STRUCTURAL EQUATION MODELING, *11*(4), 535–559

Bias in the Correlated Uniqueness Model for MTMM Data

James M. Conway
Central Connecticut State University

Filip Lievens
Ghent University

Steven E. Scullen
North Carolina State University

Charles E. Lance
The University of Georgia

This simulation investigates bias in trait factor loadings and intercorrelations when analyzing multitrait-multimethod (MTMM) data using the correlated uniqueness (CU) confirmatory factor analysis (CFA) model. A theoretical weakness of the CU model is the assumption of uncorrelated methods. However, previous simulation studies have shown little bias in trait estimates even when true method correlations are large. We hypothesized that there would be substantial bias when both method factor correlations and method factor loadings were large. We generated simulated sample data using population parameters based on our review of actual MTMM results. Results confirmed the prediction; substantial bias occurred in trait factor loadings and correlations when both method loadings and method correlations were large.

Requests for reprints should be sent to James M. Conway, Department of Psychology, Central Connecticut State University, 1615 Stanley Street, New Britain, CT 06050-4010. E-mail: conwayj@ccsu.edu

Steven E. Scullen is now in the Department of Management, Drake University.

Campbell and Fiske's (1959) introduction of the multitrait-multimethod (MTMM) matrix approach to establishing convergent and discriminant validity is one of the most highly cited articles in the history of psychology (Sternberg, 1992). Still today, the MTMM methodology is one of the cornerstone approaches to establishing the construct validity of psychological measures (Messick, 1995). Although Campbell and Fiske's article broke new and important methodological ground, it has been pointed out that their criteria for establishing convergent and discriminant validity were qualitative and subjective (e.g., Schmitt & Stults, 1986). As a result, numerous, more quantitatively based approaches to the MTMM matrix have been proposed, including (a) analysis of variance (ANOVA; e.g., Kavanagh, MacKinney, & Wolins, 1971), (b) path analysis (e.g., Kalleberg & Kluegel, 1975), (c) linear regression (e.g., Lehman, 1988), (d) exploratory factor analysis (e.g., Lomax & Algina, 1979), (e) covariance components analysis (e.g., Wothke, 1996), and (f) confirmatory factor analysis (CFA; e.g., Widaman, 1985).

Current consensus is that CFA provides the most general and flexible quantitative approach to the analysis of MTMM data (e.g., Marsh & Grayson, 1995; Millsap, 1995; Schmitt & Stults, 1986). Even here, however, there is debate over which paramaterization of the CFA model best characterizes the structure of MTMM data. Four main variants of the CFA model are (a) a class of linear and additive CFA models (e.g., Widaman, 1985, 1992), (b) a class of hierarchical second-order factor (SOF) CFA models (e.g., Marsh & Hocevar, 1988), (c) a correlated uniqueness (CU) parameterization of method effects (Kenny, 1976; Marsh, 1989), and (d) direct product (DP) models of multiplicative trait and method effects (Campbell & O'Connell, 1967; Cudeck, 1988). Our main focus in this article is on the CU model because (a) it has been recommended as a desirable option for MTMM analysis (e.g., Lievens & Conway, 2001; Marsh, 1989) and (b) there is reason to be concerned about bias in CU model estimates (Lance, Noble, & Scullen, 2002). To describe the potential for bias we use the linear CFA (also known as the correlated trait–correlated method, CTCM, model) as a point of comparison.

Lance et al. (2002) critiqued the CTCM and CU models along several lines, noting that both models have their strengths and weaknesses. Two of the issues discussed by Lance et al. are most critical for our study. The first issue is that the CU model returns convergent and proper solutions far more often than does the CTCM model (Conway, 1996; Marsh, 1990; Marsh & Bailey, 1991; Marsh, Byrne, & Craven, 1992). Because of this advantage, the CU model has gained in popularity relative to the CTCM model. The second issue and the focus of this article is the potential bias in estimated trait factor correlations and trait factor loadings under the CU model. As Lance et al. (2002) pointed out, this weakness can potentially lead to inaccurate inferences regarding construct validity.

POTENTIAL BIAS IN ESTIMATED TRAIT-RELATED PARAMETERS UNDER THE CU MODEL

Trait factor loadings and trait correlations are critical parameter estimates for MTMM data. Large trait factor loadings support inferences about convergent validity, and relatively low trait factor correlations support inferences about discriminant validity (Widaman, 1985). Bias in these parameter estimates could lead to unwarranted conclusions about the construct validity of measures. Lance et al. (2002) showed how the potential for these biases incurs. Assuming the analysis of a MTMM correlation matrix, under the CTCM model monotrait-heteromethod (MTHM), heterotrait-heteromethod (HTHM), and heterotrait-monomethod (HTMM) correlations are modeled as functions of estimated model parameters, respectively, as follows:

$$\text{MTHM} = \lambda_{Tij}\lambda_{Tij'} + \lambda_{Mij}\lambda_{Mij'}\phi_{MjMj'} \tag{1a}$$

$$\text{HTHM} = \lambda_{Tij}\lambda_{Ti'j'}\phi_{TiTi'} + \lambda_{Mij}\lambda_{Mi'j'}\phi_{MjMj'} \tag{1b}$$

$$\text{HTMM} = \lambda_{Tij}\lambda_{Ti'j}\phi_{TiTi'} + \lambda_{Mij}\lambda_{Mi'j} \tag{1c}$$

where λ_{Tij} refers to the standardized loading of the *ij*th trait-method unit (TMU; i.e., a measure of the *i*th trait as measured by the *j*th measurement method) on the *i*th latent trait factor, λ_{Mij} is the *ij*th TMU's loading on the *j*th method factor, $\phi_{TiTi'}$ refers to the correlation between *different* trait factors, and $\phi_{MjMj'}$ refers to the correlation between *different* method factors. Thus, correlations among TMUs in a MTMM matrix are seen as arising from the effects of trait and method factors and the trait and method factors' intercorrelations.

Under the CU model, these correlations are modeled as follows:

$$\text{MTHM} = \lambda_{Tij}\lambda_{Tij'} \tag{2a}$$

$$\text{HTHM} = \lambda_{Tij}\lambda_{Ti'j'}\phi_{TiTi'} \tag{2b}$$

$$\text{HTMM} = \lambda_{Tij}\lambda_{Ti'j}\phi_{TiTi'} + \theta_{\delta(ij,i'j)} \tag{2c}$$

where $\theta_{\delta(ij,i'j)}$ refers to covariances between TMUs uniquenesses that are fixed equal to zero for pairs of TMUs measured by a different method (i.e., $j \neq j'$) but that are estimated as free parameters as part of the CU model for TMUs that share a common measurement method (i.e., $j = j'$).

