

THE ANALYSIS OF MULTITRAIT-MULTIMETHOD MATRICES VIA CONSTRAINED COMPONENTS ANALYSIS

HENK A.L. KIERS

UNIVERSITY OF GRONINGEN

YOSHIO TAKANE

MCGILL UNIVERSITY

JOS M.F. TEN BERGE

UNIVERSITY OF GRONINGEN

Multitrait-Multimethod (MTMM) matrices are often analyzed by means of confirmatory factor analysis (CFA). However, fitting MTMM models often leads to improper solutions, or non-convergence. In an attempt to overcome these problems, various alternative CFA models have been proposed, but with none of these the problem of finding improper solutions was solved completely. In the present paper, an approach is proposed where improper solutions are ruled out altogether and convergence is guaranteed. The approach is based on constrained variants of components analysis (CA). Besides the fact that these methods do not give improper solutions, they have the advantage that they provide component scores which can later on be used to relate the components to external variables. The new methods are illustrated by means of simulated data, as well as empirical data sets.

Key words: principal components analysis, common factor analysis, alternating least squares algorithms, identification, uniqueness, partitioning of fit.

Introduction

Since Campbell and Fiske's (1959) proposal for assessing convergent and discriminant validity by studying multitrait-multimethod (MTMM) correlation matrices, many proposals have been made to improve their procedure. As an alternative to Campbell and Fiske's qualitative study of the MTMM matrices, it was suggested to give a quantitative description of the MTMM matrix by using a confirmatory factor analytic (CFA) model that distinguishes trait and method factors (e.g., Browne, 1984; Schmitt & Stults, 1986; Werts & Linn, 1970; Widaman, 1985). The idea is that, if such a model fits, we can assess the contributions of the separate trait and method factors and their intercorrelations. By comparing the fit of different models, one can assess the degree of convergent and discriminant validity and the amount of method variance (Widaman, 1985).

Fitting CFA models is often problematic. On the one hand, the CFA algorithm sometimes does not converge, or converges too slowly. A particular cause for such computational problems is underidentification of a model. On the other hand, the solution is often "improper" in the sense that the parameter estimates are incompatible with the CFA model. For instance, estimates for unique variances may be negative (such cases are called "Heywood cases"), or estimates of factor correlation matrices are indefinite. In such cases

This research has been made possible by a fellowship from the Royal Netherlands Academy of Arts and Sciences to the first author. The authors are obliged to three anonymous reviewers and an associate editor for constructive suggestions on the first version of this paper.

Requests for reprints should be sent to Henk A.L. Kiers, Department of Psychology (SPA), Grote Kruisstraat 2/1, 9712 TS Groningen, THE NETHERLANDS.

the estimates cannot describe (or even approximate) true variances of unique factors, or correlations between common factors, respectively.

When CFA is used to fit MTMM models, the above two types of problems seem to be very common. Specifically, among the MTMM models, a considerable number of models seems to be unidentified. For instance, it is well-known that the unconstrained MTMM models (specifying one trait and one method loading for each variable, and a diagonal matrix of unique variances) are not identified when there are less than three traits or methods. When there are three or more trait and method factors, or when certain correlations between trait and/or method factors are constrained to be zero, it is more complicated to describe which models are identified and which are not. Constraining more correlations to be zero tends to help identifying models, but even constraining all correlations to be zero does not safeguard against problems of nonconvergence and improper solutions (e.g., Marsh, 1989, p. 344). In fact, certain classes of models have been proven to be unidentified (e.g., see Grayson & Marsh, 1994; Millsap, 1992) due to particular patterns of values for the loadings. Other classes of these models are merely conjectured to be unidentified, as implied by Widaman's (1985, p. 7) remark that models allowing for non-zero correlations between trait and method factors "are very likely not identified" (see also a discussion by Marsh, 1989, p. 339). For models where method factors are constrained to be uncorrelated with trait factors, neither proofs nor systematic empirical results on identification seem available. For instance, when Wothke (1987) tried to fit such models to 23 empirical data sets, in 11 cases the models were proven to be unidentified, and in seven of the remaining cases the algorithm failed to converge (which indicates that the model is also unidentified in these cases). Moreover, in the other five, the solution was improper (i.e., Heywood case, and/or an indefinite matrix of estimates of factor correlations). Hence, for none of the 23 empirical data sets a usable solution was obtained. Similarly poor results were obtained in the analysis of 18 empirical data sets by Brannick and Spector (1990). It can be concluded that the most important problems of CFA methods, underidentification and the risk of finding improper solutions, are frequently encountered when fitting MTMM models.

The poor results of CFA to fit the most common MTMM models have led to the proposal to fit alternative CFA types of models to MTMM data. For instance, Kenny (1979; see also Marsh, 1989) proposed to adjust the CFA model by dropping the method factors and allowing for correlations between unique factors (referred to as the correlated uniqueness model) to capture correlations between traits within methods. Alternatively, Browne (1984) proposed a direct product model (DPM) which describes the correlations in the MTMM matrix as products of correlations between traits and correlations between methods, after adjustment for differential scalings and unique variances. As shown by Wothke and Browne (1990) this model can be seen as a CFA model as well. However, these alternatives are not satisfactory in all respects: the former alternative (using correlated uniqueness) considerably complicates the interpretation of the models, and the latter (DPM) tends to fit the data relatively poorly in practice (e.g., see Bagozzi & Yi, 1990, Browne 1993), which may to a large extent be caused by its use of relatively few parameters. Also, it lacks the matrix of trait loadings, which is an important tool for a detailed interpretation of the solution. Recently, Dudgeon (1994) proposed some models that share features of the original CFA approaches and the DPM approaches. However, the model depends on choosing a number of identification constraints, which requires substantive information, and hence cannot be used in a fully exploratory sense.

In the present paper, it is proposed to avoid the problems with CFA methods in an entirely different way. Rather than searching for alternative CFA models, it is proposed to use models outside the realm of confirmatory factor analysis. Specifically, it is proposed to use constrained component analysis models (see Takane, Kiers & de Leeuw, 1995), where

the constraints on the loading matrix are chosen in the same way as in the CFA models for MTMM data. By considering component analysis (CA) models, we avoid most of the problems of factor analytic techniques. For instance, using CA models, improper solutions can never occur. This is essentially because CA models are based on models for the full data matrix rather than for the correlation matrix. Another advantage of CA models is that identification of the model is not essential for (least squares) fitting of the model. Of course, it may not be very useful to interpret estimates for unidentified MTMM models, but the fact that we can assess the degree of fit of such models is useful in the comparison of a series of increasingly complex models.

The advantages of CA models are essentially obtained by modelling and fitting the data matrix rather than the correlation matrix. This seems to have an important drawback in that such methods would be limited to situations where the full data matrix is known. However, even though CA methods fit the full data matrix to a model, it suffices for most purposes to have the correlation matrix as input, as will be seen below. In fact, only in case component scores are specifically desired, it is necessary to give the full data matrix, but that would be the case for CFA approaches as well. In fact, here we touch on yet another advantage of CA approaches over CFA. In CA methods the component scores are related intrinsically to the model and the estimates correspond to the estimated loadings and component correlations. These component scores can, for instance, be used to assess validity of components by correlating them with external variables. In CFA, factor scores are indeterminate, and the usual factor score 'estimates' do not correspond exactly to the loadings and factor correlations. For instance, when we would compute correlations between the scores for the different factors, we would not obtain the factor correlations exactly.

Above, we have described a number of theoretical advantages of CA techniques over CFA techniques. However, these do not imply that CA techniques are generally preferable over CFA techniques. In particular, nothing has been said about the quality of parameter estimates from CA techniques compared to those from CFA techniques. To start with the latter, the CFA model is based on the assumption that each observation consists of a structural part and an error part (e.g., measurement error). The structural part of a score is based on an individual's scores on a set of latent factors, the error part is usually assumed to be a *random* measurement error. The CFA model fits the observed correlation or covariance matrix to the one specified by the model. If measurement error is indeed random, and the structural part of the model is as specified, the model fit will be almost perfect for a very large sample. Hence, when the assumptions are met, the model can be expected to fit well. Problems arise, however, when the error is not random, when the sample is too small, or when the model does not hold perfectly.

The CA model, on the other hand, is not based on assumptions pertaining to latent underlying factors. CA aims at giving a good description of a data set and uses a particular model (the MTMM model in the present case) for that purpose. As such, the method cannot be compared to CFA. CFA aims at recovering an underlying structure, whereas in CA no allusions to such a structure are made. In practice, however, especially in the present context of studying MTMM matrices, results often *are* interpreted with reference to underlying processes. Therefore, it seems sensible to study the performance of CA with respect to recovering underlying structures as well, and hence, interpret CA as a kind of CFA method.

Viewing CA as a kind of CFA method, we find that the CA model does not specify measurement error, and hence will only give perfect fit if indeed there is no measurement error. In cases with measurement error, one might hope that CA still finds good solutions, by relegating the measurement error to the unfitted part. Unfortunately, in practice, CA methods usually subsume some of the measurement error under the structural part, and

hence its loadings tend to be biased upward (see Widaman, 1990). Nevertheless, in many cases, component loadings are still useful and convey essentially the same information as do factor loadings. To see to what extent this holds true for the present application of CA methods, we conducted a simulation study focussing on the recovery of contrived underlying structures by the presently proposed CA methods.

Below, we will first explain a series of CA models for MTMM data. Then, we will describe algorithms for fitting these models. Next, we will give some uniqueness results for these models. It will be seen that the uniqueness results are much more powerful than in the CFA case. Then, it is described how the fit of most models can be partitioned over different components and variables. Next, we will describe a simulation study where the performance of the CA methods (in terms of computational efficiency and local optima) is studied and the quality of the estimates is compared to that of CFA models. Finally, we will present some examples demonstrating how CA methods can be used for obtaining evidence of convergent and discriminant validity from MTMM data.

