Confirmatory Factor Analyses of Multitrait-Multimethod Data: Many Problems and a Few Solutions

Herbert W. Marsh
University of Sydney

During the last 15 years there has been a steady increase in the popularity and sophistication of the confirmatory factor analysis (CFA) approach to multitrait-multimethod (MTMM) data. This approach, however, incurs some important problems, the most serious being the ill-defined solutions that plague MTMM studies and the assumption that so-called method factors reflect primarily the influence of method effects. In three different MTMM studies, ill-defined solutions were frequent and alternative parameterizations designed to solve this problem tended to mask the symptoms instead of eliminating the problem. More importantly, so-called method factors apparently represented trait variance in addition to, or instead of, method variance for at least some models in all three studies. Further support for this counterinterpretation of method factors was found when external validity criteria were added to the MTMM models and correlated with trait and so-called method factors. This problem, when it exists, invalidates the traditional interpretation of trait and method factors and the comparison of different MTMM models. A new specification of method effects as correlated uniquenesses instead of method factors was less prone to ill-defined solutions and, apparently, to the confounding of trait and method effects. *Index terms: confirmatory factor analysis, construct validity, convergent validity, correlated uniquenesses, discriminant validity, empirical underidentification, LISREL, method effects, multitrait-multimethod analysis.*

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 13, No. 4, December 1989, pp. 335–361

0146-6216/89/040335-27$2.60

The purpose of this investigation was to demonstrate and critically evaluate recently developed applications of confirmatory factor analysis (CFA) to multitrait-multimethod (MTMM) data. Campbell and Fiske (1959) argued that construct validation requires multiple indicators of the same construct to be substantially correlated with each other but substantially less correlated with indicators of different constructs. They proposed the MTMM design, in which each of a set of multiple traits is assessed with each of a set of multiple methods of assessment, and developed guidelines for evaluating MTMM data.

Campbell and Fiske's MTMM paradigm has become perhaps the most frequently employed construct validation design, and their original guidelines continue to be the most frequently used guidelines for examining MTMM data. However, important problems with their guidelines are well known (e.g., Althauser & Heberlein, 1970; Alwin, 1974; Campbell & O'Connell, 1967; Marsh, in press; Wothke, 1984) and have led to many alternative analytic approaches (e.g., Browne, 1984; Hubert & Baker, 1978; Jackson, 1969, 1975; Marsh, in press; Marsh & Hocevar, 1983; Schmitt, Coyle, & Saari, 1977; Schmitt & Stults, 1986; Stanley, 1961; Wothke, 1984). Factor-analytic approaches (e.g., Boruch & Wolins, 1970; Jöreskog, 1974; Marsh, in press; Marsh & Hocevar, 1983; Widaman, 1985) or mathematically similar path-analytic approaches (e.g., Schmitt et al., 1977; Werts &

Linn, 1970) currently appear to be the most popular approach and are the focus of the present investigation.

A GENERAL MTMM MODEL AND A TAXONOMY OF ALTERNATIVE MODELS

The General MTMM Model

In the CFA approach to MTMM data, a priori factors defined by different measures of the same trait support the construct validity of the measures, but a priori factors defined by different traits measured with the same method imply method effects. The present investigation starts from a general MTMM model (Table 1) adapted from Jöreskog (1974; see also Marsh & Hocevar, 1983; Widaman, 1985) in which:

1. There are at least $T = 3$ traits (t_1, t_2, and t_3) and $M = 3$ methods (m_1, m_2, and m_3).
2. $T \times M$ measured variables (t_1m_1, t_1m_2, ..., t_3m_3) are used to infer $T + M$ a priori common factors.
3. Each measured variable loads on the one trait factor and the one method factor that it represents, but it is constrained so as not to load on any other factors.
4. Correlations among trait factors and among method factors are freely estimated, but correlations between trait and method factors are constrained to be 0.

For this model it is assumed that there are at least three traits and three methods, but alternatives have been proposed for studies with only two methods (Kenny, 1979) or only two traits (Marsh & Hocevar, 1983). Although some researchers have estimated correlations between trait and method factors, there are important logical, interpretive, and pragmatic reasons for fixing these correlations to be 0 (see Jackson, 1975; Marsh & Hocevar, 1983; Widaman, 1985). This constraint allows the decomposition of variance into additive trait, method, and error components, and without this constraint the solution is almost always empirically underidentified (also see Widaman, 1985; Wothke, 1984). In justifying this constraint, Jöreskog (1971) noted:

> This is our way of defining each method factor to be independent of the particular traits that the method is used to measure. In other words, method factors are what is left over after all trait factors have been eliminated. (p. 128)

In the present investigation, CFA models were fit with LISREL V (Jöreskog & Sörbom, 1981) and three design matrices from LISREL were used to define all the MTMM models. For $T = 3$ traits and $M = 3$ methods (see Table 1), the three design matrices are:

1. $\boldsymbol{\lambda}_Y$, a 9 ($M \times T$ = number of measured variables) $\times$ 6 ($M + T$ = number of factors) matrix of factor loadings;
2. $\boldsymbol{\psi}$, a 6 ($M + T$ = number of factors) $\times$ 6 factor variance-covariance matrix of relations among the factors; and
3. $\boldsymbol{\theta}$, a 9 ($M \times T$ = number of measured variables) $\times$ 9 matrix of uniquenesses in which the diagonal values are analogous to one minus the communality estimates in exploratory factor analyses.

All parameters with values of 0 or 1 are fixed and values for other parameters are estimated so as to maximize goodness of fit. Standard errors are estimated for all estimated parameters but not for parameters with fixed values. This model is easily modified to accommodate more traits or methods, to conform to other models and other parameterizations that will be described, or to incorporate unique factors for the measured variables (Rindskopf, 1983).

A Taxonomy of Alternative Models

Researchers have proposed many variations of the general MTMM model to examine inferences about trait or method variance or to test substantive issues specific to a particular study (e.g., Bagozzi, 1978; Jöreskog, 1974; Marsh, Barnes, & Hocevar, 1985; Marsh & Hocevar, 1983; Widaman, 1985; Wothke, 1984). Widaman (1985) proposed an important taxonomy of such models that systematically varied different characteristics of trait and method factors. This taxonomy was designed to be appropriate for all MTMM studies, to provide a gen-

Table 1
Design Matrices: Parameters to be Estimated for the General MTMM Model (Model 4D)

	Factor Loadings (λ_y)						Uniquenesses (θ)								
Variables	t_1	t_2	t_3	m_1	m_2	m_3	t_1m_1	t_2m_1	t_3m_1	t_1m_2	t_2m_2	t_3m_2	t_1m_3	t_2m_3	t_3m_3
t_1m_1	λ_y^a	0	0	λ_y	0	0	θ								
t_2m_1	0	λ_y	0	λ_y^a	0	0	0^b	θ							
t_3m_1	0	0	λ_y	λ_y	0	0	0^b	0^b	θ						
t_1m_2	λ_y	0	0	0	λ_y	0	0	0	0	θ					
t_2m_2	0	λ_y^a	0	0	λ_y	0	0	0	0	0^b	θ				
t_3m_2	0	0	λ_y	0	λ_y^a	0	0	0	0	0^b	0^b	θ			
t_1m_3	λ_y	0	0	0	0	λ_y^a	0	0	0	0	0	0	θ		
t_2m_3	0	λ_y	0	0	0	λ_y	0	0	0	0	0	0	0^b	θ	
t_3m_3	0	0	λ_y^a	0	0	λ_y	0	0	0	0	0	0	0^b	0^b	θ

Factor Variances and Covariances (ψ)

Factors	t_1	t_2	t_3	m_1	m_2	m_3
t_1	1^a					
t_2	ψ 1^a					
t_3	ψ ψ	1^a				
m_1	0 0	0	1^a			
m_2	0 0	0	ψ	1^a		
m_3	0 0	0	ψ	ψ	1^a	

Note. All parameters with values of 0 or 1 were fixed and not estimated, whereas all other parameters were estimated without constraint. This parameterization, with factor variances (in ψ) fixed to be 1, was the fixed factor *variance* parameterization.
[a]For the fixed factor *loading* parameterization these factor loadings would be fixed to be 1 and these factor variances would be freely estimated.
[b]For method structure E (see Table 2) these correlations between uniquenesses would be estimated, whereas the method factors and their associated parameters would be eliminated from λ_y and ψ.

eral framework for making inferences about the effects of trait and method factors, and to objectify the complicated task of formulating models and representing the MTMM data.

One purpose of the present investigation was to evaluate the taxonomy in relation to these goals and to describe an expansion of the taxonomy formulated for the present investigation. The expanded taxonomy shown in Table 2 represents all possible combinations of four trait structures (trait structures 1 through 4) and five method structures (method structures A through E). The four trait structures posit no trait factors (structure 1), one general trait factor defined by all measured variables (structure 2), T uncorrelated trait factors (structure 3), and T correlated trait factors (structure 4). The five method structures posit no method factors (structure A), one general method factor defined by all measured variables (structure B), M uncorrelated method variables (structure C), M correlated method factors (structure D), and method effects inferred on the basis of correlated uniquenesses (structure E). This taxonomy differs from Widaman's original taxonomy only in the addition of method structure E.

Models under method structure E have no method factors. Instead, method effects are inferred on the basis of correlated uniquenesses (see Table 1). This method structure, particularly when there are three traits, corresponds most closely to method structure C in which there are M uncorrelated methods. In addition to the structures shown in Table 2, an

Table 2
Taxonomy of Structural Models for MTMM Data Adapted From Widaman (1985)

Trait Structure	Method Structure				
	A	B	C	D	E
1	1A: Null Model	1B: 1 general M-factor	1C: M uncorrelated M-factors	1D: M correlated M-factors	1E: T × M correlated errors
2	2A: 1 general T-factor	2B: 2 general factors	2C: 1 general T-factor, M uncorrelated M-factors	2D: 1 general T-factor, M correlated M-factors	2E: T × M correlated errors, 1 general T-factor
3	3A: T uncorrelated T-factors	3B: 1 general M-factor, T uncorrelated T-factors	3C: T uncorrelated T-factors, M uncorrelated M-factors	3D: T uncorrelated T-factors, M correlated M-factors	3E: T × M correlated errors, T uncorrelated T-factors
4	4A: T correlated T-factors	4B: 1 general M-factor, T correlated T-factors	4C: T correlated T-factors, M uncorrelated M-factors	4D: T correlated T-factors, M correlated M-factors	4E: T × M correlated errors, T correlated T-factors

additional variation of Model 4E (in which the validity factor was correlated with trait factors but not uniquenesses) was implemented. This model, Model 4E′, posited correlations with uniquenesses.