Comparing Equations 1c and 2c, it can be seen that the CTCM and CU models account for *common* method effects in HTMM correlations by alternative parameterizations of the same covariance component: (a) the CTCM model uses the common causal effect of the measurement method shared in common (i.e.,

$\lambda_{Mij}\lambda_{Mi'j}$ in Equation 1c) and (b) the CU model uses the covariance between the uniquenesses of TMUs that share the same measurement method (i.e., $\theta_{\delta(ij,i'j)}$ in Equation 2c). However, under the CTCM model, method effects appear also for MTHM ($\lambda_{Mij}\lambda_{Mij'}\phi_{MjMj'}$ from Equation 1a) and HTHM ($\lambda_{Mij}\lambda_{Mi'j'}\phi_{MjMj'}$ from Equation 1b) correlations (i.e., some portion of the MTHM and HTHM correlations are assumed to be due to effects of correlated methods on TMUs). Note, however, that these components do *not* appear for the MTHM and HTHM correlations under the CU model (Equations 2a and 2c). In *special cases* in which one or more of the λ_{Mij}s = 0 or $\phi_{MjMj'} = 0$, Equations 1a and 1b reduce to their corresponding Equations 2a and 2b. If, however, all $\lambda_{Mij}s \neq 0$ *and* $\phi_{MjMj'} \neq 0$, that is, if TMUs' loadings on method factors are nonzero *and* correlations between different methods are nonzero, then Equations 2a and 2b are misspecified.[1] In this case, there are potential sources of covariance in the MTHM and HTHM correlations that are not accounted for in Equations 2a and 2b. This is expected to result in (likely upward, and perhaps seriously) biased estimates for the λ_{Tij}s (leading to inflated estimates of convergent validity), the $\phi_{TiTi'}$s (leading to underestimates of discriminant validity), or both. The implications of this "unmeasured variables problem" (James, 1980) are summarized by Kenny and Kashy (1992):

> The average method–method covariance is added to each element of the trait–trait covariance matrix. So, if the methods are similar to one another, resulting in positive method–method covariances, the amount of trait variance will be overestimated as will be the amount of trait–trait covariance. (pp. 169–170)

Lance et al. (2002) noted that this source of specification error was one potentially serious limitation of the CU model and presented one example of how bias in estimated trait factor loadings and correlations could result in serious misinterpretation of construct validity from a MTMM matrix. Marsh and Bailey (1991) noted that method correlations are critical parameters because the CU model assumes that method correlations are zero and, necessarily, ignores method correlations in the estimation of other model parameters.

In their comprehensive simulation study, Marsh and Bailey (1991) used method correlations as high as .49. However, they found evidence of only small positive biases in the CU model (.02 and .047 on average for trait loadings and trait correlations, respectively) and concluded that those biases were trivial. A second simula-

[1]We do not mean to imply that the CTCM model is necessarily correct or that the CU model is necessarily misspecified. Instead, our point is that *if* correlated method effects are present in MTMM data, and theory and empirical evidence suggest that they typically are, then there is a potentially important source of covariance that is accounted for by the CTCM model but not by the CU model. Although the CTCM model includes this method covariance, it is always possible that the model is misspecified in other ways (e.g., the assumption that method factors are unidimensional) and that the misspecification could lead to bias in CTCM estimates.

tion study found similarly small bias (Tomás, Hontangas, & Oliver, 2000). Tomás et al. did point out, however, that Marsh and Bailey had varied method correlations within the MTMM matrix, so the average method correlation was relatively low even when some values were as high as .49. Tomás et al. simulated data in which all method correlations for a MTMM matrix were .60 but still found little bias in trait estimates.

Despite these findings, we still believe there is the potential for substantial bias in trait estimates under the CU model, especially when there is substantial method variance shared between variables. Tomás et al. (2000) were correct to point out the importance of method correlations, but Equations 1a and 1b show that what is really critical is the *product* of a method correlation and two method factor loadings. For example, if two methods are highly correlated, but a pair of variables each has a low loading on its measurement method factor, then the correlation between those two variables will not be inflated much (there is little common method bias). Thus, the CU model should be able to estimate fairly accurately the trait parameters for those variables. Although Tomás et al. simulated high method correlations, their simulated method loadings were only as high as .35; this can explain their finding of little bias. Our prediction is that the CU model will show substantial upward bias in trait loadings and trait correlations when both method correlations *and* method loadings are reasonably high, a case that may be common in real data (e.g., Cote, & Buckley, 1987).

Our prediction is based on the size of parameters (both method correlations and method loadings); thus, it is critical that the values in a simulation study be representative of realistic research situations (Paxton, Curran, Bollen, Kirby, & Chen, 2001). In the two previous simulations, it is not clear how parameters were chosen, although we speculate that it was based on a combination of logic and previous experience of the investigators. Therefore, it is not clear how well previous results generalize to MTMM situations that researchers are likely to face when analyzing real data. This issue is addressed in our study.

In summary, given (a) Lance et al.'s (2002) conceptual (algebraic) results that suggest that CU model estimates of trait loadings and correlations are potentially biased; (b) the Marsh and Bailey (1991) and Tomás et al. (2000) simulation results suggesting that there is little, if any, bias in CU estimates; (c) the unknown generalizability of the previous simulations' results; (d) the relatively common reliance on the CU method to solve convergence problems associated with the CTCM method; and (e) the significance, for the validity of substantive research conclusions, of the assumption that CU parameter estimates are relatively free of bias, further investigation of the accuracy of CU model estimates is needed. Thus, the general purpose of this study was to determine the extent of bias in trait factor loadings and trait intercorrelations under the CU model for a variety of representative sets of MTMM model parameters, including a condition with high method correlations *and* high method loadings. Therefore, we hope to offer guidance as to conditions under which the CU model is (not) likely to yield accurate estimates of trait parameters.

METHOD

Literature Review

We undertook the following steps to design the simulation study (see Paxton et al., 2001). First, to ensure ecological validity, we examined large-scale reviews of MTMM studies in applied research. MTMM studies were defined broadly (Kenny, 1995), to include instrument-based, rater-based (so-called multitrait-multirater matrices), and temporally based methods (so-called multitrait-multioccasion matrices). Five large-scale reviews were identified and reviewed (Buckley, Cote, & Comstock, 1990; Conway, 1996; Hernandez & Gonzalez-Roma, 2002; Lievens & Conway, 2001; Williams, Cote, & Buckley, 1989). The Becker and Cote (1994) review was not included because the studies overlapped considerably with other reviews. Overall, these large-scale reviews included 147 individual MTMM matrices.

Second, based on this literature review, we made informed decisions regarding MTMM study design and the population values of the parameters of interest to this study. Table 1 presents the descriptive statistics of each variable of interest (i.e., sample size, the number of traits, the number of methods, mean trait loading, mean trait correlation, mean method loading, and mean method correlation) across the MTMM matrices retrieved by the literature review. For factor loadings and correlations, we used estimated values for the CTCM model (this is the only model that

TABLE 1
Results of Literature Review and Population Parameter Values in Simulation

	Literature Review Values							*Simulation Values*
	M	*Mdn*	*SD*	*Min*	*Max*	*25th*	*75th*	
Sample size[a]	232.14	125	309.27	27	2163	82	261	125 and 500
Number of methods	2.92	3	.95	2	8	2	3	3
Number of traits[a]	4.5	4	2.03	2	11	3	5	3 and 5
Trait loadings	.55	.57	.17	.11	.93	.42	.69	.60
Trait correlations	.41	.44	.30	–.58	1.00	.28	.61	.50
Method loadings[a]	.50	.48	.18	.00	.84	.37	.65	.35, .50, and .65
Method correlations[a]	.28	.31	.33	–1.00	.91	.16	.49	.10, 30, and .50

Note. Due to missing values, *N* varies between 109 and 147. Mdn = median; Min = minimum; Max = maximum

[a]These variables were manipulated in the simulation study.

provides all necessary information), averaged across all values within a study (e.g., the mean trait factor loading for a study). The last column of Table 1 shows the values chosen in the simulation.

