2C, 3B', and 3C, but not within Model 3A, since the latter model lies below and to the left of Model 2B'. As before, Structures 2 and 2' cannot be compared, nor can Structures B and B'.

Four features of the taxonomy of models in Table 1 should be noted. First, three of the models present interpretive or mathematical problems. Since two of the models, Models 2A and 1B, each contain one general factor on which all measures load, the two models are not differentiable on either interpretive or mathematical grounds. The third problematic model, Model 2B, contains two general factors, but in the typical situation in which each trait is measured using each method, it is not possible to identify mathematically two separate general factors.

The second feature of the taxonomy in Table 1 concerns the representativeness of the set of models included. There are a very large number of models for multitrait-multimethod data that might be proposed but that are not included in the taxonomy. For example, it would be possible to specify a model with  $t - 1$  trait factors, collapsing two of the  $t$  a priori trait factors onto a single dimension, but allowing the correlations among the  $t - 1$  trait factors to be freely estimated. Such a model would occupy an intermediate position between trait Structures 2 and 3. The taxonomy in Table 1 is offered both to serve an heuristic, organizing function and to delimit the array of reasonable a priori models for multitrait-multimethod data, with

the expectation that the set of models in Table 1 will be sufficient for most research applications. However, a researcher should entertain alternative models not included in the taxonomy in Table 1 if the researcher's theory or the data, in particular research applications, so dictate.

The third feature of the taxonomy in Table 1 that should be noted is that the models stop short of estimating parameters in  $\Phi_{MT}$ , which contains correlations among trait and method factors. Constraining  $\Phi_{MT}$  to be a null matrix leads to orthogonality of the trait and method factor spaces, a condition that was considered highly desirable by Jackson (1975) and that results in a decomposition of the variance of each measured variable into an additive combination of trait, method, and error variance. In addition, experts in structural equation modeling have stated that models containing estimates in  $\Phi_{MT}$  are very likely not identified (K. G. Jöreskog, personal communication, October 29, 1982) and present both logical and empirical estimation problems of great magnitude (P. M. Bentler, personal communication, March 4, 1982). Such estimation problems have led to unacceptably large, out-of-bound estimates of factor loadings in several studies in which correlations among trait and method factors were estimated (e.g., Kalleberg & Kluegel, 1975; Lee, 1980; Schmitt, 1978). Experience suggests that allowing trait-method correlations also results in unacceptably large standard errors for many model parameters, due to high levels of dependence among parameter estimates, that is, high correlations among estimates. For all these reasons, all models considered in the present paper have  $\Phi_{MT}$  as a fixed null matrix. It is beyond the scope of the present paper to determine conditions that would allow well-identified estimates of correlations among trait and method factors.

The fourth and final feature of the taxonomy of models in Table 1 is that none of the models incorporates the specification of a general factor in the presence of both trait and method factors. Although some researchers have successfully fit such models to data (e.g., Borch & Wolins, 1970), it appears that the interpretation of a general factor that is part of a model which also includes both trait and method factors may be rather indetermi-

nate. That is, though there may be a tendency to interpret such a general factor as a general trait factor, it is possible that method covariation alone may account for the general factor. To avoid this ambiguity, none of the models considered in the present paper allowed the specification of a general factor in a model having both trait and method factors.

*Specification of alternate series of nested models.* As discussed by Jöreskog (1971, 1974) and by Bentler and Bonett (1980), structural equation modeling is used most powerfully in situations in which it is possible, on an a priori basis, to specify a series of nested models. In addition to assessing the statistical significance and practical fit of each of two alternative models, a major benefit of employing nested models is that the difference in the  $\chi^2$  values of two nested models is itself distributed as  $\chi^2$  with degrees of freedom equal to the difference in degrees of freedom for the two models. Thus, the difference in fit between two nested models may be tested statistically, in addition to noting the difference in practical fit between the models.

There are many potential rationales for specifying series of nested models for multitrait-multimethod data. The first rationale considered here is based on parsimony; all other things being equal, simplicity of a model is preferred. Under this rationale, it is reasonable to start with a highly restricted model and relax restrictions only if a model fails to represent adequately a set of data. Thus, a researcher may begin with the null model as a null hypothesis, the null model specifying a lack of correlation among observed variables. Referring to the multitrait-multimethod factor model in Equation 4, the null hypothesis may be written as:

Model 1A: Estimate parameters in  $\Psi$ , leaving  $\Lambda_T$ ,  $\Lambda_M$ , and  $\Phi_{MT}$  as null matrices,  $\Phi_{TT} = \mathbb{I}_n$ ,  $\Phi_{MM} = \mathbb{I}_m$ .

Assuming that the null hypothesis, Model 1A, is rejectable, restrictions on various submatrices may be relaxed. As noted by Jöreskog (1971), a substantive model that seems a natural initial hypothesis is a model specifying that covariation among observed variables is due only to trait factors and

their intercorrelations. This first alternative model may be written as:

Model 3A: Estimate parameters in  $\Lambda_T$ ,  $\Phi_{TT}$ , and  $\Psi$ , leaving  $\Lambda_M$  and  $\Phi_{MT}$  as null matrices,  $\Phi_{MM} = \mathbf{I}_m$ .

If Model 3A was rejectable on statistical and/or practical grounds, this may indicate that method effects are present, though sources of covariation other than method effects could lead to poor fit of a model such as that represented by Model 3A. However, perhaps the most obvious way to respecify the Model 3A would be to add method factors. Campbell and Fiske (1959) urged researchers to use methods that are as independent as possible. To allow estimation of a model in which methods have independent effects, orthogonal method factors could be added to Model 3A, resulting in the following model:

Model 3B': Estimate parameters in  $\Lambda_T$ ,  $\Phi_{TT}$ ,  $\Lambda_M$ , and  $\Psi$ , leaving  $\Phi_{MT}$  as a null matrix,  $\Phi_{MM} = \mathbf{I}_m$ .

If Model 3B' was rejectable, a final, theoretically reasonable respecification would be to allow correlations among the method factors. This would lead to the following model:

Model 3C: Estimate parameters in  $\Lambda_T$ ,  $\Phi_{TT}$ ,  $\Lambda_M$ ,  $\Phi_{MM}$ , and  $\Psi$ , leaving  $\Phi_{MT}$  as a null matrix.

A second rationale for formulating a series of nested models is one based on the partialling of irrelevant sources of variance. This second rationale is a realistic alternative to the first, due to the prevalence of method effects in many areas of psychological inquiry (see, e.g., Campbell & Fiske, 1959). Under this rationale, assuming that the null model was rejectable, the next model considered would be Model 1C, which includes  $m$  freely correlated method factors. If Model 1C was not rejectable, the data would be adequately represented by a model that had estimates of trait-related variance constrained at zero, a theoretically disconcerting, but empirically plausible, type of model.

If Model 1C was rejectable, a researcher could then proceed to either Model 2C or Model 2'C; if the selected model, 2C or 2'C, was rejectable, the remaining option would be to relax the restrictions on the trait factor correlations, resulting in Model 3C. Regardless of the model selected as the best

representation of the data, under the second rationale the effects of trait factors would be estimated only after partialling out the effects of the irrelevant, method-related sources of variance.

A third possible rationale is based on determining the proportion of covariation among observed measures uniquely explained by each set of parameters. Following this rationale, a researcher might fit many or all of the models listed in Table 1, and then engage in a wide array of nested model comparisons to investigate the covariation explained by a particular set of estimates under a variety of conditions. Although following this rationale might result in a multitude of tests between nested models that may be difficult to summarize, this rationale might lead to the selection of a best-fitting model overlooked by following either of the first two rationales.

In the model fitting reported in later sections of this paper, the first rationale, based on parsimony, was explicitly followed. But, in the interest of completeness, a large number of models in the taxonomy in Table 1 were fit to the data sets. Specifically, all models in Table 1 were fit, except those involving trait Structure 2' or method Structure B. Typically, researchers would like to demonstrate the discriminant validity of trait constructs; this would be attained by showing a significant difference in fit between a model with trait Structure 3 and a model with the identical method structure but trait Structure 2, as the latter trait structure has a complete absence of discriminant validity. Since researchers employing multitrait-multimethod approaches typically have little interest in verifying that the trait construct intercorrelations depart significantly from zero, models with trait Structure 2', proposed in order to allow such comparisons, were not fit to data.