Component Analysis Models for MTMM Data

A component analysis (CA) model for MTMM data can be defined as follows. Let \mathbf{Z} denote an $n \times mt$ data matrix with unit standardized (i.e., with zero means and sums of squares equal to 1 rather than n) scores of n subjects on mt variables measuring t different traits by m different methods. The variables are assumed to be ordered as $\mathbf{z}_{11}, \mathbf{z}_{21}, \dots, \mathbf{z}_{t1}, \mathbf{z}_{12}, \mathbf{z}_{22}, \dots, \mathbf{z}_{t2}, \dots, \mathbf{z}_{1m}, \dots, \mathbf{z}_{tm}$, where \mathbf{z}_{ij} denotes the variable measuring trait i by method j . In CA models such data are described by

$$\mathbf{Z} = \mathbf{UV}' + \mathbf{E}, \tag{1}$$

where \mathbf{U} denotes an $n \times r$ matrix with r unit standardized component scores, \mathbf{V} denotes an $mt \times r$ matrix of component loadings, and \mathbf{E} denotes an $n \times mt$ matrix of residuals. In the CA models for MTMM data proposed here, each component is restricted to refer to either one particular method or one particular trait, hence in the most complete model $r = m + t$, and \mathbf{V} is structured such that it has only two nonzero elements per row, corresponding to the positions of the trait and method component of the corresponding variable. Thus, the matrix \mathbf{V} is structured in the same way as the loading matrix in the most common CFA models for MTMM matrices (e.g., see Widaman, 1985). For instance, in case $m = 3$ and $t = 3$, we have

$$\mathbf{V} = \begin{pmatrix} v_{11} & 0 & 0 & v_{14} & 0 & 0 \\ 0 & v_{22} & 0 & v_{24} & 0 & 0 \\ 0 & 0 & v_{33} & v_{34} & 0 & 0 \\ v_{41} & 0 & 0 & 0 & v_{45} & 0 \\ 0 & v_{52} & 0 & 0 & v_{55} & 0 \\ 0 & 0 & v_{63} & 0 & v_{65} & 0 \\ v_{71} & 0 & 0 & 0 & 0 & v_{76} \\ 0 & v_{82} & 0 & 0 & 0 & v_{86} \\ 0 & 0 & v_{93} & 0 & 0 & v_{96} \end{pmatrix}. \tag{2}$$

A general description of \mathbf{V} is given by

$$\mathbf{V} = \begin{pmatrix} \mathbf{D}_1 & \mathbf{v}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \cdot & \mathbf{0} & & & \cdot \\ \cdot & \cdot & & & \mathbf{0} \\ \cdot & \cdot & & & \mathbf{0} \\ \mathbf{D}_m & \mathbf{0} & \dots & \mathbf{0} & \mathbf{v}_m \end{pmatrix}, \tag{3}$$

where $\mathbf{D}_1, \dots, \mathbf{D}_m$ denote m diagonal $t \times t$ matrices with trait loadings, and $\mathbf{v}_1, \dots, \mathbf{v}_m$ denote m t -vectors with method loadings. Several variants of the model can be obtained by dropping subsets of components (compare Widaman, 1985). For instance, models can be used that incorporate only the t trait components, or that incorporate only the m method components.

In ordinary CA, the component matrix \mathbf{U} is often constrained to be orthogonal, as can be done without affecting the fit of the model. This is because \mathbf{U} can always be transformed into an orthogonal matrix $\tilde{\mathbf{U}}$ by applying a nonsingular transformation (N say). If the associated \mathbf{V}' is premultiplied by the inverse of this transformation we get a $\tilde{\mathbf{V}} = \mathbf{V}(\mathbf{N}')^{-1}$ such that $\tilde{\mathbf{U}}\tilde{\mathbf{V}}' = \mathbf{U}\mathbf{N}\mathbf{N}^{-1}\mathbf{V}' = \mathbf{U}\mathbf{V}'$. This shows that the model is not affected by such nonsingular transformations of \mathbf{U} and \mathbf{V} . However, in models where \mathbf{V} is required to have zeros at specified positions, such transformations usually cannot be applied without affecting the fit. Therefore, constraining \mathbf{U} to be orthogonal would usually affect the fit of the model. Nevertheless, there may be theoretical reasons to constrain \mathbf{U} to be orthogonal, or partially orthogonal. For instance, in certain situations it is attractive to assume that correlations between method and trait components are zero. Alternatively, it may be attractive to assume that correlations between trait components are zero and/or correlations between method components are zero. All these assumptions can be used as constraints on the model parameters.

To sum up, we can describe the class of models under study as the models described by (1) with \mathbf{V} as in (3), in which there are t trait components and/or m method components (giving three different types of models). The different models can further be constrained by requiring that the trait or method components have mutual correlations of zero, and/or by requiring that the correlations between trait and method components are zero. In the next section, it will be described how these models can be fitted to the data (in the least squares sense) subject to any of these constraints. In a subsequent section, it will be demonstrated that the parameter estimates are unique under certain mild conditions.

Algorithms for Fitting Component Analysis Models to MTMM Data

In the present section, a class of algorithms will be described for fitting the above described models to a data set in the least squares sense. For each model a different algorithm is needed, but each algorithm has the same basic set up. Specifically, we propose to use alternating least squares algorithms in which the matrices \mathbf{U} and \mathbf{V} are updated alternately, until convergence of the loss function value. The updates are chosen such that they minimize the loss function over the parameter set at hand, considering the other parameters fixed. The differences of the algorithms come about by differences in the sets of components that are taken (only trait components, only method components, or both) and by differences in the constraints imposed on the correlations between the components. Below, we will first discuss how \mathbf{V} can be updated in the different situations, and next how \mathbf{U} can be updated, subject to the different constraints.

The procedures for updating \mathbf{V} , considering \mathbf{U} fixed, are straightforward. Denote the j -th column of \mathbf{Z} by \mathbf{z}_j and the j -th row of \mathbf{V} by \mathbf{v}_j' . For each situation, the problem of updating \mathbf{V} comes down to minimizing

$$\sigma(\mathbf{V}) = \|\mathbf{Z} - \mathbf{U}\mathbf{V}'\|^2 = \sum_{j=1}^{mt} \|\mathbf{z}_j - \mathbf{U}\mathbf{v}_j'\|^2, \quad (4)$$

subject to the constraint that certain elements of \mathbf{V} are zero. Minimizing σ over \mathbf{V} comes down to mt independent minimizations of the functions

$$\sigma_j(\mathbf{v}_j) = \|\mathbf{z}_j - \mathbf{U}\mathbf{v}_j\|^2, \quad (5)$$

$j = 1, \dots, mt$. In these problems, some of the elements of \mathbf{v}_j are constrained to be zero (as specified in (3)). If we collect the unconstrained elements of \mathbf{v}_j in $\tilde{\mathbf{v}}_j$ and the associated columns of \mathbf{U} in $\mathbf{U}_{(j)}$, we end up with the linear regression problem of minimizing

$$\sigma_j(\tilde{\mathbf{v}}_j) = \|\mathbf{z}_j - \mathbf{U}_{(j)}\tilde{\mathbf{v}}_j\|^2, \quad (6)$$

over $\tilde{\mathbf{v}}_j$ for which the solution is given by $\tilde{\mathbf{v}}_j = (\mathbf{U}'_{(j)}\mathbf{U}_{(j)})^{-1}\mathbf{U}'_{(j)}\mathbf{z}_j$. Here (and elsewhere) the inverse is replaced by the Moore-Penrose inverse if the inverse does not exist. The solution for \mathbf{v}_j can now be derived directly from $\tilde{\mathbf{v}}_j$.

For updating \mathbf{U} we cannot give one general procedure. We consider five different cases in detail. These cases partly correspond to those mentioned in Widaman's (1985, p. 6) taxonomy of MTMM models. Here it is assumed that both types of components (traits and methods) are present. For convenience we write $\mathbf{U} = (\mathbf{U}_t \mid \mathbf{U}_m)$ to distinguish the components associated with the traits and those with the methods. The cases where only one set of components is present follow directly as special cases. In all cases we impose the inactive identification constraint that \mathbf{U} has unit column sums of squares. This constraint is inactive because a scaling of the columns of \mathbf{U} can always be compensated by a scaling of the associated columns of \mathbf{V} . The five cases to be considered in detail are:

- Case 1. No active constraints on \mathbf{U} (with $\text{Diag}(\mathbf{U}'\mathbf{U}) = \mathbf{I}$ for identification).
- Case 2. $\mathbf{U}'_t\mathbf{U}_m = \mathbf{0}$ (with $\text{Diag}(\mathbf{U}'\mathbf{U}) = \mathbf{I}$ for identification).
- Case 3. $\mathbf{U}'_t\mathbf{U}_m = \mathbf{0}$ and $\mathbf{U}'_m\mathbf{U}_m = \mathbf{I}$ (with $\text{Diag}(\mathbf{U}'_t\mathbf{U}_t) = \mathbf{I}$ for identification).
- Case 4. $\mathbf{U}'_t\mathbf{U}_m = \mathbf{0}$ and $\mathbf{U}'_t\mathbf{U}_t = \mathbf{I}$ (with $\text{Diag}(\mathbf{U}'_m\mathbf{U}_m) = \mathbf{I}$ for identification).
- Case 5. $\mathbf{U}'_t\mathbf{U}_m = \mathbf{0}$, $\mathbf{U}'_t\mathbf{U}_t = \mathbf{I}$ and $\mathbf{U}'_m\mathbf{U}_m = \mathbf{I}$ (hence $\mathbf{U}'\mathbf{U} = \mathbf{I}$).