As a general rule we decided that, in case of manipulated variables (e.g., mean method correlation), the 25th percentile, 50th percentile (the median), and the 75th percentile were chosen as population parameter values. If a variable was held constant (e.g., mean trait loading), the median was chosen as population parameter value. However, we deviated slightly from these general rules for some variables because we had to make trade-offs among ecological validity, feasibility, and answering our research questions (Paxton et al., 2001). In particular, to keep the simulation manageable, we included only two values for the sample size variable, N = 125 (i.e., the median value from our review) and N = 500. Although the latter value clearly exceeded the 75th percentile value from our review, we chose this value because it replicated the sample size value chosen by the two earlier simulation studies (Marsh & Bailey, 1991; Tomás et al., 2000). We also included only two values for the matrix size variable, a matrix with three traits and three methods and a MTMM matrix with five traits and three methods. As shown in Table 1, the number of methods chosen in our simulation conformed to the median value from our review and the two values chosen for the number of traits conformed to the 25th and 75th percentiles.

Design

Four variables were manipulated: the mean method factor loading for a matrix, the mean method factor correlation for a matrix, the sample size, and the matrix size. Mean method loading and mean method correlation each had three levels. Mean method loadings were .35, .50, and .65 (the 25th, 50th, and 75th percentiles were actually .37, .48, and .65, see the last column in Table 1). Mean method correlations were .10, .30, and .50 (the 25th, 50th, and 75th percentiles were actually .16, .31, and .49, see the last column in Table 1). Method parameters varied within matrix. For example, within-matrix method loadings varied by ±.20 around the mean value for the matrix (.20 is approximately the method loading standard deviation in Table 1, .18). See the Appendix for a complete description of parameters and an example of values for the 3 × 3, high method loading, high method correlation condition. Sample size had two levels. These levels were N = 125 and N = 500. Matrix size had two levels: a MTMM matrix with three traits and three methods and a MTMM matrix with five traits and 3 methods. Crossing the levels of these four variables yielded a 3 × 3 × 2 × 2 factorial design.

The other matrix characteristics were held constant, namely mean trait factor loading and mean trait factor correlation. Mean trait loadings were held constant at .60 (the median value was actually .57) and mean trait correlations at .50 (the median value was actually .44). As with method parameters, the trait parameters varied within matrix. For example, trait loadings varied from .40 to .80 or the median ±.20.

Simulation and Analysis of Data

For each of the 36 cells (three method loading conditions × three method correlation conditions × two sample size conditions × two matrix size conditions) we used PRELIS 2 (Jöreskog & Sörbom, 1996) to generate 100 sample MTMM matrices with normally distributed variables, yielding a total of 3,600 matrices. We then used LISREL 8 (Jöreskog & Sörbom, 1993) to fit the CU model to each matrix using maximum likelihood estimation. Although not directly related to the study's purpose, we also fitted the CTCM model to each matrix.

RESULTS

Estimation Problems

We considered a solution to be appropriate if it had no estimation problems, that is, if the solution converged and had no out-of-range estimates (standardized factor loadings or factor correlations greater than 1.00 in absolute value or negative unique variances). Percentages of appropriate solutions by condition appear in Table 2. The CU model produced appropriate solutions for 3,494 of the 3,600 solutions (more than 97%). There was a tendency for more inappropriate solutions with $N = 125$ than with $N = 500$ (not surprising, due to greater sampling error for $N = 125$), but unexpectedly this only tended to happen when method correlations were low. This result was surprising because the CU model should be more appropriate (and show less bias) when method correlations are low, because the CU model implicitly assumes orthogonal method factors. We do not have an explanation for this finding.

Goodness of Fit

We based our decisions regarding model fit on Hu and Bentler's (1999) criteria. They concluded that relatively stringent fit criteria should be used and recommended the standardized root mean squared residual (SRMSR) along with at least one other fit index such as the Tucker-Lewis Index (TLI), the root mean square error of approximation (RMSEA), or both. Their proposed cutoffs for good fit were .08 or lower for the SRMSR, .95 or higher for the TLI, and .06 or lower for the RMSEA. We defined a well-fitting model as one that met all three of these criteria.

TABLE 2
Percentages of CU Solutions That Were Appropriate and That Showed Good Fit

	Method Correlations		
Method Loadings	*.50 (High)*	*.30 (Medium)*	*.10 (Low)*
High Method Loading (.55)			
3 × 3			
N = 125			
% Appropriate	100	98	94
% Good fit	*1*	11.2	*83*
N = 500			
% Appropriate	100	100	100
% Good fit	*0*	0	*93*
5 × 3			
N = 125			
% Appropriate	97	98	96
% Good fit	*5.2*	29.6	*80.2*
N = 500			
% Appropriate	100	100	100
% Good fit	*0*	43	*100*
Overall			
% Appropriate	99.3	99	97.5
% Good fit	*1.5*	21	*89.2*
Medium Method Loading (.35)			
3 × 3			
N = 125			
% Appropriate	99	96	83
% Good fit	*31.3*	49	*84.3*
N = 500			
% Appropriate	100	100	100
% Good fit	*2*	38	*100*
5 × 3			
N = 125			
% Appropriate	99	94	85
% Good fit	*44.4*	63.8	*91.8*
N = 500			
% Appropriate	100	100	100
% Good fit	*45*	94	*100*
Overall			
% Appropriate	99.5	97.5	92
% Good fit	*30.7*	61.3	*94.6*
Low Method Loading (.15)			
3 × 3			
N = 125			
% Appropriate	99	94	89
% Good fit	*50.5*	56.4	*69.7*

(continued)

TABLE 2 (*Continued*)

	Method Correlations		
Method Loadings	*.50 (High)*	*.30 (Medium)*	*.10 (Low)*
N = 500			
% Appropriate	100	100	100
% Good fit	*40*	84	*99*
5 × 3			
N = 125			
% Appropriate	98	95	86
% Good fit	*55.1*	71.6	*77.9*
N = 500			
% Appropriate	100	100	100
% Good fit	*81*	95	*100*
Overall			
% Appropriate	99.5	97.3	93.8
% Good fit	*56.7*	77.1	*87.5*

Note. Appropriate solutions were those with no estimation problems. Percents of samples showing Good Fit are shown in italics and are calculated as the percentage out of all Appropriate solutions. CU = correlated uniqueness.