With regard to method structure, little attention has been paid to the discriminability of method effects. Campbell and Fiske (1959) urged the use of methods that were as independent as possible; this desired feature is embodied in method Structure B'. Failing this, method factors might have covarying effects, as under method Structure C. The contrast in fit between method Structures B' and C, therefore, provides a test of whether meth-

ods have correlated effects. Although a verification that method factors had significantly discriminable effects might be of interest in particular research situations, such demonstrations, which would involve models with method Structure B, are usually of minor import and thus were not pursued in the present paper.

As a final note, in analyses reported later, only the a priori models in Table 1 were fit to data, and no extra, "garbage" parameters (see Browne, 1984) were allowed, parameters which would have resulted in better fit of models to data. It is well-known that in multiple regression analysis the  $F$  test for difference between a priori nested regression models is distributed as a central  $F$ . However, because of the capitalization on chance in empirically-driven stepwise regression, the  $F$  for increase in  $R^2$  in such analyses is no longer distributed as a central  $F$ , and special tables must be used to evaluate significance (see Wilkinson, 1979). In covariance structure analyses, though the estimation problems are much more complex and the following conjecture might therefore be subject to debate, it is quite possible that the  $\chi^2$  test of difference between a priori nested models is distributed as a central  $\chi^2$ , whereas the  $\chi^2$  resulting from nested models respecified on the basis of empirical considerations (e.g., modification indices) would be distributed as a noncentral  $\chi^2$ , which would be difficult to evaluate properly (Cliff, 1983).

*Estimating and testing the degree of convergent and discriminant validity and of method variance.* In order to test the significance of a set of parameters using structural equation modeling, it is common to compare the fit of two models, one of which includes the set of parameters and the other of which excludes only the set of parameters to be tested. Comparisons of this sort may be formulated in order to test the degree of convergent and discriminant validity, as well as the degree of method variance, exhibited by a set of measures.

To begin, a structural model adequately representing a set of data must be determined; suppose that a model corresponding to Model 3C fit a hypothetical multitrait-multimethod matrix. Consider first the various submatrices contained in  $\Lambda$  and  $\Phi$  in Model 3C. The factor loading in  $\Lambda_T$  for a given

measured variable reflects the saturation of the measured variable with a particular hypothesized trait factor, and the square of the factor loading provides an estimate of the proportion of variance of the measure associated with the trait factor. In addition, the product of two factor loadings in the same column of  $\Lambda_T$  provides an estimate, based on the common factor saturations of the two measures, of the monotrait-heteromethod, or validity diagonal, correlation between the two observed measures involved. For these reasons, loadings in  $\Lambda_T$  are structural modeling analogues of, and are measures of, convergent validation of measures.

The correlations in  $\Phi_{TT}$  provide estimates of the degree of relationship between trait factors. A zero correlation between trait factors indicates that the factors are easily discriminable; and, as correlations between trait factors increase, the factors are less discriminable. Further, a perfect correlation between two trait factors would represent a complete lack of ability to discriminate the trait constructs empirically. Thus, elements of  $\Phi_{TT}$  are structural modeling analogues of discriminant validity among trait constructs, with lower interfactor correlations implying greater discriminant validity among trait constructs.

Parallel interpretations may be made of elements in  $\Lambda_M$  and  $\Phi_{MM}$  with regard to the effects of method factors. That is, loadings in  $\Lambda_M$  represent method-related variance for each measure, and correlations in  $\Phi_{MM}$  reflect the discriminability of the effects of method factors. By contrasting the fit of Model 3C with properly respecified, more restricted models, tests of convergent and discriminant validity and of the amount of method-related variance are possible.

To obtain an estimate and test of the covariance among measures uniquely explained by trait factors, which represent convergent validity, the fit of Model 3C would be compared with that of Model 1C, which has  $\Lambda_T$  as a fixed null matrix and  $\Phi_{TT}$  as an identity matrix. Such a model would attempt to fit the multitrait-multimethod data with correlated method factors only; if such a model fit the data virtually as well, on both statistical and practical grounds, as did Model 3C, then there would be little justification for retention of trait factors,

that is, little evidence of convergent validity.

To test for discriminant validity, the fit of Model 3C would be compared to the fit of Model 2C, which has fixed perfect correlations among trait factors. Because correlations among latent variables are corrected for unreliability (Bentler, 1980), it is possible for the correlations among latent variables to approach unity, even though the correlations among observed measures are substantially lower. A perfect correlation between two latent variables reflects a complete inability to discriminate empirically between the variables. The alternative model with fixed perfect correlations among trait factors attempts, in essence, to represent the correlations among observed variables with a single trait factor and  $m$  correlated method factors. If this alternative model fit the data virtually as well as did Model 3C, there would be little evidence of discriminant validity among the trait constructs.

Finally, to obtain an estimate and test of the amount of method covariance among the measures, the fit of Model 3C would be compared to a model that had  $\Lambda_M$  specified as a null matrix and  $\Phi_{MM}$  as an identity matrix—Model 3A. Recalling the series of nested models formulated following the first, parsimony-based rationale discussed above, the test for method covariance is implicit in the fitting of those models. That is, if Model 3B' fit the data better than did Model 3A, there is a significant amount of method variance in the set of measures; and if Model 3C provides better fit than does Model 3B', then the method factors display significant levels of covariation. Thus, the fit of Model 3A may be compared to the fit of Models 3B' and 3C to estimate and test the significance of method covariation in the set of observed measures.

#### Data

Three multitrait-multimethod matrices were reanalyzed in the present study, two of which were previously analyzed by both Bagozzi (1978) and Schmitt (1978). The first of these matrices is from a study by Ostrom (1969), who investigated attitudes toward the church in a sample of 189 undergraduate students. The 12-variable matrix consists of correlations among the affective, behavioral, and cognitive components of attitudes toward the

church, each measured using four different types of scales: Thurstonian equal-appearing intervals, Likert summated ratings, Guttman scalogram analytic, and self-rating scales.

The second matrix, from a study by Kothandapani (1971), contains correlations among measures of attitudes toward birth control based on a sample of 100 lower-income Black women. The Kothandapani data also consist of 12 variables, assessing the affective, cognitive (or belief), and behavioral (i.e., intention to act) components of attitudes toward birth control using the same four methods as used by Ostrom (1969).

The third matrix, previously analyzed by Lee (1980), is from a study by Lawler (1967). Lawler, studying the job performance of 113 middle- and top-level managers, asked superiors, peers, and the managers themselves to rate the managers on three traits: the quality of their job performance, their ability to perform the job, and the effort they put forth on their jobs. As a result, in the Lawler study, the methods refer to the type of rater—supervisor, peer, or self—providing the data.

The three matrices to be analyzed were described above in sufficient detail to allow interpretation of structural modeling results. Details of item and scale construction, sample selection, procedures, and so forth, are provided in the original reports (Kothandapani, 1971; Lawler, 1967; Ostrom, 1969). Due to space considerations, the three correlation matrices are not presented in the present paper. The Kothandapani (1971) and Ostrom (1969) matrices may be found in the original articles as well as in Bagozzi (1978), and the Lawler (1967) matrix may be found in the original source as well as in Lee (1980).

#### Analyses

All analyses were performed using the LISREL V program (Jöreskog & Sörbom, 1981). In all model fitting, inequality constraints were invoked to ensure nonnegativity of unique variances in  $\psi$ , using procedures outlined by Rindskopf (1983). Because the Rindskopf (1983) procedure for ensuring nonnegativity of unique variances involves incorporating unique factors in the primary factor pattern matrix, unique factor loadings (i.e., the square root

of corresponding unique factor variances) were estimated, rather than estimating unique factor variances directly. As a result, in subsequent tables in which parameter estimates are presented the estimated unique factor loadings are reported, and these values must be squared to obtain estimates of unique factor variances.

