Confirmatory factor analysis of multitrait and multimethod matrices: The effects of misspecification on estimates, solutions, and goodness-of-fit indicators.

Vinay Kanetkar, University of Guelph and
Martin G. Evans, University of Toronto.

May 16, 1996, Revised on July 22, 1997, Revised May 3, 2001

Running head: mtmm misspecification

Comments to: Martin G. Evans,
Rotman School of Management,
The University of Toronto,
105 St. George Street,
Toronto, ONT. M5S 3E6.
E-mail: evans@rotman.utoronto.ca

The authors would like to thank Karen Finlay, the journal editor and three anonymous reviewers for their comments and suggestions for improving this manuscript.

Confirmatory factor analysis of multitrait and multimethod matrices: The effects of misspecification on estimates, solutions, and goodness-of-fit indicators.

Abstract

To establish construct validity, the researcher is encouraged to gather data based upon multiple indicators for each trait based upon multiple methods. There have been, over the past twenty years continuing methodological advances to test the resulting multi-trait multi-method matrices (MTMM). Two of the most important are those of Widaman (1985) and Marsh & Bailey (1991). The latter, basing their position on earlier work by Kenny (1976), differ from Widaman in proposing that correlated uniqueness may be a better way of modeling methods effects than Widaman's suggestion of correlated methods factors which appears to result in improper solutions. In this paper we show the Marsh and Bailey result to be method bound. Marsh and Bailey use only one indicator for each trait-method combination. This it turns out is what results in the improper solutions when using Widaman's nested technique. When multiple measures of each trait-method combination are used, it is perfectly possible to get plausible solutions for Widaman's nested models. We also explore whether or not it is possible to distinguish between the Widaman and Marsh and Bailey models.

1 Introduction

Ever since Campbell & Fiske (1959) proposed their construct validation procedure based on multiple indicators, there have been growing analytical developments to test various hypotheses concerning multiple traits and multiple methods. They proposed the multitrait and multimethod (MTMM) design in which each of a set of multiple traits is assessed with each of a set of multiple methods. Their proposal required that indicators of the same construct be correlated higher than indicators of different constructs, including correlations derived from similar measurement methods, and that the patterns of intercorrelations between constructs measured in different ways be similar. Their guidelines and the MTMM design have become perhaps the most frequently employed construct validation methodology.

The problems associated with their guidelines are well known (see for example Alwin 1974, Campbell & O'Connell 1967, Marsh 1989) and have stimulated a number of alternative analytic approaches (for example, Browne 1984, Kenny & Kashy 1992, Marsh & Hocevar 1988, Widaman 1985) to provide evidence about the strength of traits or methods in a particular data set. In order to empirically evaluate alternative MTMM hypotheses one may use one or more of at least four basic analytical approaches. First, there is nested sequence of hypotheses proposed by Widaman (1985). The second analytical framework extended Widaman's (1985) framework by incorporating correlated uniqueness (for example, Kenny 1976, Kenny & Kashy 1992, Marsh 1989, Marsh & Bailey 1991); that is, models in which error terms of methods are correlated. The third analytic framework builds on the work of Campbell & O'Connell (1967) to formally represent the interaction of traits and methods (for example Browne 1984, Wothke & Browne 1990, Bagozzi & Yi 1992, Cudeck 1988, Dudgeon 1994).

The final analytic framework posits that traits and methods can be represented as second order factors especially when multiple indicators are available for every trait-method combination (see for example Marsh & Hocevar 1988).

We focus here on the first two confirmatory factor analytic (CFA) methods (that is Marsh 1989, Widaman 1985). Through simulation methodology, we intend to show that the superiority of Marsh & Bailey's (1991) approach (correlated uniquenesses) may have been an artifact of the restricted number of indicators that they used. We will demonstrate that when two or more indicators per trait-method combination are used to assess the traits and methods, Marsh & Bailey's (1991) technique of correlated uniqueness, although providing a theoretically plausible solution, provides inadequate goodness-of-fit indicators and estimated parameters.

However, simpler solutions based on Widaman (1985) are both theoretically plausible and empirically adequate. These models without correlated uniquenesses converge to satisfactory solutions, provide the opportunity to test alternative hypotheses concerning factor loadings of traits and methods, and provide the opportunity to test hypotheses about the inter-correlations among traits and the inter-correlations among methods.

2 Confirmatory factor analysis of the MTMM Matrices

Widaman (1985) proposed an important taxonomy for exploring the MTMM matrix in order to identify the strength of trait factors, the strength of methods factors and the interrelations between traits and between methods. Note that this model and Marsh's model assume the independence of traits and methods. In his taxonomy he presents sixteen models involving the following combinations of trait factors and method factors:

Traits

1. No trait factor
2. One trait factor loading all measures
3. Several orthogonal trait factors (T)
4. Several oblique trait factors (same factors as in 3, but allowed to correlate)

Methods

A. No method factors
B. A single method factor loading all measures[1]
C. Several orthogonal method factors (M)
D. Several oblique method factors (same factors as in C, but allowed to correlate)

This gives us the 16 combinations (1A, the null model, through 4D the model with unconstrained trait and method factors). These are summarized in Table 1. Many of these models are nested going from the top right to bottom left of the matrix and the null model is nested within the other models. This provides an opportunity to test whether more complex models provide better fit than simpler models as well as providing information (through factor loadings) of the strength of the trait and methods factors. To provide formal specification of these alternative models, Widaman (1985) proposed

$$\Sigma = XX' = \Lambda_{tt}\Phi_{tt}\Lambda_{tt}' + \Lambda_{mm}\Phi_{mm}\Lambda_{mm}' + \Theta_{\delta}, \qquad (1)$$

[1]Combination 2A and 1B are indistinguishable.

where

Σ is variance-covariance matrix formed from X matrix of p measured variables[2] and N observations,

Λ_{tt} is $p \times t$ matrix of factor loadings and t is number of traits,

Λ_{mm} is $p \times m$ matrix of factor loadings and m is number of methods used in the study,

Φ_{tt} and Φ_{mm} are $t \times t$ and $m \times m$ respectively matrices indicating trait and method variance-covariances, and

Θ_{δ} is $p \times p$ diagonal matrix indicating trait-method combination uniqueness.

Hypotheses about construct validity and the presence or absence of trait and methods factors can be assessed by varying the parameter structure of these matrices. These are explained briefly in the middle four columns of Table 1. There are five parameter matrices in Equation (1), namely, $\Lambda_{tt}, \Lambda_{mm}, \Phi_{tt}, \Phi_{mm}$ and Θ_{δ}. For model specifications A to D (see above) where there are no correlated uniquenesses, the Θ_{δ} matrix is set to be diagonal.

[2] Note that p is equal to $m \times t \times n$ where n is number of variables for each trait-method combination.

Although the Widaman's (1985) taxonomy provides a structure to understand competing hypotheses about traits and methods, a number of researchers (see for example Marsh 1989, Brannick & Spector 1990, Marsh & Bailey 1991) have reported that model structure 4D tended to provide parameter estimates that are not admissible. Marsh (1989) and Brannick & Spector (1990) compared a number of published MTMM matrices and concluded that improper parameter estimates (for example, negative variances and/or factor intercorrelations greater than 1.00) occurred more frequently than at chance levels. This led Marsh and his associates (Marsh 1989, Marsh & Bailey 1991) to propose an alternative method structure based on earlier work of Kenny (1976); this provided an analytic model that empirically gave reasonable estimates.

Marsh & Bailey (1991) argue that the method factor models C and D are too restricted. The implication of the constraints in these models are that the tests are congeneric. They argue for a less restrictive model: Model E in which the method effects are assessed by examining the correlated uniquenesses which represent the correlation between two variables measured with the same method after removing trait effects. This is described in the final column of Table 1. They make the following comment about the relationship between Model E and Model C:

> Method structure E corresponds most closely to method structure C, in that the methods 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 methods 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* $T > 3$. (p. 49, our italics)

Later, we shall show that this special case equivalence is also a function of the number of indicators used to identify each trait and each method. We came across this insight when undertaking analyses for our study on the impact of the number of indicators on the model fit of MTMM (Evans, Kanetkar, Cole, Kataoka & Skarlicki 1993).

Marsh & Bailey (1991) argue that model E used in both simulations and with real data provides a better fit to the data than model C which often results in improper solutions. In their arguments about the superiority of Model E, Marsh and Bailey followed the conventional practice in MTMM analysis of using a single indicator of each trait-method combination even though each trait and each method had three indicators as recommended by several authors (see for example Boomsma 1985, Anderson & Gerbing 1984, Gerbing & Anderson 1987). The recommendation for having three or more indicators are based on simple CFA simulations in which each indicator is only associated with a single factor. In the case of MTMM analyses each indicator is, by definition associated with at least two factors (one method and one trait). Accordingly the requirements of multiple indicators increase. We argue that multiple indicators of each trait-method combination is essential for the proper assessment of MTMM matrices. Consequently, Marsh & Bailey (1991) did not obtain proper solutions for their 3×3 and 4×4 matrices. We replicate some of Marsh & Bailey's (1991) results and extend the analyses to situations in which multiple indicators are used to assess each method-trait combination. The position taken here of allowing each indicator to load on its corresponding method and trait factor is an alternative to Marsh's (1993a) position of creating first order latent trait-method combinations which are then partitioned into second order trait and methods factors. However such models make rather unrealistic assumptions (see for example, Kumar & Dillon 1990, Bagozzi & Yi 1991)[3]. We believe that our method will prove to be superior in that this technique permits the comparison of models that are

[3] Marsh (1993b) has also compared multiple indicators versus a single indicator in the analysis of a simplex representing a 4-period set of observations of student ratings of their instructors.

specified with trait and method factors versus models that are specified with trait factors and correlated errors.