Table 2 shows the percentages of matrices with good CU model fit in each condition (note that these figures were computed considering only those solutions that contained no improper parameter estimates as described earlier). The most striking finding was the fact that less than 2% of appropriate solutions met the fit criteria in the high method loading, high method correlation condition. In general, as method loadings and method correlations decreased, the fit of CU solutions increased; in the low-low condition 87.5% of appropriate solutions met the stringent fit criteria. These findings suggest that the fit of the CU model is sensitive to method effects. The larger the product of mean method loading and mean method correlation, the less likely the CU model was to meet the fit criteria; in the case in which the largest bias was expected (the high-high condition) the CU model almost never met the fit criteria. The fit criteria that we used might therefore serve as one indication of whether the CU model is likely to produce substantially biased trait estimates (see Discussion).

Accuracy of CU Model Trait Estimates

ANOVA: trait factor loadings. As one approach to investigating the accuracy of CU model trait estimates, we conducted two separate ANOVAs: one for trait factor loadings and one for trait factor correlations. Results from all appropriate CU solutions (those containing no improper parameter estimates) were included. We conducted the ANOVAs with the individual parameter estimate as the unit of analysis; thus, each matrix contributed multiple cases. The dependent variables were deviations calculated as the CU parameter estimate minus its corresponding population value. We refer to these scores as errors. Positive errors represent overestimations and negative errors represent underestimations. For the trait loading ANOVA, there were 41,979 cases, and for the trait correlation ANOVA, there were 22,733 cases.

The independent variables were mean method loading (three levels; .65, .50, and .35), mean method correlation (three levels; .50, .30, and .10), *N* (two levels; 500 and 125), matrix size (two levels; 3 × 3 and 5 × 3), and parameter value. Parameter value took on five levels for trait loadings (3 × 3 matrices had values of .4, .6, and .8; 5 × 3 matrices had values of .4, .5, .6, .7, and .8) and 10 levels for trait correlations. We estimated all main effects, two-way interactions, and three-way interactions. We did not evaluate statistical significance because of the unusually high power afforded by the sample sizes. Instead, we used η^2 values of .01 or greater (1% of the variance accounted for) as our criterion for interpreting an effect that was practically significant.

Results of the ANOVA for trait factor loadings appear in Table 3. Three main effects showed η^2 of at least .01: mean method loading (η^2 = .032), mean method correlation (η^2 = .081), and parameter (η^2 = .362). As expected, positive bias in trait factor loadings was highest when the mean method loading was .65 (mean error = .063) and decreased when the mean method loadings were .50 (mean error = .049) and .35 (mean error = .023). In addition, as expected, positive bias in trait factor loadings was highest when the mean method correlation was .50 (mean error = .080) and decreased when the mean method correlations were .30 (mean error = .043) and .10 (mean error = .012). The large effect for parameter value was not expected, and we explore it in more detail later. In addition, three two-way interactions had relatively large effect sizes. As we expected, the method loading-method correlation interaction (η^2 = .012) showed that bias was greatest when both method loadings and method correlations were high. There were unexpected interactions for mean method correlation × parameter (η^2 = .042) and parameter × matrix size (η^2 = .120).

Mean errors by conditions: trait factor loadings. To interpret the trait factor loading effects in more detail, we examined mean errors for the various conditions. Table 4 contains mean errors for all conditions except different levels of parameter values. There are so many levels of parameter values that we chose to address this independent variable separately (see later).

The overall values in Table 4 represent the mean errors for each combination of method loading and method correlation. These means are also shown graphically in Figure 1. The pattern is consistent with our hypothesis that the combination of high method factor loadings and high method correlations would produce substantial bias; the high-high condition showed the highest mean error .117—a

TABLE 3
ANOVA Results for Trait Factor Loadings and Trait Factor Correlations

	Trait Factor Loadings		*Trait Factor Correlations*	
Source of Variance	*F*	η^2	*F*	η^2
A. Mean method loading	686.04	*.032*	465.27	*.039*
B. Mean method correlation	1845.65	*.081*	1860.87	*.138*
C. Parameter value	5951.30	*.362*	40.71	*.016*
D. Sample size	1.38	.000	8.06	.000
E. Matrix size	27.24	.001	74.66	.003
A × B	129.08	*.012*	70.64	*.012*
A × C	12.39	.002	4.61	.004
A × D	1.41	.000	.25	.000
A × E	24.94	.001	10.95	.001
B × C	227.59	*.042*	8.67	.007
B × D	.02	.000	4.57	.000
B × E	6.24	.000	2.24	.000
C × D	.68	.000	.96	.000
C × E	3005.17	*.126*	2.81	.000
D × E	.11	.000	.47	.000
A × B × C	4.67	.002	1.20	.002
A × B × D	.95	.000	1.98	.000
A × B × E	2.95	.000	1.89	.000
A × C × D	.23	.000	1.47	.001
A × C × E	8.12	.001	1.56	.000
A × D × E	.65	.000	.56	.000
B × C × D	.86	.000	.79	.001
B × C × E	38.65	.004	1.56	.000
B × D × E	.04	.000	1.12	.000
C × D × E	.17	.000	1.03	.000

Note. Italicized η^2 values exceeed .01, our criterion for interpreting an effect. ANOVA = analysis of variance.

substantial amount of overestimation. There was also substantial overestimation in the medium method loading, high method correlation condition: a mean error of .083. Mean errors decreased as both method loading and method correlation decreased, to a low of .005 (very slight overestimation) for the low-low condition.

Results for different parameter values are shown in Table 5. Note that parameter had a main effect and also interactions with mean method correlation and matrix size. Rather than present all means necessary to fully explicate these effects, we present results for the most extreme case to show what *can* happen and then comment on differences by method correlation and matrix size. Table 5 presents mean errors by parameter value for the high method correlation, 3 × 3

TABLE 4
Mean Errors (Standard Deviations in Parentheses) for Trait Factor Loadings by Condition

	Method Correlations		
Method Loadings	*.50 (High)*	*.30 (Medium)*	*.10 (Low)*
High method loading (.55)			
3 × 3			
N = 125	.136 (.168)	.068 (.173)	.021 (.176)
N = 500	.132 (.156)	.067 (.160)	.017 (.165)
5 × 3			
N = 125	.109 (.098)	.049 (.076)	.016 (.068)
N = 500	.104 (.073)	.051 (.049)	.012 (.037)
Overall	.117 (.121)	.057 (.114)	.016 (.113)
Medium method loading (.35)			
3 × 3			
N = 125	.084 (.167)	.048 (.170)	.014 (.180)
N = 500	.084 (.154)	.048 (.157)	.014 (.164)
5 × 3			
N = 125	.085 (.107)	.052 (.094)	.015 (.089)
N = 500	.081 (.083)	.046 (.060)	.014 (.043)
Overall	.083 (.124)	.048 (.118)	.014 (.118)
Low method loading (.15)			
3 × 3			
N = 125	.041 (.173)	.024 (.179)	.003 (.187)
N = 500	.041 (.153)	.023 (.160)	.007 (.167)
5 × 3			
N = 125	.037 (.097)	.022 (.090)	.004 (.093)
N = 500	.040 (.064)	.023 (.053)	.005 (.047)
Overall	.039 (.119)	.023 (.119)	.005 (.122)

matrix condition (collapsed across the two *N*'s); this was the condition with the most extreme results.

