Confirmatory Factor Analyses of Multitrait-Multimethod Data: A Comparison of Alternative Models

Herbert W. Marsh
University of Western Sydney, Australia

Michael Bailey
University of Sydney, Australia

Alternative models for confirmatory factor analysis of multitrait-multimethod (MTMM) data were evaluated by varying the number of traits and methods and sample size for 255 MTMM matrices constructed from real data (Study 1), and for 180 MTMM matrices constructed from simulated data (Study 2). The correlated uniqueness model converged to proper solutions for 99% (Study 1) and 96% (Study 2) of the MTMM matrices, whereas the general model typically used converged to proper solutions for only 24% (Study 1) and 22% (Study 2) of the MTMM matrices. The general model was usually ill-defined (100% in Study 1, 90% in Study 2) for small MTMM matrices with small *N*s, but performed better when the size of the MTMM matrix and *N* were larger. Even when both models converged to proper solutions, however, parameter estimates for the correlated uniqueness model were more accurate and precise in relation to known population parameters in Study 2. *Index terms: confirmatory factor analysis, construct validity, discriminant validity, LISREL, method effects, multitrait-multimethod analysis, underidentified models.*

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 multitrait-multimethod (MTMM) design, in which each of a set of multiple traits is assessed with each of a set of multiple methods of assessment, and they developed guidelines for evaluating MTMM data. Their MTMM design has become perhaps the most frequently employed construct validation design, and their original guidelines continue to be used frequently.

However, important problems with their guidelines are well known (e.g., Althauser & Heberlein, 1970; Alwin, 1974; Campbell & O'Connell, 1967; Marsh, 1989a; Wothke, 1984) and have stimulated alternative analytic procedures. Confirmatory factor analysis (CFA) models (e.g., Boruch & Wolins, 1970; Jöreskog, 1974; Marsh, 1988, 1989a; Marsh & Hocevar, 1983; Schmitt & Stults, 1986; Widaman, 1985) or mathematically similar path-analytic models (e.g., Schmitt, Coyle, & Saari, 1977; Werts & Linn, 1970) currently appear to be the most popular approach and were the focus of the present investigation.

The CFA Approach to MTMM Data

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 argue for method effects. The emphasis here was on a general MTMM model adapted from Jöreskog (1971, 1974, 1981; see also Marsh, 1988, 1989a; Widaman, 1985) in which (1) there are at least three traits (T = 3) and three methods (M = 3); (2) T × M measured variables are used to infer T + M a priori common factors; (3) each

APPLIED PSYCHOLOGICAL MEASUREMENT
Vol. 15, No. 1, March 1991, pp. 47-70
© Copyright 1991 Applied Psychological Measurement Inc.
0146-6216/91/010047-24$2.45

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

measured variable loads on the one trait factor and the one method factor that it represents, but is constrained so as not to load on any other factors; and (4) correlations among the trait factors and among the method factors are freely estimated, but correlations between trait and method factors are constrained to be 0. This model allows the decomposition of variance into additive trait, method, and error components.

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. Widaman (1985) developed an important taxonomy of such models that systematically varied different characteristics of the trait and method factors. This taxonomy was designed to be appropriate for all MTMM studies, to provide a general 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.

Marsh (1988, 1989a) expanded the taxonomy to include all possible combinations of the following four trait structures:

1. No trait factors,
2. One general trait factor defined by all measured variables,
3. T uncorrelated trait factors, and
4. T correlated trait factors;

and the following five method structures:

A. No method factors,
B. One general method factor defined by all measured variables
C. M uncorrelated method variables,
D. M correlated method factors, and
E. Method effects inferred on the basis of correlated uniqueness (E).

This taxonomy differs from Widaman's original taxonomy only in the addition of method structure E. For purposes of the present investigation, primary emphasis was placed on the general model (4D) and the correlated uniqueness model (4E).

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 effects associated with variables assessed by the same method. The effects of a particular method 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 (see Wothke, 1984).

Alternatively, method effects can be represented as correlated uniquenesses (method structure E), and this representation does not assume either the unidimensionality of effects associated with a particular method or a congeneric structure. The correlated uniquenesses in method structure E merely represent correlations between pairs of variables measured with the same method after removing trait effects. If these are statistically significant and sufficiently large, there is evidence for method effects. Thus the correlated uniquenesses in method structure E apparently represent a natural interpretation of method effects.

Kenny (1976, 1979; see also Marsh & Hocevar, 1983) argued that the general MTMM model cannot be applied when there are only two traits, and he first proposed what here is called method structure E for this special case. Kenny did not consider this method structure for the more general situation, though, when there are more than two traits; instead, he used an MTMM model in which methods are posited as separate factors (e.g., Kenny, 1979). Marsh (1988, 1989a) noted, however, that method structure E is also reasonable when there are more than two traits. He generalized the application of this method structure to apply to all MTMM matrices, and expanded the typology proposed by Widaman (1985) to include this method structure.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Method structure E corresponds most closely to method structure C, in that the method effects associated with one method are assumed to be uncorrelated with those associated with other methods. When there are three 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 claimed by Marsh (1988, 1989a) is that it results typically in proper solutions without limiting the solution space or forcing parameter estimates to the boundaries of the permissible space, even when models 4C or 4D result in improper solutions.

Goodness of Fit and Poorly-Defined Solutions

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 to decide whether the predicted and observed results are sufficiently alike to justify support of a model. A variety of fit indices has been developed to aid this process, but a discussion of goodness of fit and fit indices is beyond the scope of the present investigation (see Bentler & Bonett, 1980; Mulaik, James, Van Alstine, Bennett, Lind, & Stilwell, 1989). In an evaluation of the most frequently used fit indices based on real data, simulated data, and mathematical derivations, the Tucker-Lewis Index (TLI; Tucker & Lewis, 1973) was the only widely used index that was relatively independent of sample size and did not vary substantially when superfluous parameters known to be 0 in the population were estimated (Marsh, 1989b; Marsh, Balla, & McDonald, 1988; McDonald & Marsh, 1990). For this reason, the TLI was used as one basis for assessing fit in the present investigation, although the Bentler-Bonett Index (BBI; Bentler & Bonett, 1980) was also used.

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).

Marsh (1988, 1989a) suggested that such problems are apparently more likely when the sample size is small, there are few indicators of each latent factor, measured variables are allowed to load on more than one factor, measured variables are highly correlated, there is much missing data and covariance matrices are estimated with pairwise deletion for missing data, and the model is misspecified. Anderson and Gerbing (1984; Gerbing & Anderson, 1985, 1987; see also Dillon, Kumar, & Mulani, 1987) discussed three potential causes of improper solutions: (1) sampling fluctuations in relation to true population parameters close to the limits of the permissible range of values (e.g., variances close to 0), (2) model misspecification, and (3) an indefiniteness or underidentification of the model. Such problems are usually ignored, and the implications of this practice have not been explored for MTMM studies.

There are many rules of thumb about the absolute sample size, sample sizes in relation to the number of measured variables or estimated parameters, and the number of indicators per factor that are needed to obtain well-defined solutions. The usefulness of such guidelines has been evaluated on the basis of monte carlo studies. Tanaka (1987) suggested that the appropriate

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

sample size is a function of the ratio of examinees to the number of variables or, particularly for CFA models, the ratio of examinees to the number of parameter estimates.

In apparently related research based on the stability of solutions in exploratory factor analyses, however, Guadagnoli and Velicer (1988) found that the ratio of examinees to number of variables was not important. Instead, solution stability varied primarily with absolute sample size, component saturation, and to a lesser extent, the number of indicators per factor. Boomsma (1982) recommended that at least 100 (and preferably 200) cases were required. Similar results were reported by Anderson and Gerbing (1984), although larger samples were required when there were only two indicators per factor. Anderson and Gerbing (Gerbing & Anderson, 1985, 1987) found that nonconvergence and improper solutions were more likely for smaller sample sizes and models with fewer indicators per latent trait. Their research suggests that solutions with only two indicators per factor are particularly likely to produce poorly defined solutions.

Most of these monte carlo studies are based on population models in which each measured variable is defined by one and only one latent construct (i.e., what McDonald, 1985, calls an independent clusters structure). Real data, however, will never exactly conform to this simple structure, and the deviations will always be statistically significant if the sample size is large enough (see Marsh, Balla, & McDonald, 1988; McDonald & Marsh, 1990). For this reason, solutions based on monte carlo studies are likely to be better behaved than those based on real data.

The general MTMM model posits each measured variable to be defined by two different latent constructs—a trait factor and a method factor. Because this structure is more complicated than the independent clusters structure typically used in monte carlo studies, CFA solutions for MTMM data are likely to be less well-behaved than suggested by monte carlo studies. Larger sample sizes may be needed to achieve stable, well-defined solutions. More indicators per factor also may be required. For independent cluster structures, the number of indicators per factor is equal to the ratio of measured variables to factors, but this is not the case for MTMM data. In the 3T × 3M MTMM design, for example, there are three indicators for each factor, but six factors are inferred from only nine measured variables. For the most frequently used MTMM designs (e.g., 3T × 3M, 4T × 3M, 3T × 4M, 4T × 4M, 5T × 3M, 3T × 5M), the ratio of measured variables to factors is two or less.

Purpose

LISREL V (Jöreskog & Sörbom, 1981) was used in this research to derive maximum likelihood estimates for MTMM data. The behavior of selected MTMM models was evaluated in two different studies: the first based on 255 MTMM matrices constructed from real data, and the second based on 180 MTMM matrices constructed from simulated data. In both studies, the number of traits, the number of methods, and sample sizes were systematically varied.