Case 1. No constraints on \mathbf{U} .

In this case, the function σ (the symbol σ is not only used to refer to a function of \mathbf{V} , as in (4), but also to a function of \mathbf{U}) is minimized over \mathbf{U} for given \mathbf{V} by $\mathbf{U} = \mathbf{Z}\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}$ since this is a multiple regression problem. The inactive constraint $\text{Diag}(\mathbf{U}'\mathbf{U}) = \mathbf{I}$ can be effected by normalizing the optimal \mathbf{U} columnwise. Although this changes the optimal \mathbf{U} by a rescaling, it does not affect the optimality of the $\mathbf{U}\mathbf{V}'$ after updating \mathbf{V} , because the update for \mathbf{V} compensates for the rescaling of \mathbf{U} .

Case 2. $\mathbf{U}'_t\mathbf{U}_m = \mathbf{0}$.

Every \mathbf{U} for which $\mathbf{U}'_t\mathbf{U}_m = \mathbf{0}$ can be written as $\mathbf{U} = \mathbf{X}\mathbf{B}$, where \mathbf{X} is a columnwise orthonormal matrix and $\mathbf{B} = \begin{pmatrix} \mathbf{B}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{B}_m \end{pmatrix}$ for arbitrary matrices \mathbf{B}_t and \mathbf{B}_m . This is because we can always write $\mathbf{U}_t = \mathbf{X}_t\mathbf{B}_t$ for a columnwise orthonormal \mathbf{X}_t and $\mathbf{U}_m = \mathbf{X}_m\mathbf{B}_m$ for a columnwise orthonormal \mathbf{X}_m , where \mathbf{X}_t is a basis for \mathbf{U}_t and \mathbf{X}_m is a basis for \mathbf{U}_m ; that \mathbf{X}_t can be taken orthogonal to \mathbf{X}_m (and hence \mathbf{X} is fully columnwise orthonormal) follows from $\mathbf{U}'_t\mathbf{U}_m = \mathbf{0}$. Rather than minimizing σ over \mathbf{U} , we will alternately minimize it over the constituents of \mathbf{U} , that is, over \mathbf{X} , \mathbf{B}_1 and \mathbf{B}_2 , as follows.

The function σ can be written as

$$\begin{aligned} \sigma(\mathbf{X}, \mathbf{B}_t, \mathbf{B}_m) &= \|\mathbf{Z} - \mathbf{X}\mathbf{B}\mathbf{V}'\|^2 \\ &= \|\mathbf{Z} - (\mathbf{X}_t\mathbf{B}_t \mid \mathbf{X}_m\mathbf{B}_m) \begin{pmatrix} \mathbf{V}'_t \\ \mathbf{V}'_m \end{pmatrix}\|^2, \\ &= \|\mathbf{Z} - \mathbf{X}_t\mathbf{B}_t\mathbf{V}'_t - \mathbf{X}_m\mathbf{B}_m\mathbf{V}'_m\|^2, \end{aligned} \quad (7)$$

where V_t and V_m contain the first t and the last m columns of V , respectively. We first consider the minimization of σ over X , subject to $X'X = I$, considering B and V fixed. This problem reduces to maximizing

$$f(X) = \text{tr } Z'XBV' = \text{tr } ZVB'X' \tag{8}$$

subject to $X'X = I$. The maximum of (8) is given by $X = PQ'$, where P and Q are obtained from the singular value decomposition (SVD) of ZVB' given as $ZVB' = PDQ'$. Next, the problem of minimizing σ over B_t and B_m given X and V is to be considered. For this purpose, we elaborate (7) as

$$\sigma(B_t, B_m) = \|Z - X_t B_t V_t' - X_m B_m V_m'\|^2. \tag{9}$$

The function σ is minimized over B_t by

$$B_t = X_t' Z V_t (V_t' V_t)^{-1}, \tag{10}$$

as follows from Penrose (1956) and the fact that $X_t' X_m = 0$. Similarly, σ is minimized over B_m by

$$B_m = X_m' Z V_m (V_m' V_m)^{-1}. \tag{11}$$

The resulting matrix U can now be computed as $U = (X_t B_t \mid X_m B_m)$, and if desired, normalized to unit length columnwise.

Case 3. $U_t' U_m = 0$ and $U_m' U_m = I$.

In this case, we can use the same procedure as in Case 2 to effectuate the constraint $U_t' U_m = 0$. That is, we write $U = XB$ and update X and B_t and B_m alternately. For updating B_t we can still use (11), and if desired, normalize XB_t to unit length columnwise. For updating B_m , we should now take the constraint $U_m' U_m = I$ into account. This constraint amounts to $B_m' X_m' X_m B_m = B_m' B_m = I$. Hence, for updating B_m , we have to minimize $\|Z - X_m B_m V_m'\|^2$ over B_m subject to $B_m' B_m = I$. This is equivalent to maximizing

$$g(B_m) = \text{tr } \underline{Z}' X_m B_m V_m' = \text{tr } X_m' Z V_m B_m' \tag{12}$$

over B_m , subject to $B_m' B_m = I$. This maximum is found for $B_m = P_m Q_m'$ with P_m and Q_m from the SVD $X_m' Z V_m = P_m D_m Q_m'$, see Cliff (1966). Again, U is found as $U = (X_t B_t \mid X_m B_m)$.

Case 4. $U_t' U_m = 0$ and $U_t' U_t = I$.

This case is analogous to Case 3, with the roles of traits and methods reversed. The updates for X and B_m are found as in Case 2, the update for B_t is given by $P_t Q_t'$ with P_t and Q_t from the SVD $X_t' Z V_t = P_t D_t Q_t'$, see Cliff (1966). Again, U is found as $U = (X_t B_t \mid X_m B_m)$.

Case 5. $U_t' U_m = 0$, $U_t' U_t = I$ and $U_m' U_m = I$ (hence $U'U = I$).

In this case $U = X$ (because $B = I$), hence the update for U can be found as the update for X in Case 2, with $B = I$.

Obviously, other combinations, like $U_t' U_t = I$ and $U_t' U_m$ free are conceivable alternatives. Such cases can be handled in a slightly different way. For this purpose we write the function σ as

$$\sigma(U, V) = \|Z - U_t V_t' - U_m V_m'\|^2 \tag{13}$$

and we update U_i (based on Cliff, 1966) and U_m (using multiple regression) alternately, considering the other parameters fixed.

This completes the description of the most pertinent situations for updating U . With all the above ingredients we can easily construct alternating least squares algorithms to fit the most relevant models to a data matrix Z . The algorithms should be initialized by matrices U and V that can either be chosen randomly (but satisfying the constraints), or in a very simple, but probably more rational way, as follows: Compute the principal component analysis solution of the data matrix; rotate the loading matrix orthogonally towards the indicator matrix that specifies to which component a variable belongs according to the MTMM model; rotate the component scores matrix by the same rotation; use the resulting component matrix and the resulting loading matrix (after setting to zero all loadings that are to be constrained to zero) as the "rational" start for the algorithm. A PCMATLAB program that performs all these analyses is available from the first author upon request.

The above described algorithm takes the full data matrix as a starting point. However, in practice we often only have the correlation matrix. In the next section, it will be shown that the above procedures can also be used when only the correlation matrix is given.

Sufficiency of the Correlation Matrix

In the previous section, we have described algorithms for updating U and V . In all cases, it turns out that U contains linear combinations of the columns of Z . This can be seen as follows. In Case 1 it is obvious from the updating formula. In Cases 2 through 5, U contains linear combinations of the columns of X . The update for X , in turn, is obtained as PQ' , where P and Q are taken from the SVD $ZVB' = PDQ'$. From the SVD it follows that P is in the column space of Z , hence so is $X = PQ'$, from which it follows at once that U contains linear combinations of the columns of Z . This result implies that the optimal U is in the column space of Z .

The fact that the optimal U is in the column space of Z can be exploited as follows. Rather than minimizing σ over arbitrary U , we can just as well minimize σ over matrices U that are in the column space of Z . Let $Z = Q_Z R_Z$ denote the QR-decomposition of Z . Then Q_Z gives an orthonormal basis for the column space of Z , and we can rewrite U as $U = Q_Z W$. Then the problem of minimizing σ over U in the column space of Z reduces to minimizing

$$\begin{aligned} \bar{\sigma}(W, V) &= \|Q_Z R_Z - Q_Z W V'\|^2 \\ &= \|Q_Z (R_Z - W V')\|^2 \\ &= \|R_Z - W V'\|^2. \end{aligned} \tag{14}$$

This function is to be minimized over W and V subject to the constraints on V and the constraints that $W'W = W'Q_Z'Q_Z W = U'U = \Phi$, where Φ is constrained to have a unit diagonal and possibly zeros in certain blocks, as implied by the constraints on U . The matrix R_Z in (14) can be obtained from the Cholesky decomposition of $Z'Z$.

From the above reasoning it follows that minimizing (14) over W and V gives the same minimal loss function value as for σ , and is attained for the same solution for V as would be obtained by minimizing σ . The solution for W corresponds to that of U by $U = Q_Z W$. In case we are only interested in the correlations between the factors, it suffices to compute $\Phi = W'W$. It follows that all iterative computations can be based on the correlation matrix $Z'Z$. In fact, the algorithm can be applied to the matrix R_Z instead of Z , and, if desired, the rational start can be computed by analogously replacing Z by R_Z .