Finally, a single model is typically identified as the best representation of patterns in a given set of data. Nested tests of convergent and discriminant validation and of method variance are also provided for several alternate models as well.

### Results

#### Ostrom Data

*Structural modeling of the 12-variable matrix.* Indices of fit of an array of structural models to the Ostrom (1969) data are presented in Table 2, and nested comparisons of the fit of these models are presented in Table 3. Limiting attention for the present to the top halves of Tables 2 and 3, results

of modeling the 12-variable matrix are given. Models 1A through 3C correspond to the series of models of increasing complexity that was discussed above, and the remaining models—Models 1B' through 2C (see Table 1)—were specified to allow the estimation of the degree of convergent and discriminant validity shown by the measures.

Although statistically rejectable, Model 3A attained fairly high levels of practical fit, with rho of .939 and delta of .928. Although warning of the attendant dangers, Bagozzi (1978) examined the residuals from the present Model 3A. Finding the three residual correlations among the self-rating measures to be relatively large (ranging from .112 to .116), Bagozzi (p. 18) claimed that the large residual correlations indicated "an excessive amount of error" in the self-rating measures; Bagozzi then dropped these variables from all subsequent analyses and proceeded to fit models only to the 9-variable matrix containing the Thurstone-, Likert-, and Guttman-scaled measures. But, the Bagozzi

Table 2  
Goodness-of-Fit Indices for Covariance Structure Models  
Estimated from Ostrom (1969) Multitrait-Multimethod Data

Matrix and Model	Likelihood ratio test			rho	delta
	Chi-square	df	Prob		
<b>12-Variable Matrix</b>					
Model 1A	1871.92	66	.000	---	---
Model 3A	135.48	51	.000	.939	.928
Model 3B'	53.87	39	.057	.986	.971
Model 3C	28.86	33	.673	1.005	.985
Model 1B'	775.62	54	.000	.512	.586
Model 1C	74.87	48	.008	.980	.960
Model 2A	141.15	54	.000	.941	.925
Model 2B'	72.66	42	.002	.973	.961
Model 2C	39.38	36	.321	.997	.979
<b>9-Variable Matrix</b>					
Model 1A	1271.53	36	.000	---	---
Model 3A	27.08	24	.301	.996	.979
Model 3B'	13.99	15	.526	1.002	.989
Model 3C	10.16	14	.750	1.008	.992
Model 1B'	516.29	27	.000	.472	.594
Model 1C	43.44	26	.017	.980	.966
Model 2A	43.45	27	.024	.982	.966
Model 2B'	37.17	18	.005	.969	.971
Model 2C	17.32	17	.433	.999	.986

Note. Sample size = 189.

Table 3  
 Indices of Difference Between Nested Covariance Structure Models  
 Based on Ostrom (1969) Multitrait-Multimethod Data

Matrix and Comparison	Difference in			Difference in	
	Chi-square	df	Prob	rho	delta
<b>12-Variable Matrix</b>					
Tests of added components					
Model 1A vs Model 3A	1736.44	15	.000	.939	.928
Model 3A vs Model 3B'	81.61	12	.000	.047	.043
Model 3B' vs Model 3C	25.61	6	.000	.019	.014
Tests of convergent validity					
Model 3A vs Model 1A	1736.44	15	.000	-.939	-.928
Model 3B' vs Model 1B'	721.75	15	.000	-.474	-.385
Model 3C vs Model 1C	46.01	15	.000	-.025	-.025
Tests of discriminant validity					
Model 3A vs Model 2A	5.67	3	.129	.002	-.003
Model 3B' vs Model 2B'	18.79	3	.000	-.013	-.010
Model 3C vs Model 2C	10.52	3	.015	-.008	-.006
<b>9-Variable Matrix</b>					
Tests of added components					
Model 1A vs Model 3A	1244.45	12	.000	.996	.979
Model 3A vs Model 3B'	13.09	9	.159	.006	.010
Model 3B' vs Model 3C	3.83	1	.051	.006	.003
Tests of convergent validity					
Model 3A vs Model 1A	1244.45	12	.000	-.996	-.979
Model 3B' vs Model 1B'	502.30	12	.000	-.530	-.395
Model 3C vs Model 1C	33.28	12	.001	-.028	-.026
Tests of discriminant validity					
Model 3A vs Model 2A	16.37	3	.001	-.014	-.013
Model 3B' vs Model 2B'	23.18	3	.000	-.033	-.018
Model 3C vs Model 2C	7.16	3	.067	-.009	-.006

Note. Sample size = 189.

(1978) conclusion regarding the self-rating variables appears to be a misinterpretation of the residual correlations, which may simply indicate the need for specification of a method factor for the self-rating measures. Indeed, fitting such a model (Model 3B') with orthogonal method factors resulted in a highly significant improvement in statistical fit, improved substantially both measures of practical fit, as shown in Table 3, and in the process led to greatly reduced residual correlations.

Although Model 3B' was not rejectable statistically, the borderline level of fit ( $p = .057$ ) and the presence of an additional a priori respecification, allowing correlations among the method factors, led to consideration of Model 3C. As shown in Table 3, Model 3C, which is clearly not reject-

able, provided both a highly significant improvement in statistical fit and meaningful increases in levels of practical fit over those shown by Model 3B'.

Assuming that Model 3C is the appropriate representation of the Ostrom (1969) 12-variable matrix, tests of convergent and discriminant validity and of method variance may be performed. The test of convergent validity of Model 3C involved the comparison of Models 3C and 1C, the latter model identical to Model 3C except that  $\Lambda_T$  was a fixed null matrix and  $\Phi_{TT}$  a fixed identity matrix. This comparison revealed that, though Model 1C resulted in significantly worse statistical fit to the Ostrom data, both measures of practical fit showed quite small, though not unimportant, decreases in

level of fit. This finding suggests that the proportion of covariation among observed measures uniquely representable as convergent validation (i.e., uniquely accounted for by trait factors), though significant, is of only very modest degree

To test for discriminant validity, Model 3C was compared to Model 2C, in which trait factors had fixed perfect intercorrelations. This comparison resulted in a statistically significant difference in fit ( $p = .015$ ) favoring the less restricted model, Model 3C. However, because neither measure of practical fit revealed an important difference between Models 3C and 2C and because its levels of practical fit ( $\rho = .997$ ,  $\delta = .979$ ) were quite adequate, Model 2C may be preferred over Model 3C. This latter conclusion is supported by several considerations: (1) Model 2C is more restricted than Model 3C, having three fewer parameter estimates and one, rather than three, trait factors; (2) it is a non-rejectable model ( $p = .321$ ); and (3) it fits the data virtually as well as did Model 3C.

The estimation and test of unique covariance in Model 3C attributable to method factors is obtained by comparing Model 3C to Model 3A. The covariance represented by method factors is highly significant statistically,  $\chi^2(18) = 107.22$ ,  $p < .0001$ , and is also important practically, as  $\rho$  and  $\delta$  changed .066 and .057, respectively.

Three patterns emerging from the difference tests reported in Table 3 may be noted. First, the amount of covariance among observed measures uniquely explained by trait factors (i.e., convergent validation) is highly dependent on the model accepted as a representation of the data, whether Model 3A, 3B', or 3C. Once correlated method factors are allowed, as in Model 3C, unique convergent covariation is reduced to a rather modest figure, .025 for each measure of practical fit. Second, the degree of discriminant validity among trait factors was not large at all regardless of which model was chosen to represent the Ostrom (1969) data, with differences in measures of practical fit ranging from +.002 to -.013, depending on the model considered. Third, the covariation among measures uniquely attributable to method factors was rather larger than that uniquely attributable to trait factors:

deleting method factors and their intercorrelations from Model 3C led to decreases in practical fit, .066 in  $\rho$  and .057 in  $\delta$ , that were more than twice as large as the decreases observed when deleting trait factors and their intercorrelations from Model 3C.