The general model used in the CFA approach to MTMM data typically fails to provide a well-defined solution. Both studies were designed to determine under what conditions the general model was most likely to result in a proper solution.

MTMM matrices such as those considered by Wothke (1984) typically suffer three important limitations:

1. The sample sizes in studies are typically small. The 21 MTMM matrices examined by Wothke, for example, were based on a median of 124 examinees. This is well below the 200 examinees recommended by Boomsma (1983). In addition, Boomsma's recommendation was based on fitting independent cluster solutions, so even larger sample sizes may be required for the general MTMM model.
2. Most studies consider small MTMM matrices. The ratio of the measured variables to latent factors was greater than 2 in only 1 of the 21 MTMM matrices considered by Wothke.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

3. The general MTMM model implicitly assumes that trait effects associated with a particular trait, and method effects associated with a particular method, are unidimensional. Although this assumption is probably reasonable for most trait factors, researchers typically have insufficient information about the underlying nature of method effects to determine if this assumption is reasonable for method factors.

Both studies 1 and 2 also were designed to examine the behavior of the correlated uniqueness model (Marsh, 1988, 1989a) as an alternative to the general models for MTMM data. Marsh claimed that this alternative model typically resulted in well-defined solutions even when the general model did not, and that this model provided a test of the unidimensionality of the method factors. Marsh provided empirical support for his claims, but the results were based on the examination of only three MTMM matrices. The present investigation was based on results of 435 MTMM matrices and examined a much wider variety of MTMM matrices.

There are many important advantages to working with real data; however, the true underlying factor structure and associated parameter values can never be known. An important advantage of using simulated data with known population parameters is that estimated parameter values can be compared to the known population values. Thus, Study 2 was designed to compare the ability of different models to correctly estimate known population parameters.

The final purpose of the present investigation was to evaluate the accuracy of rules of thumb about the minimum conditions needed to achieve stable well-defined solutions. Tests of these guidelines are typically based on simulated data constructed and tested with independent clusters structures in which each measured variable is associated with only one latent factor. This situation may be unrealistically simple in relation to real data and the models needed to fit that data, so generalizations based on such studies must be evaluated cautiously. If these guidelines do work for MTMM data, there is support for their generality. Otherwise, the MTMM results provide counter instances to these generalizations.

The focus of the present investigation was on the CFA approach to MTMM data, but the findings have broader implications as well. MTMM data are just one example of a two-facet test design in which one facet refers to traits and the other to methods. MTMM data are also a special case of the three-mode data considered by Tucker (1964), or the multimode data considered by McDonald (1985). Furthermore, method structure E from this study also resembles McDonald's multimode analysis and Browne's (1980) multiple battery analysis. Thus, the present findings have implications for the factor analysis of other two-facet test designs or multimode designs that may not be directly related to MTMM data.

STUDY 1: EVALUATION OF MTMM MODELS ON THE BASIS OF REAL DATA

Method

Data

Data for this study came from the Youth in Transition project (Bachman, 1970, 1975; Bachman & O'Malley, 1977). The project used a multistage probability sample, clustered by school, of all tenth grade boys in public high schools in 1966 in the 48 contiguous states of the U.S.A. Only variables common to four waves of data collection (1966—10th grade, 1968—11th grade, 1969—12th grade, and 1970—one year after graduation) were used. The 1966 data were based on responses from 2,213 boys attending 87 different high schools, and the sample sizes were 1,886, 1,799, and 1,620 for each successive wave. Responses were used from only the 1,556 boys who participated in all four data waves.

A set of 84 variables was selected, representing 21 variables (see Table 1) collected in each of the four waves. The major criteria of selection were that the variable was contained in the commercially available longitudinal database, that data were available for all four waves, and that there was no overlap in the items used to define

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 1
Number of Items per Scale (K), Scale Title, and a Sample Item for the Three Sets of Variables Used in Study 1

K		Scale Title and Sample Item
Dataset 1		
10	1.	Self Esteem: "I take a positive attitude toward myself."
5	2.	Political Knowledge Test: "Who are the two U.S. Senators from your state?"
7	3.	Honesty Values: "Is this a good thing for people to do? 'Telling a lie to spare someone's feelings.' "
3	4.	Trust in Government: "Do you feel that the people running the government are smart people who usually know what they are doing?"
13	5.	Job Ambitions: "Preferences for a job with good chances of getting ahead?"
7	6.	Interpersonal Aggression: "How many times have you 'Hurt someone badly enough to need bandages or a doctor'?"
7	7.	General Anxiety: "I worry about what other people think of me."
Dataset 2		
6	1.	Happiness: "I feel happy."
6	2.	Emotional Dependency: "I like to have people to lean on when things go badly for me."
7	3.	Irritability: "Even small things irritate me."
5	4.	Physical Development Value: "Is this a good thing for people to do? 'Being good in some form of sport.' "
12	5.	Internal Versus External Locus of Control: " 'What happens to me is my own doing' versus 'Sometimes I feel that I don't have enough control over the direction my life is taking.' "
5	6.	Stability of Self Esteem: "My opinion of myself tends to change a good deal instead of always remaining the same."
25	7.	Job Information Test: "A sociologist works with people more than a social worker."
Dataset 3		
6	1.	Level of Independence: "I go my own way in spite of what others think."
6	2.	Depression: "I feel depressed."
5	3.	Guilt: "I blame myself when things go wrong."
18	4.	Physical Health or Somatic Symptoms: "How often would you say you had a cold during the year?"
4	5.	Religiousness Value: "Being devout in one's religious faith."
9	6.	Theft/Vandalism: "Taken something not belonging to you worth over $50."
1	7.	Occupational Plans: "What sort of work do you think you might do for a living?"

the different variables. The variables were then divided into three sets of 28 variables representing seven variables (the multiple traits) and four data waves (the multiple methods). For each of the three sets of variables, a 28 × 28 covariance matrix was based on pairwise deletion for missing data after first selecting examinees (N = 1,556) who participated in all four data waves.

MTMM approaches have been used frequently with time as the method variable. Werts, Jöreskog, and Linn (1972; Werts & Linn, 1970) specifically proposed the MTMM approach for studying growth, and others (e.g., Browne, 1984, in press; Cudeck & Browne, 1983; Marsh & Butler, 1984; Marsh & Hocevar, 1988) have considered multiple occasions as the multiple methods. Browne (in press) also noted that the multiple methods in an MTMM study may refer to either the different ways used to measure traits or the multiple occasions of measurement, but proposed that both should be

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

referred to as methods.[1]

MTMM Matrices

The three sets of variables were then used to construct fifteen 28 × 28 covariance matrices based on sample sizes of 125, 250, 500, 1,000, and 1,556 (the entire sample). The examinees at each sample size were a proper subset of the examinees at each subsequent sample size, in order to maximize the similarity of examinees used at each sample size.

From each 28 × 28 covariance matrix, eight different 9 × 9 and eight different 16 × 16 covariance matrices were selected to represent 3T × 3M and 4T × 4M designs, in addition to the original 7T × 4M design. For the 3T × 3M and 4T × 4M matrices, 240 covariance matrices were constructed: 2 (size of matrix) × 3 (sets of variables) × 8 (subsets of variables) × 5 (sample sizes). For the 7T × 4M design, only 15 covariance matrices were constructed: 3 (sets of variables) × 5 (sample sizes). The general model (4D) and the correlated uniqueness model (4E) were fit to each of these 255 covariance matrices.

Results

Preliminary Analyses

Three 28 × 28 matrices (one for each of the three sets of variables) were fit to selected models from Marsh's (1988, 1989a) taxonomy that posited seven correlated trait factors (4A), four correlated method factors (1D), seven correlated trait factors and four uncorrelated method factors (4C), seven correlated traits and four correlated method factors (4D), and seven correlated trait factors with method variance represented by correlated uniquenesses (4E). All solutions were estimated by fixing factor variances to 1 (so that the trait factor loadings were standardized) and then estimating parameters for a subset of the models in the MTMM taxonomy.

The goodness-of-fit indicators shown in Table 2 indicate that the general model (4D), the correlated uniqueness model (4E), and the uncorrelated method factor model (4C) all provided apparently good fits to the data (all TLI > .94). The fit for the trait-only model (4A) was modestly poorer (TLI of .86 to .89), and the fit for the method-only model was very poor (TLI of .30 to .35). This implies that there was substantial trait variance in each set of variables and a smaller amount of method variance.

Solutions for models 4C, 4E, and 4D in Table 3 all fit the data fairly well. The three models differed in how they represent method effects, but all three provided comparable estimates of the unknown population trait factor loadings. The individual trait factor loadings in Table 3 indicate that the parameter estimates are remarkably consistent across different models applied to the same data. Trait factor loadings are marginally higher, however, for Models 4C (Mean = .682) and 4E (Mean = .681) than they are for Model 4D (Mean = .660) across all the trait factor loadings for each MTMM matrix. The size of trait factor loadings is typically interpreted as an indication of convergent validity, so support for convergent validity is marginally better for Models 4C and 4E than for Model 4D. Because the underlying population parameters are unknown, however, there is no way to determine which set of estimates is most accurate. The standard errors of the factor loadings were approximately .03 (.022 to .045).