3 Methodology

Our purpose in this paper is to replicate[4] the findings of Marsh & Bailey (1991) and to extend them to situations in which more than one indicator was used to tap each trait-method combination.

To replicate Marsh & Bailey's (1991) simulated data set we generated multiple data sets based upon three underlying models: Model 4C which involved three correlated traits, three uncorrelated methods; Model 4D which involved three correlated traits, three correlated methods; Model 4E which involved three correlated traits and correlated uniquenesses for the methods. For each of these, we generated 100 data sets. In each data set there was one indicator for each trait-method combination. To extend the Marsh & Bailey (1991) analysis all of this was repeated in different data sets in which each trait-method combination was tapped by two indicator (measured) variables.

[4] We were particularly alarmed by Marsh & Bailey's (1991) conclusions that even when the data set was generated by model 4D, and the same model was estimated, they were able to recover parameters for only a small fraction of times. We think that the problem may have been with the choice of parameters as well as the data generating algorithm.

So the replication analyses of Marsh & Bailey (1991) had 9 measured variables that tapped three traits and three methods, each trait was tapped by three indicators, each method was tapped by three indicators, each indicator tapped a single trait-method combination, and each trait-method combination was tapped by a single indicator. In the extension analyses there were 18 measured variables that tapped the three traits and the three methods, each trait was tapped by six indicators, each method was tapped by six indicators. Each indicator tapped a single trait-method combination, and each trait-method combination was tapped by two indicators. A sample size of 200 was used for all analyses[5]. We then estimated all nineteen models[6] described in Table 1. We, however, report estimation results for six models, namely, 3C, 4C, 3D, 4D, 3E and 4E. In summary, we generated 600 different samples, with sample size of 200 and estimated 19 different models for each sample. Our summary here is partial and summarizes estimates based on 3,600 different estimation efforts.

To summarize the important findings from our study, we will do two things. First we report how well each of the fit indicators identifies the correct model;

[5]Data generation details are given in Appendix A

[6]Recall our earlier footnote that model combinations 1B and 2A are indistinguishable. This resulted in one less model.

that is how many correct hits do we get from the 100 replications when the data are generated by one model and estimated by each of the others. This gives a percentage hit rate. The second is to report the average parameter estimates (and their standard deviation) and the fit indices over the 100 replications. These can be compared to the population parameters.

Although there are number of goodness-of-fit indices that could be reported, we will restrict ourselves to four of these indices. These are χ^2, Bayesian information criteria (BIC), Bentler's comparative fit index (CFI) and root mean square error of approximation (RMSEA) developed by Steiger (1990). If the sample is multivariate normally distributed, then the expected value of χ^2 is equal to the degrees of freedom (t) and its variance is equal to two times the degrees of freedom. The χ^2 indicator of best fit has a number of limitations: one is the fact that with a large sample size it is easy to get a significant difference between the observed variance covariance matrix (S) and one based on the estimated parameters (Σ). The other limitation is that it lacks any penalty for increasing the number of parameters estimated in the model. In the kind of work we are doing here, comparing models with different numbers of parameters, this is an important limitation. We will also use BIC as a criterion[7]. This has the advantage of taking the number of parameters into account. Note that the expected value of BIC is $t - t\log(N)$ and variance is $2t$. This indicator is unbounded; that is it varies between $-\infty$ and ∞. The model with the lowest BIC value is the best fitting model. CFI is a comparative fit index; this ranges between 0 (poorly fitting model) and 1 (best fitting model). Finally, the value of RMSEA for a well fitting model should be 0 with a maximum of 1 for poorly fitting models. Additional rationales for the use of these indices can be found in (see for example Jaccard & Wan 1995, Browne & Cudeck 1993).

For our first analysis, the declaration of a hit was determined by a significant difference in χ^2 or BIC and by an interval of .01 for the CFI or RSMEA; when there was a tie, each model was given equal weight.

4 Results

In the single indicator case (see Table 2), χ^2 consistently selected the overparameterized model, 4D, no matter which model was used to generate the data (4C, 4D, 4E). BIC selected the underparameterized models (3D, 4E and 4C).

[7] BIC is $\chi^2 - t\log(N)$ where t is degrees of freedom and N is sample size.

RMSEA was similar to χ^2 in its choice of the overparameterized 4D model. Surprisingly, CFI was almost indifferent between the four tested models: 4C, 3D, 4D, 4E. All had hit rates of about 25% no matter which model was used to analyze the data. Put another way, the way in which the data is generated has little impact on which model is chosen by the fit index. A very discouraging finding for the single indicator model. In the two indicator case (see Table 3), despite our expectations, our findings are not much better. χ^2 always selects model 4E. BIC nearly always selects model 4C, especially for the 4C data generation model – the only other case in which any of these indicators identifies the correct model; the other being χ^2 and model generator 4E. Both RMSEA and CFI are indifferent between 4C, 4D, 4E. Let us turn to the aggregated data to see if a more informative result emerges.

The results for the second part of the simulation are presented in Tables 4 through 9. Each Table is set up with two parts: the upper portion of each Table reports the important parameter estimates; the bottom portion of each Table reports the average Goodness-of-Fit Indices for the corresponding model estimates. We will discuss first the extent to which the simulations produced inadmissible solutions, then talk about the parameter estimates, including within traits and within methods variances and covariances and the intercorrelations between the traits and between the methods.

Correlated Traits and Uncorrelated Methods

Table 4 shows the summary results for data generated with Model 4C (correlated traits and uncorrelated methods) with one indicator for each trait-method combination. First, models 3C, 3D, and 4D all have at one parameter providing an inadmissible estimate with models 3D and 4D having this problem over half the time.

Trait loadings are reproduced correctly by both 4C and 4E; 3C, 3D, and 3E tend to overestimate one pair of loadings; 4D underestimates one pair. Method loadings are fairly well estimated by the generating model, 4C, and by models 3C, 3D, and 4D; of course there are no such loadings for 3E and 4E. Trait variances and covariances are correctly estimated by models 4C and 4E, but the other models (3C, 3D, 4D, 3E) all show a downward bias. On the other hand, method variance, correctly estimated by 4C, shows an upward bias for models 3C, 3D, and 4D; there are no such estimates for 3E and 4E.

It is of interest to know whether models 4C, 4D and 4E recover trait correlations correctly. Since variances across models differ, estimated covariances cannot be used to reach such a conclusion. Using estimated trait variance-covariances, we would conclude that models 4C and 4E gave correlations between traits T_1 and T_2, T_1 and T_3 and T_2 and T_3 to be 0.8, 0.7 and 0.6 respectively which were exactly the same values used to simulate the data. On the other hand, the same correlations estimated using model 4D were 0.74, 0.56 and 0.47. This represents a substantial under estimation of trait correlations.

Table 5 shows the summary results for data generated with Model 4C (correlated traits and uncorrelated methods) with two indicators for each trait-method combination. First, there are few inadmissible solutions. These are found only in models 3D (2) and 4D (3).

Trait loadings are reproduced correctly by both 4C and 4E; the estimates for 4D are quite close, but those for 3C and 3D produce some overestimates and some underestimates for the trait loadings. Method loadings are fairly well estimated by the generating model, 4C, and by model 4D; the pattern of over and underestimation is found for models 3C and 3D; of course there are no such loadings for 3E and 4E. Trait variances and covariances are correctly estimated by models 4C and 4E; model 4D is close, but the other models (3C, 3D, 3E) all show a strong downward bias. On the other hand, method variance, correctly estimated by models 4C and 4D, shows an upward bias for models 3C and 3D; there are no such estimates for 3E and 4E. The recovery of correlations between traits is similar when the original model, 4C, and models

4D, and 4E are estimated.

To summarize, we find that use of two indicators for each trait-method combination provides superior factor loadings and factor correlations than the situation when one indicator is used. Although misspecified models produced factor loadings that are biased, the use of two indicators for each trait-method combination resulted in unbiased estimates for correlations among traits and among methods.

Correlated Traits and Correlated Methods

Tables 6 and 7 show the results of analysis when data were generated with Model 4D (correlated methods and correlated traits). In the data presented in Table 6 we used one indicator for the trait-method combination; in the data presented in Table 7 we used two indicators. The first thing to notice is that with one indicator, only models 3E and 4E had a complete set of admissible solutions. Even the correct model (4D) has more than 50% of its solutions inadmissible!

With more than half the samples generating solutions with problems using model 4D, there are severe effects on the parameter estimates and the variation across simulations. Model 4D does not reproduce the trait loadings at all well: one pair of parameters is downwardly biased. Models 4C and 4E do a reasonable job of reproducing the original parameter estimates; model 3C overestimates one parameter and underestimates another. We should also note that in 4D (and to a lesser extent 3C), the estimates tend to have high standard deviations compared to those produced by models 4C and 4E. The method loadings are correctly reproduced by all models (3C, 4C, 3D, 4D).

The estimated trait variances, however, are underestimated for model 4D and badly underestimated for models 3C, 3D, and 3E, and are overestimated for model 4E. In comparing method variances, we find overestimates for model 4D, 3C, and 3D; underestimates for model 4C, the reverse occurs. Estimated trait correlations (derived from variances and covariances) are slightly higher than population correlations for models 4C and 4E. We would conclude that 4C and 4E gave correlations between traits T_1 and T_2, T_1 and T_3 and T_2 and T_3 to be 0.85, 0.77 and 0.69 respectively while those produced by 4D model are 0.74, 0.55 and 0.47. While trait correlations are slightly understated by model 4D, method correlations are estimated to be 0.53, 0.48 and 0.35 where population values were set at 0.4, 0.3 and 0.2. This is a substantial overestimation. The reason for the poor performance of model 4D in this simulation is due to the large number of inadmissible solutions produced.