The latter two observations made on the basis of overall measures of goodness-of-fit are also apparent when examining the estimates and associated standard errors for parameters of Model 3C, which are presented in Table 4. First, the relative importance of method over trait variance is evident in the magnitude of the respective factor loadings; loadings on method factors tend to be rather larger than loadings on trait factors. Furthermore, only 4 of the 12 trait factor loadings differ significantly from zero, whereas all 12 loadings on method factors are highly significant. Second, the lack of clear discriminant validity is shown by the trait factor intercorrelations, which, though differing significantly from zero, are each approximately one standard error from unity, thus differing nonsignificantly from the latter value. By comparison, four of the six method factor intercorrelations are two or more standard errors from unity, supporting an interpretation that method factors have more discriminable effects than do trait factors in the 12-variable matrix.

The preceding considerations seem to point to a common conclusion: Model 2C, a model with trait factors displaying a lack of discriminant validity (i.e., correlating unity) and with four correlated method factors, appeared to be a more appropriate representation of the Ostrom (1969) 12-variable matrix than were any of the remaining eight models reported in Table 2.

*Structural modeling of the 9-variable matrix.* The bottom halves of Tables 2 and 3 report, respectively, measures of fit and measures of difference in fit of structural models fit to the Ostrom (1969) 9-variable matrix. As mentioned above, Bagozzi (1978) claimed that the self-rating measures contained excessive amounts of error variability, and thus he concentrated his attempts at modeling the Ostrom data on the 9-variable matrix containing

Table 4  
Factor Pattern and Factor Intercorrelation Matrices from Model 3C  
Estimated from Ostrom (1969) 12-Variable Multitrait-Multimethod Data

Variable	Trait				Factor				Unique
	1	2	3		I	II	III	IV	
					Primary Factor Pattern				
Thurstone									
Affect	.27(.25)	.0*	.0*	.79(.10)	.0*	.0*	.0*	.0*	.55(.05)
Behavior	.0*	.57(.18)	.0*	.55(.19)	.0*	.0*	.0*	.0*	.62(.04)
Cognition	.0*	.0*	.52(.20)	.67(.17)	.0*	.0*	.0*	.0*	.54(.04)
Likert									
Affect	.35(.27)	.0*	.0*	.0*	.84(.13)	.0*	.0*	.0*	.44(.03)
Behavior	.0*	.48(.24)	.0*	.0*	.80(.15)	.0*	.0*	.0*	.38(.04)
Cognition	.0*	.0*	.41(.24)	.0*	.82(.13)	.0*	.0*	.0*	.41(.03)
Guttman									
Affect	.19(.21)	.0*	.0*	.0*	.0*	.0*	.64(.09)	.0*	.75(.05)
Behavior	.0*	.59(.19)	.0*	.0*	.0*	.0*	.55(.19)	.0*	.60(.05)
Cognition	.0*	.0*	.36(.23)	.0*	.0*	.0*	.73(.12)	.0*	.60(.04)
Self-rating									
Affect	.68(.16)	.0*	.0*	.0*	.0*	.0*	.0*	.74(.19)	.00(1.0)
Behavior	.0*	.28(.23)	.0*	.0*	.0*	.0*	.0*	.83(.09)	.50(.06)
Cognition	.0*	.0*	.39(.23)	.0*	.0*	.0*	.0*	.75(.13)	.55(.04)
					Factor Intercorrelations				
Trait 1	1.0*								
Trait 2	.82(.19)	1.0*							
Trait 3	.85(.17)	.81(.18)	1.0*						
Method I	.0*	.0*	.0*	1.0*					
Method II	.0*	.0*	.0*	.94(.03)	1.0*				
Method III	.0*	.0*	.0*	.96(.04)	.95(.04)	1.0*			
Method IV	.0*	.0*	.0*	.80(.08)	.82(.07)	.85(.07)	1.0*		

Note. All asterisked parameters were fixed at the reported values. All parameter estimates differing significantly from zero are underscored, and parenthesized values are standard errors of associated parameters.

Thurstone-, Likert-, and Guttman-scaled measures.

Model 3A, with correlated trait factors only, fit the 9-variable matrix quite well, both statistically ( $p = .30$ ) and practically ( $\rho = .996$ ,  $\delta = .979$ ). Bagozzi (1978) tested the discriminant validity of the Model 3A trait factors, finding a significant degree of discriminant validity,  $\chi^2(3) = 16.37$ ,  $p < .001$ . The present analysis adds the observation that the fitting of one, rather than three, trait factors (see Model 2A), though producing a statistically significant decrement in fit, is not a respecification of great practical importance (change in  $\rho = -.014$  and in  $\delta = -.013$ ; see Table 3).

Bagozzi (1978) stated that, finding a significant degree of discriminant validity, "the next step in the analyses should introduce method factors" (p. 20). It must be noted that the preceding statement by Bagozzi is neither dictated by the test of discriminant validity, as he implied, nor necessarily consistent with the overall test of significance. A researcher could terminate the fitting of models with Model 3A, a model that is quite nonrejectable ( $p = .30$ ), has three strong trait factors, and has a statistically significant, though practically unimpressive, degree of discriminant validity among the trait constructs. Indeed, it is easy to defend the position that a researcher should only proceed to add method factors if the data dictate, that is, if a model without method factors is statistically rejectable; under such a position, Model 3A would be accepted as the appropriate representation for the 9-variable matrix.

Because Bagozzi (1978) fit additional models to the Ostrom (1969) 9-variable matrix, models incorporating method factors were evaluated. Consider Model 3C with three correlated trait factors and two correlated method dimensions, as the Thurstone and Likert method factors correlated perfectly and therefore correlated equally with the Guttman method factor. Accepting a model identical to the present Model 3C, Bagozzi (p. 20) claimed that "one may conclude that convergent and discriminant validity have been achieved" for the 9-variable matrix. In contrast, the present pa-

per's position is that only now, after accepting a model as a representation for the 9-variable matrix, may convergent and discriminant validity and the degree of method variance be evaluated properly, and this requires a series of comparisons among appropriately specified nested models.

A first issue to be addressed is the degree of method variance; how necessary was the addition of method factors and their intercorrelations? Comparing Model 3C with Model 3A, the difference in fit is of borderline significance statistically,  $\chi^2(10) = 16.92$ ,  $p = .076$ , and the changes in  $\rho$  (.012) and  $\delta$  (.013) were rather small, barely satisfying the criterion for practical importance. Thus, the covariation among observed measures uniquely due to method factors is of quite modest magnitude.

The test of convergent validity requires comparison of Models 3C and 1C; this contrast was statistically highly significant,  $\chi^2(12) = 33.28$ ,  $p < .001$ . However, the differences in  $\rho$  (.028) and  $\delta$  (.026) approximated the corresponding figures from the 12-variable analyses, and represent quite small proportions of covariance attributable uniquely to convergent validation.

The final contrast of interest was that between Model 3C and Model 2C, estimating and testing the degree of discriminant validity. This comparison was neither statistically significant nor practically important, suggesting, contrary to Bagozzi's (1978) conclusions, that there is essentially no evidence of discriminant validity among trait factors from the 9-variable matrix if Model 3C is chosen as the model best representing the data.

As with the 12-variable matrix, consideration of the indices of model difference reported in Table 3 for the 9-variable matrix reveals three trends: (1) that the estimate of the degree of convergent validity shown by the measures is heavily dependent on the model accepted; (2) that none of the tests of discriminant validity are of notable practical importance; and (3) that neither trait (i.e., convergent) nor method factors uniquely represent sizable proportions of covariance among observed measures.