Table 5 shows that overestimation of trait factor loadings was greatest for parameter values of .40 and least for parameter values of .80. The three loadings with parameter values of .40 had an average bias of .25, which is very substantial. The loadings with parameter values of .60 had an average bias of .15 (still substantial), whereas loadings with parameter values of .80 had an average bias of only .01. The pattern of greater bias for lower parameters holds across conditions not shown in Table 5, although the differences are smaller. Regarding the interaction between parameter value and mean method correlation: for the high method correlation condition (collapsed across method loading, matrix size, and *N*) the difference in bias for .40 parameter values versus .80 parameter values was relatively large at .18 versus –.05, respectively, whereas for the low method correlation condition the dif-

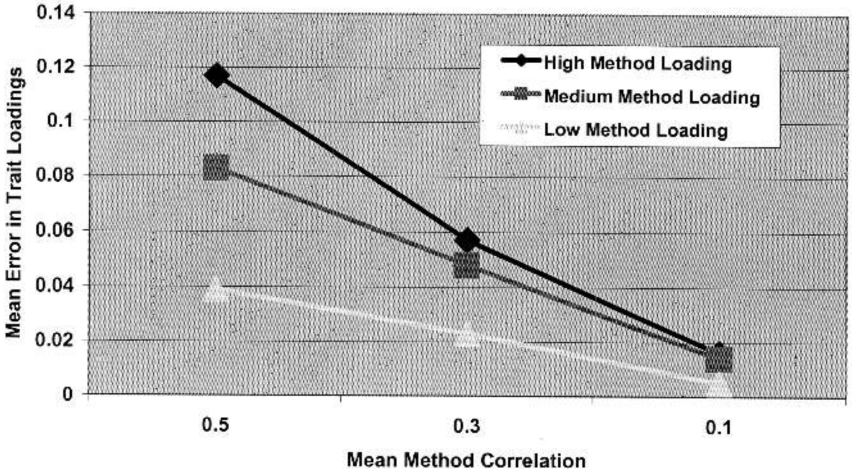

FIGURE 1 Mean errors for trait factor loadings.

ference was .05 versus –.07. Findings were similar for the interaction between parameter value and matrix size: 3 × 3 matrices had a larger difference (.15 for .40 parameter values vs. –.11 for .80 values) than did 5 × 3 matrices (.09 vs. –.01, respectively). Results for parameter value are important because they show that the CU model (a) tends to homogenize parameter estimates relative to their population values and (b) can return parameter estimates that are substantially biased.

TABLE 5
Trait Loading Parameters, Mean Sample Estimates, and Mean Errors for the High-Method Loading, High-Method Correlation Condition (Collapsed Across N's)

	Trait 1	*Trait 2*	*Trait 3*
Method 1			
Parameter	.40	.60	.80
Mean Estimate	(.700)	(.809)	(.852)
Mean Error	+.300	+.209	+.052
Method 2			
Parameter	.80	.40	.60
Mean Estimate	(.820)	(.649)	(.744)
Mean Error	+.020	+.249	+.144
Method 3			
Parameter	.60	.80	.40
Mean Estimate	(.692)	(.748)	(.590)
Mean Error	+.092	–.052	+.190

ANOVA: trait factor correlations. As shown in Table 4, the ANOVA for trait factor correlations showed three substantial main effects: mean method loading (η^2 = .039), mean method correlation (η^2 = .138), and parameter value (η^2 = .016). As expected, positive bias in trait factor correlations was highest when the mean method loading was .65 (mean error = .082) and decreased when the mean method loadings were .50 (mean error = .075) and .35 (mean error = .039). In addition, as expected, positive bias in trait factor correlations was highest when the mean method correlation was .50 (mean error = .112) and decreased when the mean method correlations were .30 (mean error = .063) and .10 (mean error = .019). There was only one interaction with a substantial effect size: the predicted two-way interaction between mean method loading and mean method correlation (η^2 = .012). As expected, bias in estimated trait factor correlations was greatest when both mean method loadings and mean method correlations were high.

Mean errors by conditions: trait factor correlations. Next, we examined mean errors in greater detail for our different conditions. Table 6 contains mean errors for all conditions except different levels of parameters (again discussed separately). These means are shown graphically in Figure 2. As with trait factor loadings, the overall values in Table 6 supported our hypothesis: the combination of high method loading and high method correlation produced the greatest bias. The mean bias for the high-high condition was .148, whereas mean bias for the low-low condition was only .013. Also similar to trait loading results, there was substantial overestimation in the medium method loading, high method correlation condition, with a mean error of .124.

Differences in bias as a function of the parameter value (which ranged from .330 to .720) were much smaller than for trait factor loadings. The mean bias for parameter of .330 (collapsed across all other independent variables) was .075, and for the parameter of .720 it was .048.

Supplementary Analyses: The CTCM Model

As a supplementary set of analyses, we fit the CTCM model to each matrix to parallel the CU model analyses. The CTCM model's tendency to produce appropriate solutions varied dramatically depending on N and matrix size, as shown in Table 7. Results for 3 × 3 matrices were poor and consistent with previous findings; for N = 125, only about 6% of solutions were appropriate, containing no improper parameter estimates, and for N = 500, about 30% were appropriate. These findings are consistent with previous research (e.g., Marsh & Bailey, 1991). The results for 5 × 3 matrices were markedly better: for N = 125, 59% were appropriate and for N = 500, 88% were appropriate. These findings are consistent with Lance et al.'s (2002) prediction that empirical underidentification problems should be less threatening in MTMM matrices that are larger than the minimum 3 trait × 3

TABLE 6
Mean Errors (Standard Deviations in Parentheses) for Trait Factor Correlations by Condition

	Method Correlations		
Method Loadings	*.50 (High)*	*.30 (Medium)*	*.10 (Low)*
High method loading (.55)			
3 × 3			
N = 125	.179 (.079)	.097 (.081)	.029 (.100)
N = 500	.173 (.050)	.095 (.047)	.029 (.050)
5 × 3			
N = 125	.143 (.088)	.069 (.089)	.010 (.098)
N = 500	.137 (.045)	.077 (.049)	.019 (.045)
Overall	.148 (.071)	.078 (.071)	.018 (.077)
Medium method loading (.35)			
3 × 3			
N = 125	.132 (.099)	.086 (.099)	.033 (.107)
N = 500	.125 (.048)	.083 (.052)	.026 (.048)
5 × 3			
N = 125	.126 (.090)	.066 (.102)	.023 (.109)
N = 500	.119 (.048)	.071 (.051)	.025 (.052)
Overall	.124 (.073)	.072 (.080)	.025 (.083)
Low method loading (.15)			
3 × 3			
N = 125	.069 (.111)	.035 (.114)	.017 (.108)
N = 500	.075 (.050)	.043 (.049)	.014 (.055)
5 × 3			
N = 125	.063 (.107)	.032 (.118)	.015 (.118)
N = 500	.062 (.054)	.041 (.054)	.010 (.055)
Overall	.064 (.085)	.037 (.090)	.013 (.089)

method design. Fit for the CTCM model was also best for 5 × 3, *N* = 500. Other conditions often failed to meet the fit criteria.