When we turn to the two indicator data set (Table 7), we find that nearly all models produced admissible solutions, except for model 3D with several problematic runs. There are substantial differences in the solutions produced by the two indicator data sets compared with the one indicator data sets. With the two indicator data set, we are not only able to produce estimates that are admissible but also estimates that are capable of distinguishing between alternative models. For models 4D and 4C, estimated factor loadings are well within one standard deviation of population parameters. The method loadings are correctly reproduced by the generating model 4D and by model 4C, the other two models do slightly less well. Model 4D reproduces the trait variances; models 3C, 3D, and 3E all greatly underestimate these variances, models 4C and 4E overestimate them. For method variances, 4D reproduces them well, 4C provides underestimates, and 3C and 3D overestimates them.

Turning to correlations among traits and among methods, we note that model 4D estimated trait correlations to be $\rho_{T_1,T_2} = 0.80$, $\rho_{T_1,T_3} = 0.69$, and $\rho_{T_2,T_3} = 0.58$. Moreover method correlations were $\rho_{M_1,M_2} = 0.42$, $\rho_{M_1,M_3} = 0.32$, and $\rho_{M_2,M_3} = 0.22$. All of these estimates are close to the population parameters. Estimated trait correlations are close to the population parameters for other models: both model 4C and 4E indicate that $\rho_{T_1,T_2} = 0.84$,

$\rho_{T_1,T_3} = 0.76$, and $\rho_{T_2,T_3} = 0.68$. This is encouraging because, even though the estimated models are misspecified, we would not draw false conclusions about the trait correlations.

Our conclusion that a one indicator data set generally may not provide valid estimates compared to two indicator data sets is reinforced in Tables 6 and 7. When population parameters are based on the 4D model with a single indicator and we estimated the 4D model, our estimates of the factor loadings have larger variances than those found in models 4C and 4E. This is particularly, alarming when we examine the implied trait correlations. Our population parameter ρ_{T_1,T_3} is 0.7 while the average estimated value for 100 replications is 0.53. This represents a substantial bias! This does not seem to be a problem when a two indicator data set is used (see Table 7).

Correlated Traits and Correlated Method Uniquenesses

Tables 8 and 9 show a summary of estimates from data generated with Model 4E (correlated uniquenesses and correlated traits). In the data presented in Table 8, we used one indicator for the trait-method combination; in Table 9 we used two indicators. The results are surprising. When the data are generated using model 4E with one indicator (Table 8), only models 3E and 4E have no inadmissible solutions. All other models may produce solutions with inadmissible estimates. Taking a closer look at trait and method factor loadings, we note that the trait loadings estimated using model 4E reproduce the population parameters; those for 3C, 3D, and 3E have several parameters with an upward bias, those for 4C and 4D are close to the population parameters. All models except 3E force the method effects (correlated uniquenesses) to be estimated by methods factors. The loadings for all except model 3D show high variation across our simulation runs. Model 4E and 4C reproduce the trait variances. All other models show underestimates. Turning to correlations among traits and among methods, we note that models 4E and 4C estimated trait correlations correctly when one indicator is used to measure the trait-method combination (Table 8). Trait correlations estimated by model 4D are $\rho_{T_1,T_2} = 0.73$, $\rho_{T_1,T_3} = 0.57$, and $\rho_{T_2,T_3} = 0.40$. Moreover method correlations are $\rho_{M_1,M_2} = 0.66$, $\rho_{M_1,M_3} = 0.68$, and $\rho_{M_2,M_3} = 0.64$.

Looking at the simulation based on two indicators again demonstrated the increased robustness of the solutions (Table 9). Although several models have more than 10 percent inadmissible solutions, they are not as bad as the single inidcator case. Model 4E reproduces the population parameters correctly. The parameter estimates produced by model 3E are close to the original parameter estimates. Model 3D produces a couple of trait estimates that are downwardly biased, an effect that is even stronger in Model 4D. The trait parameters produced by Model 3C and 4C show some major discrepancies from the original population parameters with many estimates changing sign. For all models that estimated method factors, the parameters show inconsistent patterns with a mix of positive and negative signs and high variances Despite this, when we look at trait correlations estimated by 4C, 4D and 4E with two indicators for each trait-method combination (Table 9), we find that estimated correlations are very close to population correlations.

The implications of the analyses of these six simulated data sets is clear. There is little information to be gained from the fit indices. The type of model chosen to be superior depends upon the fit index used rather than on the type of data used to generate the model. More information can be gathered from the parameter estimates themselves and by comparing them with the parameters of the model used to generate the data. This is a luxury only afforded to the simulation researcher. It is unavailable to the research with real data.

From our simulation, we see that the symmetry between the 4C and 4E models depends upon two conditions: first that the underlying data really conforms to the 4C model; second that only one indicator is used to assess a trait-method combination. If either of these conditions is not met the symmetry noted by Marsh & Bailey (1991) breaks down. When the data conforms to a correlated uniquenesses model (4E), models 4C and 4E fit the data when there is only one indicator; however when there are two indicators, only model 4E fits the data well (for model 4C, the trait loadings do not match those in the generated data. When the data conforms to the 4C model (correlated traits, uncorrelated methods) either a 4C or a 4E model will fit the data when either one or two indicators are used for each trait-method pair.

Unfortunately it is difficult to undertake this test with real data. Assuming that one had two indicators. One could only distinguish between models 4C and 4E by a careful examination of the trait loadings. If the true model was 4C, then both 4C and 4E will fit the data and the trait loading will be identical and will correspond to one's initial expectations about the sign of the loading. If the true model was 4E then both 4C and 4E might be selected on the basis of different fit indices, however the trait loadings would differ between the two models. Those in model 4E would be in line with prior expectations, those from model 4C would show uninterpretable and unexpected patterns. Only on this basis could model 4C be ruled out and model 4E accepted.

5 Conclusion and Discussion

In this paper we have built on the work of Widaman and of March and Bailey to explore the appropriateness of various ways of assessing MTMM matrices. Specifically we pitted the Marsh and Bailey method of estimating models with correlated uniquenesses against the earlier models of Widaman. We find that Marsh and Bailey were premature in advocating the abandonment of Widaman's uncorrelated methods models in favor of correlated uniquenesses. Their conclusion was dependent upon the fact that they only used one indicator for each trait-method combination. When additional indicators are used for each method-trait combination there is no problem in a) getting proper solutions for correlated methods models, and b) the estimates are accurate reproductions of the initial data on which the model was built.

Second, we find that fit indices are not very useful in distinguishing between the appropriateness of different models. The favored model is dependent more on the fit index than it is on the model used to generate the data.

Finally we provide the following advice for researchers. First you must use more than one indicator for each trait-method combination. This is merely a generalization of the advice to have multiple indicators of each factor, a generalization that must be fulfilled. When multiple indicators are used, the important correlations between the traits are estimated reasonably well by the 4C, 4D, and 4E models, so that a failure to assess the correct model is not fatal from the point of view of one's substantive research endeavours. When multiple indicators are used, and when the researcher has a reasonable understanding of the expected loadings of her/his indicators on the trait variables, it is possible to identify the model that best fits the data. We cannot do this on statistical grounds alone, the researcher's knowledge and judgment comes into play.

Finally, one reviewer asked what advice we would give to a researcher reanalyzing previously gathered data in which there was only one indicator per trait-method pair. This could arise in two situations. First from analysis of archival data and second from poor planning of once primary research. In such cases, Model 4E is attractive as it produces no inadmissable solutions. However, the results should be interpreted with caution: Model 4E reproduces correlations accurately if data are generated with models 4C or 4E; however they are overestimated if the data were generated by Model 4D.

Appendix A: Simulation Details

The data for equation (1) were generated as follows. Let M be a $p \times N$ matrix of independently distributed random normal variates with mean of zero and variance of one, where N is the sample size (in our case 200) and p number of variables to be simulated. Note that $p = m \times t \times n$ where m is number of methods, t is number of traits and n is number indicators for trait and method combination. In order to obtain the population variance-covariance matrix (Σ), trait loadings (Λ_{tt}), method loadings (Λ_{mm}), trait correlations (Φ_{tt}) and method correlations (Φ_{mm}) matrices were assumed as follows. For models 4C and 4D, matrix Λ_{tt} was

$$\Lambda_{tt} = \begin{bmatrix} \lambda_{1,T_1} = 1 & 0 & 0 \\ \lambda_{2,T_1} = .8 & 0 & 0 \\ \lambda_{3,T_1} = .8 & 0 & 0 \\ 0 & \lambda_{4,T_2} = 1 & 0 \\ 0 & \lambda_{5,T_2} = .7 & 0 \\ 0 & \lambda_{6,T_2} = .7 & 0 \\ 0 & 0 & \lambda_{7,T_3} = 1 \\ 0 & 0 & \lambda_{8,T_3} = .6 \\ 0 & 0 & \lambda_{9,T_3} = .6 \end{bmatrix},$$

and rows indicate measured indicators. Population parameters λ_{1,T_1}, λ_{4,T_2} and λ_{7,T_3} were fixed to 1 for identification purpose. Λ_{mm} matrix was

$$\Lambda_{mm} = \begin{bmatrix} \lambda_{1,M_1} = 1 & 0 & 0 \\ 0 & \lambda_{2,M_2} = 1 & 0 \\ 0 & 0 & \lambda_{3,M_3} = 1 \\ \lambda_{4,M_1} = .8 & 0 & 0 \\ 0 & \lambda_{5,M_2} = .7 & 0 \\ 0 & 0 & \lambda_{6,M_3} = .6 \\ \lambda_{7,M_1} = .8 & 0 & 0 \\ 0 & \lambda_{8,M_2} = .7 & 0 \\ 0 & 0 & \lambda_{9,M_3} = .6 \end{bmatrix},$$

Note that λ_{1,M_1}, λ_{2,M_2} and λ_{3,M_3} were all fixed to 1 for identification purpose.