Rotational Uniqueness

Above, we have described algorithms for the CA methods for fitting the MTMM models. These algorithms do not depend on identification of the models. Nevertheless, from a substantive point of view it is interesting to know whether or not a model is identified, and hence has unique estimates. In the present section, it will first be shown that sufficient conditions for “rotational” (non)uniqueness of CFA models are also sufficient for (non)uniqueness of corresponding CA models. Hence, all results that are available on rotational nonuniqueness of CFA models are directly relevant to the (non)uniqueness issue of CA models. We will discuss some of these results in the second part of this section.

Identification of CFA models is a very complicated issue, especially because of the presence of the unique variance parameters in Ψ , in the CFA model: $\Sigma = \Lambda\Phi\Lambda' + \Psi$, where Σ denotes the population covariance matrix, and Λ and Φ denote the loading and factor correlation matrices, respectively. In fact, we need conditions for identification of the unique variances, as well as for Λ and Φ . A sufficient condition for identification of the unique variances has been described by Anderson and Rubin (1956), but since it requires $mt > 2(m + t)$ it is often impractical with MTMM data (where e.g., $m \leq 2$, $t \leq 2$, $m = t = 3$ and $m = t = 4$ are common cases). Most results on uniqueness of CFA models pertain to necessary conditions for uniqueness. In particular, most results pertain to the uniqueness of the $\Lambda\Phi\Lambda'$ part of the model.

The $\Lambda\Phi\Lambda'$ part of the model is unique when no alternative matrices $\tilde{\Lambda}$ and $\tilde{\Phi}$ exist that satisfy the same constraints as Λ and Φ , and for which $\Lambda\Phi\Lambda' = \tilde{\Lambda}\tilde{\Phi}\tilde{\Lambda}'$. If Λ and Φ are unconstrained, any nonsingular T could be used to produce $\tilde{\Lambda} = \Lambda T$ and $\tilde{\Phi} = T^{-1}\Phi(T')^{-1}$ such that $\Lambda\Phi\Lambda' = \tilde{\Lambda}\tilde{\Phi}\tilde{\Lambda}'$. Hence, in the absence of constraints on Λ and Φ , (oblique) *rotations* exist that transform one solution into an equivalent one, and therefore such models are called “rotationally nonunique”. Algina (1980) has given necessary and sufficient conditions for rotational uniqueness in CFA. Millsap (1992) reformulated Algina’s uniqueness conditions for the MTMM model in which Φ is unconstrained. Grayson & Marsh (1994) gave a more general treatment of (non)uniqueness of the $\Lambda\Phi\Lambda'$ part, by considering certain MTMM models in which certain elements of Φ are constrained to zero (as in Cases 2 through 5).

As announced in the introduction of this section, the above mentioned results can be used for assessing uniqueness of the CA model as well. In CA models, the data are described as $Z = UV' + E$, where UV' forms the interesting part of the model. The model is *nonunique* if there is a set of alternative matrices \tilde{U} and \tilde{V} (differing from U and V by more than a trivial reflection) that satisfy the same constraints as imposed on U and V , and that yield the same loss function value; the model is *unique* if such alternative \tilde{U} and \tilde{V} cannot be found. It will now be shown that conditions which entail rotational (non)uniqueness in CFA, entail (non)uniqueness in CA.

Let Λ and Φ (assumed to be positive definite) be matrices that satisfy certain model constraints, and suppose matrices $\tilde{\Lambda}$ and $\tilde{\Phi}$ exist that satisfy the same constraints as Λ and Φ , respectively, and for which $\Lambda\Phi\Lambda' = \tilde{\Lambda}\tilde{\Phi}\tilde{\Lambda}'$. Hence, for this Λ and Φ the CFA model at hand is rotationally nonunique. Then for $V = \Lambda$ and any U chosen such that $U'U = \Phi$ (as can be done, e.g., by using the Cholesky decomposition of Φ), the corresponding CA model is also nonunique, as will now be proven. Let a (truncated) eigendecomposition of $\Lambda\Phi\Lambda' = \tilde{\Lambda}\tilde{\Phi}\tilde{\Lambda}'$ be given as $K_r D_r^2 K_r'$, where D_r^2 is the diagonal matrix with nonzero eigenvalues, and K_r is the columnwise orthonormal matrix with corresponding eigenvectors of $\Lambda\Phi\Lambda'$. Furthermore, let the Cholesky decomposition of $\tilde{\Phi}$ be given by $\tilde{\Phi} = C'C$, and let $N_1 \equiv UA'K_r D_r^{-1}$ and $N_2 \equiv C\tilde{\Lambda}'K_r D_r^{-1}$. From $N_1'N_1 = D_r^{-1}K_r'AU'UA'K_r D_r^{-1} = D_r^{-1}K_r'K_r D_r^2 K_r'K_r D_r^{-1} = I$, it follows that the $r \times r$ matrix N_1 is orthonormal, and similarly, from $N_2'N_2 = I$, it follows that N_2 is orthonormal. Hence, if we take $\tilde{U} = N_1 N_2' C$ and $\tilde{V} =$

$\bar{\Lambda}$, we have $\bar{U}'\bar{U} = C'C = \bar{\Phi}$, and $\bar{U}\bar{V}' = N_1N_2C\bar{\Lambda}' = UA'K_rD_r^{-2}K_r'\bar{\Lambda}C'\bar{\Lambda}' = UA'K_rD_r^{-2}K_r'\bar{\Lambda}\bar{\Phi}\bar{\Lambda}' = UA'K_rD_r^{-2}K_r'D_r^2K_r' = UA'K_rK_r' = UA' = UV'$, where it is used that $K_rK_r'\Lambda = \Lambda$, because K_r spans the column space of $\Lambda\Phi\Lambda'$, and Φ is positive definite. Thus it has been proven that in conditions where the CFA model is rotationally nonunique, the corresponding CA model is nonunique as well.

Conversely, suppose the CA model is nonunique, that is, suppose that, for given matrices U and V that satisfy certain constraints, there exist nontrivially different matrices \bar{U} and \bar{V} , that satisfy the same constraints, and for which $\bar{U}\bar{V}' = UV'$. Then the corresponding CFA model is nonunique as well, as can be proven as follows. Let $\Phi = U'U$ and $\Lambda = V$, then $\bar{\Phi} = \bar{U}'\bar{U}$ and $\bar{\Lambda} = \bar{V}$, $\bar{\Phi}$ and $\bar{\Lambda}$ satisfy the same constraints as Φ and Λ , and $\bar{\Lambda}\bar{\Phi}\bar{\Lambda}' = \bar{V}\bar{U}'\bar{U}\bar{V}' = VU'UV' = \Lambda\Phi\Lambda'$. Thus it follows from nonuniqueness of a CA model that the corresponding CFA model is rotationally nonunique. By implication, we have that, if a CFA model is unique, the corresponding CA model must be unique as well.

As mentioned above, sufficient conditions for uniqueness have already been given by Algina, and one of these was reformulated for the MTMM case by Millsap (1992, p. 128). This condition consists of two parts that both should hold. The first condition, the “nullity condition” (stating that at least $m + t - 1$ elements of each column of the loading matrix should be constrained to 0) is satisfied if and only if $t > 2$ and $m > 2$. The second condition is a special case of the uniqueness conditions provided by Kiers and Takane (1993). Because they prove that this condition is not only sufficient but also necessary, their result is somewhat stronger than Millsap’s. Applied to the MTMM models, this result can be formulated as Theorem 1 below.

Theorem 1. Let V_l be defined as the submatrix of V obtained by deleting the l -th column of V and the $(t$ or $m)$ rows of V corresponding to the unconstrained elements of the l -th column of V , $l = 1, \dots, r$. Let U and V have full column rank. If U is constrained such that $\text{Diag}(U'U) = I$ and V is constrained as in (3), then $UV' = \bar{U}\bar{V}'$ implies $\bar{U} = U\tau$ and $\bar{V} = V\tau$ (where τ is a sign matrix) if and only if all matrices V_l ($l = 1, \dots, r$) have full column rank.

The uniqueness conditions in Theorem 1 can only be verified after having obtained a solution for V , because it depends on the values in V . Kiers and Takane (1993) also offered a sufficient condition that can be evaluated in advance (depending only on the type of constraints, not on the specific values in V), but this condition is *never* satisfied in the MTMM models under study here. Millsap (1992) gave a number of corollaries to Theorem 1, that may be used to facilitate assessment of the uniqueness conditions. He also described a number of cases in which the MTMM models are rotationally nonunique, and hence in which the corresponding CA models are nonunique as well.

As is readily verified, the condition in Theorem 1 has a simple implication: If $m < 3$ or $t < 3$ the rank condition can never be satisfied. This can be seen as follows. The matrices V_1, \dots, V_t are of order $(mt - m) \times (m + t - 1)$ and the matrices V_{t+1}, \dots, V_{t+m} are of order $(mt - t) \times (m + t - 1)$. When $t < 3$, the former matrices have more columns than rows, and when $m < 3$, the latter have more columns than rows. In both cases certain matrices cannot have full column rank, and hence the rank condition is not satisfied. Therefore, models where U is unconstrained and $m < 3$ or $t < 3$ are always underidentified, which implies that for unconstrained U , the “nullity condition” is also necessary for uniqueness.

Grayson and Marsh (1994) also described conditions for rotational nonuniqueness, which are particularly interesting for constrained MTMM models (among which Cases 2 through 5). For instance, they have shown that the Case 2 model is rotationally nonunique when the nonzero elements of Λ can be written as $\lambda_{ij} = \alpha_i\beta_j$, $i = 1, \dots, mt$ and $j = 1, \dots,$

$m + t$. In fact, they have shown that this condition is equivalent to having all unconstrained elements of Λ nonzero and Λ being rank deficient.