In structure 1, general factors are defined to be uncorrelated with other factors in the model. Although general factors are posited to represent either trait variance or method variance, this assumption may not be accurate and may be difficult to test. In this taxonomy Models 2A and 1B are equivalent, and it is generally not possible to determine whether the one general factor reflects trait variance, method variance, or some combination of trait and method variance. Model 2B requires additional constraints that may be arbitrary and may not provide equivalent solutions. Hence its usefulness may be dubious unless there is an a priori basis for the constraints.

The general factors posited in method structure B and trait structure 2 may present interpretive or estimation problems. Widaman (1985) avoided some problems by constraining each general factor to be uncorrelated with all other factors, and this constraint is used here. The rationale for this constraint is consistent with the requirement that trait and method factors be uncorrelated. Models 1B and 2A, however, are the same, whereas Model 2B requires one additional, perhaps arbitrary, zero constraint to assure rotational identification. Finally, even for models that contain a general method factor in combination with *T* trait factors, or a general trait factor in combination with *M* method factors, the interpretation of the general factor may be problematic.

POTENTIAL PROBLEMS IN THE ESTIMATION AND INTERPRETATION OF MTMM MODELS

Goodness of Fit

An important unresolved problem in CFA is the assessment of goodness of fit. To the extent that a hypothesized model is identified and is able to fit the observed data, there is support for the model. The problem of goodness of fit is how to decide whether the predicted and observed results are sufficiently alike to warrant support of a model. Although χ^2 values can be used to test whether these differences are statistically significant, there is a growing recognition of the inappropriateness of this classical hypothesis-testing approach. Because hypothesized models are only designed to approximate reality, all such restrictive models are a priori false and will be shown to be false with a sufficiently large sample size (Cudeck & Browne, 1983; Marsh, Balla, & McDonald, 1988; McDonald, 1985). Hence a variety of fit indices have been derived to aid in this decision process. These include the ratio of χ^2 to degrees of freedom and the Tucker-Lewis index (TLI; Tucker & Lewis, 1973), both used here. In simulation studies of more than 30 such indices, Marsh et al. (1988) and Marsh, McDonald, and Balla (1989) found that both the χ^2/df and TLI indices imposed apparently appropriate penalty functions for the inclusion of additional parameters that controlled for capitalizing on chance, whereas TLI was the only widely used index that was also relatively independent of sample size. TLI is emphasized below, but values for other fit indices such as the Bentler and Bonett (1980) index can easily be computed from the results.

Model selection must be based on subjective evaluation of substantive issues, inspection of parameter values, model parsimony, and a comparison of the performances of competing models, as well as goodness of fit. In the application of CFA to MTMM data there is an unfortunate tendency to underemphasize the examination of parameter estimates and to overemphasize goodness of fit. If a solution is ill-defined, then further interpretations must be made very cautiously if at all. If the parameter estimates for a model make no sense in relation to the substantive a priori model, then fit may be irrelevant.

As described by Bentler and Bonett (1980), when two models are nested, the statistical significance of the difference in the χ^2s can be tested relative to the difference in their *df*. Widaman (1985) emphasized this feature in developing his taxonomy of MTMM models and in comparing the fit of different models. However, the problems associated with the application of the classical hypothesis-testing approach also apply to this test of χ^2 differences. When sample size is sufficiently large, the saturated model (i.e., a model with $df = 0$) will perform significantly better than any restricted model (see Cudeck & Browne, 1983) such as those in Table 2, thus making problematic the interpretation of tests between any two restricted models. Furthermore, many important comparisons are not nested and hence cannot be made with this procedure [e.g., the trait-only (4A) and method-only (1D) models in Table 2]. Because of these problems with the χ^2 difference test, a perhaps more useful test is to simply compare the TLIs for competing models.

Poorly Defined Solutions

Another serious unresolved problem for CFA is that of poorly defined solutions. This problem is particularly prevalent in MTMM studies. Poorly defined solutions refer to underidentified or empirically underidentified models (Kenny, 1979; Wothke, 1984), failures in the convergence of the iterative procedure used to estimate parameters, parameter estimates that are outside their permissible range of values (e.g., negative variance estimates called Heywood cases), or standard errors of parameter estimates that are excessively large. Each of these problems is an indication that the empirical solution is poorly defined, even if the model is apparently identified otherwise and even if goodness of fit is adequate (Jöreskog & Sörbom, 1981).

Such problems are apparently more likely when sample size is small, when there are few indicators of each latent factor, when measured variables are allowed to load on more than one factor, when measured variables are highly correlated, when many data are missing and covariance matrices are estimated with pairwise deletion for missing data, and when the model is misspecified. Knowingly or unknowingly, such problems are usually ignored, and the implications of this practice have not been explored for MTMM studies. Although there is no generally appropriate resolution for such problems, alternative parameterizations of the MTMM model (see below) may eliminate some improper parameter estimates.

There are apparent ambiguities about the identification status of MTMM models. Some researchers (e.g., Alwin, 1974; Browne, 1984; Jöreskog, 1974; Schmitt, 1978) have suggested that models with correlations between traits and methods are permissible, and Long (1983, p. 55) claimed to prove the identification for this model. However, Bollen and Jöreskog (1985) demonstrated that the criteria used by Long were not sufficient to demonstrate identification, and Widaman (1985) explicitly eliminated such models from his taxonomy, claiming that they "are very likely not identified" (p. 7).

In order to test the identification status of a model with correlated traits and methods, D. A. Kenny (personal communication, January 23, 1987) used simulated data "to see if LISREL could recover loadings for your Model 4D with traits and methods correlated. It did so, but not exactly. It was not clear whether the difference was due to under-identification or rounding error."

An attempt was also made to fit Model 4D with correlations between method and trait factors to the simulated population covariance matrix published by Cole and Maxwell (1985), in which the population correlations between trait and method factors were simulated to be 0. Using their population covariance matrix, the population values proved recoverable, but more than 500 iterations were required. For their sample matrices that included random error, however, the solutions failed to converge after more than 1,000 iterations. It appears that although the model with correlated trait and method factors may technically be identified, it is unlikely to result in a proper solution for actual data and hence is of little practical use. Because models in Table 2 do not posit trait/method correlations and because all studies considered here have at least three traits and three methods, these ambiguities will not be examined here; however, they illustrate that the issue of identification has not been resolved.

Different Parameterizations: Potential Cures for Poorly Defined Solutions

The standard parameterizations. In order for the models in Table 2 to be identified, one parameter for each latent factor must be fixed—typically at a value of 1.0 (see Jöreskog & Sörbom, 1981; Long, 1983). This is usually done either by fixing the factor loading of one measured variable for each latent factor to be 1.0 and estimating the factor variance, or by fixing the factor variance of each latent factor to be 1.0 (so that the factor variance/covariance matrix is a correlation matrix) and estimating all the factor loadings.

These will be denoted the *fixed factor loading* and *fixed factor variance* parameterizations respectively, and collectively they will be referred to as the *standard parameterizations*. For well-defined CFA solutions both parameterizations are equivalent, but fixing the factor variances introduces an implicit inequality constraint that restricts the factor variances to be non-negative. Thus, fixing factor variance estimates may lead to a proper solution when fixing factor loadings does not.

The Rindskopf parameterization. Rindskopf (1983) proposed a solution for negative uniqueness estimates by using $M \times T$ additional factors—one unique factor for each of the $M \times T$ measured variables—to define each uniqueness. Because the factor loading on each unique factor is the square root of the uniqueness, the uniqueness is implicitly constrained to be non-negative.

Jöreskog (1981), commenting on the merits of imposing inequality constraints, noted that if a solution is inadmissible LISREL will find a solution outside the permissible parameter space, whereas the imposition of inequality constraints will produce a solution on the boundary of the parameter space. Jöreskog concluded: "In both cases the conclusion will be that the model is wrong or that the sample size is too small" (p. 91). Similarly, Dillon, Kumar, and Mulani (1987) noted that in their research the Rindskopf parameterization always resulted in the offending parameter estimate taking on a zero value that resulted in the same solution as simply fixing the parameter to be 0.

Method structure E—an alternative conceptualization of method variance. Method variance is an undesirable source of systematic variance that distorts correlations between different traits measured with the same method. As typically depicted in MTMM models (i.e., method structures C and D), a single method factor is used to represent the method effect associated with variables assessed by the same method. The effects of a particular method of assessment are implicitly assumed to be unidimensional and the sizes of the method factor loadings provide an estimate of its influence on each measured variable. Hence, method structures C and D restrict method covariance components to have a congeneric-like structure (but see Wothke, 1984).

Alternatively, method effects can be represented as correlated uniquenesses (method structure E);

this representation assumes neither the unidimensionality of effects associated with a particular method nor a congeneric structure. Kenny (1979; see also Marsh, in press; Marsh & Hocevar, 1983) proposed this method structure for the special case in which there are only two traits, but it is also reasonable when there are more than two traits. Method structure E also resembles McDonald's (1985) multimode analysis and Browne's (1980) multiple battery analysis.

Method structure E corresponds most closely to method structure C (Table 2) in that the method effects associated with one method are assumed to be uncorrelated with those of other methods. When there are $T = 3$ traits and the solutions are well-defined, method structures C and E are merely alternative parameterizations of the same model. When $T > 3$, however, the number of correlated uniquenesses in method structure E ($M \times [T \times (T - 1)/2]$) is greater than the number of factor loadings used to define method factors in method structure C ($T \times M$). Thus method structure C is a special case of method structure E in which each method factor is required to be unidimensional, and this assumption is testable when $T > 3$.