All traits and methods variances were kept at one. Thus for model 4C, matrix Φ_{tt} was

$$\Phi_{tt} = \begin{bmatrix} 1 & .8 & .7 \\ .8 & 1 & .6 \\ .7 & .6 & 1 \end{bmatrix},$$

while matrix Φ_{mm} was identity. For model 4D, matrix Φ_{tt} was same as that for model 4C and matrix Φ_{mm} was

$$\Phi_{mm} = \begin{bmatrix} 1 & .4 & .3 \\ .4 & 1 & .2 \\ .3 & .2 & 1 \end{bmatrix}.$$

Finally, matrix Θ_δ was kept diagonal with all entries along diagonal were equal to 0.2.

For model 4E, matrices Λ_{mm} and Φ_{mm} are both zero and Θ_δ was set to

$$\Theta_\delta = \begin{bmatrix} .2 & 0 & 0 & .12 & 0 & 0 & .12 & 0 & 0 \\ 0 & .2 & 0 & 0 & .1 & 0 & 0 & .1 & 0 \\ 0 & 0 & .2 & 0 & 0 & .08 & 0 & 0 & .08 \\ .12 & 0 & 0 & .2 & 0 & 0 & .12 & 0 & 0 \\ 0 & .1 & 0 & 0 & .2 & 0 & 0 & .1 & 0 \\ 0 & 0 & .08 & 0 & 0 & .2 & 0 & 0 & .08 \\ .12 & 0 & 0 & .12 & 0 & 0 & .2 & 0 & 0 \\ 0 & .1 & 0 & 0 & .1 & 0 & 0 & .2 & 0 \\ 0 & 0 & .08 & 0 & 0 & .08 & 0 & 0 & .2 \end{bmatrix},$$

with the intention that uniqueness measurement error correlations were higher for the first method (.6) and the third method correlations were low (.4). By using various components of parameters, we computed the population variance-covariance matrix, Σ which is $p \times p$ symmetric matrix. This matrix was then decomposed such that $UU' = \Sigma$ where matrix U is a lower triangular matrix. To generate multivariate normal variates, we then computed matrix $X = UM$. This procedure of generating multivariate normal variates is consistent with the procedure suggested by Devroye (1986).

In situation where trait-method combination was tapped by two measured variables, we modified Λ_{tt} and Λ_{mm} matrices. For Λ_{tt}, the earlier population matrix was appended with three additional rows as the third row, three additional rows as the sixth and three additional rows as the ninth row. Similarly, Λ_{mm} was also appended by nine additional rows that replicated each of the fourth, fifth and sixth rows. All other matrices remained unchanged for model 4C and 4D.

All data generations, analyses, estimation and summary were carried out using SAS version 6.10. PROC IML was used to generate the sample of observations. These observations were then used by PROC CALIS for estimation purposes. A summary of the parameters and of goodness-of-fit indicators was carried out using PROC TABULATE. As one might imagine, the SAS code to accomplish all these tasks is extensive and this code is available from the first author.

References

Alwin, D. F. (1974), Approaches to the interpretation of relationship and the multitrait-multimethod matrix, in H. L. Costner, ed., 'Sociological Methodology 1973–74', Jossey-Bass, San Francisco, CA., pp. 79–105.

Anderson, J. C. & Gerbing, D. W. (1984), 'The effects of sampling error on convergence, improper solutions, and goodness-of-fit indices for maximum likelihood confirmatory factor analysis', Psychometrika **49**, 155–73.

Bagozzi, R. P. & Yi, Y. (1991), 'Multitrait-multimethod matrices in consumer research', The Journal of Consumer Research **17 (March)**, 426–39.

Bagozzi, R. P. & Yi, Y. (1992), 'Testing hypotheses about methods, traits, and communalities in the direct-product model', Applied Psychological Measurement **16 (December)**, 373–80.

Boomsma, A. (1985), 'Nonconvergence, improper solutions, and starting values in lisrel maximum likelihood estimation', Psychometrika **50**, 229–42.

Brannick, M. T. & Spector, P. E. (1990), 'Estimation problems in the block-diagonal model of the multitrait-multimethod matrix', Applied Psychological Measurement **14 (Dec.)**, 325–339.

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

Browne, M. W. & Cudeck, R. (1993), Alternative ways to assessing model fit, in K. A. Bollen & J. S. Long, eds, 'Testing Structural Equation Models', A Sage Focus Edition, Newbury Park, CA., pp. 136–62.

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. (1988), 'Multiplicative models and mtmm matrices', Journal of Educational Statistics **13**, 131–47.

Devroye, L. (1986), *Non-Uniform Random Variate Generation*, Springer-Verlag, New York:NY.

Dudgeon, P. (1994), 'A reparameterization of the restricted factor analysis model for multitrait-multimethod matrices', *British Journal of Mathematical and Statistical Psychology* **47**, 283–308.

Evans, M. G., Kanetkar, V., Cole, N., Kataoka, H. & Skarlicki, D. (1993), Confirmatory factor analysis of multitrait multimethod matrices: The effect of unbalanced trait and method indicators, Working Paper, University of Toronto.

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.

Jaccard, J. & Wan, C. K. (1995), 'Measurement error in the analysis of interactions effects between continuous predictors using multiple regression: Multiple indicator and structural equation approaches', *Psychological Bulletin* **117 (2)**, 348–57.

Kenny, D. A. (1976), 'An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix', *Journal of Experimental Social Psychology* **12**, 247–52.

Kenny, D. A. & Kashy, D. A. (1992), 'Analysis of multitrait-multimethod matrix by confirmatory factor analysis', *Psychological Bulletin* **112 (July)**, 165–72.

Kumar, A. & Dillon, W. R. (1990), 'On the use of confirmatory measurement models in the analysis of multiple-informant reports', *Journal of Marketing Research* **28(Feb.)**, 102–11.

Marsh, H. W. (1989), 'Confirmatory factor analyses of of multitrait-multimethod data: Many problems and a few solutions', *Applied Psychological Measurement* **13**, 335–61.

Marsh, H. W. (1993a), 'Multitrait-multimethod analyses: Inferring each trait-method combination with multiple indicators', *Applied Measurement in Education* **6**, 49–81.

Marsh, H. W. (1993b), 'Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models', *Journal of Educational Measurement* **30**, 157–83.

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

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 (Feb.)**, 107–17.

Steiger, J. H. (1990), 'Structural model evaluation and modification: An interval estimation approach', *Multivariate behavioral Research* **25**, 173–80.

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

Wothke, W. & Browne, M. W. (1990), 'The direct product model for the mtmm matrix parameterized as a second order factor analysis model', *Psychometrika* **55**, 255–62.

Table 1

Taxonomy and Model Specification for MTMM Data Adapted From Marsh

Trait Structure	Method Structure: A	B	C	D
1	1A: Null Model Θ_δ diagonal all other matrices 0	1B: 1 method factor Λ_{mm} is p × 1 Φ_{mm} is 1 × 1 $\Lambda_{tt} = 0$ $\Phi_{tt} = 0$	1C: M uncorrelated method factors Λ_{mm} is p × m Φ_{mm} is diagonal $\Lambda_{tt} = 0$ $\Phi_{tt} = 0$	1D: M correlated method factors Λ_{mm} is p × m Φ_{mm} is m × m $\Lambda_{tt} = 0$ $\Phi_{tt} = 0$
2	2A: 1 trait factor $\Lambda_{mm} = 0$ $\Phi_{mm} = 0$ Λ_{tt} is p × 1 Φ_{tt} is 1 × 1	2B: 2 general factors Λ_{mm} is p × 1 Φ_{mm} is 1 × 1 Λ_{tt} is p × 1 Φ_{tt} is 1 × 1	2C: 1 general trait and M uncorrelated method factors Λ_{mm} is p × m Φ_{mm} is diagonal Λ_{tt} is p × 1 Φ_{tt} is 1 × 1	2D: 1 general trait and M correlated method factors Λ_{mm} is p × m Φ_{mm} is m × m Λ_{tt} is p × 1 Φ_{tt} is 1 × 1
3	3A: T uncorrelated trait factors $\Lambda_{mm} = 0$ $\Phi_{mm} = 0$ Λ_{tt} is p × t Φ_{tt} is diagonal	3B: T uncorrelated trait factors and 1 method factor Λ_{mm} is p × 1 Φ_{mm} is 1 × 1 Λ_{tt} is p × t Φ_{tt} is diagonal	3C: T uncorrelated trait and M uncorrelated m-factors Λ_{mm} is p × m Φ_{mm} is diagonal Λ_{tt} is p × t Φ_{tt} is diagonal	3D: T uncorrelated trait and M correlated m-factor Λ_{mm} is p × m Φ_{mm} is m × m Λ_{tt} is p × t Φ_{tt} is diagonal
4	4A: T correlated trait factors $\Lambda_{mm} = 0$ $\Phi_{mm} = 0$ Λ_{tt} is p × t Φ_{tt} is t × t	4B: T correlated trait factors and 1 method factor Λ_{mm} is p × 1 Φ_{mm} is 1 × 1 Λ_{tt} is p × t Φ_{tt} is t × t	4C: T correlated trait and M uncorrelated m-factors Λ_{mm} is p × m Φ_{mm} is diagonal Λ_{tt} is p × t Φ_{tt} is t × t	4D: T correlated trait and M correlated m-factor Λ_{mm} is p × m Φ_{mm} is m × m Λ_{tt} is p × t Φ_{tt} is t × t

Model specifications A to D require that matrix Θ_δ be diagonal.