[1]An implicit assumption in fitting models considered here to MTMM data is that covariation among the multiple indicators of a given trait can be represented as a unidimensional trait factor. When time is the method effect, however, the classic simplex pattern is a plausible alternative in which the multiple indicators of each trait are specifically proposed not to be unidimensional. In order to explore the unidimensional representation in the present investigation, a set of 21 4 × 4 covariance matrices was constructed in which the four measured variables were the indicators of the same trait on four occasions. A single factor model was fit to each of the 21 covariance matrices. The χ^2 test statistic associated with each of the 21 single-factor models was statistically significant, due in part to the large sample size. The TLI goodness of fit statistics varied from .974 to .999 (median = .996), suggesting that most of the covariation among multiple indicators of the same trait could be explained by a single trait factor. In contrast, the classic simplex model provided much poorer fit to the data. Thus, the assumption of MTMM models considered here—that the multiple indicators of each trait are unidimensional—appeared to be reasonable.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 2
Summary of Goodness of Fit in Study 1 for the Tucker-Lewis Index (TLI) and the Bentler-Bonnett Index (BBI)

Model	χ^2	*df*	χ^2/df Ratio	TLI	BBI
Dataset 1					
0[a]	12664.20	431	33.50	.000	.000
1D[b]	8145.31	344	23.68	.302	.357
4A	1579.25	329	4.80	.883	.875
4C[c]	831.91	301	2.76	.946	.934
4D	712.88	295	2.42	.956	.944
4E	652.25	245	2.66	.949	.949
Dataset 2					
0[a]	12681.29	378	33.55	.000	.000
1D	7648.86	344	22.24	.348	.400
4A	1773.59	329	5.39	.865	.860
4C	826.52	301	2.75	.946	.935
4D	730.79	295	2.48	.955	.942
4E	525.86	245	2.15	.965	.959
Dataset 3					
0[a]	13439.57	378	35.55	.000	.000
1D	8525.22	344	24.78	.312	.366
4A	1599.38	329	4.86	.888	.881
4C	711.62	301	2.36	.961	.947
4D	640.88	295	2.17	.966	.952
4E	487.30	245	1.99	.971	.964

[a]The null model posited 28 uncorrelated factors and was only used in computing TLI and BBI.
[b]This solution converged to an improper solution, with a factor correlation greater than 1.0.
[c]This solution converged to an improper solution, with a negative uniqueness.

Perhaps the most remarkable finding of these preliminary results is that solutions based on the general model (4D) are well-defined for all three covariance matrices. This is in marked contrast to findings such as those of Wothke (1984), who found that solutions based on the general model were ill-defined for all of the 21 MTMM matrices he considered. The three MTMM matrices considered here, however, differ substantially from those considered by Wothke in two ways: (1) the median number of variables in the MTMM matrices considered by Wothke was only nine (i.e., three traits and three methods), and none had more than 20 variables, as compared to the 28 variables considered here; and (2) the median sample size considered by Wothke was 124 and none was greater than 750, compared to the sample size of 1,556 considered here. Hence, the primary purpose of Study 1 was to determine if the size of the MTMM matrix and/or the sample size contributed systematically to the production of proper solutions for Model 4D.

Behavior of Models 4D and 4E as a Function of *N* and the Size of the MTMM Matrix

In Study 1, 255 MTMM matrices were analyzed that systematically varied the sample size (125, 250, 500, 1,000, and 1,556) and the number of variables (9 = 3T × 3M, 16 = 4T × 4M, and 28 = 7T × 4M). The 7T × 4M matrices for N = 1,556 were the same as those used in the preliminary analyses, and all other matrices were based on subsets of these data. The general model (4D) and the correlated uniqueness model (4E) were fit to each of these 255 covariance matrices. The general model was typically well-defined for the 7T × 4M design, but not for the 3T × 3M or the 4T × 4M designs (see Figure 1). A similar pattern was observed for different sample sizes, although smaller sample sizes produced more ill-defined solutions. In contrast to the general MTMM model (4D), the correlated uniqueness model (4E) was well defined for 252 of 255 matrices; for model 4E there was 100% convergence to proper solutions for all three designs with $N \geq 250$. For the 3T × 3M design, 3 of 24 solutions based on Model 4E were not well-defined at N = 125, but the remaining solutions based on Model 4E were all well-defined.

The finding that the general MTMM model was frequently ill-defined is consistent with previous research (e.g., Wothke, 1984). What is important to note, however, is that it was frequently well-defined, particularly when based on large MTMM matrices. The finding that solutions were more likely to be well-defined for larger sample sizes is also consistent with previous research. Thus, it appears that both large sample sizes and a large number of variables contribute positively to the production of proper solutions with Model 4D. More importantly, the data show that

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 3
Trait Factor Loadings For Models 4C, 4D, and 4E
(Decimal Points Omitted)

	Dataset 1			Dataset 2			Dataset 3		
Variable	4C	4E	4D	4C	4E	4D	4C	4E	4D
Factor 1									
T_1M_1	61	60	61	63	62	64	57	57	56
T_1M_2	74	73	75	76	75	77	75	75	73
T_1M_3	80	80	74	75	76	68	76	76	76
T_1M_4	72	72	65	68	69	61	76	76	76
Factor 2									
T_2M_1	68	68	68	56	55	55	57	57	47
T_2M_2	64	64	64	62	61	61	73	73	62
T_2M_3	70	70	70	62	63	62	78	78	76
T_2M_4	74	74	74	63	64	63	69	69	57
Factor 3									
T_3M_1	56	56	53	65	65	56	58	57	57
T_3M_2	74	75	72	78	78	66	64	64	63
T_3M_3	79	79	80	76	76	72	66	66	64
T_3M_4	65	64	65	69	69	64	62	62	57
Factor 4									
T_4M_1	51	52	49	50	48	51	60	60	56
T_4M_2	65	65	63	61	59	62	76	75	71
T_4M_3	68	67	67	73	75	71	77	77	68
T_4M_4	55	55	57	66	65	64	75	75	69
Factor 5									
T_5M_1	55	55	48	54	53	55	62	62	61
T_5M_2	68	69	63	74	73	75	74	74	73
T_5M_3	70	69	68	82	82	79	82	82	82
T_5M_4	61	61	61	81	81	78	72	73	72
Factor 6									
T_6M_1	53	54	53	52	52	53	60	60	59
T_6M_2	53	53	51	69	69	68	71	71	69
T_6M_3	53	53	52	76	76	70	66	66	64
T_6M_4	54	53	54	73	73	64	63	64	65
Factor 7									
T_7M_1	65	65	67	73	72	73	70	69	70
T_7M_2	77	77	80	85	84	85	80	80	80
T_7M_3	83	83	77	82	83	82	78	78	78
T_7M_4	72	72	67	80	81	80	69	68	69
Mean	657	656	642	694	693	671	695	694	668

Note. The standard errors of factor loadings were approximately .03 (.022 to .045).

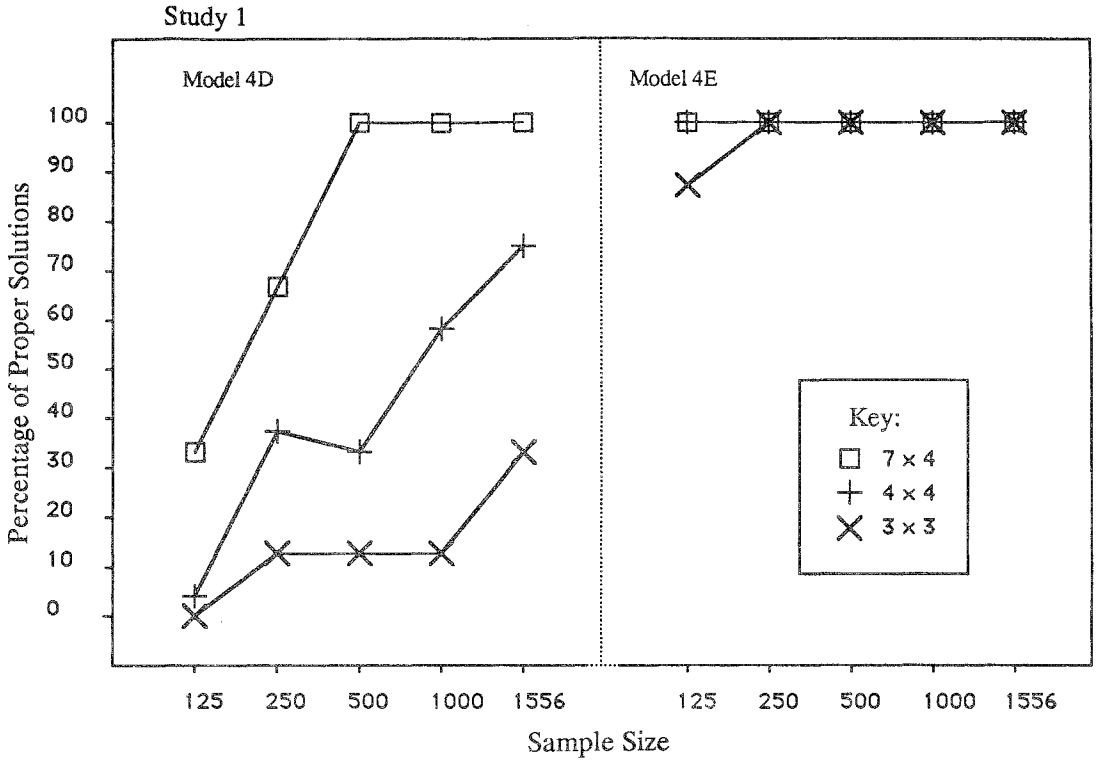
Model 4E results in proper solutions for almost all of the 255 MTMM matrices considered here. These findings provide strong support for Marsh's (1988, 1989a) claims about the usefulness of Model 4E.