The estimates and associated standard errors of parameters from Model 3C for the 9-variable ma-

trix are presented in Table 5. Interestingly, Bagozzi (1978, Table 4) presented the trait and method factor loadings, but neither presented nor discussed either the correlations among factors or the standard errors (or significance) of any of the parameter estimates. When both estimates and standard errors of all parameters are considered, the problematic status of Model 3C is apparent. As shown in Table 5, only three of the nine loadings on trait factors differed significantly from zero, with the third factor having not one significant loading. The method factors fared even less well; not one of the nine method factor loadings differed significantly from zero. Thus, when Bagozzi (1978, Table 5) presented squared factor loadings as a decomposition of variance of each measure into trait-related, method-related, and error variance, Bagozzi should have noted (but failed to do so) that 15 of the 18 trait- and method-related variance components did not differ reliably from zero, a particularly damaging admission.

The lack of discriminant validity is evident from the fact that the estimates of correlations among the trait factors were each less than one standard error from 1.00, thus differing nonsignificantly from unity. This finding supports the contention that affective, behavioral, and cognitive components of attitude toward the church are not empirically distinguishable constructs in the Ostrom (1969) 9-variable matrix.

In summary, it appears that Bagozzi (1978) misinterpreted the overall test of fit of the present Models 3A and 3C; the nonsignificant chi-square values and very large measures of practical fit only imply that these models fit the data well, not that convergent and discriminant validation had been demonstrated. In fact, and contrary to Bagozzi's conclusions, when appropriate model comparisons are performed (as shown in Table 3), there are nonsignificant proportions of method variance and discriminant validity, and statistically significant but practically very small portions of covariance related uniquely to trait factors and thus to convergent validation of measures in the 9-variable matrix.

#### Kothandapani Data

*Structural modeling of the 12-variable matrix.* The indices of fit of the set of structural

models to the 12-variable matrix from Kothandapani (1971) are presented in the top half of Table 6; nested comparisons among these models are presented in the top half of Table 7.

The fit of Model 3A to the Kothandapani (1971) data was very poor on both statistical and practical grounds. As with the Ostrom (1969) data, Bagozzi (1978) inspected the residuals from the present Model 3A. Finding the residual correlations among the self-rating measures to be rather large (ranging from .389 to .439) and attributing the magnitude of these covariances to the errorful nature of the ratings, Bagozzi (1978) again eliminated the self-rating measures from consideration and fit structural models to the 9-variable matrix containing Thurstone-, Likert-, and Guttman-scaled measures. But the large residual correlations for the self-rating measures appear to point to the need for a method factor for these measures, rather than to the level of error variance present. Fitting a model that included orthogonal method factors (Model 3B') and then a model adding correlations among the method factors (Model 3C) revealed improvements in indices of fit over those shown by Model 3A that were highly significant statistically and important practically.

The resultant model, Model 3C, with correlated traits and correlated method factors, appears to be the most acceptable representation for the 12-variable matrix of all of the alternative models in Table 6. Although Model 3C is rejectable statistically, it should be noted that the lack of statistical fit is not extreme, that is,  $.01 < p < .05$ . In addition, though the levels of practical fit of Model 3 are not impressively large, the rho (.945) and delta (.933) values are clearly acceptable according to the Bentler and Bonett (1980) guidelines. Undoubtedly, respecification of Model 3C, such as allowing correlated unique factors, would result in a statistically nonrejectable model that had even higher levels of practical fit. However, such respecifications would be made to a model that is presently largely acceptable, and would be made on empirical (rather than a priori theoretical) bases, resulting in chi-square difference measures that would be difficult to evaluate properly. For these reasons and because Model 3C had rather acceptable levels of statistical and practical fit, Model 3C may be considered an

Table 5  
Factor Pattern and Factor Intercorrelation Matrices from Model 3C  
Estimated from Ostrom (1969) 9-Variable Multitrait-Multimethod Data

Variable	Trait			Factor			Method		
	I	2	3	I	II	III	II	III	Unique
Primary Factor Pattern									
Thurstone									
Affect	.76(.29)	.0*	.0*	.35(.68)	.0*	.0*	.0*	.0*	.55(.07)
Behavior	.0*	.73(.27)	.0*	.28(.70)	.0*	.0*	.0*	.0*	.63(.05)
Cognition	.0*	.0*	.71(.38)	.40(.68)	.0*	.0*	.0*	.0*	.58(.04)
Likert									
Affect	.66(.60)	.0*	.0*	.0*	.62(.62)	.0*	.0*	.0*	.43(.05)
Behavior	.0*	.68(.60)	.0*	.0*	.65(.60)	.0*	.0*	.0*	.34(.09)
Cognition	.0*	.0*	.66(.61)	.0*	.62(.65)	.0*	.0*	.0*	.42(.04)
Guttman									
Affect	.53(.35)	.0*	.0*	.0*	.0*	.38(.49)	.0*	.0*	.76(.04)
Behavior	.0*	.74(.27)	.0*	.0*	.0*	.31(.69)	.0*	.0*	.61(.06)
Cognition	.0*	.0*	.62(.47)	.0*	.0*	.51(.56)	.0*	.0*	.60(.05)
Factor Intercorrelations									
Trait 1	1.0*								
Trait 2	.86(.18)	1.0*							
Trait 3	.94(.10)	.90(.12)	1.0*						
Method I	.0*	.0*	.0*	1.0*					
Method II	.0*	.0*	.0*	1.0*	1.0*				
Method III	.0*	.0*	.0*	.94(.19)	.94(.19)	1.0*			

Note. All asterisked parameters were fixed at the reported values. All parameter estimates differing significantly from zero are underscored, and parenthesized values are standard errors of associated parameters.

Table 6  
Goodness-of-Fit Indices for Covariance Structure Models  
Estimated from Kothandapani (1971) Multitrait-Multimethod Data

Matrix and Model	Likelihood ratio test			rho	delta
	Chi-square	df	Prob		
12-Variable Matrix					
Model 1A	785.54	66	.000	---	---
Model 3A	367.25	51	.000	.431	.532
Model 3B'	73.96	39	.001	.918	.906
Model 3C	52.71	33	.016	.945	.933
Model 1B'	500.57	54	.000	.241	.363
Model 1C	416.20	48	.000	.296	.470
Model 2A	489.87	54	.000	.260	.376
Model 2B'	273.15	42	.000	.495	.652
Model 2C	179.56	36	.000	.634	.771
9-Variable Matrix					
Model 1A	504.81	36	.000	---	---
Model 3A	202.97	24	.000	.427	.598
Model 3B'	29.76	15	.013	.924	.941
Model 3C	24.20	12	.019	.922	.952
Model 1B'	326.92	27	.000	.147	.352
Model 1C	279.03	24	.000	.184	.447
Model 2A	319.60	27	.000	.168	.367
Model 2B'	160.43	18	.000	.392	.682
Model 2C	116.31	15	.000	.481	.770

Note. Sample size = 100.

adequate representation of the Kothandapani (1971) 12-variable matrix.

The test of convergent validity, obtained by comparing Models 3C and 1C, showed a highly significant difference between models in the level of statistical fit (see Table 7). The very large differences in rho ( $-.649$ ) and delta ( $-.463$ ) demonstrate that the difference is extremely important practically as well, with convergent validation explaining uniquely a large portion of covariation among observed measures.

The degree of discriminant validity among trait factors was also highly significant on both statistical and practical grounds. As shown in Table 7, levels of both rho and delta would fall precipitously,  $-.311$  and  $-.162$ , respectively, if only a single trait dimension was allowed.

As noted above, covariation among observed measures uniquely explained by method factors was estimated by comparing Model 3C with Model 3A.

This comparison revealed a difference between models that was both significant statistically,  $\chi^2(18) = 314.54$ ,  $p < .0001$ , and important practically, with change in rho and delta of  $.514$  and  $.401$ , respectively.