A particularly important advantage of method structure E is that it apparently eliminates some improper solutions without limiting the solution space or forcing parameter estimates to the boundaries of the permissible space. Because method variance is one source of uniqueness, uniqueness is reflected in both method factors and uniquenesses. When improper solutions occur, they tend to be due to either negative method factor variances or negative uniquenesses, but not both. In method structure E all sources of uniqueness are contained in the diagonal of $\boldsymbol{\theta}$, and in many cases—as demonstrated in the present investigation—this combined influence will not be negative even when method factor variances or uniquenesses are negative for other parameterizations.

Thus even when there are three traits, so that method structures C and E are equivalent when the solutions are well defined, it is possible that method structure C will result in poorly defined solutions whereas method structure E will not. When there are more than three traits it is possible for method structure E to fit the data better than method structures C or D, thus calling into question the assumed unidimensionality of method effects in structures C and D.

Problems in the Interpretation of Trait and Method Factors

Widaman's taxonomy and the MTMM models in Table 2 implicitly assume that method factors represent method variance, trait factors represent trait variance, a general factor in combination with trait factors represents method variance, and a general factor in combination with method factors represents trait variance. For the present purposes these assumptions will be referred to as the traditional interpretation of the MTMM models.

These assumptions are probably reasonable when correlations among trait factors and among method factors are small. But in the more common situation when correlations among trait factors and among method factors are substantial, these assumptions may not be reasonable. Examined below is the possibility that so-called method factors actually reflect trait variance, but the problem might also occur with respect to so-called trait factors that actually reflect method variance.

In most MTMM studies the multiple traits are correlated; this may produce a general trait factor that makes ambiguous the interpretation of so-called general method factors or even correlated method factors. When traits are substantially correlated, the so-called general method factor (method structure B) may represent trait variance instead of, or in addition to, method variance. Also under these circumstances, each of the so-called correlated method factors (method structure D) may represent this general trait factor and correlations among method factors may represent the convergence of this general trait across the methods of assessment. If this problem exists, the traditional interpretation of MTMM models and the comparison of alternative models is unjustified. Hence, tests of this plausible counterinterpretation of method factors must be examined.

Results to be discussed here suggest that the traditional interpretation of method factors may be unjustified in the following cases:

1. Interpretations based on the Campbell-Fiske guidelines and an examination of the MTMM matrix differ substantially from those based on the CFA approach. (There are, of course, problems with the Campbell-Fiske approach, but if both the Campbell-Fiske and CFA approaches lead to consistent conclusions, confidence in these conclusions is increased.)
2. Substantive theory dictates an expected pattern of correlations among trait factors that is not supported.
3. The substantive nature of the data dictates an expected pattern of correlations among method factors that is not supported (although a priori hypotheses of relations among method factors may be difficult to formulate).
4. Models 4A (trait factors only) and 1D (method factors only) both fit the data reasonably well and Model 4D provides only a modest improvement.
5. The amount of variance explained by trait factors is substantially reduced by the inclusion of method factors.
6. External validity criteria collected in addition to the MTMM variables are more substantially correlated with so-called method factors than with trait factors, and there is an a priori basis for assuming the external criteria to be more strongly related to trait factors than method factors. (It may be impossible to obtain external validity criteria that are free of all method effects, hence the aim is to ensure that any method effects associated with the external validity criteria are unrelated to those associated with the MTMM data. However, there is still a danger that, unknown to the researcher, the external validity criteria are affected by the same method effects as the original MTMM variables.)

Although each of these indications of potential problems with the interpretation of method effects is fallible, taken together they provide a stronger basis for evaluating these interpretations than do typical applications of the CFA approach. They also require that more emphasis be placed on the substantive interpretation of results than is typical in the CFA approach to MTMM data.

APPLICATION OF THE CFA APPROACH IN THREE MTMM STUDIES

The purpose of the present investigation was to evaluate the application of the MTMM taxonomy (Table 2), the problems of poorly defined solutions and parameterizations designed to eliminate them, the merits of method structure E, and the validity of traditional interpretations of trait and method factors. Data were from three MTMM studies: Ostrom (1969), Byrne and Shavelson (1986), and Marsh and Ireland (1984). In all three studies there were at least three traits and three methods, and there was at least one external validity criterion in addition to the MTMM data. In the present analysis of each of the studies, models in the taxonomy were fit to the MTMM data, the behavior of the solutions was examined, the substantive nature of the data and the parameter estimates was used to evaluate alternative interpretations of the method and trait factors, and external validity criteria were added to the MTMM models in order to test alternative interpretations of trait and method factors.

The Ostrom (1969) Study

Description of the Study and Data

Ostrom (1969) examined the distinction between affective, behavioral, and cognitive components (t_1 through t_3) of attitudes toward the church assessed with four different methods of scale construction (m_1 through m_4). Ostrom also collected additional "overt behavioral indices" and hypothesized that these should be most highly correlated with the

[1]Campbell and Fiske (1959) stated that "the presence of method variance is indicated by the difference in level of correlations between parallel values of the monomethod block and the heteromethod block, assuming comparable reliabilities among the tests" (p. 85). Marsh (in press) operationalized this statement to provide estimates of the relative size of method effects associated with each method of assessment and discussed limitations in the inferences based upon it.

behavioral trait component. For purposes of the present investigation, one of these—responses to the item "How many days out of the year do you attend church services?"—was used. Ostrom presented the MTMM matrix based on responses by 189 people, as well as a more detailed account of the theoretical rationale, the 12 MTMM variables, and the external validity criterion.

Application of the Campbell-Fiske guidelines suggested strong support for convergent validity. However, support for discriminant validity was problematic and there appeared to be method variance associated with at least m_2 and m_4.[1] The substantive nature of the data indicated that the traits should be substantially correlated, but there is no a priori basis for positing the relative size of these different correlations. Finally, for models in which the external validity criterion was added, the criterion should be (1) more correlated with specific and general trait factors than with specific and general method factors, and (2) most highly correlated with the behavioral trait component.

CFA models similar to those considered here have been applied to these data by Bagozzi (1978), Schmitt (1978), and Widaman (1985). Schmitt (1978) excluded one of the methods and estimated trait/method correlations, hence his results are not comparable. Bagozzi (1978) fit Model 4A to the 12 variables considered here, but an inspection of correlations between the uniquenesses led him to eliminate one of the methods from subsequent analyses. (Note that such correlated uniquenesses are indicative of a method effect as depicted in method structure E.) Widaman (1985) also noted this apparent misinterpretation of method effects and was critical of other conclusions by Bagozzi. Widaman fit many of the models used here and chose to represent the MTMM data with Model 4D. However, his solution for Model 4D was poorly defined in that a uniqueness was estimated to be 0 and had a large standard error.[2] None of these previous CFAs of the Ostrom data incorporated the external validity criterion included here.

[2]Standard errors of estimated parameters that were extremely large were indicated to be 1.0 by Widaman (1985); however, the footnote indicating this was mistakenly omitted from the published article (K. F. Widaman, personal communication, September 3, 1987).

Behavior of the Solutions Under Different Parameterizations

The behavior of models in the taxonomy for these data is summarized in Table 3. All models were first tested with the fixed factor loading and the fixed factor variance parameterizations. The Rindskopf parameterization was used when both standard parameterizations produced poorly defined solutions, as well as for subsequent models that incorporated validity criteria.

For the fixed factor loading parameterization, seven of the 19 models were poorly defined as indicated by a failure to converge or improper solutions. For the fixed factor variance parameterization, five of these seven models were still poorly defined but the symptoms of this problem were not always the same. When these five models were tested with the Rindskopf parameterization, one solution was improper and the remaining four had uniqueness estimates close to 0 with extremely large standard errors. In Model 1D there were factor correlations greater than 1.0 for all three parameterizations, demonstrating that none of the parameterizations protects against this type of improper solution. Although the different parameterizations varied in their behavior and manifest symptoms, none eliminated the poorly defined solutions.

Method Structure E

In method structure E correlated uniquenesses are used to represent method effects, and in support of this structure all four solutions based on it are well defined. When there are three traits, method structure E is equivalent to method structure C if the models are well defined. For the Ostrom data this was demonstrated for Models 1C and 1E, but Models 2C, 3C, and 4C were poorly defined for all three parameterizations. Even though Model 2C failed to converge for either of the standard parameterizations, the parameter estimates for trait factors and overall fit were nearly the same as for

Table 3
Summary of Goodness of Fit Based on χ^2 and the TLI, and Solution Behavior for the Ostrom Data Under Three Parameterizations

	Parameterization														
	Fixed Factor Loadings					Fixed Factor Variances					Rindskopf Parameterization				
Model	χ^2	df	χ^2/df Ratio	TLI	Problem[a]	χ^2	df	χ^2/df Ratio	TLI	Problem	χ^2	df	χ^2/df Ratio	TLI	Problem
Models Without Validity Factors															
1A	1872	66	28.36	.000	-										
1B/2A	141	54	2.61	.941	-										
1C	776	54	14.36	.512	-										
1D	75	48	1.56	.980	2	75	48	1.56	.980	2	75	48	1.56	.980	2
1E	776	54	14.36	.512	-										
2B	187	44	4.26	.881	-										
2C	73	42	1.73	.973	1	73	42	1.73	.973	1	73	42	1.73	.973	5
2D	40	36	1.11	.996	1	39	36	1.09	.997	-					
2E	73	42	1.73	.973	-										
3A	722	54	13.37	.548	-										
3B	112	42	2.68	.939	1	109	42	2.60	.942	-					
3C	608	42	14.47	.508	1	607	42	14.46	.508	4	607	42	14.46	.508	5
3D	44	36	1.23	.991	-										
3E	607	42	14.46	.508	-										
4A	135	51	2.66	.939	-										
4B	57	39	1.47	.983	-										
4C	54	39	1.38	.986	3	54	39	1.38	.986	1	54	39	1.38	.986	5
4D	29	33	0.87	1.005	4	22	33	0.66	1.012	1	29	33	0.87	1.005	5
4E	54	39	1.38	.986	-										
Models With Validity Factors															
1A											2041	78	26.16	.000	-
1D											127	57	2.22	.951	2
2D											53	44	1.20	.992	-
4A											164	61	2.69	.933	-
4B											89	48	1.85	.966	-
4C											57	45	1.27	.989	5
4D											38	39	.97	1.001	5
4E[b]											79	49	1.61	.976	-
4E′[b]											58	46	1.26	.990	-

[a]Problems: 1 = failure to converge; 2 = factor correlation > 1.0; 3 = factor variance < 0; 4 = uniqueness < 0; 5 = excessively large standard errors. Problems 2, 3, and 4 were only examined if the solution converged, and problem 5 was examined only if the solution converged to a proper solution.