STUDY 2: EVALUATION OF MTMM MODELS ON THE BASIS OF SIMULATED DATA

Because Study 1 was based on real data, it was not possible to compare the accuracy of parameter estimates from different MTMM models to known population parameters. Thus, Study 2 was designed to construct MTMM matrices with known population parameters from simulated data, and to compare the ability of different models to accurately estimate the known population parameters.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Figure 1
Convergence Behavior in Study 1 as a Function of the MTMM Model (4D or 4E), the Size of the MTMM Matrix (7T × 4M, 4T × 4M, or 3T × 3M), and Sample Size (125, 250, 500, 1,000, or 1,556)



Method

180 MTMM matrices were simulated to represent three sizes of matrices (3T × 3M, 4T × 4M, and 6T × 6M), three *N*s (160, 400, and 1,000), and two types of data—Type D and Type E—based on the method structures defined above. Each of the 18 = 3 × 3 × 2 combinations was represented by 10 different MTMM matrices. Models 4C, 4D, and 4E were fit to each of the 180 MTMM matrices, and results of these 540 solutions were evaluated.

All 180 MTMM matrices were based on six known population matrices representing the three sizes of MTMM matrices and the two types of data. Variables for individual cases were generated on the basis of known population parameters and a normally distributed random error component (details of the construction of the MTMM matrices are presented in the Appendix). For each of the six population matrices, 15,600 cases were generated to construct 30 matrices—10 for each of the three sample sizes.

All 180 MTMM matrices were generated from population models in which trait factor loadings varied from .3 to .8, and correlations among the traits varied from .25 to .49. Type D and Type E data differed in the construction of method effects. For Type D data, consistent with assumptions of method structure D, the method effects associated with variables having the same method were unidimensional (i.e., they could be accounted for by a single method factor), and the different method factors were correlated. The method factor loadings for Type D data varied from .25 to .50, and method factor correlations varied from .25 to .49.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

For Type E data, consistent with the assumptions of method structure E, method effects were represented as correlations among the uniquenesses for variables having the same method. For the 4T × 4M and the 6T × 6M matrices with the Type E data, two underlying method factors were associated with each method. For example, in the 6T × 6M design, there were nonzero correlations among uniquenesses associated with the first three traits having the same method, and among uniquenesses associated with the remaining three traits having the same method, but there were 0 correlations between the uniquenesses associated with these two subsets of traits. In this respect, Type E data did not conform with the assumption that method effects associated with a given method were unidimensional. For the 3T × 3M matrices constructed with Type E data,[2] uniquenesses associated with two of the traits having the same method were correlated with each other but uncorrelated with the uniqueness for the third trait.

Results

Behavior of the Different Solutions

52% of the 540 solutions converged to a proper solution; for model 4E there was 100% convergence to proper solutions for the 6 × 6 and 4 × 4 designs at all sample sizes. Figure 2 shows that this behavior varied substantially with the design of the MTMM matrix and the model used to fit the data. In order to supplement Figure 2, a 3 (sample size) × 3 (size of MTMM matrix) × 2 (data type) × 3 (model) ANOVA was used to estimate the sizes of the different effects. Although the nominal tests of statistical significance are not valid, due to the heterogeneity of variances and the fact that the dependent variable was dichotomous (i.e., the solution was proper or not), the variance explained by different effects provides a useful descriptive summary.

Most of the variance in the proportion of proper solutions (η^2 = .71) could be explained by the various effects. The largest effect (η^2 = .41) was for the model. Model 4E resulted in proper solutions for 96% of the matrices, whereas Models 4C and 4D resulted in proper solutions for only 33% and 22% of the matrices, respectively.

The next largest effect (η^2 = .10) was due to data type, although this effect depended on the model (η^2 = .05). For Type E data, Models 4C and 4D nearly always resulted in improper solutions, whereas Model E almost always resulted in proper solutions. For Type D data, Model 4E again resulted in well-defined solutions most of the time, whereas Models 4C and 4D provided more proper solutions than for the Type E data.

There were small effects of sample size (η^2 = .02) and the size of the MTMM matrix (η^2 = .02), but these interacted with data-type by model-type effect (both η^2s = .01). Solutions were more likely to be proper when the sample size and the size of the MTMM matrix were larger, but this occurred primarily for the Type D data when fit with Models 4C and 4D. The only other effect to account for more than 1% of the variance was the data type by sample size interaction (η^2 = .02), but this effect interacted with model type, as already discussed.

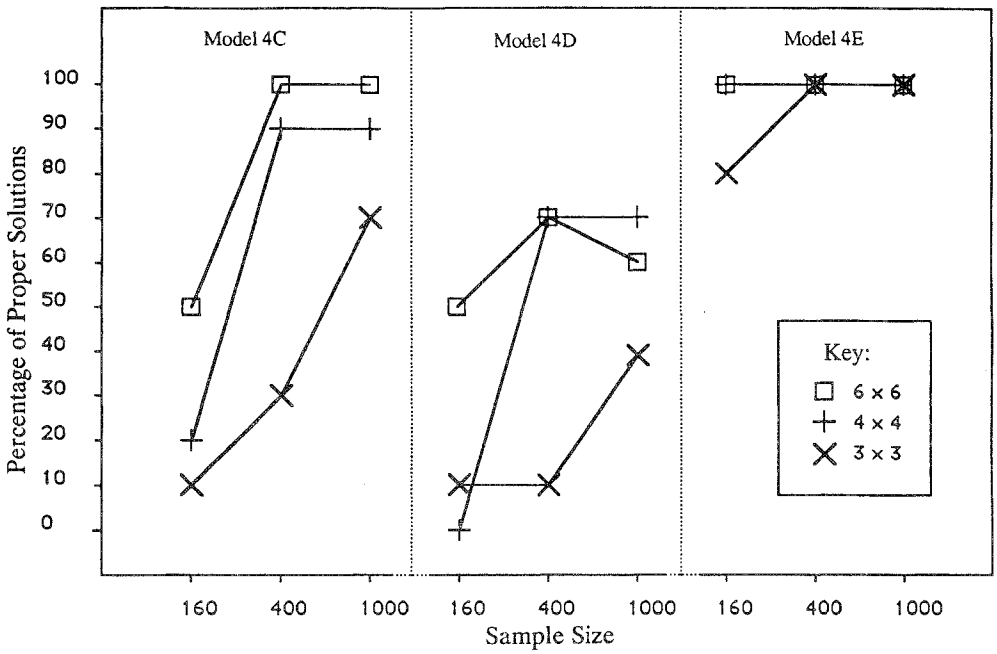
Consistent with the results of Study 1, Model 4E almost always resulted in proper solutions, whereas Models 4C and 4D frequently did not. In Study 1, Model 4D was most likely to result in proper solutions when the sample size and size of the MTMM matrix were large. A similar effect was observed here for Models 4C and 4D for the Type D data, but not for the Type E data. The structure of method effects in the Type E data was not consistent with the unidimensional

[2]Technically, the 3T × 3M matrices constructed with Type E data are consistent with the assumption of unidimensional method effects associated with each method of assessment underlying Models 4C and 4D. As noted earlier, as long as the solutions are well-defined, Models 4E and 4C are merely alternative parameterizations of the same model for the 3T × 3M design. The assumption of the unidimensionality of method effects is only testable if the number of traits is greater than 3. Logically, this should bias the results in favor of Model 4D and against Model 4E. In practice, however, Model 4D rarely resulted in a proper solution for 3T × 3M matrices in either study, so that this potential bias was of no practical consequence.

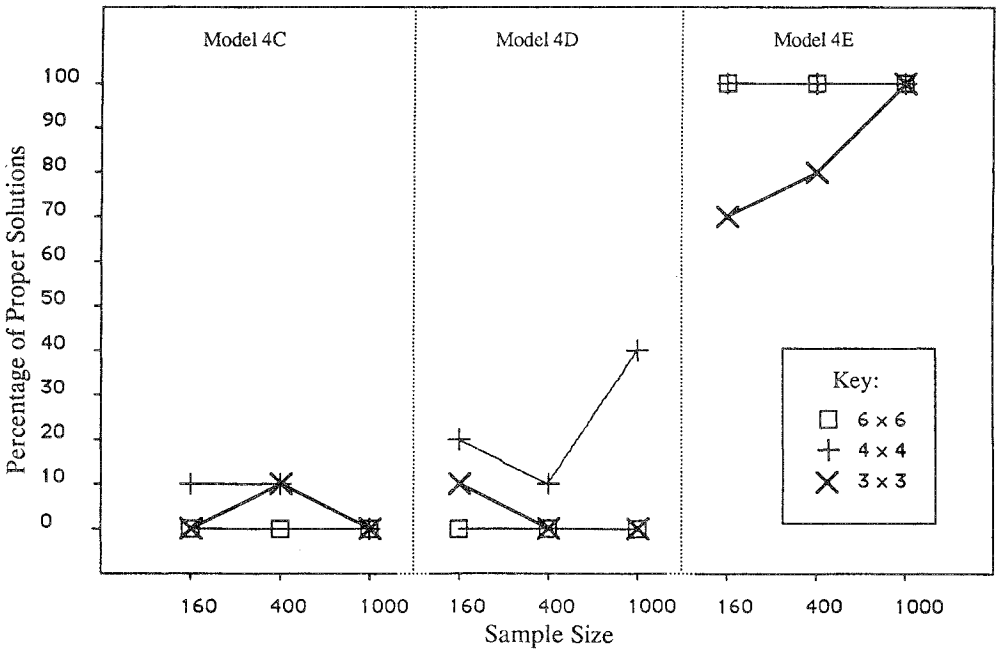
Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Figure 2
Convergence Behavior in Study 2 as a Function of the MTMM Model (4C, 4D, or 4E), the Data Type (4D or 4E), the Size of the MTMM Matrix (6T × 6M, 4T × 4M, or 3T × 3M), and Sample Size (160, 400, or 1,000)

(a) Type D Data



(b) Type E Data



Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

method factor posited in Models 4C and 4D, which apparently accounts for why these models so frequently resulted in improper solutions. The unidimensionality of method effects in Type D was, however, only a special case of the more general method effects posited by Model 4E, and apparently accounts for why Model 4E consistently resulted in proper solutions for both types of data.