Considering the model difference tests given in the top half of Table 7, three trends are clear, each contrasting greatly with trends evident from analyses of the Ostrom (1969) 12-variable matrix. First, the proportion of covariation among observed measures uniquely representable as convergent validation (i.e., due to trait factors) was very large and was unrelated to the model chosen as a representation of the data. That is, the deletion of trait factors from Models 3A, 3B', or 3C resulted in large and unacceptable decreases in fit of the covariance structure model to the data. Second, the degree of discriminant validity was also pronounced and relatively unrelated to the chosen model. Although there were some fluctuations in indices

Table 7  
Indices of Difference Between Nested Covariance Structure Models  
Based on Kothandapani (1971) Multitrait-Multimethod Data

Matrix and Comparison	Difference in		Prob	Difference in	
	Chi-square	df		rho	delta
<b>12-Variable Matrix</b>					
Tests of added components					
Model 1A vs Model 3A	418.29	15	.000	.431	.532
Model 3A vs Model 3B'	293.29	12	.000	.487	.374
Model 3B' vs Model 3C	21.25	6	.002	.027	.027
Tests of convergent validity					
Model 3A vs Model 1A	418.29	15	.000	-.431	-.532
Model 3B' vs Model 1B'	426.61	15	.000	-.677	-.543
Model 3C vs Model 1C	363.49	15	.000	-.649	-.463
Tests of discriminant validity					
Model 3A vs Model 2A	122.62	3	.000	-.171	-.156
Model 3B' vs Model 2B'	199.19	3	.000	-.423	-.254
Model 3C vs Model 2C	126.85	3	.000	-.311	-.162
<b>9-Variable Matrix</b>					
Tests of added components					
Model 1A vs Model 3A	301.84	12	.000	.427	.598
Model 3A vs Model 3B'	173.21	9	.000	.497	.343
Model 3B' vs Model 3C	5.56	3	.135	-.002	.011
Tests of convergent validity					
Model 3A vs Model 1A	301.84	12	.000	-.427	-.598
Model 3B' vs Model 1B'	297.16	12	.000	-.777	-.589
Model 3C vs Model 1C	254.83	12	.000	-.738	-.505
Tests of discriminant validity					
Model 3A vs Model 2A	116.63	3	.000	-.259	-.231
Model 3B' vs Model 2B'	130.67	3	.000	-.532	-.259
Model 3C vs Model 2C	92.11	3	.000	-.441	-.182

Note. Sample size = 100.

of fit across models, the proportion of covariation represented uniquely as discriminant validation was always highly significant statistically and important practically. Third, the covariation explained uniquely by method factors was quite large, resulting in large differences in both rho and delta between Model 3C and Model 3A.

These trends may also be observed when considering the estimates and associated standard errors of parameters of Model 3C, presented in Table 8. The unique importance of both trait and method factors is shown by the fact that all 12 of the trait factor loadings and 11 of the 12 method factor loadings differed significantly from zero, a notable contrast to results from analyses of the Ostrom

(1969) data. The level of significance of the factor loadings is revealed by noting that 20 of the 23 significant loadings differ from zero by at least five standard errors.

The discriminant validity of the trait factors is shown in the correlations among the trait factors, only one of which differed appreciably and significantly from zero. The degree of discriminant validity between the two most highly correlated traits, the affective and cognitive aspects of attitude, is quite acceptable, as the estimated correlation between the two factors is approximately five standard errors from unity. Finally, the method factors appear to have relatively independent effects; this was presaged by the rather small, though

Table 8  
Factor Pattern and Factor Intercorrelation Matrices from Model 3C  
Estimated from Kothandapani (1971) 12-Variable Multitrait-Multimethod Data

Variable	Trait				Method				Unique
	1	2	3	I	II	III	IV		
	Primary Factor Pattern								
Thurstone									
Affect	.71(.09)	.0*	.0*	.17(.09)	.0*	.0*	.0*	.0*	.70(.06)
Behavior	.0*	.65(.09)	.0*	.47(.11)	.0*	.0*	.0*	.0*	.57(.08)
Cognition	.0*	.0*	.71(.11)	.75(.15)	.0*	.0*	.0*	.0*	.00(1.0)
Likert									
Affect	.91(.08)	.0*	.0*	.0*	.48(.08)	.0*	.0*	.0*	.18(.14)
Behavior	.0*	.69(.08)	.0*	.0*	.73(.09)	.0*	.0*	.0*	.16(.11)
Cognition	.0*	.0*	.61(.09)	.0*	.81(.09)	.0*	.0*	.0*	.13(.20)
Guttman									
Affect	.65(.09)	.0*	.0*	.0*	.0*	.56(.10)	.0*	.0*	.50(.08)
Behavior	.0*	.58(.11)	.0*	.0*	.0*	.55(.11)	.0*	.0*	.67(.07)
Cognition	.0*	.0*	.74(.09)	.0*	.0*	.38(.11)	.0*	.0*	.52(.08)
Self-rating									
Affect	.53(.08)	.0*	.0*	.0*	.0*	.0*	.80(.08)	.0*	.38(.08)
Behavior	.0*	.43(.09)	.0*	.0*	.0*	.0*	.76(.09)	.0*	.54(.06)
Cognition	.0*	.0*	.53(.08)	.0*	.0*	.0*	.68(.09)	.0*	.49(.05)
	Factor Intercorrelations								
Trait 1	1.0*								
Trait 2	.61(.08)	1.0*							
Trait 3	.04(.14)	.24(.15)	1.0*						
Method I	.0*	.0*	.0*	1.0*					
Method II	.0*	.0*	.0*	.17(.16)	1.0*				
Method III	.0*	.0*	.0*	-.23(.22)	.17(.17)	1.0*			
Method IV	.0*	.0*	.0*	.23(.14)	.51(.10)	.33(.14)	1.0*		

Note. All asterisked parameters were fixed at the reported values. All parameter estimates differing significantly from zero are underscored, and parenthesized values are standard errors of associated parameters.

important, increases in practical fit accompanying the estimation of intercorrelations of method factors (see the difference between Models 3B' and 3C in Table 7). Of the six correlations among the four method factors, only two differed significantly from zero, and both of these involved the self-rating method factor (IV).

In summary, Model 3C appears to be a quite adequate representation for the Kothandapani (1971) 12-variable matrix. The measures of fit and the parameter estimates each reflect the rather strong influence of both trait and method factors, each of which appears to have relatively independent effects.

*Structural modeling of the 9-variable matrix.* As noted above, Bagozzi (1978) concentrated his modeling attempts on the 9-variable matrix that excluded the supposedly highly errorful self-rating measures. The fit of the array of structural models to the 9-variable matrix are given in the lower half of Table 6, and indices of difference in fit from nested comparisons among these models are presented in the lower half of Table 7. These results of overall fit are provided for completeness and to allow comparisons of values with those reported by Bagozzi (1978). The results for the 9-variable matrix are, however, not discussed in detail; because the outcomes of significance tests and estimates of parameters were very similar to those reported for the 12-variable analyses, in-depth presentation of results would be redundant.

The major way in which results for the 9-variable matrix differed from those for the 12-variable matrix was the specification of method factor intercorrelations, which were neither statistically significant nor practically important for the 9-variable matrix. Thus, Model 3B' appears to be preferable to Model 3C for the latter matrix. But, as with the 12-variable analyses, results for the 9-variable matrix showed that the degree of both trait-related (i.e., convergent) and discriminant validation were highly significant on both statistical and practical grounds and were relatively independent of whether Model 3A, 3B', or 3C was chosen to represent the data. In addition, method factors explained statistically and practically important portions of covariation among observed variables.

#### Lawler Data

Results of structural modeling of the Lawler (1967) data are briefly presented and are given primarily to illustrate an earlier assertion. This assertion held that a less restricted model may be preferred over a more restricted model, regardless of differences in level of statistical and practical fit, if parameter estimates common to the two models differ greatly across models, as this would often imply that parameters were estimated in less biased fashion in the less restricted model.

The fit of Model 3A, with correlated trait factors only, to the Lawler data was not acceptable on either statistical,  $\chi^2(24) = 100.94$ ,  $p < .0001$ , or practical grounds,  $\rho = .630$  and  $\delta = .710$ . Adding orthogonal method factors to Model 3A resulted in Model 3B', which was not rejectable statistically,  $\chi^2(15) = 20.01$ ,  $p = .172$ , and had rather adequate levels of practical fit,  $\rho = .961$  and  $\delta = .943$ . The estimates of parameters in Model 3B' are presented in the left half of Table 9. These estimates for Model 3B' imply (1) that trait factors represent fairly important sources of covariation, especially for supervisor and peer ratings, suggesting substantial agreement across raters in individual differences among managers on the three traits; (2) that, except for one loading on the self-rating method factor, rather modest amounts of covariation are associated with method factors; and (3) that the three trait factors are approximately equally and moderately highly correlated, but correlated at a level which suggests that the three trait factors represent easily differentiable constructs.