Table 2
Hit Rate for Alternative Models
for Simulated Data Based a Single Indicator
Number of replications 100

Simulated Model	Fit Indicator	Estimated using specification 3C	4C	3D	4D	3E	4E
4C	χ^2	–	24.50	14.00	37.00	–	24.50
4D	χ^2	–	19.75	18.25	42.25	–	19.75
4E	χ^2	–	26.25	2.75	44.25	–	19.75
4C	BIC	0.20	30.20	32.70	6.50	0.20	30.20
4D	BIC	–	26.92	36.58	10.08	–	26.42
4E	BIC	–	35.10	10.10	19.70	–	35.10
4C	RMSEA	–	22.33	15.83	39.50	–	22.33
4D	RMSEA	–	13.75	17.75	55.25	–	13.25
4E	RMSEA	–	22.33	2.00	54.33	–	21.33
4C	CFI	0.16	25.42	23.42	25.42	0.17	25.42
4D	CFI	–	25.33	24.00	25.33	–	25.33
4E	CFI	–	26.83	19.50	26.83	–	26.83

Table 3
Hit Rate for Alternative Models
for Simulated Data Based Two Indicators
Number of replications 100

Simulated Model	Fit Indicator	Estimated using specification 3C	4C	3D	4D	3E	4E
4C	χ^2	–	–	–	1.50	–	98.50
4D	χ^2	–	–	–	8.50	–	91.50
4E	χ^2	–	–	–	–	–	100.0
4C	BIC	–	93.00	2.00	5.00	–	–
4D	BIC	–	61.50	4.50	34.00	–	–
4E	BIC	–	69.17	1.16	26.17	–	3.50
4C	RMSEA	–	30.42	0.25	32.92	–	36.42
4D	RMSEA	–	22.17	2.00	54.16	–	21.67
4E	RMSEA	–	25.25	0.25	27.25	–	47.25
4C	CFI	–	30.58	8.25	29.58	–	31.58
4D	CFI	–	27.25	17.75	27.25	–	27.75
4E	CFI	–	25.92	6.75	29.92	–	37.42

Table 4
Population Parameters and their estimates
for Simulated Data Based on Model 4C with a Single Indicator
and alternative model specifications of MTMM
Based on 100 replications and sample size of 200
(standard deviations in parentheses)

Para-meter	Popul-ation	Estimated using specification 3C	4C	3D	4D	3E	4E
λ_{1,T_1}	1.[1]	1.	1.	1.	1.	1.	1.
λ_{2,T_1}	.8	1.07 (1.10)	0.80 (.11)	1.13 (3.31)	0.56 (1.93)	1.43 (2.76)	0.80 (.11)
λ_{3,T_1}	.8	0.88 (1.26)	0.79 (.11)	1.30 (3.14)	0.66 (1.14)	1.23 (2.26)	0.79 (.11)
λ_{4,T_2}	1.[1]	1.	1.	1.	1.	1.	1.
λ_{5,T_2}	.7	0.64 (.65)	0.70 (.07)	0.70 (.14)	0.70 (.12)	0.70 (.14)	0.70 (.07)
λ_{6,T_2}	.7	0.66 (.70)	0.70 (.08)	0.72 (.13)	0.70 (.13)	0.73 (.14)	0.70 (.08)
λ_{7,T_3}	1.[1]	1.	1.	1.	1.	1.	1.
λ_{8,T_3}	.6	0.58 (.12)	0.60 (.08)	0.59 (.12)	0.60 (.19)	0.57 (.12)	0.60 (.08)
λ_{9,T_3}	.6	0.59 (.11)	0.60 (.07)	0.60 (.11)	0.62 (.16)	0.59 (.11)	0.60 (.07)
λ_{1,M_1}	1.[1]	1.	1.	1.	1.	–	–
λ_{2,M_2}	1.[1]	1.	1.	1.	1.	–	–
λ_{3,M_3}	1.[1]	1.	1.	1.	1.	–	–
λ_{4,M_1}	.8	0.80 (.05)	0.82 (.10)	0.81 (.05)	0.81 (.10)	–	–
λ_{5,M_2}	.7	0.70 (.04)	0.70 (.06)	0.70 (.04)	0.70 (.05)	–	–
λ_{6,M_3}	.6	0.63 (.05)	0.60 (.06)	0.65 (.05)	0.61 (.06)	–	–
λ_{7,M_1}	.8	0.80 (.05)	0.81 (.09)	0.81 (.05)	0.83 (.10)	–	–
λ_{8,M_2}	.7	0.67 (.04)	0.70 (.06)	0.66 (.04)	0.69 (.05)	–	–
λ_{9,M_3}	.6	0.60 (.04)	0.61 (.06)	0.61 (.04)	0.60 (.06)	–	–
ϕ_{T_1,T_1}	1.	0.16 (.12)	1.03 (.22)	0.10 (.12)	0.67 (.66)	0.27 (.50)	1.03 (.22)
ϕ_{T_2,T_2}	1.	0.34 (.09)	1.01 (.19)	0.33 (.08)	0.78 (.46)	0.35 (.09)	1.01 (.19)
ϕ_{T_3,T_3}	1.	0.47 (.10)	1.02 (.19)	0.48 (.10)	0.75 (.49)	0.48 (.10)	1.02 (.19)
ϕ_{M_1,M_1}	1.	1.79 (.23)	1.00 (.21)	1.94 (.24)	1.37 (.67)	–	–
ϕ_{M_2,M_2}	1.	1.49 (.21)	1.01 (.16)	1.62 (.21)	1.27 (.42)	–	–
ϕ_{M_3,M_3}	1.	1.49 (.19)	1.01 (.15)	1.60 (.19)	1.29 (.47)	–	–
ϕ_{T_1,T_2}	.8	–	0.81 (.17)	–	0.51 (.53)	–	0.81 (.17)
ϕ_{T_1,T_3}	.7	–	0.72 (.16)	–	0.40 (.53)	–	0.72 (.16)
ϕ_{T_2,T_3}	.6	–	0.61 (.14)	–	0.36 (.44)	–	0.61 (.14)
ϕ_{M_1,M_2}	–	–	–	0.74 (.15)	0.32 (.48)	–	–
ϕ_{M_1,M_3}	–	–	–	0.76 (.16)	0.32 (.54)	–	–
ϕ_{M_2,M_3}	–	–	–	0.61 (.15)	0.28 (.40)	–	–

Comparison of Goodness-of-Fit Indicators

	3C	4C	3D	4D	3E	4E
χ^2	95.4 (22.8)	15.8 (6.1)	20.0 (6.7)	11.5 (4.9)	94.9 (22.2)	15.8 (6.1)
df	18	15	15	12	18	15
BIC	−1.4 (22.6)	−63.8 (6.1)	−62.7 (7.4)	−55.1 (5.0)	−0.6 (22.2)	−63.7 (6.1)
RMSEA	.14 (.02)	.02 (.02)	.03 (.03)	.01 (.02)	.15 (.02)	.02 (.02)
CFI	.945 (.01)	.998 (.00)[2]	.996 (.00)	.999 (.00)	.945 (.01)	.998 (.00)
NSB[3]	24	1	56	52	4	0

[1] Parameters fixed for identification purpose.
[2] Standard deviations less than 0.005 are reported as .00
[3] Number of samples out of 100 with at least one parameter estimate on the boundary.

Table 5
Population Parameters and their estimates for Simulated Data
Based on Model 4C with Two Indicators and alternative model specifications of MTMM
Based on 100 replications and sample size of 200
(standard deviations in parentheses)