For the models where \mathbf{U} is constrained as in Cases 2 through 5, the condition in Theorem 1 is still sufficient for uniqueness, but no longer necessary. In fact, it has been found (to be reported elsewhere) that Case 2 models with $t = 3$ and $m = 2$ are usually unique, hence the condition in terms of ranks of \mathbf{V}_l matrices of Theorem 1 (which is violated here) is clearly not necessary. Unfortunately, necessary and sufficient conditions for identification in these cases do not seem to be available. For Case 5, we can formulate a sufficient condition that is weaker than the one from Theorem 1. This condition is based on Algina's (1980) Theorem 3. As shown by Bekker (1986, p. 610), however, this condition is, despite Algina's claim, not necessary for rotational uniqueness. An alternative sufficient condition for uniqueness in Cases 3, 4 and 5 is given by Grayson and Marsh (1994, Theorem 5).

Although the present section does not offer general results on identification of CA methods for MTMM data, some conclusions can be made now: It has been seen that CA models are more often identified than CFA models, simply because they do not depend on identification of unique variances. In addition, identification of CA models is established easier than that of CFA models (due to the availability of simple sufficient conditions).

A final remark on the above uniqueness results for CA models is in order. The results all imply uniqueness of \mathbf{U} and \mathbf{V} , given that the error terms in $\mathbf{Z} = \mathbf{UV}' + \mathbf{E}$ are fixed. However, it is conceivable that different solutions for $\mathbf{E} = \mathbf{Z} - \mathbf{UV}'$ produce the same loss function value. In fact, the above uniqueness theorem only shows that the model \mathbf{UV}' is identified, given a set of error terms. This is similar to identification results in factor analysis, where even for identified models, it is conceivable that different parameter estimates yield the same (maximum) likelihood function value.

Partitioning of Fit

For cases where a solution is unique, it is of interest to see how much the different parts of the model contribute to the total fit of the data. In particular, it is interesting to distinguish contributions from the trait and the method components. Such a partitioning can readily be made for the Cases 2 through 5, as follows.

In (13), the loss function has been written as $\sigma(\mathbf{U}, \mathbf{V}) = \|\mathbf{Z} - \mathbf{U}_t \mathbf{V}'_t - \mathbf{U}_m \mathbf{V}'_m\|^2$. We define the "fit", or equivalently, the "explained variance" of the model, as the sum of the variances of the variables minus the unexplained variance. Hence the fit is

$$f(\mathbf{U}, \mathbf{V}) = \|\mathbf{Z}\|^2 - \|\mathbf{Z} - \mathbf{U}_t \mathbf{V}'_t - \mathbf{U}_m \mathbf{V}'_m\|^2. \tag{15}$$

For Cases 2 through 5, this function can be simplified, by using $\mathbf{U}'_t \mathbf{U}_m = \mathbf{0}$ as

$$\begin{aligned} f(\mathbf{U}, \mathbf{V}) &= -2 \operatorname{tr} \mathbf{Z}' \mathbf{U}_t \mathbf{V}'_t - 2 \operatorname{tr} \mathbf{Z}' \mathbf{U}_m \mathbf{V}'_m + \|\mathbf{U}_t \mathbf{V}'_t\|^2 + \|\mathbf{U}_m \mathbf{V}'_m\|^2 \\ &= \|\mathbf{Z}\|^2 - \|\mathbf{Z} - \mathbf{U}_t \mathbf{V}'_t\|^2 + \|\mathbf{Z}\|^2 - \|\mathbf{Z} - \mathbf{U}_m \mathbf{V}'_m\|^2, \end{aligned} \tag{16}$$

which is a sum of the variance explained by the trait components and that explained by the method components. A further simplification can be found by noticing that the optimal \mathbf{V}_t and the optimal \mathbf{V}_m can be obtained independently. Specifically, given \mathbf{U}_t , the matrix \mathbf{V}_t minimizes $\|\mathbf{Z} - \mathbf{U}_t \mathbf{V}'_t\|^2$ subject to the constraint that each row of \mathbf{V}_t has at most one nonzero element (at a prespecified position), and similarly, given \mathbf{U}_m , the matrix \mathbf{V}_m minimizes $\|\mathbf{Z} - \mathbf{U}_m \mathbf{V}'_m\|^2$ subject to the constraint that each row of \mathbf{V}_m has at most one nonzero element (at a prespecified position). The optimal \mathbf{V}_t , given \mathbf{U}_t , in fact solves a series of separate univariate regression problems: The unconstrained element in the j -th row of \mathbf{V}_t (say the l_j -th, hence the element v_{j,l_j}) is found by regressing the j -th column

of \mathbf{Z} on (only) the l_t -th column of \mathbf{U}_t , hence $v_{jl_t}^t = \mathbf{u}_{l_t}'\mathbf{z}_j$. It follows that $\|\mathbf{z}_j - \mathbf{U}_t\mathbf{v}_j^{t'}\|^2 = \|\mathbf{z}_j - \mathbf{u}_{l_t}v_{jl_t}^t\|^2 = \mathbf{z}_j'\mathbf{z}_j - (\mathbf{z}_j'\mathbf{u}_{l_t})^2$, and, analogously, $\|\mathbf{z}_j - \mathbf{U}_m\mathbf{v}_j^{m'}\|^2 = \mathbf{z}_j'\mathbf{z}_j - (\mathbf{z}_j'\mathbf{u}_{l_m})^2$. Hence we can partition the fit of a single variable as

$$\begin{aligned} f_j &= 2\mathbf{z}_j'\mathbf{z}_j - \|\mathbf{z}_j - \mathbf{U}_t\mathbf{v}_j^{t'}\|^2 + \|\mathbf{z}_j - \mathbf{U}_m\mathbf{v}_j^{m'}\|^2 \\ &= (\mathbf{z}_j'\mathbf{u}_{l_t})^2 + (\mathbf{z}_j'\mathbf{u}_{l_m})^2. \end{aligned} \tag{17}$$

The fit partitioning in (17), which holds for all models in which trait and method components are constrained to be mutually uncorrelated, implies that for each variable the contribution of the associated trait component and of the associated method component can be added to find the total contribution to the fit of this variable. From (17) we find that the total fit of all variables can be partitioned as

$$\begin{aligned} f(\mathbf{U}, \mathbf{V}) &= \|\mathbf{Z}\|^2 - \|\mathbf{Z} - \mathbf{U}_t\mathbf{V}_t'\|^2 + \|\mathbf{Z}\|^2 - \|\mathbf{Z} - \mathbf{U}_m\mathbf{V}_m'\|^2 \\ &= \sum_{j=1}^t w_{jl_t}^t (\mathbf{z}_j'\mathbf{u}_{l_t})^2 + \sum_{j=1}^m w_{jl_m}^m (\mathbf{z}_j'\mathbf{u}_{l_m})^2, \end{aligned} \tag{18}$$

where $\mathbf{W} = (\mathbf{W}_t \mid \mathbf{W}_m)$ is an indicator matrix with zeros indicating the constrained elements of \mathbf{V} , and unit elements the unconstrained elements; clearly, \mathbf{W} has two unit elements per row, one in \mathbf{W}_t , indicating the trait factor to which a variable belongs, and one in \mathbf{W}_m , indicating the associated method factor. It can be concluded that, for all models belonging to Cases 2 through 5, the squared component loading gives the contribution of the particular component to the fit.

Performance

Above, several theoretical advantages of the CA methods for MTMM matrices have been discussed. For instance, CA methods have no problems of finding improper solutions of any kind, they have a straightforward way of obtaining component scores, and the fitted models are more often identified than CFA models. In the present section, it is studied whether or not the CA methods perform well in practice. We focus on two aspects. First, the performance of the algorithms is discussed. Second, it is studied to what extent the CA method is able to recover a known underlying structure, and this ability will be compared to that of the corresponding CFA methods.

Data Construction

The above issues are studied by applying the methods to 90 constructed data sets. Each data set was constructed as a sample from a population where the CFA version of the MTMM model holds, and for which the parameters of the CFA model are known. We considered cases with three trait and three method factors only. Each data set consists of n realisations of the (nine-dimensional) random variable

$$\mathbf{y} = \Lambda\mathbf{u} + \mathbf{e},$$

where \mathbf{u} is a six-dimensional random variable with a multivariate normal distribution $N(\mathbf{0}, \Phi)$, \mathbf{e} is a nine-dimensional random (noise) variable with a multivariate normal distribution $N(\mathbf{0}, \mathbf{D}_e^2)$, Λ and Φ are a prespecified loading and factor correlation matrix, respectively, and $\mathbf{D}_e^2 \equiv \text{Diag}(\mathbf{I} - \Lambda\Phi\Lambda')$, thus arranging that the population covariance matrix has unit diagonal elements, and can be considered a correlation matrix. To construct the 90 data sets, two different sample sizes ($n = 100$ and $n = 300$), three different loading matrices

and three different factor correlation matrices were used. The three different loading matrices were

$$\Lambda_1 = \begin{pmatrix} .45 & 0 & 0 & .65 & 0 & 0 \\ 0 & .55 & 0 & .55 & 0 & 0 \\ 0 & 0 & .65 & .45 & 0 & 0 \\ .55 & 0 & 0 & 0 & .55 & 0 \\ 0 & .65 & 0 & 0 & .45 & 0 \\ 0 & 0 & .45 & 0 & .65 & 0 \\ .65 & 0 & 0 & 0 & 0 & .45 \\ 0 & .45 & 0 & 0 & 0 & .65 \\ 0 & 0 & .55 & 0 & 0 & .55 \end{pmatrix}; \Lambda_2 = \begin{pmatrix} .55 & 0 & 0 & .75 & 0 & 0 \\ 0 & .65 & 0 & .65 & 0 & 0 \\ 0 & 0 & .75 & .55 & 0 & 0 \\ .65 & 0 & 0 & 0 & .65 & 0 \\ 0 & .75 & 0 & 0 & .55 & 0 \\ 0 & 0 & .55 & 0 & .75 & 0 \\ .75 & 0 & 0 & 0 & 0 & .55 \\ 0 & .55 & 0 & 0 & 0 & .75 \\ 0 & 0 & .65 & 0 & 0 & .65 \end{pmatrix};$$

$$\Lambda_3 = \begin{pmatrix} .6 & 0 & 0 & .6 & 0 & 0 \\ 0 & .6 & 0 & .6 & 0 & 0 \\ 0 & 0 & .6 & .6 & 0 & 0 \\ .6 & 0 & 0 & 0 & .6 & 0 \\ 0 & .6 & 0 & 0 & .6 & 0 \\ 0 & 0 & .6 & 0 & .6 & 0 \\ .6 & 0 & 0 & 0 & 0 & .6 \\ 0 & .6 & 0 & 0 & 0 & .6 \\ 0 & 0 & .6 & 0 & 0 & .6 \end{pmatrix}.$$