Although the fit of Model 3B' to the Lawler (1967) data was adequate statistically and practically, the improvements in fit to the Ostrom (1969) and Kothandapani (1971) data associated with Model 3C led to consideration of the fit of Model 3C to the Lawler data. Model 3C, with correlated trait and correlated method factors, was quite nonrejectable,  $\chi^2(12) = 4.67$ ,  $p = .968$ , and attained very high levels of practical fit,  $\rho = 1.070$  and  $\delta = .987$ . Despite the improvements in indices of fit, there is an indication that Model 3C may not be a truly adequate final model for the Lawler data. Specifically, the rather large loading of the super-

Table 9  
 Factor Pattern and Factor Intercorrelation Matrices from Models 3B' and 3C  
 Estimated from Lawler (1967) 9-Variable Multitrait-Multimethod Data

Variable	Factors in Model 3B'									Factors in Model 3C												
	Trait			Method			Trait			Method			Trait			Method						
	1	2	3	I	II	III	Unique	1	2	3	I	II	III	Unique	1	2	3	I	II	III	Unique	
Primary Factor Pattern																						
Supervisor																						
Quality	.87	.0*	.0*	.27	.0*	.0*	.40	.24	.0*	.0*	.84	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.49	
Ability	.0*	.76	.0*	.29	.0*	.0*	.58	.0*	.77	.0*	.63	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.00	
Effort	.0*	.0*	.78	.38	.0*	.0*	.50	.0*	.0*	.44	.67	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.59	
Peer																						
Quality	.73	.0*	.0*	.0*	.58	.0*	.35	.49	.0*	.0*	.0*	.0*	.0*	.0*	.87	.0*	.0*	.0*	.0*	.0*	.00	
Ability	.0*	.67	.0*	.0*	.37	.0*	.64	.0*	.28	.0*	.0*	.0*	.0*	.0*	.66	.0*	.0*	.0*	.0*	.0*	.70	
Effort	.0*	.0*	.69	.0*	.38	.0*	.61	.0*	.0*	.48	.0*	.0*	.0*	.0*	.66	.0*	.0*	.0*	.0*	.0*	.59	
Self																						
Quality	.10	.0*	.0*	.0*	.0*	1.00	.00	-.28	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.0*	.00	
Ability	.0*	.13	.0*	.0*	.0*	.42	.90	.0*	.17	.0*	.0*	.0*	.0*	.0*	.46	.0*	.0*	.0*	.0*	.0*	.88	
Effort	.0*	.0*	.35	.0*	.0*	.38	.85	.0*	.0*	.57	.0*	.0*	.0*	.0*	.41	.0*	.0*	.0*	.0*	.0*	.71	

Factor Intercorrelations

Trait 1	1.0*																				
Trait 2	.68	1.0*																			
Trait 3	.65	.53	1.0*																		
Method I	.0*	.0*	.0*	1.0*																	
Method II	.0*	.0*	.0*	.0*	1.0*																
Method III	.0*	.0*	.0*	.0*	.0*	1.0*															

Note. All asterisked parameters were fixed at the reported values. All parameter estimates differing significantly from zero are underscored.

visor rating of ability on the Ability trait factor does not differ significantly from zero, due to its very large standard error, and this suggests empirical identification problems attendant with Model 3C. But, because the intent of the present paper is to discuss the general trend of differences in parameter estimates common to Models 3B' and 3C, no further model fitting was pursued to improve the empirical identification of Model 3C.

Although Model 3C provided rather better fit to the data than did Model 3B', of greater interest here are the great changes in estimates of parameters common to Models 3B' and 3C. The estimates of parameters in Model 3C are presented in the right half of Table 9, and the interpretations of patterns inherent in these estimates differ greatly from those in Model 3B'. First, trait factors no longer appear to represent highly important sources of covariation; the Quality and Ability trait factors have no significant loadings, and, though all three of the loadings on the Effort factor are significant, the Effort factor loadings are only moderate in magnitude. Second, method factors represent rather large determiners of covariation, since all nine loadings were moderate to large and all were statistically significant. Third, the patterns of correlation among trait and among method factors reveal interesting trends. The correlations among trait factors fell from substantial values in Model 3B' to essentially zero in Model 3C; and, when method factors were allowed to correlate in Model 3C, only the Supervisor and Peer factors were highly related, with the Self factor virtually orthogonal to the preceding two.

A reasonable interpretation of the Model 3C estimates is the following: Neither supervisors, peers, nor the managers themselves could easily distinguish among the three traits to be rated; given the rating a manager received on one of the three trait rating scales by a particular rater, there was a high likelihood that the manager would receive similar scores from that rater on the remaining two scales. This led to the prominence of the three method factors, which in this case represent halo effects associated with each type of rater. It is also clear that others' ratings of managers were rather similar and that others' and self ratings were rather dis-

cordant. Thus, the Supervisor and Peer method factors were highly correlated (.73), but neither of these factors correlated significantly with the Self factor. Finally, to the extent that all three types of raters could agree on differentiable trait constructs, this happened only with regard to the effort managers exhibited on the job. But even here, the method (or halo) factors tended to explain rather more variance in effort ratings than did the Effort trait factor.

#### Discussion

The aim of the present paper was to develop a general approach for specifying structural models for multitrait-multimethod data. The result is a two-part procedure. The first part involves specification of a series of hierarchically nested structural models. The three approaches described for specifying such series of models were: (1) beginning with the simplest reasonable substantive model and increasing the complexity of the model as required by the data; (2) beginning with method factors and determining whether trait factors explain important proportions of covariation beyond that explained by method factors; and (3) fitting each model in the taxonomy of potential models and comparing the fit of a wide variety of structural models. Of the approaches presented, the one preferred on grounds of parsimony was that involving specification of models of increasing complexity, models which are formed by adding successive sets of parameters and which are therefore hierarchically nested. The latter series of models was designed to enable the determination, by comparing the fit of the several models, of which sets of parameters to include in the final model. Applying the series of models to the 9-variable Kothandapani (1971) matrix, for example, led to the finding that allowing correlations among method factors failed to increase levels of either statistical or practical fit. Thus, Model 3B', as compared with Model 3C, provides a more parsimonious and equally well-fitting representation of relations among measures in the 9-variable Kothandapani matrix.

The second aspect of the general approach is predicated on selection of a single model as the most acceptable or best-fitting structural model for a set of data. Then, by comparing the fit of the

best-fitting model with that of appropriately specified alternative models, it is possible to obtain estimates and tests of the degree of convergent validation, discriminant validation, and method variance exhibited by the set of measures.

The major contribution of the present paper is the formulation of model comparisons that provide estimates of convergent and discriminant validation and the amount of method variance. Many authors have previously fit many of the models used in the present study to various sets of multitrait-multimethod data. Indeed, Schmitt (1978) presented the results of fitting most of the models proposed in the present paper, as well as several additional models, to the 9-variable Ostrom (1969) and Kothandapani (1971) matrices, though Schmitt did report some incorrect chi-square statistics for these models (cf. values in Table 6 of Schmitt, 1978, against values in Tables 3 and 6 and the text in Bagozzi, 1978, and Tables 2 and 6 of the present paper). However, Schmitt (1978) did not formulate nested comparisons among models to test for convergent and discriminant validity. As a result, after reading Schmitt (1978), a researcher is left with an array of models with varying levels of fit and little guidance on how to decide which model is most appropriate. In the present paper, appropriate nested comparisons among models were performed, and these comparisons were evaluated with respect to both statistical significance and practical importance. Once this was done, the decision as to the appropriate structural representation for each matrix and the estimation of the degree of convergent validation, discriminant validation, and method variance were easily made.