We calculated errors for the CTCM model as we did for the CU model, by taking the difference between an estimate and its parameter. For trait loadings, the mean error across all conditions was –.01, indicating almost no bias. For trait correlations, the mean error was –.02. We conducted ANOVAs like those for the CU model but no main effect or interaction effect showed a η^2 value of .01 or above. This indicated that none of the independent variables had any meaningful effects on the degree of bias, and we do not present mean errors separately by condition.

One intriguing finding was the relatively high percentage of appropriate solutions for the CTCM model in the 5 × 3, *N* = 500 condition. The fact that performance was considerably better in this condition than for either 3 × 3, *N* = 500 or 5 ×

TABLE 7
Percentages of CTCM Solutions that are Appropriate and that Show Good Fit

	Method Correlations		
Method Loadings	*.50 (High)*	*.30 (Medium)*	*.10 (Low)*
High Method Loading (.55)			
3 × 3			
N = 125			
% Appropriate	4	6	9
% Good fit	*100*	33.3	*77.8*
N = 500			
% Appropriate	32	33	65
% Good fit	*100*	0	*93.8*
5 × 3			
N = 125			
% Appropriate	78	74	73
% Good fit	*97.4*	28.4	*78.1*
N = 500			
% Appropriate	98	95	99
% Good fit	*63.3*	43.2	*100*
Overall			
% Appropriate	53	52	61.5
% Good fit	*82.1*	30.8	*91.1*
Medium method loading (.35)			
3 × 3			
N = 125			
% Appropriate	9	5	10
% Good fit	*44.4*	40	*90*
N = 500			
% Appropriate	23	22	49
% Good fit	*0*	50	*100*
5 × 3			
N = 125			
% Appropriate	66	55	73
% Good fit	*54.5*	74.5	*91.8*
N = 500			
% Appropriate	91	91	97
% Good fit	*42.9*	94.5	*100*
Overall			
% Appropriate	47.25	43.25	57.25
% Good fit	*41.8*	80.9	*96.9*
Low method loading (.15)			
3 × 3			
N = 125			
% Appropriate	5	4	5
% Good fit	*40*	50	*80*

(continued)

TABLE 7 (*Continued*)

	Method Correlations		
Method Loadings	*.50 (High)*	*.30 (Medium)*	*.10 (Low)*
N = 500			
% Appropriate	14	11	19
% Good fit	*42.9*	81.8	*100*
5 × 3			
N = 125			
% Appropriate	34	33	42
% Good fit	*55.9*	75.8	*81*
N = 500			
% Appropriate	70	82	73
% Good fit	*78.6*	95.1	*100*
Overall			
% Appropriate	30.75	32.5	34.75
% Good fit	*66.7*	87.7	*93.5*

Note. Appropriate solutions are those with no estimation problems. Percents of samples showing good fit are shown in italics and are calculated as the percentage out of all appropriate solutions. CTCM = correlated trait–correlated method.

3, *N* = 125, suggests that the combination of a large number of variables and a large *N* facilitates good CTCM performance. It seems likely that the relatively large number of degrees of freedom due to the large number of variables, coupled with the small amount sampling error due to the large *N,* results in better estimation of CTCM parameters. To explore this finding further we generated data for an additional condition (100 additional samples). This condition maintained *N* = 500 but increased the number of variables, using a 7 trait × 4 method matrix (this condition is similar to Marsh & Bailey's 1991 analysis of a 7 × 4 matrix based on real data with a large *N*). There were 28 observed measures and 295 *df* for the CTCM model, as compared to the 12 *df* in the 3 × 3 condition and 62 *df* in the 5 × 3 condition. We specified trait factor loading parameters varying from .45 to .75 (in increments of .05), with a mean of .60. Method factor loading parameters varied from .35 to .65 (in increments of .05), with a mean of .50. Trait factor correlation parameters varied from .33 to .68 with a mean of .488; method factor correlation parameters varied from .20 to .42 with a mean of .298.

We fit the CTCM model to each of the 100 correlation matrices; 99 showed appropriate solutions and all 99 met the fit criteria. Mean estimates were almost exactly the same as parameters: the mean errors were –.004 for trait factor loadings and –.01 for trait factor correlations (showing no evidence of bias). This finding, consistent with Marsh and Bailey's (1991) and Lance et al.'s (2002) prediction, indicates that the CTCM model can perform quite well but that the study design needs to go well beyond the typical matrix size and *N*.

DISCUSSION

Our simulation confirmed empirically what our algebraic argument suggested should be true—that when the *product* of the method loadings and method correlations is relatively large, the CU model is likely to yield biased estimates of trait factor loadings and trait factor correlations. This finding extends what previous simulation studies have shown. Other simulations (Marsh & Bailey, 1991; Tomás et al., 2000) have concluded that even relatively large method loadings and intercorrelations have only minor biasing effects on CU estimates of trait parameters. Our results are consistent with that conclusion, if the method loadings and method correlations are not *both* in their upper ranges. However, if the two types of method parameters simultaneously take on fairly large, but realistic, values, our findings show that CU estimates of the trait parameters can be substantially biased. Previous research had examined the effects of relatively large method factor loadings and the effects of relatively large method factor correlations but had not considered the two simultaneously.

Our study would be theoretically interesting, but of little practical importance, if the parameter values in our simulation represented extreme conditions that seldom arise in empirical research. We emphasize that this is not the case. All of the parameter values used in this research were based on our survey of the ranges of values that have appeared in published studies employing the MTMM approach (see Table 1). The high values in this study, representing approximately the 75th percentiles of those ranges, are hardly extreme. Therefore, it is reasonable to assume that the conditions we simulated in this study are realistic representations of the conditions present in MTMM research.

The most significant question this raises is whether the biases that we have demonstrated are serious enough to affect substantive research conclusions. The possibility that it has is demonstrated by a pair of articles by Lievens and Conway (2001) and Lance, Lambert, Gewin, Lievens, and Conway (2004). Lievens and Conway (2001) used the CU model to measure dimension- and exercise-related variance in a large number of assessment center (AC) studies. They concluded that, on average, AC ratings exhibit approximately equal amounts of dimension and exercise variance. This finding was somewhat at odds with conventional wisdom based on prior qualitative and quantitative AC research. A subsequent reanalysis (Lance et al., 2004), which used most of the same data but was based on other theoretically defensible CFA models (e.g., one-dimension correlated exercises and/or zero-dimension correlated exercises), supported the conventional thinking that exercise variance generally dominates over dimension variance.