These loading matrices were selected to cover one well-behaved model (based on Λ_2) and two potentially problematic models with relatively high amounts of noise (Λ_1) and under-identification (Λ_3). Note that Λ_3 has deficient rank, and therefore would, in an error free population, lead to nonuniqueness for both the Case 1 and the Case 2 model (see Grayson & Marsh's, 1994, Theorem 4), but to uniqueness in the Case 5 model (see Grayson & Marsh's, 1994, Theorem 5).

The three different factor correlation matrices were

$$\Phi_1 = \begin{pmatrix} 1 & .4 & .4 & .15 & .15 & .15 \\ .4 & 1 & .4 & .15 & .15 & .15 \\ .4 & .4 & 1 & .15 & .15 & .15 \\ .15 & .15 & .15 & 1 & .4 & .4 \\ .15 & .15 & .15 & .4 & 1 & .4 \\ .15 & .15 & .15 & .4 & .4 & 1 \end{pmatrix}; \Phi_2 = \begin{pmatrix} 1 & .4 & .4 & 0 & 0 & 0 \\ .4 & 1 & .4 & 0 & 0 & 0 \\ .4 & .4 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & .4 & .4 \\ 0 & 0 & 0 & .4 & 1 & .4 \\ 0 & 0 & 0 & .4 & .4 & 1 \end{pmatrix}; \Phi_5 = \mathbf{I}.$$

As the subscripts indicate, these covered Cases 1, 2 and 5, respectively. In Φ_1 the correlations between method and trait factors were taken considerably smaller than those among trait or among method factors. This seems reasonable from a substantive point of view, and was needed to ensure that the matrix \mathbf{D}_e^2 had positive values for all choices of Λ .

For each cell of the above described $3 \times 3 \times 2$ design, 5 data sets were generated as follows. Given a specified matrix Φ , we calculated $\Phi^{1/2}$ and drew n random samples ($\mathbf{x}_i, i = 1, \dots, n$, collected in the rows of \mathbf{X}) from the multivariate normal distribution $N(\mathbf{0}, \mathbf{I})$, and n random samples ($\mathbf{e}_i, i = 1, \dots, n$, collected in \mathbf{E}) from the multivariate normal distribution $N(\mathbf{0}, \mathbf{D}_e^2)$. With the $n \times (m + t)$ matrix \mathbf{X} we computed

$$\mathbf{Y} = \mathbf{X}\Phi^{1/2}\Lambda' + \mathbf{E},$$

which represents the raw sample. Since MTMM methods are usually applied to correlation matrices, we standardized \mathbf{Y} columnwise, to obtain \mathbf{Z} , and computed the corresponding correlation matrix \mathbf{R} . To all 90 data sets, we applied the CA methods for Case 1, Case 2 and Case 5 (denoted as Methods 1, 2 and 5, respectively).

Local Optima

The first purpose of the present simulation study is to study the sensitivity of our algorithms to local optima. Therefore, for each analysis we used the rational start described in the algorithm section and five random starts. We considered the algorithm converged as soon as the loss function value decreased by less than .0001%. To optimize the chance of finding the global minimum in this simulation study, we added a start based on the 'true values' contained in the matrices Λ and Φ that were used in the construction of the data. A solution was considered locally optimal if the function value differed more than .01 from the best function value (allegedly the global minimum) in the seven runs of an analysis.

It was found that the rational start landed in a local optimum 17 times (out of 90) for Method 1, 25 times for Method 2, and 16 times for Method 5 (without systematic differences between construction conditions). Hence the rational start seems to work fairly well. Not surprisingly, the start based on true values worked considerably better (frequencies of hitting local optima were 1, 12 and 7, respectively). The random starts, on the other hand, led to 34 local minima (averaged over 5 random starts) for Method 1, 25.8 for Method 2 and 13.8 for Method 5. Using random starts, the chance of hitting local optima seems to decrease as the number of model constraints (orthogonality constraints) is increased. Furthermore, it can be observed that, for Method 1, the rational start performs better than the random starts, for Method 2 it performs about as well as the random starts, and for Method 5 the rational start is worse than the random starts. It can be concluded that the algorithms are not very sensitive to hitting local optima, but relying on the rational start alone is not warranted. Using a number of random starts will diminish the chance of missing the global minimum. Besides, efforts to develop a better rational start could be undertaken, especially towards improving the rational start for Methods 2 and 5, since, for these Methods, the algorithm is less sensitive to local optima when started randomly.

Model Choice

To study whether or not comparison of the model fit can be used to find the correct model for a particular data set, we compared the differences in fit values (explained variances) for the three different methods. Because the Methods 1, 2 and 5 fit increasingly more restricted models, the explained variances for Method 1 will be highest and those for Method 5 lowest. However, fit differences can be expected to be large or small depending on the model underlying the data. When the data are based on Φ_1 , the underlying model is of the form in Case 1, and the data are denoted as "Model 1 data". For Model 1 data, we expect that the variance explained by Method 1 is considerably higher than that explained by Method 2 and Method 5, because the latter try to fit models that are too restrictive for the data at hand. In fact, in these analyses, compared to Method 1, Method 2 imposes 9 additional constraints, and Method 5 imposes 15 additional constraints. When the data are based on Φ_2 ("Model 2 data"), the variance explained by Method 1 should no longer be much larger than that explained by Method 2, because the model fitted by Method 2 is just as suitable to find the underlying structure as the model fitted by Method 1. We do expect, however, that the variance explained by Method 2 is considerably higher than that explained by the too restrictive Method 5. Finally, for Model 5 data (based on Φ_5) we expect that Methods 1 and 2 account for only slightly more variance than Method 5 does, because all models can account for the structure underlying Model 5 data.

In Table 1 average fit values (in percentages of the total variance, which is 9 throughout) are reported for Methods 5, 2 and 1, respectively. In addition, differences between explained variances of Methods 1 and 2, and of Methods 2 and 5 are reported. From Table 1, it is clear that Model 1 data can be fitted rather well by Method 5, but considerably

TABLE 1

Average Percentages of Explained Variance for Models 1, 2 and 5

Model 1 Data (Φ_1)		Explained Variance (in %)				
		Method 5	Method 2	Method 1	M2 - M5	M1 - M2
<i>n</i>	Λ					
100	Λ_1	84.4	90.4	91.0	6.0	0.6
100	Λ_2	89.5	99.2	99.4	9.7	0.2
100	Λ_3	86.7	94.5	94.9	7.8	0.4
300	Λ_1	84.9	89.8	90.1	4.8	0.3
300	Λ_2	90.8	99.3	99.3	8.5	0.1
300	Λ_3	87.5	94.1	94.5	6.6	0.4
Model 2 Data (Φ_2)		Explained Variance (in %)				
		Method 5	Method 2	Method 1	M2 - M5	M1 - M2
<i>n</i>	Λ					
100	Λ_1	83.7	87.4	88.3	3.7	0.9
100	Λ_2	89.3	95.1	95.5	5.8	0.4
100	Λ_3	86.4	91.1	91.8	4.7	0.7
300	Λ_1	84.0	86.6	87.2	2.7	0.5
300	Λ_2	90.2	94.9	95.0	4.6	0.2
300	Λ_3	86.7	90.4	91.1	3.7	0.7
Model 5 Data (Φ_5)		Explained Variance (in %)				
		Method 5	Method 2	Method 1	M2 - M5	M1 - M2
<i>n</i>	Λ					
100	Λ_1	86.3	86.9	88.4	0.5	1.6
100	Λ_2	94.3	94.9	95.6	0.6	0.7
100	Λ_3	89.7	90.3	91.8	0.6	1.5
300	Λ_1	86.2	86.3	87.5	0.1	1.2
300	Λ_2	94.7	94.8	95.2	0.1	0.4
300	Λ_3	89.6	89.8	91.2	0.1	1.5

better by Method 2. The increase in fit when using Method 1 is only modest. Hence, on the basis of explained variances alone, it seems hard to distinguish these Model 1 data from Model 2 data. Because Model 2 is a more parsimonious model than Model 1, one would tend to model these data by Model 2. It does seem clear for these data that Model 5 is incorrect, since Methods 1 and 2 lead to substantially higher fit values.

The results for Model 2 data are rather similar to those for Model 1 data. The main difference is that the fit percentages for the Model 2 data are smaller than for the Model 1 data. The increase in fit by choosing Method 1 instead of Method 2 is again small (as it now should be indeed, because Model 2 is the correct model), and the increase in fit by choosing Method 2 rather than Method 5 is relatively large. Hence, it seems that the trade-off between explained variance and parsimony of the model leads us to choose Model 2 to model these Model 2 data.