Accuracy of Parameter Estimates

An important feature of Study 2 was that the population parameters for the simulated data were known, which allowed for evaluation of the accuracy of the parameter estimates. For MTMM data, the critical parameter estimates are the trait factor loadings and correlations among the trait factors. Models 4C, 4D, and 4E differ substantially in how they account for method effects, but they provide comparable estimates of the population trait factor loadings and the trait factor correlations.

Trait factor loadings. The deviation between each estimated factor loading and the corresponding (known) population parameter was computed, and the mean deviation across all the factor loadings within each solution was obtained. The number of deviations averaged in each solution was 9, 16, and 36 for the 3T × 3M, 4T × 4M, and 6T × 6M matrices, respectively.

Deviations between estimated and known population trait factor loadings (i.e., mean errors in Table 4) are summarized in Table 4 for each cell of the design, according to whether or not the solution was proper. In order to supplement the description of these results, a 3 (sample size) × 3 (size of MTMM matrix) × 2 (data type) × 3 (model) ANOVA was conducted on the average deviations for the 279 proper solutions. The cells of the design varied according to the size of the MTMM matrix (3 × 3, 4 × 4, or 6 × 6), the type of data (D or E), the model (4C, 4D, or 4E), and the sample size (160, 400, or 1,000).

The main purpose of this analysis was to provide a convenient summary of the sizes of the different effects, although the tests of statistical significance are probably reasonable. The substantial and systematic differences in the occurrence of proper solutions described previously, however, complicate these analyses. The design was balanced in relation to the manipulated characteristics of the study, but it was not balanced in relation to proper and improper solutions. Each of the 54 cells was represented by 10 solutions, but 13 cells had no proper solutions, and another eight cells had only one proper solution. Because the design was not balanced for analyses of only the proper solutions, only main effects and two-way interactions were examined.

In addition, unadjusted marginal values (e.g., the average deviation for Model 4D averaged across all levels of the other variables in the design) did not equal the marginal means adjusted for the other variables in the design. Although there is no fully satisfactory solution to this problem, the adjusted and unadjusted marginal values, shown in Table 4, generally do not vary substantially, and they result in similar interpretations.

For the 279 proper solutions, the average deviation was .017—that is, the average of estimated trait factor loading was .017 larger than the average of the known population trait factor loadings.[3] More than half of the variance in these average deviation scores was explained by the design characteristics (η^2 = .54) of the study. Most of this variance was explained by the main effects of data type (η^2 = .22) and model (η^2 = .27). The interaction of these two effects was also statistically significant, but it accounted for much less variance (η^2 = .02). No other effects were statistically significant.

The average deviations (mean errors in Table 4) were positive for Models 4C (.035) and 4E (.02), but negative for Model 4D (–.015). Deviations were close to 0 for Type E data (–.002), but

[3]A preliminary inspection of the factor loadings and factor correlations for proper solutions indicated that the size of deviations within each solution did not vary systematically, depending on the particular factor loading or factor correlation that was estimated.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 4

Average Deviation Between the Parameter Estimate and the Known Population Parameter (Mean Error), Within-Solution Standard Deviation (WSSD), Standard Deviation of the Mean Deviation Scores Within Each Cell (SD), and Number of Solutions for Trait Factor Loadings and Trait Factor Correlations, as a Function of the MTMM Matrix Type (Size of Matrix and Data Type), MTMM Model Used to Fit the Data, *N* the Sample Size, and Whether the Solution Was Proper or Improper

		Proper Solutions							Improper Solutions						
			Loadings			Correlations				Loadings			Correlations		
Model	*N*	No.	Mean Error	WSSD	SD	Mean Error	WSSD	SD	No.	Mean Error	WSSD	SD	Mean Error	WSSD	SD
3 × 3 Matrix, Type D															
4C	160	1	.045	.086	–	.095	.177	–	9	.053	.126	.030	.109	.120	.115
4C	400	2	.047	.073	.026	.049	.104	.082	8	.039	.086	.018	.115	.089	.055
4C	1000	5	.028	.068	.016	.063	.077	.019	5	.045	.069	.013	.096	.081	.049
4D	160	1	.063	.086	–	.133	.159	–	9	.084	.515	.112	–.064	.168	.153
4D	400	0	–	–	–	–	–	–	10	–.002	.129	.075	–.083	.205	.466
4D	1000	5	–.007	.094	.025	.001	.099	.081	5	.001	.107	.021	–.097	.107	.064
4E	160	9	.049	.122	.029	.124	.135	.102	1	.074	.116	–	–.031	.053	–
4E	400	10	.041	.083	.018	.097	.090	.063	0	–	–	–	–	–	–
4E	1000	10	.037	.069	.016	.079	.090	.041	0	–	–	–	–	–	–
3 × 3 Matrix, Type E															
4C	160	0	–	–	–	–	–	–	10	.013	.160	.037	.012	.131	.015
4C	400	1	–.012	.083	–	–.171	.148	–	9	.006	.091	.015	–.312	.122	.043
4C	1000	0	–	–	–	–	–	–	10	–.001	.081	.017	.020	.094	.048
4D	160	1	–.070	.200	–	.304	.269	–	9	.126	.625	.130	–.129	.123	.099
4D	400	2	–.029	.200	.026	–.091	.133	.113	8	.049	.354	.083	–.079	.130	.087
4D	1000	0	–	–	–	–	–	–	10	–.006	.088	.026	.137	.103	.050
4E	160	7	.002	.117	.033	.079	.160	.146	3	.040	.261	.036	–.105	.109	.127
4E	400	10	.000	.090	.017	–.048	.113	.067	0	–	–	–	–	–	–
4E	1000	10	–.002	.081	.018	.182	.094	.046	0	–	–	–	–	–	–
4 × 4 Matrix, Type D															
4C	160	3	.025	.065	.015	.119	.083	.026	7	.028	.092	.030	.053	.060	.069
4C	400	9	.035	.051	.009	.073	.054	.039	1	.051	.038	–	.118	.044	–
4C	1000	10	.044	.043	.011	.094	.047	.024	0	–	–	–	–	–	–
4D	160	1	–.049	.176	–	–.252	.305	–	9	.071	.437	.099	.024	.141	.196
4D	400	7	–.016	.079	.039	–.095	.096	.014	3	.082	.060	.026	.191	.056	.056
4D	1000	7	–.003	.070	.025	–.009	.069	.074	3	–.217	.118	.051	–1.83	1.334	.701
4E	160	10	.028	.083	.026	.075	.067	.066	0	–	–	–	–	–	–
4E	400	10	.036	.050	.010	.079	.055	.041	0	–	–	–	–	–	–
4E	1000	10	.045	.043	.011	.093	.047	.025	0	–	–	–	–	–	–
4 × 4 Matrix, Type E															
4C	160	1	–.008	.046	–	–.062	.136	–	9	.010	.129	.026	–.024	.054	.091
4C	400	1	–.004	.037	–	.092	.141	–	9	.031	.054	.010	–.006	.070	.030
4C	1000	0	–	–	–	–	–	–	10	–.005	.046	.011	–.004	.060	.046
4D	160	0	–	–	–	–	–	–	10	.101	.435	.148	–.061	.067	.133
4D	400	1	–.010	.070	–	.012	.097	–	9	.002	.062	.014	–.012	.083	.044
4D	1000	0	–	–	–	–	–	–	10	–.007	.050	.012	–.012	.061	.049
4E	160	8	.003	.076	.019	.001	.063	.079	2	.025	.291	.046	–.106	.039	.070
4E	400	10	–.001	.052	.010	.000	.078	.040	0	–	–	–	–	–	–
4E	1000	10	–.006	.063	.011	–.006	.060	.046	0	–	–	–	–	–	–
6 × 6 Matrix, Type D															
4C	160	7	.030	.040	.014	.096	.048	.049	3	.043	.046	.011	.098	.031	.042
4C	400	9	.035	.033	.009	.076	.032	.031	1	.042	.029	–	.137	.059	–
4C	1000	10	.046	.028	.006	.096	.037	.165	0	–	–	–	–	–	–
4D	160	4	.018	.081	.037	.040	.055	.119	6	–.002	.253	.209	–2.694	.776	6.710
4D	400	7	–.054	.056	.079	–.023	.075	.287	3	.044	.456	.073	.127	.057	.127
4D	1000	6	–.018	.042	.034	–.064	.049	.097	4	.099	.435	.109	–.886	.055	.459
4E	160	10	.034	.042	.014	.096	.043	.046	0	–	–	–	–	–	–

continued on the next page

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 4, continued

Average Deviation Between the Parameter Estimate and the Known Population Parameter (Mean Error), Within-Solution Standard Deviation (WSSD), Standard Deviation of the Mean Deviation Scores Within Each Cell (SD), and Number of Solutions for Trait Factor Loadings and Trait Factor Correlations, as a Function of the MTMM Matrix Type (Size of Matrix and Data Type), MTMM Model Used to Fit the Data, *N* the Sample Size, and Whether the Solution Was Proper or Improper