[b]In Model 4E with the validity factor, method factors were uncorrelated with the validity factor. In Model 4E′ selected uniquenesses were allowed to be correlated with the validity factor.

Model 2E. Model 3C converged to an improper solution for the fixed factor variance parameterization, but the parameter estimates for trait factors and overall fit were the same as for Model 3E. Model 4C converged to an improper solution for the fixed factor loading parameterization, but parameter estimates for trait factors and overall fit were the same as for Model 4E. The Rindskopf parameterization eliminated improper solutions for Models 2C, 3C, and 4C, but resulted in uniqueness estimates of 0 with large standard errors. These findings suggest that method structure E is a better representation of method effects than method structure C.

Substantive Interpretation of Trait and Method Factors

Interpretations based on the MTMM data. Model 4D provides an exceptionally good fit to the data, but there are problems with the solution. First, it is poorly defined for all three parameterizations. Second, trait factors are very weak (7 of the 12 factor loadings are not statistically significant) and this contradicts conclusions based on the Campbell-Fiske guidelines. Models 4C, 4E, and 3D also fit the data very well (TLIs > .98), and Models 4A, 1D, and even 1B/2A explain most of the variance (TLIs > .9). In contrast to Model 4D, Models 4E, 4C, and 4A have strong trait factors for which all factor loadings are significant. As noted by Widaman, it may be problematic to compare trait and method variance for these data because most of the variance can be explained by either trait or method factors, and neither trait nor method factors uniquely explain much variance. Also, because trait factor loadings are so much lower when correlated method factors are included, these so-called method factors may reflect trait variance.

Widaman (1985) chose Model 4D to represent these MTMM data on the basis of fit. However, Model 4E (Table 4) also provides a good fit and has important advantages over Model 4D. First, it is well defined whereas Model 4D is not. Second, the strong trait factors in Model 4E more accurately reflect conclusions from the Campbell-Fiske guidelines. Third, the weak method effects for m_1 and m_3 are consistent with an inspection of the MTMM matrix. Because interpretations based on Models 4D and 4E are so different, further consideration of the two models is important.

Interpretations based on correlations between MTMM factors and the validity factor. A 13th variable, the behavioral index of days of church attendance, was added to the 12 MTMM variables. This variable defined a single-item validity factor that was correlated with each of the MTMM factors in selected models (Table 5). This variable is qualitatively different from the 12 MTMM variables, but it is not ideal as an external validity criterion because it is also based on self-reports. Ostrom noted that an external validity criterion would be better, but decided that the self-report measure could be used "without serious risk of distortion" (p. 20). For purposes of illustration this perhaps suspect criterion was used. Because such a criterion is likely to be affected by method effects similar to those affecting the MTMM variables (although perhaps to a lesser extent), the criterion was expected to correlate with method factors but to correlate with trait

Table 4
Parameter Estimates for Model 4E for the Ostrom Data

	Factor Loadings			Uniquenesses (θ)											
Variables	t_1	t_2	t_3	t_1m_1	t_2m_1	t_3m_1	t_1m_2	t_2m_2	t_3m_2	t_1m_3	t_2m_3	t_3m_3	t_1m_4	t_2m_4	t_2m_4
t_1m_1	.80*	0	0	.36*											
t_2m_1	0	.74*	0	.01	.46*										
t_3m_1	0	0	.81*	−.01	.06	.34*									
t_1m_2	.87*	0	0	0	0	0	.25*								
t_2m_2	0	.90*	0	0	0	0	.05*	.20							
t_3m_2	0	0	.88*	0	0	0	.04	.06*	.22*						
t_1m_3	.64*	0	0	0	0	0	0	0	0	.59*					
t_2m_3	0	.76*	0	0	0	0	0	0	0	−.03	.42				
t_3m_3	0	0	.80*	0	0	0	0	0	0	−.01	.01	.36*			
t_1m_4	.84*	0	0	0	0	0	0	0	0	0	0	0	.29*		
t_2m_4	0	.74*	0	0	0	0	0	0	0	0	0	0	.14*	.44*	
t_3m_4	0	0	.74*	0	0	0	0	0	0	0	0	0	.16*	.15*	.45*
Factor Variance/Covariances (ψ)															
Factors	t_1	t_2	t_3												
t_1	1														
t_2	.96*	1													
t_3	.98*	.94*	1												

*$p < .05$.

Table 5
Correlations Between Trait (t_1-t_3), Method (m_1-m_4), General Method (g_1), General Trait (g_2), and Validity (v_1) Factors for Selected MTMM Models Based on Ostrom Study

Models

Model 4E′ (below main diagonal) and Model 4E (above main diagonal)

Factors	t_1	t_2	t_3	v_1
t_1	1	.96*	.97*	.69*
t_2	.96*	1	.94*	.77*
t_3	.97*	.94*	1	.61*
v_1	.68*	.74*	.60*	1

Model 4D

Factors	t_1	t_2	t_3	m_1	m_2	m_3	m_4	v_1
t_1	1							
t_2	.91*	1						
t_3	.95*	.91*	1					
m_1	0	0	0	1				
m_2	0	0	0	.91*	1			
m_3	0	0	0	.89*	.86*	1		
m_4	0	0	0	.75*	.70*	.74	1	
v_1	.33*	.38*	.25*	.65*	.65*	.66*	.46*	1

Model 4B (below main diagonal) and Model 4A (below main diagonal)

Factors	t_1	t_2	t_3	g_1	v_1
t_1	1	.98*	.99*	-	.67*
t_2	.96*	1	.97*	-	.76*
t_3	.97*	.94*	1	-	.60*
g_1	0	0	0	1	-
v_1	.69*	.76*	.58*	−.21*	1

Model 2D (below main diagonal) and Model 1D (above main diagonal)

Factors	m_1	m_2	m_3	m_4	g_2	v_1
m_1	1	.97*	[a]	.89*	-	.70*
m_2	.97*	1	.99*	.87*	-	.68*
m_3	.96*	.95*	1	.93	-	.73*
m_4	.88*	.87*	.88*	1	-	.58*
g_2	0	0	0	0	1	-
v_1	.59*	.62*	.59*	.51*	.59*	1

*$p < .05$.

[a]This correlation (1.01) was greater than 1.0 so that the solution was improper. A similar result was found when the validity criterion was not included (see Table 3).

factors more substantially.

The solution for Model 4D with the added validity factor was again poorly defined (Table 3). Furthermore, the validity factor correlated more substantially with each of the method factors than with any of the trait factors (Table 5). Method factors were also highly correlated with the validity factor in Model 1D that contained only correlated method factors. Models 4C and 4D were poorly defined whereas Model 4E was well defined. In contrast to Model 4D, the trait factors in Model 4E were substantially correlated with the validity factors, and, as posited by Ostrom (1969), the largest correlation was for t_2 (behavioral components).

Correlations between the validity and trait factors were similar for Models 4A and 4B. These results imply that the so-called method factors in Models 4D, 1D, and 2D actually reflect trait variance, and that the high correlations among these so-called method factors apparently represent convergence on this general trait across the methods of assessment.

In contrast to models with method factors, the validity factor was constrained to be uncorrelated with method effects in Model 4E. Although it is a simple matter to allow method factors to be correlated with the validity factor, the method effects are represented as correlated uniquenesses in Model 4E. When selected uniquenesses were allowed to be correlated with the validity factor (Model 4E′) the goodness of fit was modestly improved and correlations between trait and validity factors were slightly lower. Because there is little substantive difference between Models 4E and 4E′, the more parsimonious Model 4E may be preferable.

Summary of Analyses of the Ostrom Data

None of the three parameterizations eliminated problems of poorly defined solutions for the Ostrom data. The fixed factor loading parameterization was most prone to improper solutions. The Rindskopf parameterization was more likely to converge to proper solutions, but only at the expense of uniqueness estimates of 0 with extremely large standard errors. In contrast, solutions for method structure E were always well defined, suggesting that it might be a more appropriate formulation of method effects.

For the Ostrom data most of the variance can be explained in terms of either method factors or trait factors, whereas the inclusion of both trait and method factors produced only a small improvement in fit. Because relatively little variance was uniquely due to either trait or method factors, any conclusions about their relative importance are problematic. Even more serious problems exist in the interpretation of the correlated method factors. These so-called method factors were more substantially correlated with an external validity criterion than were the trait factors, and apparently reflect trait variance instead of, or in addition to, method effects.

The solution for Model 4E apparently provides a better representation of the MTMM data than the solution for Model 4D selected by Widaman, even though the fit of Model 4D is slightly better. The assumption of uncorrelated method effects in Model 4E is problematic, but the traditional interpretation of the method factors in Model 4D is clearly unjustified and undermines any comparisons between it and other models. This illustrates the problems associated with using fit, rather than substantive interpretations of the parameter estimates, as the primary basis for selecting from among alternative models.

The Byrne and Shavelson (1986) Study

Description of the Study and Data

Byrne and Shavelson (1986) examined the relations between three academic self-concept traits (Math, Verbal, and School self-concepts) measured by three different self-concept instruments (m_1 through m_3). School performance measures were also available for English and mathematics. Marsh and Shavelson (1985) reported Math and Verbal self-concepts to be nearly uncorrelated with each other even though both were substantially correlated with School self-concept. They posited two higher-order academic facets—verbal/academic and math/academic self-concept—to explain specific facets of academic self-concept. Their research posited a specific pattern of correlations among the trait factors, and suggested that two general trait factors may provide a reasonable fit to the Byrne and Shavelson data.

For the expanded MTMM models that include validity factors, each validity factor should be substantially correlated with the trait factors (particularly the trait factor in the matching content area and, to a lesser extent, the school factor) and relatively uncorrelated with method factors. The Byrne and Shavelson study is unusual because there is a good a priori basis for predicting the structure of the trait factors, and because two of the trait factors are relatively uncorrelated.