Finally, for the Model 5 data, it can be seen that the variances explained by Methods 1 and 2 are not much higher than those explained by Method 5, hence, on the basis of the trade-off between explained variance and parsimony, one would correctly choose Model 5 to model these data.

In conclusion, it can be said that, on the basis of comparison of explained variances alone, the Model 2 and Model 5 data tend to be correctly recognized as such, but that Model 1 data tend to be recognized as Model 2 data. It is tempting to attribute the present finding to the fact that the correlations between method and trait factors in the Model 1 data were only small (.15). However, some other experimentation showed that also when Model 1 data were constructed on the basis of considerably larger correlations, Method 2 performed almost as well as Method 1, and one would tend to conclude even for those data that Model 1 is the best model. Related to this is the finding that for Model 1 data the difference in fit between Method 1 and Method 2 was smaller than the difference in fit between Method 1 and Method 2 observed for the Model 2 data, whereas the reverse could be expected (and was indeed found for error free data, where Methods 1 and 2 both fit Model 2 data perfectly and hence lead to a zero fit difference). Some experimentation suggested that other data construction procedures tend to produce similar results, except when the error distributions have very small or widely different variances. Further study into an explanation of this peculiar phenomenon has not been undertaken because it will not affect the main conclusion that Method 1 and Method 2 tend to yield very similar fit values even for Model 1 data, and hence by the present procedure of comparing only fit values, it is hard to distinguish Model 1 data from Model 2 data. Inspection of the loadings may in such cases further help choosing the model, as will be illustrated in the first exemplary analysis in the next section.

Recovery of the Underlying Structure

To study the performance of our methods in terms of recovery of the underlying structure, we compared the resulting loadings and component correlations (associated with the highest function value, out of seven runs) to the loadings and factor correlations on which the data construction was based. For these comparisons we used two recovery measures. Let Λ and Φ denote the loading and factor correlation matrix, respectively, on which the data construction for a particular data set was based. Then, we have chosen as measures for recovery:

$$\text{REC}_\Lambda = \frac{1}{18} \sum_{i=1}^9 \sum_{j=1}^6 |\lambda_{ij} - v_{ij}|,$$

which is the average difference between unconstrained observed and population loadings, and

$$\text{REC}_{\Phi} = \frac{1}{q} \sum_{i=1}^6 \sum_{j>i}^6 |\phi_{ij} - \mathbf{u}'_i \mathbf{u}_j|,$$

where q denotes the number of different unconstrained elements in Φ , which is 15 for Method 1 and 6 for Method 2, so that REC_{Φ} denotes the average absolute difference between unconstrained observed and population values of Φ ; in Method 5 all values in Φ are constrained, hence REC_{Φ} can only be computed for Methods 1 and 2. Widaman (1993) demonstrated that components analysis will tend to give loadings that are biased upward, and intercomponent correlations that are biased downward. To see if this tendency was observed in our studies, we checked for each loading and for each component correlation, whether it was biased upward or downward.

Clearly, Method 1 performed well only when data were constructed on the basis of Λ_2 . Apparently, high error level (for data based on Λ_1) or near nonuniqueness (data based on Λ_3) may cause Method 1 to yield results that differ very much from the input Λ and Φ . For data based on Λ_2 , Method 1 was able to reconstruct the original Λ and Φ reasonably well, as reflected in the small REC_{Λ} and REC_{Φ} values.

As far as the direction of the differences between obtained and 'true' loadings is concerned, we found the expected tendency of loadings being biased upward. Specifically, in the conditions based on Λ_1 and Λ_3 , on the average more than 15 of the 18 loadings were biased upward; in the conditions based on Λ_2 this average was 13.4. It should be noted that in these conditions, at most 15 loadings were biased upward, and in one case only 11 upward biases were found, and hence 7 downward biases. This corroborates the results that, in the conditions based on Λ_2 , Method 1 performs quite reasonably.

As far as biases in component correlations are concerned, except in one of the conditions with Λ_2 , at least nine of the fifteen correlations were biased downward, which sustains Widaman's findings. Upon closer inspection of the Method 1 solutions, it could be seen that in many of the poorly recovered Λ_1 and Λ_3 cases, the solution was of a degenerate nature. That is, the correlations between components tended to values near 1 (among traits or among methods) or near -1 (between traits and methods). Such phenomena were observed in many instances, except when data were based on Λ_2 . Such solutions, of course, are irrelevant from a data analytical point of view and can be discarded as degenerate. In fact, the phenomenon is similar to the degenerate factors phenomenon encountered with PARAFAC (see Kruskal, Harshman & Lundy, 1989).

Like Method 1, Methods 2 and 3 performed best in the conditions based on Λ_2 . In the other conditions, they performed considerably better than Method 1, with average absolute differences in the interval [.10, .20]. These deviations from the true values may to some extent distort the interpretations of the solutions, but are acceptable. This result is rather surprising, especially for Method 2: For the conditions based on Λ_1 , it implies that even with relatively large error terms, the present component models can recover the original structure fairly well; for the conditions based on Λ_3 , it implies that even when the model is underidentified in the population (as is the case for the Model 2 data, based on Λ_3 , see Theorem 4 by Grayson & Marsh, 1994), component analytic results for a sample are not meaningless, and to some extent approximate the true (population) values. Moreover, when analyzing the population correlation matrix, CA did find a unique solution, which indeed approximated the population loadings and correlations to some extent.

As far as biases are concerned with Methods 2 and 3, it was found that on the average between 10 and 12 loadings were biased upward, and, of the six intercomponent correlations in Method 2, on the average between 4 and 6 were biased downward. Hence, these results sustain Widaman's (1993) results on bias of loadings and component intercorrelations. On the other hand, it should be noted that the number of upward biases of the

TABLE 2

Recovery of the Original Loadings and Component Correlations
by the CA Methods

Average Results for Model 1 Data, Based on Application of Method 1				
	REC- Λ		REC- Φ	
	$n=100$	$n=300$	$n=100$	$n=300$
Λ_1	0.48	0.54	.37	.56
Λ_2	0.05	0.03	.10	.06
Λ_3	1.84	2.42	.69	.84

Average Results for Model 2 Data, Based on Application of Method 2				
	REC- Λ		REC- Φ	
	$n=100$	$n=300$	$n=100$	$n=300$
Λ_1	0.16	0.14	.13	.14
Λ_2	0.08	0.06	.08	.08
Λ_3	0.17	0.15	.11	.10

Average Results for Model 5 Data, Based on Application of Method 5		
	REC- Λ	
	$n=100$	$n=300$
Λ_1	0.15	0.14
Λ_2	0.08	0.06
Λ_3	0.15	0.14

loadings was on the average lower than for Method 1, which is in line with the better performance of Methods 2 and 3.

Table 1 also gives information on the effect of difference in sample size. Whereas for common factor analysis, sample size is usually an important determinant of the performance, it can be seen that sample size hardly affected the reproductive capacity of the CA methods.

For comparative purposes, we also fitted the three CFA versions of the MTMM models to the data (using LISREL7; Jöreskog & Sörbom, 1989). The default LISREL start

was used at first. It turned out that, in many cases, the program gave an improper solution, or did not converge. In cases where the problems were severe, the program failed to write the solution to the requested file. To give the method a better chance of finding a proper solution, we reran all analyses using the 'true' structure as start. Almost the same number of improper solutions was encountered. Specifically, fitting Model 1 led to improper solutions in 57 (out of 90) times and nonconvergence occurred in three cases. Only 45 solutions were denoted as "admissible" (and were consequently written to file). Fitting Model 2 yielded 47 improper solutions, but they converged throughout, and 63 solutions were deemed admissible. Model 5 was far less problematic, leading to only 16 improper solutions (all Heywood cases), and all 90 solutions were considered admissible by LISREL.

The abundance of inadmissible solutions makes comparison with the CA methods hard. Therefore, we compared REC values only for data sets where LISREL solutions were admissible. A stronger selection where all improper solutions were discarded was not deemed necessary; on the other hand, taking into account nonadmissible solutions was deemed inappropriate because LISREL's warnings clearly suggest that these solutions should not be used in practice. For the cases where LISREL gave an admissible solution, we simply checked whether CA or CFA led to the best recovery value at hand. These counts are reported in Table 3.

Of course, the prevalence of improper CFA solutions by itself illustrates an important advantage of CA versus CFA, since CA never leads to improper solutions. From Table 3, it can be seen that, even in cases where CFA does give a proper (or at least "admissible") solution, it is fairly often outperformed by the CA solution. This occurred especially for Models 1 and 2, with $n = 100$, but also occasionally in other cases. Hence, the expected larger bias in CA loadings was not consistently found. Apparently, in cases where CA gives considerable biases, the CFA loadings are not recovered very well either. On the other hand, in cases where CFA does perform well, notably those where Model 5 is fitted, but also in instances of fitting Models 1 and 2, with $n = 300$, CA tended to lag behind.

The present simulation study reveals several things. First of all, it can be concluded that fitting Model 1, which is well known to be problematic for CFA, is often problematic for CA methods too. The problems carry over in situations where the population loadings are poorly identified, or where the error terms are large. However, fitting data based on fairly well identified population loadings, and on relatively small error terms did not cause the CA method to run into major problems, and hardly ever did for the CFA method. In these cases, however, the CA method almost always outperformed CFA in terms of recovering the underlying structure, *even though this underlying structure conforms exactly to the CFA model*.