The implications of the Lievens and Conway study for construct validity and for the best practical use of AC ratings are quite different from those of the Lance et al. study. Theoretically, the Lance et al. study suggests a lack of construct validity for AC dimension ratings (at least in terms of the traditional view that ACs are designed to measure stable dimensions), whereas the Lievens and Conway study provides more reason for optimism about that construct validity. On a practical level, the Lievens and Conway study suggests that, in some cases, it is appropriate to use dimensional AC ratings to provide assessees with feedback about their strengths and weaknesses. The Lance et al. results cast doubt on that, suggesting instead that AC ratings should only be used as indicators of overall performance.

The important point here is that the CU model led researchers to a different substantive conclusion than the other models did. The same kind of thing could happen in other areas of MTMM research anytime the CU model is used. Perhaps it already has. That leads to another important question. How should the MTMM researcher proceed, given the estimation problems associated with the CTCM model and the possible bias associated with the CU model?

Our results lend empirical support to the recommendation made by Lance et al. (2002) that researchers should prefer, on a theoretical basis, the CTCM model or some variant thereof (see Lance et al., 2002; Widaman, 1985) over the CU model. We reemphasize that the CTCM model is just one of a family of models, any one of which might be the most appropriate for a given situation. As discussed by Lance et al. (2002), other members of that family (e.g., the one-dimension correlated exercises model in Lance et al., in press) might well be more theoretically and empirically appropriate than is the CTCM model. Estimation problems associated with the CTCM model are sometimes eliminated by use of a different model from that group.

Our results also indicate that, all else being equal, larger MTMM matrices (i.e., those with more traits and methods) based on larger samples tend to converge and to yield admissible CTCM solutions more frequently than do smaller matrices, matrices based on relatively small samples, or both. Therefore, we suggest that researchers incorporate more traits and methods into their studies and that, if possible, they gather samples considerably larger than the typical size of about 125. Performance in our $N = 500$ condition was clearly superior to the $N = 125$.

Of course, it is still possible, even with larger matrices, larger samples, and more theoretically appropriate models, that the researcher will still be plagued with estimation problems. If the researcher encounters persistent estimation problems, or if the CTCM solution contains unduly large standard errors, the CU model might well provide a plausible solution. The question then becomes one of whether or not that solution is accurate.

Our study suggests that when the true method factor loadings *or* method factor correlations are small to moderate, CU estimates of the trait parameters are likely to exhibit only relatively small positive biases. In those cases, it would certainly be reasonable to argue that having a solution with minor biases is superior to having no solution at all. Unfortunately, of course, outside of simulation studies the true values of the method parameters are never known.

Assuming that the CTCM and related models fail and that the CU model yields a convergent and admissible solution, we suggest one criterion that researchers might use to assess the potential accuracy of the estimated trait parameters in a CU solution. We have shown that the CU model is theoretically most appropriate (or perhaps least inappropriate) in cases in which intercorrelations among the method factors or method factor loadings are low. Although the percentage of cases (see Table 2) in which the CU model yields appropriate solutions does not vary greatly according to the magnitudes of the method correlations, the percentage of admissible solutions that show good fit increases dramatically as the method correlations move down toward zero. This is true for all matrix sizes and for both sample sizes. This means that the great majority of CU solutions in our simulation that exhibited both admissible parameter estimates *and* good overall fit statistics came from the low method correlations condition. We believe this fact can be used as follows in assessing the likely accuracy of a CU solution.

If the CTCM and related models fail, but the CU model succeeds in returning an admissible solution, then the researcher should check the goodness-of-fit statistics (SRMSR, TLI, and RMSEA). An admissible CU solution that also has acceptable levels of each of these goodness-of-fit statistics is likely to be the product of a situation in which the intermethod correlations were low (i.e., a situation in which the CU method is most likely to yield the least biased results).

Limitations and Suggestions for Future Research

One limitation of our simulation study is that we do not know precisely how well our results generalize to real research settings. For example, generalizability will depend on the correctness of our assumption that the CTCM model holds for the population. We generated our data based on a CTCM population model, and our results are valid to the extent that this actually occurs in real data. However, just as we never know the true levels of method correlations or method loadings in real data, we never know exactly what population model is true. For example, Lance et al. (2004) argued that more restricted models are theoretically defensible for AC data. Still, we believe that whenever there are correlated method factors (which we believe is likely in real data) the CU model will tend to overestimate trait parameters.

Another issue is our choice of parameter values. We believe the mean values generalize well, but different research situations will vary in how the loadings and correlations are distributed around the means. For example, we observed the highest bias in loadings with the highest parameters, and that effect may decrease if the spread of the loadings is less than in our simulation. Nevertheless, we are confident that our overall conclusion about bias in CU trait estimates will generalize to actual research settings.

We did not look at bias in method factor loadings or correlations. This was because our theoretical rationale focused on upward bias in trait parameter estimates. However, we could also speculate on downward bias in method estimates, and this could be a useful path for future research. Scullen's (1999) technique for estimating the proportion of method variance for CU solutions could be used to see how often out-of-range estimates appear and how biased the method loadings are.

SUMMARY

Although prior simulation studies have suggested that the CU model yields only trivially biased estimates of trait factor loadings and trait factor correlations, our simulation study shows that this is not true in all cases. In cases in which method factors are both fairly substantial and highly intercorrelated, the trait parameters are likely to be positively biased. This has the potential to lead to inappropriate conclusions regarding construct validity. It is important for researchers to be wary of CU model estimates of trait parameters.

We urge researchers to consider the CTCM or other theoretically appropriate models from Widaman's (1985) taxonomy before turning to the CU model. Our results suggest a number of strategies that researchers can use to increase their likelihood of success with the CTCM and related models, as well as some strategies that might be useful for evaluating a solution in those cases in which the CU model is the only alternative.

REFERENCES

Becker, T. E., & Cote, J. A. (1994). Additive and multiplicative effects in applied psychological research: An empirical assessment of three models. *Journal of Management, 20,* 625–641.

Buckley, M. R., Cote, J. A., & Comstock, S. M. (1990). Measurement errors in the behavioral sciences: The case of personality/attitude research. *Educational and Psychological Measurement, 50,* 447–474.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56,* 81–106.

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

Conway, J. M. (1996). Analysis and design of multitrait-multirater performance appraisal studies. *Journal of Management, 22,* 139–162.

Cote, J. A., & Buckley, M. R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24,* 315–318.

Cudeck, R. (1988). Multiplicative models and MTMM matrices. *Journal of Educational Statistics, 13,* 131–147.

Hernandez, A., & Gonzalez-Roma, V. (2002). Analysis of multitrait-multioccasion data: Additive versus multiplicative models. *Multivariate Behavioral Research, 37,* 59–87.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

James, L. R. (1980). The unmeasured variables problem in path analysis. *Journal of Applied Psychology, 65,* 415–421.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8 user's guide.* Chicago: Scientific Software International.

Jöreskog, K. G., & Sörbom, D. (1996). *PRELIS 2: User's reference guide.* Chicago: Scientific Software International.

Kalleberg, A. L., & Kluegel, J. R. (1975). Analysis of the multitrait-multimethod matrix: Some limitations and an alternative. *Journal of Applied Psychology, 60,* 1–9.