Application of the proposed general approach to the Ostrom (1969) and Kothandapani (1971) matrices led to conclusions directly contrary to those stated by Bagozzi (1978). Bagozzi, apparently on the basis of overall tests of fit and the magnitude of the trait and method factor loadings, claimed that both convergent and discriminant validity were exhibited by the Ostrom data and that the affective, behavioral, and cognitive components of attitude each explained unique portions of variance. Bagozzi, however, had not performed nested model comparisons that would justify statements regard-

ing the degree of convergent and discriminant validity. When such tests were performed in the present paper, the results showed that very small proportions of variance were uniquely attributed to convergent validation and that practically nonsignificant levels of discriminant validation were evidenced in both the 12- and 9-variable Ostrom matrices, directly contradicting Bagozzi's claims.

Application of the procedures proposed in the present paper also belie Bagozzi's (1978) evaluation of the patterns residing in the Kothandapani (1971) data. Bagozzi (1978, pp. 27-28) claimed that in the Kothandapani data the level of convergent validation was weak and that discriminant validity had not been achieved, that is, each of the three components of attitude failed to explain unique variation unexplained by the remaining two. When appropriate nested model comparisons are made, however, the picture is much different. Both convergent and discriminant validation are much in evidence in both the 12- and 9-variable Kothandapani matrices; either dropping trait factors altogether or forcing estimation of a single trait factor led to precipitous declines in measures of statistical and practical fit.

The many previous articles on methods of analyzing multitrait-multimethod data point to one unmistakable fact: representing the trends present in such data is a complicated undertaking. The general approach to modeling multitrait-multimethod data proposed in the present paper—employing hierarchically nested series of structural models, nested comparisons for degree of convergent and discriminant validity and method variance, and attention to the significance and stability of individual parameters—is an attempt to make the modeling of such data a simpler and more objective enterprise.

### References

- Althausen, R. P. (1974). Inferring validity from the multitrait-multimethod matrix: Another assessment. In H. L. Costner (Ed.), *Sociological methodology 1973-1974* (pp. 106-127). San Francisco: Jossey-Bass.
- Althausen, R. P., & Heberlein, T. A. (1970). Validity and the multitrait-multimethod matrix. In E. F. Borgatta & G. W. Bohrnstedt (Eds.), *Sociological methodology 1970* (pp. 151-169). San Francisco: Jossey-Bass.

- Althausen, R. P., Heberlein, T. A., & Scott, R. A. (1971). A causal assessment of validity: The augmented multitrait-multimethod matrix. In H. M. Blalock, Jr. (Ed.), *Causal models in the social sciences* (pp. 374–399). Chicago: Aldine.
- Alwin, D. F. (1974). Approaches to the interpretation of relationships in the multitrait-multimethod matrix. In H. L. Costner (Ed.), *Sociological methodology 1973–1974* (pp. 79–105). San Francisco: Jossey-Bass.
- Bagozzi, R. P. (1978). The construct validity of the affective, behavioral, and cognitive components of attitude by analysis of covariance structures. *Multivariate Behavioral Research*, 13, 9–31.
- Bagozzi, R. P. (1980). *Causal models in marketing*. New York: Wiley.
- Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology*, 31, 419–456.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, 88, 588–606.
- Bentler, P. M., & Lee, S. Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, 32, 87–104.
- Boruch, R. F., & Wolins, L. (1970). A procedure for estimation of trait, method, and error variance attributable to a measure. *Educational and Psychological Measurement*, 30, 547–574.
- Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology*, 37, 1–21.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81–105.
- Cliff, N. (1983). Some cautions concerning the application of causal modeling methods. *Multivariate Behavioral Research*, 18, 115–126.
- Conger, A. J. (1971). Evaluation of multimethod factor analysis. *Psychological Bulletin*, 75, 416–420.
- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.
- Duncan, O. D. (1975). *Introduction to structural equation models*. New York: Academic Press.
- Golding, S. L. (1977). Method variance, inadequate constructs, or things that go bump in the night? *Multivariate Behavioral Research*, 12, 89–98.
- Golding, S. L., & Seidman, E. (1974). Analyses of multitrait-multimethod matrices: A two step principal components procedure. *Multivariate Behavioral Research*, 9, 479–496.
- Hubert, L. J., & Baker, F. B. (1978). Analyzing the multitrait-multimethod matrix. *Multivariate Behavioral Research*, 13, 163–179.
- Hubert, L. J., & Baker, F. B. (1979). A note on analyzing the multitrait-multimethod matrix: An application of a generalized proximity function comparison. *British Journal of Mathematical and Statistical Psychology*, 32, 179–184.
- Jackson, D. N. (1969). Multimethod factor analysis in the evaluation of convergent and discriminant validity. *Psychological Bulletin*, 72, 30–49.
- Jackson, D. N. (1971). Comments on “Evaluation of multimethod factor analysis.” *Psychological Bulletin*, 75, 421–423.
- Jackson, D. N. (1975). Multimethod factor analysis: A reformulation. *Multivariate Behavioral Research*, 10, 259–275.
- Jackson, D. N. (1977). Distinguishing trait and method variance in multitrait-multimethod matrices: A reply to Golding. *Multivariate Behavioral Research*, 12, 99–110.
- Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 34, 183–202.
- Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, 36, 109–133.
- Jöreskog, K. G. (1974). Analyzing psychological data by structural analysis of covariance matrices. In R. C. Atkinson, D. H. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 1–56). San Francisco: W. H. Freeman.
- Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relations by maximum likelihood and least squares methods (user's guide)*. Chicago: International Educational Resources.
- Kalleberg, A. L., & Kluegel, J. R. (1975). Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. *Journal of Applied Psychology*, 60, 1–9.
- Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, 12, 247–252.
- Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.
- Kothandapani, V. (1971). Validation of feeling, belief, and intention to act as three components of attitude and their contributions to prediction of contraceptive behavior. *Journal of Personality and Social Psychology*, 19, 321–333.
- Lawler, E. E., III. (1967). The multitrait-multirater approach to measuring managerial job performance. *Journal of Applied Psychology*, 51, 369–381.
- Lee, S. Y. (1980). Estimation of covariance structure models with parameters subject to functional restraints. *Psychometrika*, 45, 309–324.
- Marsh, H. W., & Hocevar, D. (1983). Confirmatory

- factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement*, 20, 231–248.
- Ostrom, T. M. (1969). The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, 5, 12–30.
- Rindskopf, D. (1983). Parameterizing inequality constraints on unique variances in linear structural models. *Psychometrika*, 48, 73–83.
- Schmitt, N. (1978). Path analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, 2, 157–173.
- Schmitt, N., Coyle, B. W., & Saari, B. B. (1977). A review and critique of analyses of multitrait-multimethod matrices. *Multivariate Behavioral Research*, 12, 447–478.
- Schmitt, N., & Saari, B. B. (1978). Behavior, situation, and rater variance in descriptions of leader behaviors. *Multivariate Behavioral Research*, 13, 483–495.
- Schwarzer, R. (1982). *Convergent and discriminant validation in personality assessment: Techniques for distinguishing trait and method variance*. Paper presented at the meeting of the American Educational Research Association, New York.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38, 1–10.
- Werts, C. E., & Linn, R. L. (1970). Path analysis: Psychological examples. *Psychological Bulletin*, 74, 193–212.
- Wilkinson, L. (1979). Tests of significance in stepwise regression. *Psychological Bulletin*, 86, 168–174.

#### Acknowledgment

The present research was supported, in part, by National Institute for Child Health and Human Development grants HD-14688 and HD-04612, and by grants from the Academic Senate and from the Academic Computing Center at the University of California at Riverside. The author thanks Robert MacCallum, Jack McArdle, and two anonymous reviewers for valuable comments and creative suggestions offered on a previous version of this paper.

#### Author's Address

Send requests for reprints or further information to Keith F. Widaman, Department of Psychology, University of California, Riverside CA 92521, U.S.A.