		Proper Solutions							Improper Solutions						
			Loadings			Correlations				Loadings			Correlations		
Model	*N*	No.	Mean Error	WSSD	SD	Mean Error	WSSD	SD	No.	Mean Error	WSSD	SD	Mean Error	WSSD	SD
4E	400	10	.036	.032	.009	.082	.034	.035	0	–	–	–	–	–	–
4E	1000	10	.046	.028	.006	.095	.031	.167	0	–	–	–	–	–	–
6 × 6 Matrix, Type E															
4C	160	0	–	–	–	–	–	–	10	.001	.056	.012	–.017	.061	.051
4C	400	0	–	–	–	–	–	–	10	.008	.032	.006	.004	.043	.027
4C	1000	0	–	–	–	–	–	–	10	–.001	.030	.005	.003	.040	.014
4D	160	0	–	–	–	–	–	–	10	.044	.425	.240	–.095	.106	.015
4D	400	4	.000	.048	.011	.012	.043	.020	6	.057	.034	.003	–.160	.515	.026
4D	1000	0	–	–	–	–	–	–	10	–.001	.032	.005	.001	.405	.015
4E	160	10	.001	.057	.012	–.006	.068	.052	0	–	–	–	–	–	–
4E	400	10	.005	.032	.007	.000	.041	.026	0	–	–	–	–	–	–
4E	1000	10	–.002	.030	.004	.000	.039	.014	0	–	–	–	–	–	–
Marginal Values: Matrix Size															
3T × 3M Matrix															
		73	.020	.078	.033	.065	.089	.100	177	.049	.290	.117	–.172	.138	1.588
Adjusted[a]			.017	.081		.059	.092								
4T × 4M Matrix															
		103	.013	.058	.036	.010	.069	.126	77	.018	.098	.045	.001	.097	.182
Adjusted			.015	.058		.017	.068								
6T × 6M Matrix															
		103	.020	.051	.029	.044	.058	.067	77	–.007	.061	.051	–.067	.118	.389
Adjusted			.019	.049		.041	.056								
Marginal Values: Data Type															
Type D Data															
		183	.027	.058	.034	.055	.064	.110	87	.030	.182	.095	–.220	.207	1.802
Adjusted			.029	.059		.060	.064								
Type E Data		96	–.002	.066	.017	.002	.081	.075	174	.019	.158	.082	–.025	.080	.086
Adjusted			.005	.064		–.006	.081								
Marginal Values: Model Type															
Model 4C															
		59	.035	.045	.016	.078	.056	.051	121	.015	.079	.026	.024	.077	.083
Adjusted			.025	.054		.062	.070								
Model 4D															
		46	–.015	.078	.045	–.055	.086	.170	134	.030	.241	.118	–.193	.161	1.448
Adjusted			–.021	.086		–.061	.097								
Model 4E															
		174	.020	.061	.025	.047	.073	.074	6	.040	.247	.034	–.093	.076	.091
Adjusted			.025	.057		.054	.063								
Marginal Values: Sample Size															
N = 160		74	.019	.094	.032	.050	.111	.100	106	.034	.208	.072	–.010	.124	.017
Adjusted			.019	.095		.050	.112								
N = 400		98	.017	.059	.028	.035	.068	.085	82	.019	.162	.085	–.068	.119	.377
Adjusted			.017	.060		.033	.069								
N = 1,000		107	.017	.039	.035	.030	.044	.117	73	.012	.109	.106	.232	.115	1.922
Adjusted			.017	.038		.031	.042								
Total		279	.017	.061	.032	.037	.070	.102	261	.023	.166	.087	–.090	.120	.104

[a]For the proper solutions, the multiple classification option of the SPSSX ANOVA procedure (SPSS, 1986) was used to adjust each marginal mean for all the other independent variables.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

positive for the Type D data (.027). For Type D data, deviations were slightly positive for Models 4C and 4E, but slightly negative for Model 4D. For Type E data, deviations were almost exactly 0 for Model 4E. (So few solutions converged properly with Models 4C and 4D that generalizations may be unwarranted.) Across all proper solutions, the marginal values show that trait factor loadings were very accurate. For proper solutions with Type D data, there is apparently little difference in the accuracy of Models 4C, 4D, and 4E, although Models 4C and 4E slightly overestimated trait factor loadings and Model 4D slightly underestimated those loadings.

WSSD in Table 4 is based on the standard deviation of the deviation scores within each solution. The number of deviations contributing to each within-solution standard deviation was 9, 16, and 36 for the 3T × 3M, 4T × 4M, and 6T × 6M matrices, respectively. The mean WSSD across all 279 proper solutions was .061. Most of the variation in these WSSDs is explained by design characteristics ($\eta^2 = .73$) of the study. Not surprisingly, a large proportion of this explained variance was due to differences in sample size ($\eta^2 = .46$). Similarly, because the average deviation is based on more deviations for the larger MTMM matrices, the size of the MTMM matrix ($\eta^2 = .15$) also accounted for much of the variance. The model used to fit the data also had a substantial effect ($\eta^2 = .11$). The WSSDs were smaller when the sample size and the size of the MTMM matrix was larger, and were smaller for Models 4C and 4E than for Model 4D.

Evaluation of the trait factor loadings focused on the proper solutions, but it was also informative to consider the improper solutions in this special situation in which the true population parameters were known. The average deviation between the estimated trait factor loading and the known population trait factor loading was .023 for the 261 improper solutions, compared to .017 for the 279 proper solutions. This difference was not statistically significant, although the systematic relation between the proper solutions and design characteristics meant that this statistical test should be interpreted cautiously. The cells of Table 4 in which there are at least three proper and three improper solutions suggest that trait factor loadings are reasonably similar for proper and improper solutions.

For Model 4C, trait factor loadings are slightly more accurate for the improper solutions than the proper solutions, although the estimates for improper solutions are somewhat more variable. For Model 4D, improper solutions provide less accurate and much more variable estimates than proper solutions. (There are too few improper solutions for Model 4E to warrant generalizations.) Parameter estimates based on improper solutions should be evaluated with extreme caution; yet the trait factor loadings based on Model 4C for these improper solutions were reasonably accurate and differed little from those based on proper solutions in the present investigation. It should be emphasized, however, that substantive interpretations of a single covariance matrix based on real data are typically unwarranted if the solution is not proper.

Trait factor correlations. The deviation between each estimated trait factor correlation and the corresponding (known) population parameter was computed, and the mean deviation across all correlations within each solution was obtained. For each solution, the number of deviations being averaged was 3, 6, and 15 for the 3T × 3M, 4T × 4M, and 6T × 6M matrices, respectively.

For the 279 proper solutions, the average deviation was .037—that is, the average of estimated trait factor correlations was .037 larger than the average of the known population trait factor correlations.[3] ANOVAs similar to those used to summarize trait factor loadings were conducted. Nearly half of the variance in the deviations could be explained in terms of the design characteristics ($\eta^2 = .47$). Six effects were statistically significant: the main effects of model type ($\eta^2 = .18$), data type ($\eta^2 = .08$), and matrix size ($\eta^2 = .03$), and two-way interactions between data type and model type ($\eta^2 = .10$), between model type and matrix size ($\eta^2 = .06$), and between matrix size and sample size ($\eta^2 = .02$).

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Table 4 indicates that the average deviations were positive for Models 4C (.078) and 4E (.047), but negative for Model 4D (-.055). Deviations were .002 for Type E data, and .055 for the Type D data. For Type D data, Models 4C and 4E overestimated the true population correlations, whereas Model 4D underestimated them. For Type E data, the average deviation between estimated and true population correlations was .002 for Model 4E. (Models 4C and 4D converged to proper solutions so infrequently for Type E data that generalizations may not be warranted.) Across all combinations of the MTMM matrices, the marginal values for proper solutions indicate that trait-factor correlations were reasonably accurate. For Type D data, there was apparently little difference in the accuracy of Models 4C, 4D, and 4E, although Model 4D slightly underestimated the true population values, and Models 4C and 4E slightly overestimated the true population values.

The mean WSSD across all 279 proper solutions was .07. Half of the variation in these WSSDs was explained by design characteristics ($\eta^2 = .50$) of the study. A large proportion of this explained variance was due to differences in sample size ($\eta^2 = .31$) and the size of the MTMM matrix ($\eta^2 = .08$). The only other variable that had a substantial effect was the model ($\eta^2 = .06$). The WSSD was smaller when the sample size and the size of the MTMM matrix were larger, and was larger for Model 4D than for Models 4C and 4E.

A comparison of the trait factor correlations for the proper and improper solutions may be unwarranted. Many of the improper solutions were the result of factor correlation matrices that were not positive definite, and estimated factor correlations greater than 1. However, estimates based on Model 4C were slightly more accurate for improper solutions than for proper solutions, even though estimates based on improper solutions were slightly more variable. In contrast, estimates based on improper solutions with Model 4D were grossly inaccurate and had very large standard deviations. (There were so few improper solutions based on Model 4E that generalizations are probably unwarranted.) Again, it should be emphasized that substantive interpretations of a single covariance matrix based on real data are typically unwarranted if the solution is not proper.

Summary

In support of Marsh's (1988, 1989a) claims, Model 4E consistently resulted in proper solutions even when Models 4C and 4D did not. For Type D data, Models 4C and 4D frequently resulted in proper solutions when the size of the MTMM matrix and the sample size were large, but not when the size of the MTMM matrix and sample size were small. These results are consistent with findings in Study 1. For Type E data, Models 4C and 4D performed poorly for all sample sizes and all matrix sizes.

Across all the different MTMM matrices, parameter estimates were reasonably accurate for trait factor loadings and trait factor correlations. For Type E data, factor loadings and factor correlations were almost exactly equal to the population values, although nearly all these proper solutions were based on Model 4E. For Type D data, however, Models 4C and 4E tended to overestimate the known factor loadings and factor correlations, whereas Model 4D tended to underestimate these known population parameters.