Kavanagh, M. J., MacKinney, A. C., & Wolins, L. (1971). Issues in managerial performance: Multitrait-multimethod analyses of ratings. *Psychological Bulletin, 75,* 34–49.

Kenny, D. A. (1976). An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology, 65,* 507–516.

Kenny, D. A. (1995). The multitrait-multimethod matrix: Design, analysis, and conceptual issues. In P. E. Shrout & S. T. Fiske (Eds.), *Personality, research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 111–124). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112,* 165–172.

Lance, C. E., Lambert, T., Gewin, A. G., Lievens, F. & Conway, J. M. (2004). Revised estimates of dimension and exercise variance components in assessment center post-exercise dimension ratings. *Journal of Applied Psychology, 88,* 50–66.

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods, 7,* 228–244.

Lehman, D. R. (1988). An alternative procedure for assessing convergent and discriminant validity. *Applied Psychological Measurement, 12,* 411–423.

Lievens, F., & Conway, J. M. (2001). Dimension and exercise variance in assessment center scores: A large-scale evaluation of multitrait-multimethod studies. *Journal of Applied Psychology, 86,* 1202–1222.

Lomax, R. G., & Algina, J. (1979). Comparison of two procedures for analyzing multitrait-multimethod matrices. *Journal of Educational Measurement, 16,* 177–186.

Marsh, H. W. (1989). Confirmatory factor analysis of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13,* 335–361.

Marsh, H. W. (1990). Confirmatory factor analysis of multitrait-multimethod data: The construct validation of multidimensional self-concept responses. *Journal of Personality, 58,* 661–692.

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15,* 47–70.

Marsh, H. W., Byrne, B. M., & Craven, R. (1992). Overcoming problems in confirmatory factor analysis of MTMM data: The correlated uniqueness model and factorial invariance. *Multivariate Behavioral Research, 27,* 489–507.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousand Oaks, CA: Sage.

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.

Messick, S. (1995). Validation of assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50,* 741–749.

Millsap, R. E. (1995). The statistical analysis of method effects in multitrait-multimethod data: A review. In P. E. Shrout & S. T. Fiske (Eds.), *Personality research, methods, and theory: A festschrift honoring Donald W. Fiske* (pp. 93–109). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling, 8,* 287–312.

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

Scullen, S. E. (1999). Using confirmatory factor analysis of correlated uniqueness to estimate method variance in multitrait-multimethod matrices. *Organizational Research Methods, 2,* 275–292.

Sternberg, R. J. (1992). *Psychological Bulletin'*s top 10 "hit parade." *Psychological Bulletin, 112,* 387–388.

Tomás, J. M., Hontangas, P. M., & Oliver, A. (2000). Linear confirmatory factor models to evaluate multitrait-multimethod matrices: The effects of number of indicators and correlation among methods. *Multivariate Behavioral Research, 35,* 469–499.

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

Widaman, K. F. (1992). Multitrait-multimethod models in aging research. *Experimental Aging Research, 18,* 185–201.

Williams, L. J., Cote, J. A., & Buckley, M. R. (1989). Lack of method variance in self-reported affect and perceptions at work: Reality or artifact? *Journal of Applied Psychology, 74,* 462–468.

Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. In G. A. Marcoulides & R. E. Schumacker (Eds.), *Advanced structural equation modeling: Issues and techniques* (pp. 7–56). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

APPENDIX

Trait factor loading parameters were chosen by specifying the constant (across matrices) mean value of .60 and then varying the loadings within-matrix by ±.20 around the mean value. For the 3 × 3 matrices (as shown later), the loadings within-matrix took on values of .40, .60, and .80. For the 5 × 3 matrices, values still ranged from .40 to .80 but in increments of .10 rather than .20. As shown later, trait loadings always averaged .60 within-trait as well as within-method.

Method factor loading parameters were specified in a similar way; loadings varied by ±.20 around the mean value for the condition (e.g., from .45 to .85 in the condition in which the mean method loading was .65). For the 3 × 3 matrices as shown later, loadings varied in increments of .20. For the 5 × 3 matrices the increments were of .10. Average loadings for a matrix were equal within-traits and within-methods. A final constraint on method loadings was that variables with high trait factor loadings (e.g., .80) were assigned relatively low method factor loadings (e.g., .45).

Specification of trait factor correlation parameters began by specifying the second-order loadings of the trait factors on a general factor to account for correlations among traits (following Marsh & Bailey, 1991). For the 3 × 3 condition we specified second-order loadings of .90, .70, and .55; these loadings produced the trait factor loadings shown later, with a mean of .505 (very close to our desired value of .50). For the 5 × 3 condition the second-order loadings were .9, .8, .7, .6, and .55, giving a mean correlation of exactly .50.

We used a similar approach to specifying method factor correlation parameters. Second-order loadings by condition were .90, .70, and .55 (high condition, mean factor correlation = .505), .80, .60, and .30 (medium condition, mean factor correlation = .30), and .50, .40, and .10 (low condition, mean factor correlation = .097).

Factor Loadings

	T1	*T2*	*T3*	*M1*	*M2*	*M3*
T1M1	.40			.85		
T2M1		.60		.65		
T3M1			.80	.45		
T1M2	.80				.45	
T2M2		.40			.85	
T3M2			.60		.65	
T1M3	.60					.65
T2M3		.80				.45
T3M3			.40			.85

Factor Correlations

	T1	*T2*	*T3*	*M1*	*M2*	*M3*
T1	1.000					
T2	.630	1.000				
T3	.495	.385	1.000			
M1	.000	.000	.000	1.000		
M2	.000	.000	.000	.630	1.000	
M3	.000	.000	.000	.495	.385	1.000

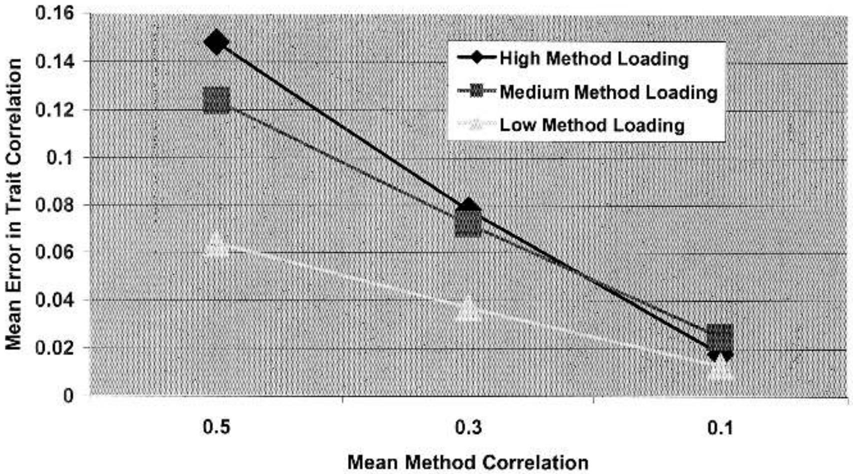

FIGURE 2 Mean errors for trait factor correlations.