Application of the Campbell-Fiske guidelines to the MTMM matrix by Marsh (in press) suggested strong support for convergent and discriminant validity; every convergent validity coefficient was substantial, was larger than every heterotrait-heteromethod coefficient, and was larger than nearly every heterotrait-monomethod coefficient (see Table 6). For all three instruments School self-concept was moderately correlated with Math and Verbal self-concepts, whereas Math and Verbal self-concepts were nearly uncorrelated with each other. There was evidence of some method effects associated with at least m_3 and perhaps m_2.

The Byrne and Shavelson study is an exemplary MTMM study because of the clear support for the Campbell-Fiske guidelines, the large sample size (817, after deleting persons with missing data), the good psychometric properties of the measures, and the a priori knowledge of the trait factor structure. All models in Table 2 were fit using the fixed factor loading parameterization; the fixed factor variance parameterization was used for models that were poorly defined with the first parameterization. The Rindskopf parameterization was used for solutions that were poorly defined by both standard parameterizations.

Behavior of the Solutions Under Different Parameterizations

Nearly half of the solutions (nine of 19) were poorly defined for the fixed factor loading parameterization (Table 7); four solutions were improper, and five solutions failed to converge. When these nine poorly defined solutions were tested with the fixed factor variance parameterization, two of the solutions were well defined but the other seven remained poorly defined. For the Rindskopf parameterization only one of the seven problem solutions was improper, but all of the offending parameter estimates were approximately 0 and had large standard errors in the other six solutions.

Other characteristics of the results were:

1. All four solutions for method structure E were well defined whereas all four corresponding models for method structure C were poorly

Table 6
MTMM Matrix and Validity Criteria From Byrne and Shavelson (1986)
(Decimal Points Omitted)

Variable	t_1m_1	t_2m_1	t_3m_1	t_1m_2	t_2m_2	t_3m_2	t_1m_3	t_2m_3	t_3m_3	v_1	v_2
t_1m_1	---										
t_2m_1	384	---									
t_3m_1	441	002	---								
t_1m_2	622	368	353	---							
t_2m_2	438	703	008	441	---						
t_3m_2	465	069	871	424	136	---					
t_1m_3	678	331	478	550	380	513	---				
t_2m_3	458	541	057	381	658	096	584	---			
t_3m_3	414	027	825	372	029	810	592	135	---		
v_1	462	229	215	329	288	170	526	541	228	---	
v_2	346	036	552	277	037	495	493	149	632	517	---

Note. The 9 MTMM variables were scale scores representing three traits corresponding to School self-concept (t_1), Verbal self-concept (t_2), and Math self-concept (t_3) assessed by three methods corresponding to three different self-concept instruments. The validity criteria were school performance in English (v_1) and mathematics (v_2).

Table 7
Summary of Goodness of Fit Based on χ^2 and the TLI, and Solution Behavior for the Byrne and Shavelson Data Under Three Parameterizations

	Parameterization														
	Fixed Factor Loadings					Fixed Factor Variances					Rindskopf Parameterization				
Model	χ^2	*df*	χ^2/df Ratio	TLI	Problem[a]	χ^2	*df*	χ^2/df Ratio	TLI	Problem	χ^2	*df*	χ_2/df Ratio	TLI	Problem
Without Validity Criteria															
1A	5272	36	146.44	.000	-										
1B/2A	2365	27	87.59	.405	-										
1C	3807	27	141.02	.037	1	3807	27	141.02	.037	1	3934	27	145.71	.005	5
1D	2302	24	95.92	.347	2	2302	24	95.92	.347	2	2302	24	95.92	.347	2
1E	3807	27			-										
2B	534	20	26.70	.823	-										
2C	1698	18	94.34	.358	1	1528	18	84.88	.423	4	1528	18	84.90	.423	5
2D	385	15	25.63	.831	1	323	15	21.52	.859	-					
2E	1528	18	84.90	.423	-										
3A	1107	27	41.01	.726	-										
3B	313	18	17.37	.887	1	310	18	17.21	.889	1	387	18	21.51	.859	5
3C	708	18	39.31	.737	1	707	18	39.26	.737	4	707	18	39.28	.737	5
3D	94	15	6.30	.964	3	135	15	9.03	.945	1	188	15	12.54	.921	5
3E	707	18	39.26	.737	-										
4A	451	24	18.81	.878	-										
4B	112	15	7.49	.955	5										
4C	65	15	4.31	.977	3,4	65	15	4.31	.977	1	66	15	4.38	.977	5
4D	28	12	2.35	.991	1	40	12	3.32	.984	-					
4E	65	12	4.31	.977	-										
With Validity Criteria															
1A						6399	45	142.21	.000	-					
1D						2699	38	71.03	.504	4					
2B						800	36	22.23	.850	-					
2D						469	27	17.35	.884	-					
4A						697	38	18.33	.877	-					
4B						268	27	9.92	.937	-					
4C						132	27	5.73	.967	1	148	23	6.45	.961	5
4D						73	20	3.65	.981	-					
4E[b]						363	29	12.52	.918	-					
4E′[b]						149	26	5.73	.966	-					

[a]See Table 3 for a description of the problems.
[b]See Table 3 for the distinction between Models 4E and 4E′.

defined. When the four solutions for method structure C converged to improper solutions with either of the standard parameterizations, the parameter estimates for trait factors and the overall fit were the same as for the corresponding solution for method structure E. For the Rindskopf parameterization, the parameter estimates varied somewhat, the fit was always somewhat poorer, and some parameters had values close to 0 with large standard errors. For this application method structure E provides a better representation of method effects than method structure C.

2. Model 1D (correlated methods) produced the same improper solution—factor correlations greater than 1.0—for all three parameterizations. Because none of these parameterizations constrains factor correlations to be less than 1.0, they provide no protection from this problem.
3. Model 3D converged to an improper solution with the fixed factor loading parameterization. Although the solution was constrained to be proper by the Rindskopf parameterization, the χ^2 value was approximately twice as large.
4. For Model 4D the fixed factor loading parameterization resulted in an improper solution whereas the fixed factor variance parameterization produced a well-defined solution. The fit of the fixed factor variance solution was

somewhat poorer, indicating that it apparently imposed a limitation on the solution space.

Substantive Interpretations of Trait and Method Factors

Interpretations based on the MTMM data. Models positing only method factors fit the data poorly. Model 2B, with one so-called general method factor and one general trait factor, did substantially better. However, partly due to the manner in which Model 2B was specified, these two general factors represent the math/academic and verbal/academic factors that were originally posited. The interpretation of either of these as a general method factor is unjustified.

Model 2D (three correlated *M* factors and one general *T* factor) provided a reasonable fit to the data, but inspection of the parameter estimates demonstrated interpretational problems. The three School measures should have loaded substantially on the general trait factor, but all three loadings were small (.10 to .22). Loadings for Verbal and Math scales were larger but in the opposite direction, suggesting that the so-called general trait factor represented a bipolar (verbal vs. math self-concept) factor. For each of the method factors all loadings were substantial and positive, and the three method factors were substantially intercorrelated (.80 to .97). However, these so-called method factors and the high correlations among them seem to represent convergence on general trait factors associated with each of the self-concept instruments. The interpretation of these factors is speculative, though substantively interesting, but the traditional interpretation of the factors is clearly unjustified.

Model 4D, particularly given the large sample size, provides a remarkably good fit to the data (TLI = .991). Parameter estimates (Table 8) indicate that each of the trait factors is well defined. Consistent with theory and previous research, the School trait factor is substantially correlated with the Verbal and Math factors whereas the Verbal and Math factors are nearly uncorrelated with each other. These trait factors are stronger than the method factors in that all nine measured variables have

Table 8
Parameters For Model 4D for the Byrne and Shavelson Data (t_1 = School Self-Concept, t_2 = Verbal Self-Concept, t_3 = Math Self-Concept)

Variables	Factor Loadings (λ_y) t_1	t_2	t_3	m_1	m_2	m_3	Uniqueness (θ)
t_1m_1	.84*	0	0	.18	0	0	.27*
t_2m_1	0	.57*	0	.56*	0	0	.35*
t_3m_1	0	0	.94*	.03	0	0	.11*
t_1m_2	.68*	0	0	0	.26*	0	.48*
t_2m_2	0	.70*	0	0	.65*	0	.07
t_3m_2	0	0	.93*	0	.19*	0	.12*
t_1m_3	.77*	0	0	0	0	.60*	.03
t_2m_3	0	.87*	0	0	0	.33*	.15*
t_3m_3	0	0	.85*	0	0	.25*	.18*
Factor Variances and Covariances (ψ)							
Factors	t_1	t_2	t_3	m_1	m_2	m_3	
t_1	1						
t_2	.59*	1					
t_3	.60*	.04	1				
m_1	0	0	0	1			
m_2	0	0	0	.80*	1		
m_3	0	0	0	.25	.22	1	

higher trait factor loadings than method factor loadings, but models without any method factors (e.g., Model 4A) provide a poorer fit to the data.

Models 4C and 4E also provide good fit to the data (TLI = .977). For Models 4C, 4D, and 4E the factor loadings are similar for the School and Math traits, but the Verbal trait factor loadings are smaller for Model 4D. Correlations among trait factors are similar for all three models. Model 4D may be preferred because it fits the data slightly better, but there is no compelling basis for rejecting Model 4E and the two models lead to similar conclusions.

Relations between MTMM factors and external validity criteria. Two validity factors defined by measured achievement in English and mathematics were added to selected MTMM models (Table 9). In contrast to the external validity criterion used with the Ostrom data, method effects associated with the self-report measures are unlikely to be related to the achievement test scores. Support for the validity of the interpretation of the MTMM solutions requires each achievement factor to be most highly correlated with its matching trait factor, less correlated with the School trait factor, substantially less correlated with the nonmatching trait factor, and relatively uncorrelated with the method factors. The validity of the method factors would also be supported if the hypothesized pattern of correlations between trait factors and validity factors is improved by the addition of method factors.

Five models contain three trait factors and two validity factors in combination with correlated method factors (Model 4D), uncorrelated method factors (Model 4C), uncorrelated method effects represented as correlated errors (Model 4E), one general method factor (Model 4B), or no method factors (Model 4A). For each of these models, there is reasonable support for the predicted pattern of correlations between validity and trait factors. The inclusion of method factors improves support for this hypothesized relation in terms of Verbal self-concept, but has little effect on predictions in relation to Math self-concept or School self-concept. These results provide clear support for the a priori interpretation of the trait factors.