The tendency for Models 4C and 4E to overestimate trait factor loadings and, in particular, trait factor correlations for Type D data, may represent a systematic bias. For Type D data, each measured variable has a component due to a method factor, and the different method factors are correlated. Because Models 4C and 4E do not allow for correlations among effects associated with different methods, these correlated method effects apparently are represented as general trait variance.

Results suggest that this occurrence has a very small positive bias on estimates of trait factor loadings, and a somewhat larger positive bias on the size of trait factor correlations. In the interpretation of MTMM matrices, higher trait factor loadings are interpreted to mean better

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

convergent validity, whereas higher trait factor correlations are interpreted to mean less distinctive traits or—when trait correlations are sufficient large—a lack of discriminant validity. Thus, the implications of this possible bias in the application of Model 4E to Type D data is to slightly increase support for convergent validity, but to slightly underestimate support for divergent validity.

DISCUSSION

Despite the growing enthusiasm for the CFA approach to MTMM data, problems demonstrated in previous research (e.g., Marsh, 1988, 1989a; Wothke, 1984) call into question its value. In particular, the CFA approach to MTMM data is plagued by technical difficulties in the estimation of parameter values. The most important criticism, perhaps, was Wothke's finding that the general MTMM model resulted in ill-defined solutions for all 21 previously published MTMM matrices that he considered.

Marsh (1988, 1989a) described Model 4E in which method effects are represented as correlated uniquenesses. He claimed that Model 4E has two important advantages over Models 4C and 4D. First, it was relatively immune to the ill-defined solutions that plagued other models. Second, when there were more than three traits, Model 4E 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 unidimensional method factor. When method effects associated with a particular method are not unidimensional, Model 4E provides a better representation of the data.

The results of this research provide strong support for both claims. The first claim is supported by the fact that across all 435 MTMM matrices in Studies 1 and 2, Model 4E resulted in proper solutions for 98% of the matrices, and Model 4D resulted in proper solutions for only 23% of the matrices.

The second claim for the advantages of Model 4E is more difficult to test. In Study 1, Model 4E fit the data as well or better than Model 4D, even though Models 4D and 4E both fit the data better than Model 4C. Model 4C posits uncorrelated, unidimensional method effects, Model 4D posits correlated, unidimensional method effects, and Model 4E posits uncorrelated method effects that are not constrained to be unidimensional. Logically, the pattern implies that the underlying method effects were correlated and not unidimensional, but that the advantage of not constraining method effects to be unidimensional (Model 4E) was larger than that of allowing the method effects to be correlated (Model 4D). This suggests indirect support for the second claim.

The claim was tested more directly in Study 2. For MTMM matrices constructed from Type E data in which the method effects were not unidimensional, solutions based on Models 4C and 4D were almost always improper, whereas solutions based on 4E were almost always proper. For this Type E data, trait factor loadings and trait factor correlations averaged across all MTMM matrices were almost exactly equal to the known population parameters. Taken together, the two studies provide good support for the second claim.

Marsh (1989a) also suggested that if there are correlated traits and correlated methods, some of what appears to be method effects in Model 4D may reflect trait effects. This claim was difficult to substantiate based on real data. His approach was to include additional variables in the MTMM design that were intended to test the validity of the traditional interpretations of trait and method effects. Results of both Studies 1 and 2 are consistent with the suggestion, even though it was not the purpose of the present investigation to test this suggestion. In Study 1, trait factor loadings based on Model 4D were smaller than those based on Model 4C and 4E. Although this finding was consistent with Marsh's suggestion, the use of real data meant that there was no way to test whether Model 4D underrepresented the size of trait effects. In Study 2, however, population parameters were known, and trait factor loadings based on Model 4D were systematically lower

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

than the known population parameter estimates. These results do support Marsh's suggestions, but the parameter estimates based on Model 4D and 4E for proper solutions did not differ enough to be substantively important in the interpretation of the MTMM matrices.

Model 4E, like Model 4C, assumes that the effects associated with each method in the MTMM design are independent. When this assumption is true, as in Type E data in Study 2, Model 4E provides very accurate estimates of trait factor loadings and trait factor correlations. Of more concern, however, is what happens when this assumption is not true, as in Type D data in Study 2. When MTMM data based on Type D data were fit with Models 4C and 4E, the estimated factor loadings and factor correlations were systematically larger—by a small amount—than the known population parameters. In the typical interpretation of the CFA approach to MTMM data, larger trait factor loadings imply better convergent validity, whereas larger trait factor correlations imply poorer divergent validity. In this respect, Model 4E may be biased toward showing stronger convergent validity but weaker divergent validity. Both of these potential biases are important, but the primary purpose of the MTMM design is to test divergent validity, and the bias for Model 4E is apparently conservative for this purpose. In evaluating the practical implications of these apparent biases, however, it is important to note that the sizes of these biases in the present investigation were trivial.

An important direction for further research is to test the generality of empirical findings based on the datasets described here. The generality of the present findings is substantially strengthened, however, by their consistency across a large number of different real and simulated MTMM matrices. To further test the generality of findings in Study 2, some additional datasets were constructed from Type D data in which correlations among traits and methods were much higher (.60 to .75) than those considered here. These additional datasets produced results similar to those reported here for both the likelihood of convergence and the accuracy of parameter estimates. Hence, there appears to be good support for the generality of the findings reported here.

The Comparison of Models 4D and 4E

Models 4D and 4E were compared according to the frequency of proper solutions, the systematic bias in the estimates, and the precision of estimates. Results indicate that Model 4E is preferable to Model 4D on all three criteria. Results based on both real and simulated data demonstrated that Model 4E almost always resulted in proper solutions for all sizes of MTMM matrices, and Model 4D typically did not, except for matrices based on a large number of variables and large sample sizes.

Model 4D was so likely to result in improper solutions that its usefulness appears to be dubious. In contrast, Model 4E generally resulted in proper solutions for all sizes of MTMM matrices and all sample sizes considered here. Model 4E may also be preferred according to the second and third criteria. Study 1, based on real data and unknown population parameters, provided a weak basis for evaluating these criteria. Even in Study 2, the evaluation of these criteria was complicated by the large number of improper solutions resulting from Model 4D. If these improper solutions are considered, then parameter estimates based on Model 4D—particularly trait factor correlation estimates—are much less accurate and much less precise than those based on Model 4E. For this reason, evaluation of these criteria is limited to proper solutions.

In Study 2, half the MTMM matrices were constructed with Type E data that were consistent with the assumptions underlying Model 4E, whereas half were constructed with Type D data that were consistent with assumptions underlying Model 4D. Not surprisingly, when Model 4E was fit to MTMM matrices constructed from Type E data, the average deviation between known population parameters and estimated trait parameters was close to 0 and had a small standard deviation. Estimates based on Model 4D were less accurate and less precise, but resulted

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

in so few proper solutions (8 of 90 solutions) that generalizations should be made cautiously.

Surprisingly, Model 4E performed about as well as Model 4D on MTMM matrices constructed from Type D data. Considering only proper solutions, Models 4D and 4E were both slightly biased; Model 4D underestimated trait factor loadings and trait factor correlations to about the same extent as Model 4E overestimated these parameters. Estimates based on Model 4E were, however, much more precise. Coupled with the large number of improper solutions for Model 4D, even the results based on Type D data appear to support Model 4E.

Logically, Model 4E should be preferred when method effects associated with the same method are not unidimensional, and the empirical results clearly supported this expectation. Logically, Model 4D should also be preferred when method effects associated with different methods are correlated, but surprisingly, the empirical results failed to support this expectation. Although these results cast doubt on the general usefulness of Model 4D, a potentially important limitation to Model 4E is its inability to incorporate correlations between method effects associated with different methods.

Recognizing this limitation, Marsh (1989a) demonstrated that such correlated methods can be represented in Model 4E by selectively freeing additional correlated uniquenesses—a model he called Model 4E′. As long as the CFA model is still identified, freeing enough additional correlated uniquenesses could account for correlated methods. Although there may be an a priori basis for selecting the correlated uniquenesses to be freed, this would typically be based on empirical decisions using the same data that are used to test the model. Such an ad hoc approach based on post hoc analyses is admittedly weak. Fortunately, the results from the present investigation suggest that parameter estimates based on Model 4E are reasonably accurate, even when the assumption of uncorrelated method effects is violated and this ad hoc approach is not employed.

It has been suggested that the primary justification of Model 4E is that it makes an available analytic tool work better, but that it is not based on substantive considerations, and that correlated uniquenesses do not have meaning in the "real world." Model 4E does work better with the CFA approach than other models considered here, but correlated uniquenesses do have substantive interpretations: In model 4E, each correlated uniqueness represents the correlation between two variables sharing the same method after removing the trait effects.

This is a natural and reasonable representation of what is typically meant by method effects. If these correlated uniquenesses are nonsignificant and sufficiently small, method effects are negligible. If they are non-negligible, the pattern of correlated uniquenesses is more likely to indicate the nature of the method effects than simply fitting a model that assumes that method effects are unidimensional. Furthermore, the unidimensionality of method effects is merely a special case of Model 4E that can be explicitly tested by comparing Model 4E with other models considered here. For these reasons, Model 4E is a better, more natural and more heuristic representation of MTMM data than other models considered here.

Evaluation of Guidelines Concerning the Minimum Requirements For CFA Studies

The present investigation appears to have important implications for some frequently cited rules of thumb about the general requirements needed to achieve stable, well-defined solutions in factor analytic studies. In particular, generalizations based on monte carlo studies may substantially underestimate the minimum requirements for studies based on real data. This problem apparently stems from the fact that most simulation studies test models and construct data based on an independent clusters structure, in which each measured variable is associated with only one latent factor.