Para-meter	Popul-ation	Estimated using specification 3C	4C	3D	4D	3E	4E
λ_{2,T_1}	.8	0.79 (.18)	0.81 (.07)	0.85 (.99)	0.81 (.08)	0.82 (.16)	0.81 (.05)
λ_{3,T_1}	.8	0.67 (.19)	0.81 (.07)	0.75 (1.25)	0.81 (.08)	0.81 (.21)	0.81 (.09)
λ_{5,T_2}	.7	0.69 (.10)	0.70 (.06)	0.69 (.11)	0.62 (.57)	0.64 (.09)	0.70 (.05)
λ_{6,T_2}	.7	0.70 (.10)	0.70 (.07)	0.70 (.11)	0.65 (.40)	0.68 (.12)	0.70 (.07)
λ_{8,T_3}	.6	0.59 (.07)	0.60 (.05)	0.62 (.07)	0.60 (.06)	0.58 (.08)	0.59 (.05)
λ_{9,T_3}	.6	0.61 (.09)	0.60 (.07)	0.62 (.08)	0.60 (.07)	0.59 (.08)	0.60 (.06)
λ_{10,T_1}	.8	0.81 (.14)	0.80 (.05)	0.93 (.95)	0.80 (.05)	1.01 (.25)	0.81 (.08)
λ_{11,T_1}	.8	1.01 (.22)	0.81 (.07)	1.22 (.95)	0.81 (.08)	0.68 (.20)	0.81 (.08)
λ_{12,T_1}	.8	0.99 (.21)	0.80 (.07)	1.17 (.70)	0.80 (.07)	0.99 (.25)	0.80 (.08)
λ_{13,T_2}	.7	0.62 (.09)	0.70 (.05)	0.65 (.10)	0.69 (.25)	0.68 (.10)	0.70 (.06)
λ_{14,T_2}	.7	0.69 (.09)	0.71 (.05)	0.69 (.10)	0.65 (.52)	0.71 (.12)	0.70 (.08)
λ_{15,T_2}	.7	0.70 (.10)	0.70 (.06)	0.70 (.11)	0.64 (.46)	0.71 (.12)	0.70 (.07)
λ_{16,T_3}	.6	0.57 (.08)	0.59 (.05)	0.61 (.08)	0.59 (.07)	0.59 (.09)	0.60 (.07)
λ_{17,T_3}	.6	0.59 (.08)	0.60 (.06)	0.62 (.08)	0.59 (.07)	0.61 (.09)	0.61 (.07)
λ_{18,T_3}	.6	0.60 (.08)	0.60 (.06)	0.61 (.08)	0.60 (.06)	0.60 (.09)	0.60 (.06)
λ_{4,M_1}	.8	0.82 (.04)	0.79 (.06)	0.83 (.04)	0.80 (.07)		
λ_{5,M_2}	.7	0.72 (.04)	0.71 (.05)	0.73 (.04)	0.71 (.05)		
λ_{6,M_3}	.6	0.64 (.04)	0.61 (.05)	0.66 (.04)	0.61 (.05)		
λ_{7,M_1}	.8	0.85 (.05)	0.79 (.06)	0.85 (.04)	0.80 (.06)		
λ_{8,M_2}	.7	0.69 (.04)	0.70 (.05)	0.69 (.04)	0.70 (.05)		
λ_{9,M_3}	.6	0.62 (.04)	0.60 (.05)	0.63 (.04)	0.60 (.05)		
λ_{10,M_1}	.8	0.80 (.04)	0.80 (.05)	0.80 (.03)	0.80 (.05)		
λ_{11,M_2}	.7	0.76 (.04)	0.70 (.05)	0.79 (.04)	0.70 (.06)		
λ_{12,M_3}	.6	0.69 (.05)	0.61 (.06)	0.73 (.05)	0.62 (.07)		
λ_{13,M_1}	.8	0.74 (.04)	0.80 (.05)	0.73 (.04)	0.80 (.05)		
λ_{14,M_2}	.7	0.72 (.04)	0.71 (.05)	0.73 (.04)	0.71 (.05)		
λ_{15,M_3}	.6	0.64 (.05)	0.60 (.05)	0.66 (.04)	0.61 (.06)		
λ_{16,M_1}	.8	0.72 (.04)	0.80 (.06)	0.69 (.04)	0.80 (.06)		
λ_{17,M_2}	.7	0.70 (.04)	0.72 (.05)	0.70 (.04)	0.72 (.06)		
λ_{18,M_3}	.6	0.62 (.04)	0.60 (.05)	0.63 (.04)	0.61 (.05)		
ϕ_{T_1,T_1}	1.	0.22 (.08)	0.99 (.20)	0.13 (.06)	0.97 (.25)	0.22 (.08)	0.99 (.21)
ϕ_{T_2,T_2}	1.	0.38 (.08)	0.99 (.17)	0.33 (.07)	0.96 (.23)	0.37 (.09)	0.99 (.18)
ϕ_{T_3,T_3}	1.	0.45 (.09)	0.99 (.16)	0.45 (.08)	0.97 (.19)	0.45 (.09)	0.98 (.17)
ϕ_{M_1,M_1}	1.	1.57 (.21)	1.01 (.15)	1.78 (.22)	1.04 (.27)		
ϕ_{M_2,M_2}	1.	1.35 (.17)	0.98 (.14)	1.50 (.18)	1.01 (.21)		
ϕ_{M_3,M_3}	1.	1.40 (.18)	1.01 (.15)	1.53 (.18)	1.02 (.23)		
ϕ_{T_1,T_2}	.8	–	0.79 (.16)	–	0.76 (.21)	–	0.79 (.16)
ϕ_{T_1,T_3}	.7	–	0.69 (.15)	–	0.66 (.20)	–	0.68 (.15)
ϕ_{T_2,T_3}	.6	–	0.59 (.14)	–	0.57 (.18)	–	0.59 (.14)
ϕ_{M_1,M_2}	–	–	–	0.66 (.16)	0.05 (.22)	–	
ϕ_{M_1,M_3}	–	–	–	0.69 (.17)	0.03 (.23)	–	
ϕ_{M_2,M_3}	–	–	–	0.60 (.14)	0.03 (.23)	–	

Comparison of Goodness-of-Fit-Indicators

	3C	4C	3D	4D	3E	4E
χ^2	237.9 (29.6)	119.4 (16.1)	165.5 (20.8)	117.1 (19.3)	208.6 (28.4)	92.0 (14.5)
df	117	114	114	111	90	87
BIC	−382.0(29.6)	−484.6(16.1)	−438.7(20.8)	−471.1(18.8)	−268.3(28.4)	−369.0(14.5)
RMSEA	.07 (.01)	.02 (.01)	.05 (.01)	.02 (.02)	.08 (.01)	.02 (.02)
CFI	.970 (.01)	.998 (.00)[2]	.987 (.01)	0.998 (.00)	.971 (.01)	.997 (.00)
NSB[3]	0	0	3	2	0	0

[1] Parameters fixed for identification purpose are omitted in this table.
[2] Standard deviations less than 0.005 are reported as .00
[3] Number of samples out of 100 with at least one parameter estimate on the boundary.

Table 6
Population Parameters and their estimates
for Simulated Data Based on Model 4D with a Single Indicator
and alternative model specifications of MTMM
Based on 100 replications and sample size of 200
(standard deviations in parentheses

Para-meter	Popul-ation	Estimated using specification 3C	4C	3D	4D	3E	4E
λ_{2,T_1}	.8	0.89 (1.51)	0.78 (.07)	2.05 (3.34)	0.46 (2.25)	1.20 (1.34)	0.78 (.07)
λ_{3,T_1}	.8	0.75 (1.57)	0.76 (.07)	1.45 (3.76)	0.44 (1.75)	1.12 (1.89)	0.76 (.07)
λ_{5,T_2}	.7	0.71 (.13)	0.68 (.06)	0.70 (.14)	0.69 (.11)	0.71 (.14)	0.68 (.06)
λ_{6,T_2}	.7	0.74 (.14)	0.64 (.06)	0.73 (.15)	0.70 (.12)	0.75 (.14)	0.64 (.06)
λ_{8,T_3}	.6	0.60 (.10)	0.61 (.06)	0.60 (.10)	0.64 (.25)	0.60 (.10)	0.61 (.06)
λ_{9,T_3}	.6	0.61 (.12)	0.57 (.06)	0.61 (.11)	0.61 (.16)	0.61 (.12)	0.57 (.06)
λ_{4,M_1}	.8	0.81 (.06)	0.81 (.18)	0.81 (.05)	0.80 (.11)	–	–
λ_{5,M_2}	.7	0.71 (.05)	0.70 (.08)	0.71 (.04)	0.70 (.08)	–	–
λ_{6,M_3}	.6	0.63 (.05)	0.61 (.07)	0.65 (.04)	0.61 (.10)	–	–
λ_{7,M_1}	.8	0.78 (.05)	0.79 (.16)	0.79 (.04)	0.84 (.18)	–	–
λ_{8,M_2}	.7	0.66 (.05)	0.69 (.07)	0.65 (.04)	0.69 (.08)	–	–
λ_{9,M_3}	.6	0.59 (.05)	0.60 (.07)	0.60 (.04)	0.60 (.09)	–	–
ϕ_{T_1,T_1}	1	0.18 (.11)	1.52 (.24)	0.08 (.10)	0.70 (.67)	0.23 (.33)	1.52 (.24)
ϕ_{T_2,T_2}	1	0.35 (.09)	1.37 (.19)	0.33 (.09)	0.82 (.43)	0.35 (.09)	1.38 (.19)
ϕ_{T_3,T_3}	1	0.48 (.12)	1.32 (.18)	0.48 (.11)	0.74 (.46)	0.48 (.12)	1.32 (.19)
ϕ_{M_1,M_1}	1	1.69 (.22)	0.48 (.16)	1.95 (.22)	1.34 (.66)	–	–
ϕ_{M_2,M_2}	1	1.40 (.20)	0.72 (.12)	1.62 (.19)	1.22 (.48)	–	–
ϕ_{M_2,M_2}	1	1.44 (.19)	0.86 (.15)	1.62 (.18)	1.26 (.52)	–	–
ϕ_{T_1,T_2}	.8	–	1.23 (.19)	–	0.56 (.51)	–	1.23 (.19)
ϕ_{T_1,T_3}	.7	–	1.10 (.17)	–	0.40 (.53)	–	1.10 (.17)
ϕ_{T_2,T_3}	.6	–	0.93 (.14)	–	0.37 (.42)	–	0.93 (.14)
ϕ_{M_1,M_2}	.4	–	–	1.16 (.17)	0.68 (.53)	–	–
ϕ_{M_1,M_3}	.3	–	–	1.08 (.16)	0.60 (.57)	–	–
ϕ_{M_2,M_3}	.2	–	–	0.81 (.13)	0.43 (.47)	–	–

Comparison of Goodness-of-Fit Indicators

χ^2	186.4 (28.0)	16.9 (5.9)	18.7 (6.5)	11.0 (4.5)	186.0 (27.7)	16.9 (5.9)
df	18	15	15	12	18	15
BIC	90.1 (27.9)	−62.8 (6.1)	−63.7 (6.7)	−57.1 (5.1)	90.6 (27.5)	−62.6 (5.9)
RMSEA	.22 (.02)	.02 (.02)	.03 (.03)	.01 (.02)	.22 (.02)	.03 (.02)
CFI	.887 (.02)	.998 (.00)[2]	.997 (.00)	.999 (.00)	.887 (.01)	.998 (.00)
NSB[3]	17	4	50	71	2	0

[1] Parameters fixed for identification purpose are omitted in this table.
[2] Standard deviations less than 0.005 are reported as .00
[3] Number of samples out of 100 with at least one parameter estimate on the boundary.