The comparison of Models 4C and 4E is informative. Model 4C is ill-defined but its TLI is much better than that of Model 4E. In Model 4C the validity factors are correlated with method factors, whereas in Model 4E no correlations were posited between validity factors and the uniquenesses that represent method effects. The results from Models 4C and 4D suggest that method factors are correlated with validity factors, and this apparently explains the poorer fit of Model 4E. The hypothesized pattern of relations between trait and validity factors is also stronger for Model 4C than Model 4E. However, when selected uniquenesses in Model 4E were correlated with the validity factors (Model 4E′ in Table 9), goodness of fit and support for the posited pattern of correlations were similar to Model 4C.

Correlations between trait and validity factors are similar in Models 4C, 4D, and 4E′ (Table 9). The only substantive difference is that English achievement is somewhat more highly correlated with School self-concept than Verbal self-concept in Model 4D, whereas English achievement is more highly correlated with Verbal self-concept than School self-concept for Models 4E′ and 4C. Support for the posited pattern of relations is somewhat weaker for Model 4D even though its fit is best. Model 4C is poorly defined. The correlations between method effects and validity factors are not easily represented in Model 4E′. There is no compelling reason for rejecting either Model 4D or Model 4E. In fact, the similar interpretations based on each of these different models suggest that the traditional interpretation of these models is probably justified.

Four models contain three method factors and two achievement factors in combination with correlated trait factors (Model 4D discussed above), uncorrelated trait factors (Model 4C discussed above), one general trait factor (Model 2D), and no trait factors (Model 1D). Neither Model 2D nor Model 1D contains specific trait factors, and their method factors are substantially and positively correlated with the validity factors. In fact, English achievement is more highly correlated with m_3 in Model 2D, and mathematics achievement is more highly correlated with m_1 and m_3 in Model 1D, than are any trait factors in any of the other models.

These results provide clear support for the earlier

Table 9
Correlations Between Trait (t_1-t_3), Method (m_1-m_3), General Method (g_1), General Trait (g_2), and Validity (v_1-v_2) Factors for Selected MTMM Models Based on Byrne and Shavelson Study

Model

Model 4E′ (below main diagonal) and Model 4E (above main diagonal)

Factors	t_1	t_2	t_3	v_1	v_2
t_1	1	.61*	.61*	.54*	.46*
t_2	.61*	1	.05	.43*	.08*
t_3	.61*	.05	1	.22*	.59*
v_1	.56*	.63*	.21*	1	.52*
v_2	.48*	.07*	.57*	.52*	1

Model 4D (below main diagonal) and Model 4C (above main diagonal)

Factors	t_1	t_2	t_3	m_1	m_2	m_3	v_1	v_2
t_1	1	.62*	.61*	0	0	0	.52*	.40*
t_2	.61*	1	.06	0	0	0	.60*	.07
t_3	.61*	.06	1	0	0	0	.21*	.57*
m_1	0	0	0	1	0	0	−.29*	.07
m_2	0	0	0	.85*	1	0	−.38*	.02
m_3	0	0	0	.22	.18	1	.21*	.31*
v_1	.59*	.57*	.21*	−.13*	−.16*	.10*	1	.52*
v_2	.41*	.06	.56*	.07	.02	.37*	.52*	1

Model 4B (below main diagonal) and Model 4A (above main diagonal)

Factors	t_1	t_2	t_3	g_1	v_1	v_2
t_1	1	.66*	.63*	-	.58*	.50*
t_2	.66*	1	.08*	-	.44*	.09*
t_3	.63*	.10*	1	-	.22*	.60*
g_1	0	0	0	1	-	-
v_1	.54*	.50*	.20*	−.32*	1	.52*
v_2	.48*	.12*	.59*	−.25*	.52*	1

Model 2B

Factors	g_1	g_2	v_1	v_2
g_1	1			
g_2	0	1		
v_1	.53*	.21*	1	
v_2	.11*	.61*	.52*	1

Model 2D (below main diagonal) and Model 1D (above main diagonal)

Factors	m_1	m_2	m_3	g_2	v_1	v_2
m_1	1	[a]	[a]	-	.36*	.66
m_2	.98*	1	.95*	-	.22*	.54*
m_3	.89*	.82*	1	-	.35*	.70*
g_2	0	0	0	1	-	-
v_1	.54*	.39*	.62*	.08	1	.52*
v_2	.38*	.30*	.48*	−.47*	.52*	1

*$p < .05$.

[a]These correlations (1.11 and 1.05) were greater than 1.0 so that the solution was improper. A similar result was found when the validity criterion was not included (see Table 7).

supposition that these so-called method factors contain substantial amounts of trait variance. In Model 2B there are just two general factors that might be interpreted to reflect a general method factor and a general trait factor. However, the correlations between these two general factors and the validity factors demonstrate that they reflect the math/academic and verbal/academic factors originally posited.

Summary of Analyses of the Byrne and Shavelson Data

The Byrne and Shavelson study is an exemplary MTMM study that should be well suited to the CFA approach. For this reason it is particularly disappointing that so many of the models from Table 2 resulted in poorly defined solutions. So long as both correlated trait factors and either correlated or uncorrelated method factors were included in the MTMM model, support for the traditional interpretation of MTMM factors appeared reasonable. However, for models with no trait factors or only one general trait factor, the interpretation of method factors as representing method variance was clearly unjustified. Instead, the so-called method factors reflected the influence of trait variance. Further support for this counterinterpretation was provided by the substantial correlations between these so-called method factors and the validity factors.

The Marsh and Ireland (1984) Study

Description of the Study and Data

Marsh and Ireland (1984) asked each of six teachers (m_1 through m_6) to evaluate 139 student essays on six single-item scales of writing effectiveness: Mechanics, Sentence Structure, Word Usage, Organization, Content/ideas, and Quality of style (t_1 through t_6). Previous research reported a large general component of writing effectiveness suggesting that trait factors should be substantially correlated. There was no a priori hypothesis about the relative size of different trait correlations, but the traits were roughly ordered from lower-order components to higher-order components according to Foley's (1971) adaptation of the Bloom taxonomy.

Unlike the first two MTMM studies, ratings for the different methods were not completed by the same person. The teachers did not know any of the students who had written the essays (nor, typically, each other), and each teacher performed the rating task independently of the others.[3] In addition to the 36 variables that constitute the MTMM data, a school performance measure of writing effectiveness was also available. Application of the Campbell-Fiske guidelines (Marsh & Ireland, 1984) suggested strong support for the convergent validity of all six traits. However, there was little support for discriminant validity and some indication of method effects associated with ratings by each of the six teachers.

Behavior of the Solutions Under Different Parameterizations

Because the number of variables in this MTMM study was large, only a subset of the MTMM models was tested with the fixed factor variance parameterization (Table 10). Nevertheless, 12 of these 13 models resulted in well-defined solutions. Model 4B was improper in that several trait correlations exceeded 1.0, but this type of improper parameter estimate is unlikely to be eliminated by any of the parameterizations. Model 4D was technically improper in that the factor correlation matrix (see Table 11) was not positive definite even though none of the correlations was greater than 1.0. Despite these problems, the solutions for the Marsh and Ireland data appear to be better behaved than for either of the first two MTMM studies.

[3]The CFA approach to MTMM data assumes that the different methods represent fixed effects. Although this limitation may be reasonable for some applications, it is probably inappropriate for the Marsh and Ireland data where the different raters more realistically constitute a random-effects facet (i.e., a sample of a potentially much larger sample of raters). However, the present author is unaware of any solution to this problem.

Table 10
Summary of Goodness of Fit Based on χ^2 and the TLI, and Solution Behavior for the Marsh and Ireland Data, as Tested With the Fixed Factor Variance Parameterization

Model	χ^2	*df*	χ^2/df Ratio	TLI	Problem[a]
Without Validity Criteria					
1A	6407	630	10.17	.000	-
1B/2A	2439	593	4.11	.660	-
1C	1991	594	3.36	.742	-
1D	1296	579	2.24	.865	-
1E	1590	540	2.94	.788	-
2C	1199	558	2.15	.875	-
2D	997	543	1.84	.909	-
2E	852	504	1.69	.925	-
4A	2294	479	3.96	.677	-
4B	1776	543	3.27	.752	2
4C	981	543	1.81	.912	-
4D	845	528	1.60	.935	-
4E	722	489	1.48	.948	-
With Validity Criteria					
1C	6593	666	9.90	.000	-
1D	1334	610	2.19	.867	-
2D	1035	573	1.83	.909	-
2E	896	540	1.66	.923	-
4A	2331	610	3.82	.683	-
4B	1814	573	3.17	.757	2
4C	1009	568	1.78	.913	-
4D	873	533	1.64	.928	-
4E	758	520	1.56	.948	-

[a]See Table 3 for a description of the problems.

Substantive Interpretation of Trait and Method Factors

Interpretations based on the MTMM data. The goodness-of-fit statistics (Table 10) demonstrated that much of the variance can be explained by either six correlated method factors (Model 1D) or six correlated trait factors (Model 4A), but that Model 1D fit the data slightly better than 4A. For Model 4D, method factor loadings were consistently much larger than trait factor loadings. A superficial inspection of these results might suggest that the ratings reflect primarily method effect, but there are problems with this interpretation. First, it contradicts conclusions based on the Campbell-Fiske guidelines. Second, trait factor loadings were substantially smaller in Model 4D than in Model 4A. This suggests that the so-called method factors represent general trait factors associated with each teacher, and that the high correlations represent agreement across teachers on this general trait.

Correlations between MTMM factors and the validity factor. In order to test the counterinterpretation of the method factors, the school performance measure was added to Model 4D. The parameters for the MTMM variables were relatively unaffected by the inclusion of this additional variable. However, the school performance factor was substantially more correlated with so-called method factors (r = .56 to .68) than with trait factors (r = .14 to .30). Because this pattern of results is so implausible, the traditional interpretation of the so-called method factors in this model must be rejected. Models 2D and 1D also posited correlated method factors, and correlations between these method factors and the validity factor were also very high (r = .53 to .75), whereas the general trait factor in Model 2D was only modestly correlated with the validity factor.