In studies reviewed earlier, it is often recommended that at least three indicators per factor are needed to achieve a stable, well-defined struc-

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

ture. This advice is apparently supported by monte carlo studies based on independent clusters data fit to independent clusters models. Real data and the models needed to describe real data are, however, more complicated than this perhaps overly simplistic simulated data, so generalizations based on these studies must be viewed cautiously. With MTMM data, each measured variable is associated with two latent constructs—a trait factor and a method factor. For example, when Model 4D was applied to data simulated to meet the assumptions of this model in Study 2, designs with three indicators per factor (i.e., the 3T × 3M design) were typically ill-defined for all sample sizes, and even designs with four indicators per factor (i.e., the 4T × 4M design) were typically ill-defined unless the sample size was large. These findings are consistent with the results of Study 1 based on real data. The problem may be in how the number of indicators per factor is defined. In the 3T × 3M design, for example, each trait and method factor had three indicators, but six factors were inferred on the basis of only nine variables—a ratio of 1.33 variables per factor. For the MTMM designs considered here, only the 6T × 6M design had three times as many measured variables (36) as latent factors (12).

The present results also call into question rules of thumb about the ratio of sample size to either the number of variables or the number of parameter estimates needed to obtain stable, well-defined solutions. These rules of thumb imply that for a fixed sample size, solutions with more measured variables and more estimated parameters are less likely to result in proper solutions. In contrast, the present results indicate an opposite pattern. Model 4E required more parameter estimates than Models 4C and 4D for all the MTMM matrices considered here—except the 3T × 3M matrices—but it was much more likely to result in proper solutions. Similarly, at each sample size the design with the larger number of measured variables (i.e., 7T × 4M and 6T × 6M versus 4T × 4M, and 4T × 4M versus 3T × 3M) was more likely to result in proper solutions. Previous research has also shown, however, that having more indicators for each factor is more likely to result in proper solutions. The present results are consistent with this generalization.

Although these findings based on MTMM data may not generalize to CFA studies based on simpler data structures, they provide convincing counter-instances that demonstrate that at least some rules of thumb may have limited generality. The problem appears to stem from overgeneralizing the results of simulation studies based on unrealistically simple factor structures and models.

APPENDIX

The variables used in Study 2 were created using the pseudo-random number generator in SPSSx (SPSS, 1986). For both model 4D and 4E data, the process began by generating an approximately unit normal random variate γ representing the second-order trait factor that accounted for the correlation between traits. The j trait factors f_j were then created by using a weighted combination of γ with random unit normal variates ε_j, such that

$$f_j = W_j\gamma + \varepsilon_j[1 - (W_j\gamma)^2]^{1/2} \quad . \qquad (1)$$

The weight w_j used for each variable was systematically varied so that for the 3 × 3 model, the weights were .5, .5, and .6 (generating population factor correlations of .25 to .30); for the 4 × 4 model .5, .5, .6, and .6 (correlations .25 to .36); and for the 6 × 6 model .5, .5, .6, .6, .7, and .7 (correlations .25 to .49). For the 4D model data, the j method factors h_j were generated in the same way, using the same set of weights. The value s_{pq}, representing the score of a person on the pth trait and the qth method, was then calculated as:

$$s_{pq} = U_p f_p + V_q h_q + \varepsilon_{pq}(1 - U_p^2 - V_q^2)^{1/2} \quad , \qquad (2)$$

with the εs again being unit normal variables. The Trait weights u were systematically varied in the cycle (.3, .4, .5, .6, .7, .3, . . .) for the 3 × 3 and

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

6 × 6 model, and in the cycle (.3, .4, .5, .6, .7, .8, .3, . . .) for the 4 × 4 model. The Method weights v were similarly varied in the cycle (.25, .3, .35, .4, .45, .25, . . .) for the 3 × 3 and 6 × 6 model, and in the cycle (.25, .3, .35, .4, .45, .5, .25, . . .) for the 4 × 4 model, with the scores generated in the order s_{11}, s_{12}, s_{13}, and so forth.

There are many ways of simulating correlated uniquenesses in Model 4E. A different method was used for each size of model. Each began by generating two independent random normal variates, η^a and η^b. For the 4 × 4 model, the same method and weights were used that were used for the 4D data, except that η^a and η^b were used as two uncorrelated Method factors, the first applying to Traits 1 and 2, and the second to Traits 3 and 4. For the 3 × 3 model, a weighted sum of η^a and η^β was used to represent the Method; the successive weights for s_{11}, s_{12}, s_{13}, and so forth were .25, .3, .35, .28, .32, .18, 0, 0, 0 for η^a and 0, 0, 0, .28, .32, .17, .3, .35, .4 for η^b.

For the 6 × 6 model, the weights used were .25, .3, .35, .4, .45, .25, .21, .25, .28, .32, .18, .21, 0, 0, 0, 0, 0, 0 for η^a and 0, 0, 0, 0, 0, 0, .21, .24, .28, .32, .17, .21, .25, .3, .35, .4, .45, .25 for η^b. These were used for Traits 1 to 3, and the same weights were applied to two other random normal deviates, η^c and η^d, for Traits 4 to 6. As in the case of the 4D model data, a suitably-weighted random normal deviate e_{pq} was added to produce a final s_{pq} having population mean 0 and variance 1.

References

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

Anderson, J. C., & Gerbing, D. W. (1984). The effect of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis. *Psychometrika, 49,* 155–173.

Bachman, J. G. (1970). *Youth in transition: The impact of family background and intelligence on tenth-grade boys* (Vol. 2). Ann Arbor MI: Institute for Social Research.

Bachman, J. G. (1975). *Youth in transition: Data file documentation* (Vol. 2). Ann Arbor MI: Institute for Social Research.

Bachman, J. G., & O'Malley, P. M. (1977). Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality and Social Psychology, 35,* 365–380.

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

Boomsma, A. (1982). The robustness of LISREL against small sample sizes in factor analysis models. In K. G. Jöreskog & H. Wold (Eds.), *Systems under indirect observation: Causality, structure, prediction* (Part I). Amsterdam: North-Holland.

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.

Browne, M. W. (in press). Relationships between an additive model and a multiplicative model for multitrait-multimethod matrices. In R. Coppi & S. Bolasco (Eds.), *Analysis of multiway data matrices.* Amsterdam: Elsevier.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the 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.

Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research, 18,* 147–267.

Dillon, W. R., Kumar, A., & Mulani, M. (1987). Offending estimates in covariance structure analysis: Comments on the causes and solutions to Heywood cases. *Psychological Bulletin, 101,* 126–135.

Gerbing, D. W., & Anderson, J. C. (1985). Effects of sampling error and model characteristics on parameter estimation for maximum likelihood confirmatory factor analysis. *Multivariate Behavioral*

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Research, 20, 255-271.
Gerbing, D. W., & Anderson, J. C. (1987). Improper solutions in the analysis of covariance structures: Their interpretability and a comparison of alternative solutions. *Psychometrika, 52,* 99-111.
Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103,* 265-275.
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. V. Krantz, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2, pp. 1-56). San Francisco CA: W. U. 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 IL: International Educational Services.
Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology, 12,* 247-252.
Kenny, D. A. (1979). *Correlation and causality.* New York: Wiley.
Marsh, H. W. (1988). Multitrait-multimethod analyses. In J. P. Keeves (Ed.), *Educational research methodology, measurement and evaluation: An international handbook.* Oxford: Pergamon Press.
Marsh, H. W. (1989a). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13,* 335-361.
Marsh, H. W. (March, 1989b). *Goodness of fit in confirmatory factor analysis: The effectiveness of penalties for capitalizing on chance.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco CA, U.S.A.
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., & 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., & Hocevar, D. (1988). A new, more powerful approach to multitrait-multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology, 73,* 107-117.
McDonald, R. P. (1985). *Factor analysis and related methods.* Hillsdale NJ: Erlbaum.
McDonald, R. P., & Marsh, H. W. (1990). Choosing a multivariate model: Noncentrality and goodness-of-fit. *Psychological Bulletin, 107,* 247-255.
Mulaik, S. A., James, L. R., Van Alstine, J., Bennett, N., Lind, S., & Stilwell, C. D. (1989). Evaluation of goodness-of-fit indices for structural equation models. *Psychological Bulletin, 105,* 430-445.
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). Methodology review: Analyses of multitrait-multimethod matrices. *Applied Psychological Measurement, 10,* 1-22.
SPSS. (1986). *SPSSx user's guide.* Chicago IL: SPSS, Inc.
Tanaka, J. S. (1987). "How big is big enough?": Sample size and goodness of fit in structural equation models with latent variables. *Child Development, 58,* 134-146.
Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Frederickson & H. Gulliksen (Eds.), *Contributions to mathematical psychology* (pp. 109-127). New York: Holt, Rinehart, & Winston.
Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1-10.
Werts, C. E., Jöreskog, K. G., & Linn, R. L. (1972). A multitrait-multimethod model for studying growth. *Educational and Psychological Measurement, 32,* 655-678.
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, Chicago IL.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/

Acknowledgments

The authors acknowledge the helpful comments of Roderick McDonald, the Editor, and an anonymous reviewer on earlier drafts of this paper.

Author's Address

Send requests for reprints or further information to Herbert W. Marsh, University of Western Sydney, Macarthur, P.O. Box 555, Campbelltown NSW 2560, Australia.

Downloaded from the Digital Conservancy at the University of Minnesota, http://purl.umn.edu/93227.
May be reproduced with no cost by students and faculty for academic use. Non-academic reproduction
requires payment of royalties through the Copyright Clearance Center, http://www.copyright.com/