Table 7
Population Parameters and their estimates for Simulated Data
Based on Model 4D with Two Indicators and alternative model specifications of MTMM
Based on 100 replications and sample size of 200
(standard deviations in parentheses)

Para-meter	Popul-ation	Estimated using specification 3C	4C	3D	4D	3E	4E
λ_{2,T_1}	.8	0.78 (.18)	0.80 (.06)	0.83 (.87)	0.79 (.08)	0.81 (.16)	0.80 (.05)
λ_{3,T_1}	.8	0.73 (.18)	0.81 (.06)	0.61 (.63)	0.80 (.09)	0.80 (.23)	0.80 (.06)
λ_{5,T_2}	.7	0.64 (.58)	0.59 (1.18)	0.69 (.11)	0.62 (.59)	0.66 (.10)	0.75 (.05)
λ_{6,T_2}	.7	0.63 (.72)	0.45 (2.31)	0.70 (.10)	0.60 (.74)	0.70 (.11)	0.71 (.07)
λ_{8,T_3}	.6	0.59 (.08)	0.62 (.05)	0.62 (.08)	0.59 (.08)	0.61 (.10)	0.68 (.05)
λ_{9,T_3}	.6	0.61 (.07)	0.60 (.05)	0.62 (.07)	0.59 (.08)	0.59 (.09)	0.62 (.06)
λ_{10,T_1}	.8	0.81 (.15)	0.80 (.04)	0.91 (.85)	0.80 (.06)	1.00 (.25)	0.74 (.06)
λ_{11,T_1}	.8	0.97 (.18)	0.74 (.06)	1.42 (1.41)	0.78 (.08)	0.75 (.21)	0.81 (.07)
λ_{12,T_1}	.8	0.96 (.22)	0.73 (.06)	1.36 (1.06)	0.80 (.09)	0.97 (.24)	0.73 (.06)
λ_{13,T_2}	.7	0.61 (.34)	0.64 (1.14)	0.67 (.10)	0.72 (.16)	0.69 (.11)	0.70 (.07)
λ_{14,T_2}	.7	0.67 (.27)	0.59 (1.16)	0.69 (.11)	0.63 (.53)	0.70 (.11)	0.68 (.06)
λ_{15,T_2}	.7	0.62 (.95)	0.48 (2.05)	0.70 (.10)	0.62 (.67)	0.71 (.11)	0.68 (.06)
λ_{16,T_3}	.6	0.60 (.09)	0.67 (.05)	0.64 (.09)	0.61 (.08)	0.59 (.08)	0.62 (.05)
λ_{17,T_3}	.6	0.59 (.08)	0.62 (.05)	0.62 (.08)	0.59 (.07)	0.61 (.08)	0.60 (.05)
λ_{18,T_3}	.6	0.60 (.08)	0.60 (.05)	0.62 (.08)	0.61 (.18)	0.61 (.09)	0.60 (.05)
λ_{4,M_1}	.8	0.82 (.05)	0.79 (.09)	0.83 (.04)	0.81 (.06)		
λ_{5,M_2}	.7	0.72 (.05)	0.70 (.05)	0.73 (.04)	0.71 (.05)		
λ_{6,M_3}	.6	0.65 (.05)	0.62 (.07)	0.67 (.04)	0.61 (.07)		
λ_{7,M_1}	.8	0.83 (.06)	0.75 (.09)	0.84 (.05)	0.79 (.08)		
λ_{8,M_2}	.7	0.70 (.05)	0.71 (.06)	0.69 (.04)	0.71 (.06)		
λ_{9,M_3}	.6	0.62 (.04)	0.61 (.05)	0.63 (.04)	0.61 (.06)		
λ_{10,M_1}	.8	0.80 (.04)	0.80 (.08)	0.80 (.03)	0.80 (.05)		
λ_{11,M_2}	.7	0.74 (.05)	0.68 (.05)	0.79 (.04)	0.71 (.05)		
λ_{12,M_3}	.6	0.68 (.05)	0.62 (.07)	0.74 (.04)	0.61 (.08)		
λ_{13,M_1}	.8	0.75 (.05)	0.85 (.11)	0.71 (.04)	0.80 (.08)		
λ_{14,M_2}	.7	0.72 (.05)	0.70 (.06)	0.73 (.04)	0.71 (.06)		
λ_{15,M_3}	.6	0.64 (.05)	0.61 (.07)	0.66 (.04)	0.60 (.07)		
λ_{16,M_1}	.8	0.72 (.05)	0.85 (.10)	0.68 (.04)	0.80 (.07)		
λ_{17,M_2}	.7	0.69 (.05)	0.70 (.07)	0.68 (.04)	0.70 (.06)		
λ_{18,M_3}	.6	0.61 (.05)	0.61 (.06)	0.62 (.04)	0.61 (.07)		
ϕ_{T_1,T_1}	1.	0.26 (.08)	1.36 (.25)	0.11 (.07)	1.00 (.29)	0.26 (.20)	1.37 (.23)
ϕ_{T_2,T_2}	1.	0.40 (.10)	1.23 (.24)	0.33 (.07)	0.98 (.28)	0.40 (.09)	1.25 (.21)
ϕ_{T_3,T_3}	1.	0.47 (.08)	1.23 (.19)	0.46 (.07)	0.99 (.23)	0.47 (.08)	1.24 (.18)
ϕ_{M_1,M_1}	1.	1.46 (.22)	0.56 (.14)	1.83 (.23)	1.01 (.31)		
ϕ_{M_2,M_2}	1.	1.28 (.20)	0.74 (.13)	1.54 (.21)	1.02 (.23)		
ϕ_{M_3,M_3}	1.	1.32 (.19)	0.81 (.14)	1.54 (.19)	1.01 (.23)		
ϕ_{T_1,T_2}	.8	–	1.09 (.23)	–	0.79 (.26)	–	1.10 (.20)
ϕ_{T_1,T_3}	.7	–	0.99 (.19)	–	0.68 (.23)	–	1.00 (.18)
ϕ_{T_2,T_3}	.6	–	0.83 (.18)	–	0.57 (.21)	–	0.84 (.16)
ϕ_{M_1,M_2}	.4	–	–	1.08 (.16)	0.42 (.22)	–	
ϕ_{M_1,M_3}	.3	–	–	1.02 (.16)	0.32 (.24)	–	
ϕ_{M_2,M_3}	.2	–	–	0.81 (.15)	0.22 (.20)	–	

Comparison of Goodness-of-Fit-Indicators

	3C	4C	3D	4D	3E	4E
χ^2	322.5 (39.9)	128.0 (29.6)	151.7 (21.1)	114.9 (17.1)	290.9 (37.0)	96.7 (16.6)
df	117	114	114	111	90	87
BIC	−297.4(39.9)	−476.0(29.6)	−452.8 (21.1)	−473.4(16.7)	−186.0(37.0)	−364.3(16.6)
RMSEA	.09 (.01)	.02 (.02)	.04 (.01)	.01 (.01)	.11 (.01)	.02 (.02)
CFI	.950 (.01)	.996 (.01)	.991 (.00)[2]	0.998 (.00)	.951 (.01)	.997 (.00)
NSB[3]	0	0	10	5	1	0

[1] Parameters fixed for identification purpose are omitted in this table.
[2] Standard deviations less than 0.005 are reported as .00
[3] Number of samples out of 100 with at least one parameter estimate on the boundary.

Table 8
Population Parameters and their estimates
for Simulated Data Based on Model 4E with a Single Indicator
and alternative model specifications of MTMM
Based on 100 replications and sample size of 200
(standard deviations in parentheses)

Para-meter	Popul-ation	Estimated using specification 3C	4C	3D	4D	3E	4E
λ_{2,T_1}	.8	.86 (.07)	.80 (.04)	.99 (.42)	.81 (.10)	.86 (.06)	.80 (.04)
λ_{3,T_1}	.8	.88 (.07)	.81 (.05)	.89 (.28)	.80 (.09)	.88 (.07)	.81 (.05)
λ_{5,T_2}	.7	.73 (.05)	.70 (.04)	.68 (.09)	.69 (.08)	.73 (.05)	.70 (.04)
λ_{6,T_2}	.7	.74 (.05)	.70 (.04)	.69 (.09)	.69 (.08)	.75 (.05)	.70 (.04)
λ_{8,T_3}	.6	.62 (.05)	.60 (.03)	.59 (.07)	.60 (.21)	.62 (.05)	.60 (.03)
λ_{9,T_3}	.6	.64 (.05)	.61 (.04)	.60 (.06)	.61 (.15)	.64 (.05)	.61 (.04)
λ_{4,M_1}	–	.85 (1.27)	1.01 (.47)	.93 (.07)	.95 (1.01)	–	–
λ_{5,M_2}	–	.93 (.79)	1.01 (.19)	.83 (.06)	.98 (.18)	–	–
λ_{6,M_3}	–	1.04 (.29)	1.06 (.32)	.82 (.07)	1.18 (1.15)	–	–
λ_{7,M_1}	–	.88 (1.32)	.99 (.77)	.88 (.06)	1.05 (.61)	–	–
λ_{8,M_2}	–	.88 (1.29)	1.01 (.21)	.69 (.06)	1.00 (.27)	–	–
λ_{9,M_3}	–	1.04 (.27)	1.06 (.30)	.68 (.07)	1.07 (.60)	–	–
ϕ_{T_1,T_1}	1.	.73 (.12)	1.00 (.11)	.12 (.06)	.75 (.28)	.72 (.11)	1.00 (.11)
ϕ_{T_2,T_2}	1.	.76 (.11)	1.02 (.11)	.30 (.06)	.74 (.33)	.76 (.11)	1.02 (.12)
ϕ_{T_3,T_3}	1.	.76 (.11)	.99 (.10)	.46 (.08)	.64 (.39)	.75 (.10)	.99 (.10)
ϕ_{M_1,M_1}	–	.20 (.07)	.12 (.04)	.98 (.12)	.37 (.28)	–	–
ϕ_{M_2,M_2}	–	.12 (.04)	.10 (.03)	.66 (.09)	.28 (.17)	–	–
ϕ_{M_3,M_3}	–	.10 (.04)	.08 (.03)	.65 (.10)	.26 (.18)	–	–
ϕ_{T_1,T_2}	.8	–	.81 (.09)	–	.55 (.29)	–	.81 (.09)
ϕ_{T_1,T_3}	.7	–	.69 (.08)	–	.39 (.31)	–	.69 (.08)
ϕ_{T_2,T_3}	.6	–	.59 (.08)	–	.28 (.34)	–	.59 (.08)
ϕ_{M_1,M_2}	–	–	–	.66 (.09)	.21 (.21)	–	
ϕ_{M_1,M_3}	–	–	–	.67 (.10)	.21 (.21)	–	
ϕ_{M_2,M_3}	–	–	–	.53 (.08)	.17 (.16)	–	