In contrast to models with correlated methods, Model 4C posited method factors to be uncorrelated. For Model 4C, correlations between trait factors and the validity factor varied between r = .71 and .89, whereas five of the six correlations between method factors and the validity factor were nonsignificant. Even though the fit for Model 4C was somewhat poorer than for Model 4D, the substantive interpretation of the solution implies that it better reflected the MTMM data.

Method Structure E

The first two MTMM studies both contained three trait factors, and for such studies the results for method structure E are equivalent to those for method structure C if the solutions are well defined. However, when there are more than three traits, as here, the two structures are not equivalent. For T = 6, method structure C uses six parameters to define each method factor, whereas there are 15 $[T \times (T-1)/2]$ correlations among the uniquenesses associated with each method.

Table 11
Correlations Between Trait (t_1-t_6), Method (m_1-m_6), General Method (g_1), General Trait (g_2) and Validity (v_1) Factors for Selected MTMM Models Based on Marsh and Ireland Data

Model

Model 4E

Factors	t_1	t_2	t_3	t_4	t_5	t_6	v_1
t_1	1						
t_2	.98*	1					
t_3	.92*	.96*	1				
t_4	.93*	.94*	.93*	1			
t_5	.90*	.91*	.93*	.96*	1		
t_6	.94*	.96*	.95*	.98*	.98*	1	
v_1	.81*	.80*	.77*	.83*	.76*	.80*	1

Model 4D (below main diagonal) and Model 4C (above main diagonal)

Factors	t_1	t_2	t_3	t_4	t_5	t_6	m_1	m_2	m_3	m_4	m_5	m_6	v_1
t_1	1	.99*	.92*	.92*	.89*	.94*	0	0	0	0	0	0	.77*
t_2	.99*	1	.95*	.94*	.89*	.95*	0	0	0	0	0	0	.75*
t_3	.64*	.76*	1	.92*	.94*	.94*	0	0	0	0	0	0	.71*
t_4	.70*	.74*	.61*	1	.97*	.98*	0	0	0	0	0	0	.79*
t_5	.57*	.58*	.62*	.84*	1	.99*	0	0	0	0	0	0	.73*
t_6	.78*	.79*	.68*	.93*	.98*	1	0	0	0	0	0	0	.73*
m_1	0	0	0	0	0	0	1	0	0	0	0	0	.08
m_2	0	0	0	0	0	0	.71*	1	0	0	0	0	−.10
m_3	0	0	0	0	0	0	.72*	.71*	1	0	0	0	.06
m_4	0	0	0	0	0	0	.72*	.73*	.79*	1	0	0	.12
m_5	0	0	0	0	0	0	.66*	.65*	.79*	.77*	1	0	.08
m_6	0	0	0	0	0	0	.72*	.72*	.78*	.67*	.65*	1	.16*
v_1	.28*	.25*	.14	.30*	.17	.22*	.66*	.56*	.69*	.68*	.66*	.67*	1

Model 4B (below main diagonal) and Model 4A (above main diagonal)

Factors	t_1	t_2	t_3	t_4	t_5	t_6	g_1	v_1
t_1	1	.104*	.96*	.96*	.94*	.98*	-	.81*
t_2	.106*	1	.99*	.98*	.95*	.99*	-	.80*
t_3	.64*	.80*	1	.98*	.101*	.100*	-	.76*
t_4	.66*	.84*	.77*	1	.103*	.104*	-	.83*
t_5	.45*	.59*	.101*	.100*	1	.107*	-	.77*
t_6	.71*	.82*	.81*	.110*	.110*	1	-	.80*
g_1	0	0	0	0	0	0	1	-
v_1	.27*	.26*	.20*	.35*	.26*	.29*	.74*	1

Model 2D (below main diagonal) and Model 1D (above main diagonal)

Factors	m_1	m_2	m_3	m_4	m_5	m_6	g_2	v_1
m_1	1	.77*	.75*	.78*	.73*	.77*	-	.73*
m_2	.73*	1	.75*	.79*	.73*	.76*	-	.61*
m_3	.71*	.70*	1	.82*	.81*	.81*	-	.75*
m_4	.74*	.77*	.79*	1	.80*	.72*	-	.74*
m_5	.67*	.69*	.78*	.78*	1	.70*	-	.71*
m_6	.72*	.72*	.77*	.68*	.65*	1	-	.72*
g_2	0	0	0	0	0	0	1	-
v_1	.65*	.53*	.67*	.67*	.63*	.64*	.33*	1

Note. All parameters with values of 1 or 0 were fixed. Factor correlations greater than 1.00 are improper estimates.
$*p < .05$.

Insofar as method structures C and E are both well defined and fit the data equally well, the more parsimonious method structure C is preferable. However, the fit of models based on method structure E was much better than that of models based on method structure C for these data. This suggests that the 15 correlated uniquenesses associated with each method effect cannot be explained by a single method factor, and that method effects do not have a congeneric-like structure. This is very important in that all method factors in the entire taxonomy are based on these assumptions. This also explains why Model 4E (TLI = .948) fits the data better than Model 4D (TLI = .935), even though Model 4E posits uncorrelated method effects whereas Model 4D posits correlated method factors.

The superiority of Model 4E over 4C is also shown in the expanded models containing the validity factor. Correlations between trait factors and the validity factor are substantial for Models 4E and 4C, but are higher for Model 4E. As noted for the first two MTMM studies, correlations between method effects and the validity factor are not easily incorporated into Model 4E. However, correlations between validity and method factors were small and generally nonsignificant for Model 4C. Similarly, inspection of the modification indices provided by LISREL (see Jöreskog & Sörbom, 1981) indicated that uniquenesses in Model 4E were essentially uncorrelated with the validity factor. For this reason, no alternative model corresponding to Model 4E′ in the first two studies was proposed.

SUMMARY AND IMPLICATIONS

What motivates the kinds of analyses discussed here? One perception[4] is that MTMM analyses are motivated by the desire to establish specific trait representations in measures. Method variance is seen as contaminating that representation. The CFA approach, as traditionally applied, has modeled trait and method factors as if they were equally important. The approach advocated here places greater emphasis on the interpretation of trait representations.

[4]This perspective was expressed by an anonymous reviewer.

This is accomplished by comparing different models to determine if the introduction of method factors substantially alters the interpretation of trait representations, by introducing an alternative method structure (method structure E) that apparently provides a more accurate representation of the trait representation, and by demonstrating how external validity criteria can be used to test the validity of the traditional interpretations of different models. Although the conventional approach was due at least in part to early work by Jöreskog, the perspective taken here is consistent with Jöreskog's statement that "method factors are what is left over after all trait factors have been eliminated" (1971, p. 128).

Despite the growing enthusiasm for the CFA approach to MTMM data, problems demonstrated here call into question its value, the traditional interpretation of MTMM factors, and the validity of previous applications of CFA to MTMM data. The most important of these problems are the technical difficulties in estimating parameters and the interpretation of so-called method effects that apparently represent the effects of trait variance in addition to, or instead of, method variance. So long as problems as basic as these remain unresolved, the promise of the CFA approach to MTMM data cannot be fulfilled.

Method structures in Widaman's (1985) taxonomy and those used in most applications of CFA to MTMM data posit a separate method factor associated with each method of assessment. An alternative conceptualization, method structure E, was formulated in which method effects are represented as correlated uniquenesses.

Method structure E has three important advantages over method structures C and D. First, models with method structures C and D were frequently ill-defined no matter what parameterization was used, whereas models based on method structure E were always well defined in the present applications. Second, when there were more than three traits, method structure E provided a test of the implicit assumption that all the correlated uniquenesses associated with a single method of assessment could be explained in terms of a single method factor. The importance of this second advantage was demonstrated for the Marsh and Ireland data in that Model 4E provided a better fit than the corresponding Models 4C and 4D. Third, method structure E apparently provided a more accurate interpretation of trait variance than alternative models when these interpretations were evaluated in relation to external validity criteria. In this respect, the use of external validity criteria to validate interpretations of the method and trait effects is an important conceptual innovation.

The most serious potential problem with MTMM models is the implicit assumption that so-called method factors represent primarily the effects of method variance. If this assumption is violated, then the interpretation of trait and method factors in most CFA studies and the detailed comparison of nested models proposed by Widaman (1985) may be unjustified. Results from the MTMM studies considered here suggest that this assumption is often implausible. In all three MTMM studies, the so-called method factors for at least some of the MTMM models apparently represented trait variance in addition to or instead of method variance (see also Marsh & Butler, 1984, for another compelling example). When there actually are distinct traits that are at least moderately correlated, this phenomenon is most likely in models that posit correlated method factors (method structure D). Using method structure D, the problem is likely to be most severe in models that posit no trait factors (1D) and to become less severe as the trait structure proceeds from 1 to 4. This problem will apparently be least likely to occur when method factors are required to be uncorrelated, as in method structures C and E.

The emphasis of the present investigation has been on potential problems in the interpretation of so-called method factors that really reflect variance that should be attributed to a general trait effect. This is consistent with Jöreskog's (1971) conceptualization of method effects (i.e., what remains after trait factors have been removed) as well as the present author's perspective on the intent of MTMM analyses. It is important to note, however, that the converse phenomenon may also exist. That is, it is possible that so-called trait effects really reflect variance that should be attributed to a general method effect.

If an appropriate method structure is not employed, then so-called trait factors may represent method variance in addition to, or instead of, trait variance. An unresolved conceptual and technical problem is how to discriminate between method and trait factors when both are highly correlated. In the extreme, it is easy to imagine the case where a MTMM matrix of correlations produced by highly correlated trait and method factors could be explained by a single factor (Model 1B/2A). Although this situation would clearly indicate a lack of discriminant validity, there would be little basis for determining whether the single factor represented a general trait effect, a general method effect, or a combination of the two.

The taxonomy of MTMM models in Table 2 was based in large part on Widaman's (1985) taxonomy. Widaman used essentially the same CFA approach and many of the same MTMM models, and even analyzed one of the same MTMM studies. Because Widaman's evaluation of the CFA approach was much more optimistic, it is informative to critically evaluate his findings in relation to the criteria used here.

Widaman did not provide a detailed report of the behavior of his CFA solutions, but results reported here indicate that poorly defined solutions occurred for the Ostrom data considered in both studies. Widaman chose to present five MTMM solutions as the most appropriate representations of his MTMM analyses. However, four of these solutions had uniquenesses of 0 in conjunction with large standard errors, whereas the fifth solution required a correlation between two method factors to be 1.0. Wothke (1984) also reported that 21 MTMM matrices—including the three analyzed by Widaman—resulted in poorly defined solutions when he fit Model 4D.

Apparently, none of the solutions chosen by Widaman was well defined according to criteria used here, suggesting that Widaman was also plagued by poorly defined solutions. Widaman did not report a critical evaluation of alternative interpretations of his method factors, but results reported here suggest that this was a problem for the Ostrom data. Using criteria described earlier, there is reason to suspect that so-called method factors in at least some of Widaman's results of other MTMM matrices may have also represented trait variance in addition to, or instead of, method variance. Thus, a critical evaluation of Widaman's results provides little basis for optimism about the application of CFA to MTMM data. His results suggest the same sort of problems that were identified here.

RECOMMENDATIONS

Problems with the CFA approach to MTMM data appear to be most serious for MTMM studies in which method effects are substantially correlated and for MTMM models that posit correlated method factors. Campbell and Fiske (1959) originally stressed that the multiple methods should be as distinct as possible, and this advice seems appropriate for the CFA studies as well.

The choice of method effects is, however, often dictated by the nature of the study, and the pattern of correlations among method factors may be difficult to determine a priori. Particularly when both traits and methods are substantially correlated, the researcher must critically evaluate the MTMM solutions for alternative interpretations. Because the traditional interpretation of trait and method factors may frequently be unjustified, the burden of proof lies with the researcher to demonstrate that the interpretations are justified. This requires that more emphasis be given to substantive interpretations than has typically been the case in CFA studies.

The use of uncorrelated traits may also be helpful, though this is unusual in MTMM studies. Byrne and Shavelson (1986), however, did consider two trait factors that were nearly uncorrelated. This is important because the application of Model 4D to their data resulted in a well-defined solution that was substantively meaningful. Even though Model 4D is the basis of most studies, this is one of its few successful applications. Wothke (1984), for example, reported that Model 4D resulted in poorly defined solutions for all 21 MTMM matrices in his study. Perhaps the inclusion of the two uncorrelated traits was part of the reason that Model 4D worked for the Byrne and Shavelson data. Because there was not just one general trait factor underlying the specific trait factors, trait and method factors were not so easily confounded. This suggests that it may be useful for researchers to include at least some traits that are relatively uncorrelated in MTMM studies.

Widaman proposed that the interpretation of MTMM studies should be based in part on a series of comparisons between nested models in his taxonomy. The rigor of this approach is laudable, but it is overly restrictive because many comparisons do not involve nested models and because classical hypothesis testing may be inappropriate. Furthermore, even when models are nested, it is important to establish the validity of the traditional interpretations of trait and method effects for both nested models. Models 1D and 2D were particularly important in Widaman's comparisons of nested models, but results presented here suggest that the traditional interpretation is particularly problematic for these models. Because of these problems, the nested comparisons emphasized by Widaman should be used cautiously, and only after substantive support for the interpretation of the models being compared has been established.

The choice of the model that best represents the MTMM data should be based primarily on substantive interpretations of parameter estimates and secondarily on goodness of fit. The purpose of MTMM models is not necessarily to provide the best fit to the data, but rather to make accurate inferences about trait factors and, perhaps, method factors. If a solution is poorly defined or parameter estimates are substantively unreasonable, it is better to infer their effects from different models even if the fit is marginally poorer. However, if parameter estimates are substantively unreasonable for most of the models in the taxonomy, if the fit of most of these models is poor, or if solutions are consistently ill-defined, then the CFA approach may be inappropriate for the particular MTMM data.

Although 19 different models are posited in Table 2, MTMM data can typically be evaluated with fewer models. The particular models to be considered will vary depending on the application, but models positing trait and method factors (4D and 4C) and method effects represented as correlated uniquenesses (4E) seem most useful, supplemented perhaps by those positing one general factor (1B/2A), only trait factors (4A), and only method factors (1D). Particularly when Models 4D and 4E are both well-defined and lead to similar conclusions, as with the Byrne and Shavelson (1986) data, the traditional interpretation of these models is probably justified. In this case it may be reasonable to base inferences about method and trait effects on these models—alone—dispensing with other models altogether. However, other models from the taxonomy, or models idiosyncratic to particular substantive issues, may provide useful supplemental information about the data.

In the CFA approach to MTMM data, the number of estimated parameters increases rapidly with the number of traits and methods. Rules of thumb (e.g., Tanaka, 1987) suggest that the behavior of CFA solutions may be related to the ratio of the number of estimated parameters to sample size, so that MTMM designs with large numbers of traits and methods may require very large sample sizes. Note, however, that three-trait/three-method designs attempt to infer six factors from only nine measured variables—a ratio of 1.33 measures per factor. For a six-trait/six-method design, the ratio of measured variables (36) to factors (12) is 3. This suggests the possibility that for a given sample size, solutions might actually be better behaved for larger MTMM designs with more traits and methods and more parameters to be estimated. Further research is needed to determine how the behavior of MTMM solutions varies with sample size and with the number of traits and methods.

Model 4E represents an apparently important innovation in the CFA approach to MTMM data—particularly if claims about this new model are justified in subsequent research. Model 4E has not been widely applied elsewhere, and its apparent advantages should be further examined in additional studies. A particularly useful evaluation would be to apply various models—including method structure E—to simulated data in which the underlying factor structure is known. Subject to the results of this further research, Model 4E should be included in the MTMM models examined in all applications of CFA to MTMM data.

References

Althauser, R. P., & Heberlein, T. A. (1970). Validity and the multitrait-multimethod matrix. In E. F. Borgatta & W. Bohrnstedt (Eds.), *Sociological methodology 1970* (pp. 151–169). San Francisco: Jossey-Bass.

Alwin, D. F. (1974). Approaches to the interpretation of relationships and the multitrait-multimethod matrix. In H. L. Costner (Ed.), *Sociological methodology 1973–4* (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.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Bollen, K. A., & Jöreskog, K. G. (1985). Uniqueness does not imply identification: A note on confirmatory factor analysis. *Sociological Methods & Research, 14,* 155–163.

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. (1980). Factor analysis of multiple batteries by maximum likelihood. *British Journal of Mathematical and Statistical Psychology, 33,* 184–199.

Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology, 37,* 1–21.

Byrne, B. M., & Shavelson, R. J. (1986). On the structure of adolescent self-concept. *Journal of Educational Psychology, 78,* 474–481.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–105.

Campbell, D. T., & O'Connell, E. J. (1967). Method factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research, 2,* 409–426.

Cole, D., & Maxwell, S. E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analysis approach. *Multivariate Behavioral Research, 4,* 389–418.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 2,* 147–167.

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis:

Comments on the causes and solutions to Heywood cases. *Psychological Bulletin, 101,* 126–135.

Foley, J. J. (1971). Evaluation of learning in writing. In B. S. Bloom, J. T. Hastings, & G. F. Madaus (Eds.), *Handbook on formative and summative evaluation of student learning*. New York: McGraw-Hill.

Hubert, L. J., & Baker, F. B. (1978). Analyzing the multitrait-multimethod matrix. *Multivariate Behavioral Research, 13,* 163–179.

Jackson, D. N. (1969). Multimethod factor analysis in the evaluation of convergent and discriminant validity. *Psychological Bulletin, 72,* 30–49.

Jackson, D. N. (1975). Multimethod factor analysis: A reformulation. *Multivariate Behavioral Research, 10,* 259–275.

Jöreskog, K. G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika, 36,* 409–426.

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. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics, 8,* 65–92.

Jöreskog, K. G., & Sörbom, D. (1981). *LISREL V: Analysis of linear structural relations by the method of maximum likelihood*. Chicago: International Educational Services.

Kenny, D. A. (1979). *Correlation and causality*. New York: Wiley.

Long, K. S. (1983). *Confirmatory factor analysis: A preface to LISREL*. Beverly Hills CA: Sage.

Marsh, H. W. (in press). Multitrait-multimethod analysis. In J. P. Keeves (Ed.), *Educational research methodology, measurement, and evaluation: An international handbook*. Oxford: Pergamon.

Marsh, H. W., Balla, J., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Marsh, H. W., Barnes, J., & Hocevar, D. (1985). Self-other agreement on multidimensional self-concept ratings: Factor analysis and multitrait-multimethod analysis. *Journal of Personality and Social Psychology, 49,* 1360–1377.

Marsh, H. W., & Butler, S. (1984). Evaluating reading diagnostic tests: An application of confirmatory factor analysis to multitrait-multimethod data. *Applied Psychological Measurement, 8,* 307–320.

Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement, 20,* 231–248.

Marsh, H. W., & Ireland, R. (1984). *Multidimensional evaluations of writing effectiveness*. (ERIC Document Reproduction Service No. ED 242 785)

Marsh, H. W., McDonald, R. P., & Balla, J. (1989). *Goodness of fit in confirmatory factor analysis: The effect of sample size and model parsimony*. Manuscript submitted for publication.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20,* 107–125.

McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale NJ: Erlbaum.

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, 13,* 447–478.

Schmitt, N., & Stults, D. M. (1986). Methodological review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement, 10,* 1–22.

Stanley, J. C. (1961). Analysis of unreplicated three-way classification with application to rater bias and trait independence. *Psychometrika, 26,* 205–219.

Tanaka, J. S. (1987). "How big is big enough?": Sample size goodness of fit in structural equation models. *Child Development, 58,* 134–146.

Tucker, L. R., & Lewis, C. (1973). The reliability index 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.

Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement, 9,* 1–26.

Wothke, W. (1984). *The estimation of trait and method components in multitrait-multimethod measurement*. Unpublished doctoral dissertation, University of Chicago.

Acknowledgments

The author thanks Barbara Byrne and Richard Shavelson for providing data used in the present investigation, and Michael Browne, Dennis Hocevar, David Kenny, Jack McArdle, Keith Widaman, and Werner Wothke for comments on earlier drafts of this paper. The contents, however, represent the views of the author.

Author's Address

Send requests for reprints or further information to Herbert W. Marsh, Faculty of Education, University of Sydney, NSW 2006, Australia.