Comparison of Goodness-of-Fit-Indicators

χ^2	283.2 (31.9)	15.8 (5.9)	28.0 (7.9)	10.7 (5.0)	281.7 (31.9)	15.7 (5.7)
df	18	15	15	12	18	15
BIC	187.2 (31.8)	−64.0 (5.9)	−52.1 (8.1)	−57.4 (5.5)	186.3 (31.9)	−63.8 (5.7)
RMSEA	.27 (.02)	.02 (.02)	.06 (.02)	.01 (.02)	.27 (.02)	.02 (.02)
CFI	.844 (.01)	.998 (.00)[2]	.992 (.00)	1.00 (.00)	.845 (.01)	.998 (.00)
NSB[3]	9	7	12	71	0	0

[1] Parameters fixed for identification purpose are omitted in this table.
[2] Standard deviations less than 0.005 are reported as .00
[3] Number of samples out of 100 with at least one parameter estimate on the boundary.

Table 9
Population Parameters and their estimates for Simulated Data
Based on Model 4E with Two Indicators and alternative model specifications of MTMM
Based on 100 replications and sample size of 200
(standard deviations in parentheses)

Para-meter	Popul-ation	Estimated using specification 3C	4C	3D	4D	3E	4E
λ_{2,T_1}	.8	0.78 (.08)	0.78 (.05)	0.75 (.16)	0.77 (.06)	0.78 (.04)	0.80 (.03)
λ_{3,T_1}	.8	0.79 (.09)	0.78 (.06)	0.77 (.17)	0.77 (.07)	0.82 (.05)	0.80 (.04)
λ_{5,T_2}	.7	0.04 (2.91)	-0.78 (5.22)	0.68 (.50)	0.31 (2.27)	0.67 (.03)	0.70 (.03)
λ_{6,T_2}	.7	0.01 (2.99)	-0.84 (5.28)	0.62 (.39)	0.18 (2.81)	0.71 (.04)	0.70 (.03)
λ_{8,T_3}	.6	0.60 (.04)	0.60 (.04)	0.60 (.06)	0.60 (.05)	0.57 (.03)	0.60 (.03)
λ_{9,T_3}	.6	0.60 (.04)	0.60 (.04)	0.59 (.05)	0.58 (.05)	0.60 (.05)	0.60 (.04)
λ_{10,T_1}	.8	0.77 (.04)	0.80 (.03)	0.75 (.13)	0.80 (.04)	0.82 (.05)	0.80 (.04)
λ_{11,T_1}	.8	0.78 (.08)	0.79 (.05)	0.76 (.14)	0.78 (.06)	0.83 (.05)	0.79 (.04)
λ_{12,T_1}	.8	0.79 (.09)	0.79 (.06)	0.76 (.15)	0.78 (.06)	0.83 (.05)	0.80 (.04)
λ_{13,T_2}	.7	0.38 (1.44)	-0.12 (3.59)	0.71 (1.14)	0.66 (1.52)	0.71 (.04)	0.69 (.04)
λ_{14,T_2}	.7	0.01 (2.91)	-0.77 (5.31)	0.57 (.62)	0.33 (2.15)	0.71 (.04)	0.70 (.04)
λ_{15,T_2}	.7	-0.01 (3.03)	-0.80 (5.32)	0.61 (.68)	0.24 (2.27)	0.71 (.04)	0.70 (.03)
λ_{16,T_3}	.6	0.59 (.04)	0.61 (.03)	0.62 (.04)	0.62 (.05)	0.60 (.04)	0.60 (.04)
λ_{17,T_3}	.6	0.59 (.04)	0.60 (.04)	0.59 (.05)	0.59 (.05)	0.61 (.05)	0.60 (.03)
λ_{18,T_3}	.6	0.60 (.04)	0.60 (.04)	0.60 (.06)	0.59 (.05)	0.61 (.05)	0.60 (.04)
λ_{4,M_1}	–	-1.75 (6.82)	-0.96 (6.28)	-1.30 (5.62)	-0.19 (4.33)		
λ_{5,M_2}	–	1.00 (.13)	1.01 (.13)	0.99 (.26)	0.99 (.16)		
λ_{6,M_3}	–	1.02 (.22)	1.01 (.20)	0.98 (.27)	0.98 (.17)		
λ_{7,M_1}	–	-0.98 (4.22)	0.40 (2.22)	-0.79 (5.26)	0.47 (2.67)		
λ_{8,M_2}	–	0.97 (.14)	0.98 (.13)	0.86 (.16)	0.96 (.13)		
λ_{9,M_3}	–	0.99 (.19)	0.99 (.19)	0.86 (.18)	0.97 (.16)		
λ_{10,M_1}	–	-0.45 (2.98)	0.08 (2.25)	-0.39 (3.08)	-0.04 (2.08)		
λ_{11,M_2}	–	1.01 (.13)	1.00 (.13)	1.00 (.08)	1.00 (.11)		
λ_{12,M_3}	–	1.02 (.15)	1.01 (.14)	1.00 (.09)	1.01 (.14)		
λ_{13,M_1}	–	-1.23 (5.04)	-0.26 (3.36)	-1.04 (4.24)	-0.56 (3.14)		
λ_{14,M_2}	–	1.01 (.16)	1.01 (.15)	0.99 (.26)	1.00 (.16)		
λ_{15,M_3}	–	1.01 (.24)	1.00 (.24)	0.97 (.25)	0.99 (.21)		
λ_{16,M_1}	–	-1.00 (3.79)	0.05 (2.49)	-0.83 (3.91)	-0.53 (2.89)		
λ_{17,M_2}	–	0.96 (.14)	0.96 (.14)	0.86 (.16)	0.94 (.13)		
λ_{18,M_3}	–	0.97 (.19)	0.97 (.17)	0.85 (.18)	0.94 (.17)		
ϕ_{T_1,T_1}	1.	0.86 (.21)	1.02 (.18)	0.53 0.47	0.97 0.26	0.78 (.12)	1.01 (.13)
ϕ_{T_2,T_2}	1.	0.79 (.25)	0.93 (.32)	0.48 (.38)	0.92 (.35)	0.80 (.11)	1.01 (.11)
ϕ_{T_3,T_3}	1.	0.83 (.14)	1.00 (.14)	0.64 (.29)	0.95 (.25)	0.81 (.11)	1.00 (.12)
ϕ_{M_1,M_1}	–	0.14 (.09)	0.11 (.09)	0.55 (.43)	0.16 (.24)		
ϕ_{M_2,M_2}	–	0.14 (.05)	0.12 (.06)	0.43 (.19)	0.19 (.13)		
ϕ_{M_3,M_3}	–	0.12 (.06)	0.11 (.06)	0.41 (.19)	0.17 (.13)		
ϕ_{T_1,T_2}	.8	–	0.75 (.27)	–	0.73 (.30)	–	0.81 (.11)
ϕ_{T_1,T_3}	.7	–	0.70 (.14)	–	0.65 (.23)	–	0.70 (.11)
ϕ_{T_2,T_3}	.6	–	0.56 (.21)	–	0.53 (.27)	–	0.59 (.09)
ϕ_{M_1,M_2}	–	–	–	0.36 (.29)	0.07 (.16)	–	
ϕ_{M_1,M_3}	–	–	–	0.35 (.29)	0.07 (.16)	–	
ϕ_{M_2,M_3}	–	–	–	0.32 (.18)	0.08 (.12)	–	

Comparison of Goodness-of-Fit-Indicators

	3C	4C	3D	4D	3E	4E
χ^2	438.9 (101.2)	168.5 (133.7)	287.1 (176.5)	138.8 (86.5)	368.6 (40.6)	88.1 (14.4)
df	117	114	114	111	90	87
BIC	−181.9(100.6)	−436.3(132.7)	−318.6 (174.8)	−451.5(85.7)	−108.2(40.6)	−372.9(14.4)
RMSEA	.12 (.02)	.03 (.04)	.08 (.04)	.02 (.03)	.12 (.01)	.01 (.01)
CFI	.931 (.02)	.988 (.03)	.963 (.04)	0.993 (.02)	.940 (.01)	.999 (.00)[2]
NSB[3]	16	15	28	41	0	0

[1] Parameters fixed for identification purpose are omitted in this table.
[2] Standard deviations less than 0.005 are reported as .00
[3] Number of samples out of 100 with at least one parameter estimate on the boundary.