As far as fitting Model 2 is concerned, it turned out that CA and CFA both performed fairly well as far as models based on well-identified population loadings and small errors are concerned. Especially in cases with large error, but also in cases with unidentified population loadings, CFA failed fairly often, whereas CA tended to perform acceptably in most of these cases as well. Where CFA did not fail, it usually performed worse than CA in the small sample cases, and better in the larger sample cases. Finally, in the case of fitting Model 5, CA always performed reasonably well, but was typically outperformed by CFA.

These results together suggest that CA can be used as a reasonable alternative to CFA for recovering an underlying structure, which will be somewhat less adequate than CFA in well identified situations, but will be particularly useful in relatively poorly identified situations.

TABLE 3

Comparison of CA and CFA in Terms of Frequencies of Finding the Best Recovery

Model 1 Data Fitted to Model 1				
	REC- Λ		REC- Φ	
CA/CFA *	$n=100$	$n=300$	$n=100$	$n=300$
Λ_1	2/0	0/2	2/0	0/2
Λ_2	4/0	1/4	2/2	1/4
Λ_3	1/0	0/2	1/0	0/2

Model 2 Data Fitted to Model 2				
	REC- Λ		REC- Φ	
CA/CFA	$n=100$	$n=300$	$n=100$	$n=300$
Λ_1	1/0	0/2	1/0	0/2
Λ_2	4/0	1/4	4/0	2/3
Λ_3	3/0	0/5	3/0	4/1

Model 5 Data Fitted to Model 5		
	REC- Λ	
CA/CFA	$n=100$	$n=300$
Λ_1	0/5	0/5
Λ_2	1/4	0/5
Λ_3	0/5	0/5

*) The entries in this table should be read as follows: a/b indicates that in $a+b$ cases LISREL found an admissible solution; in a of these cases CA performed best, in the other b cases CFA performed best.

Empirical Examples

In the present section, two empirical data sets are analyzed by the CA procedures for fitting MTMM models. The first data set is Lawler's (1967) MTMM matrix on managerial job performance. The data pertain to three Traits (Quality of job performance, Ability to

TABLE 4

Percentages of Explained Variance and Numbers of Local Optima
for the Lawler Data

Model	Constraints	Percentage of Explained Variance	Number of Local Optima
Traits Only	zero correlations	50.1 %	0
	unconstrained	55.3 %	0
Methods Only	zero correlations	57.8 %	0
	unconstrained	63.3 %	0
Traits and Methods	Model 5	83.9 %	3
	Model 4	87.6 %	2
	Model 3	85.1 %	2
	Model 2	88.0 %	3
	Model 1	88.9 %	2

perform the job, and Effort put forth on the job) measured by three Methods (Superior ratings, Peer ratings, and Self ratings). This MTMM matrix has been used in many exemplary analyses. Problems of nonconvergence and improper solutions were encountered in fitting this matrix by CFA models (e.g., see Brannick & Spector, 1990; Browne, 1993; Wothke, 1987). In the present section, we report the application of all discussed CA variants for fitting this MTMM matrix. These variants include fitting models for Trait Factors Only (with or without zero correlations constraints), Method Factors Only (with or without zero correlations constraints), and the five aforementioned models for Trait and Method Factors, with different sets of constraints. All analyses were based on a rational start and five random restarts. Table 4 reports the percentages of explained variance, and the numbers of local optima encountered.

As is clear from Table 4, the algorithm did not find many local optima for the simplest models, and even for the most complex ones, the number of local optima was acceptable, and gives confidence that the global optimum has indeed been found. Inspecting the percentages of explained variance, it is clear that the "Traits Only" and "Methods Only" models do not fit the data very well. The data seem to be based on substantial trait and method components. Of the models employing both trait and method components, the unconstrained model (Model 1) trivially fits best, but the Models 2 and 4 fit almost as well. Among these, Model 4 (using orthogonal trait components, and correlated method components, while method and trait components are mutually orthogonal) is the most parsimonious, and, moreover, seems to make intuitive sense: It is, in fact, desirable that Superior, Peer and Self ratings are positively correlated; at the same time, it is conceivable that quality, ability and effort are uncorrelated aspects of job performance. Thus, by

comparison of model fits, we have come to a sensible model, which, moreover, explains almost 90% of the variance. Note that we could make this fit comparison only because we could compute solutions for *all* models, in contrast to what is the case in CFA applied to these data (where nonconvergence or improper solutions preclude comparison of fit values).

Besides fit comparison, we can use substantive interpretation to choose between models. For this purpose, loadings and component correlations for Models 1 and 4 are reported in Table 5. We also inspected the Model 2 solution, but observed that it was quite similar to the Model 4 solution, with nonzero but small correlations between the trait components. We prefer to report Model 4, which admittedly exaggerates one of the main observations for these data, namely that correlations between trait components are small. Clearly, the Model 4 solution gives more reasonable loadings than the Model 1 solution. Especially the $-.87$ loading of v_8 on "Ability" is anomalous, and is an extra indication for choosing Model 4 rather than the slightly better fitting Model 1. Moreover, the Model 1 solution has a rather irregular, and therefore hard to interpret, pattern with fairly large positive and negative correlations between method and trait components. This is reminiscent of the impropriety of the solution found when fitting this model by CFA.

The original purpose of analyzing MTMM matrices was to study convergent and discriminant validity of a number of measures, and to assess the effects of Method variance. The fact that the "Methods Only" model does not account for a large amount of the variance indicates at least some degree of convergent validity. As described by Widaman (1985, p. 9), convergent validity is indicated more precisely by the size of the loadings on the trait components. These are fairly high for Quality and Effort (related to two and three variables, respectively), but not for Ability (which is related to only one variable). Hence Quality and Effort seem to have reasonable convergent validity, but Ability does not. Low correlations between trait components can be seen as indicative of discriminant validity. Since in the present model, these correlations are zero, the traits clearly have discriminant validity. As far as the method effects are concerned, it can be concluded that method effects are rather strong. This follows from the fact that the "Traits Only" models fit very poorly and that the loadings on Method components are often higher than on trait components. Furthermore, two of the method effects are related: Superior and Peer judgments are correlated rather strongly (.61).

The second data set is the second MTMM matrix reported by Marsh, Byrne and Craven (1992, p. 507), henceforth called the MBC data. The matrix pertains to four traits and three methods. In addition, the correlations of all variables with three external variables, GPA (grade point average), English achievement and Mathematics achievement, are available. The reported 15×15 correlation matrix was not positive definite (possibly because it was based on pairwise deletion). Therefore, before the analyses, we used a positive definite approximation to this matrix, which approximated this matrix very well. We applied the Cholesky decomposition to the full 15×15 matrix, to obtain a 15×15 pseudo data matrix \mathbf{R}_Z for *all* variables. The actual, unknown, scores on the 15 variables can be expressed as $\mathbf{Z} = \mathbf{Q}_Z \mathbf{R}_Z$, where \mathbf{Q}_Z is an unknown columnwise orthonormal matrix. It follows that $\mathbf{Z}_{(12)}$, the matrix with the scores on the first twelve variables (the actual MTMM variables), can be written as $\mathbf{Z}_{(12)} = \mathbf{Q}_Z \mathbf{R}_{Z(12)}$, where $\mathbf{R}_{Z(12)}$ is the matrix with the first twelve columns of \mathbf{R}_Z . The matrix $\mathbf{R}_{Z(12)}$ was used as input in the MTMM analyses. The last three columns of \mathbf{Z} and hence of \mathbf{R}_Z were used for an external validation of the components. As has been explained in the introduction, an important advantage of CA over CFA is the possibility of computing correlations of components scores with external variables. This advantage was put to use in the present example.

As with the Lawler data, we fitted nine models to the present data set. Table 6 reports the percentages of explained variance, and the numbers of local optima encountered. The

TABLE 5

Loadings and Component Correlations for Models 4 and 1
Fitted to the Lawler Data.

Loadings												
	Model 4						Model 1					
	Traits			Methods			Traits			Methods		
	Qual.	Abil.	Eff.	Sup.	Peer	Self	Qual.	Abil.	Eff.	Sup.	Peer	Self
v1	.65	0	0	.70	0	0	.75	0	0	.88	0	0
v2	0	.16	0	.91	0	0	0	.46	0	1.01	0	0
v3	0	0	.57	.68	0	0	0	0	.75	.82	0	0
v4	.64	0	0	0	.71	0	.64	0	0	0	.87	0
v5	0	-.01	0	0	.92	0	0	.72	0	0	1.02	0
v6	0	0	.61	0	.66	0	0	0	.79	0	.85	0
v7	-.12	0	0	0	0	.97	.39	0	0	0	0	.88
v8	0	.90	0	0	0	.42	0	-.87	0	0	0	1.13
v9	0	0	.73	0	0	.52	0	0	.80	0	0	.39

Component Correlations												
	Model 4						Model 1					
	Qual.	Abil.	Eff.	Sup.	Peer	Self	Qual.	Abil.	Eff.	Sup.	Peer	Self
Quality	1						1					
Ability	0	1					.24	1				
Effort	0	0	1				.51	.31	1			
Superior	0	0	0	1			-.31	-.38	-.34	1		
Peer	0	0	0	.61	1		-.24	-.44	-.36	.71	1	
Self	0	0	0	.06	.13	1	-.07	.53	.06	-.08	-.11	1

number of local optima encountered by the algorithm was small, except for Model 1. Inspection of the Model 1 results revealed that the algorithm found a degenerate solution here. Among the remaining models the "Method Only" models are inferior to the rest; the "Traits Only" models give a reasonable fit now. It can be concluded that there is relatively little Method variance in this data. Nevertheless, the effect of the methods does seem nonnegligible, as evidenced by the considerably higher fit by the Models 2 and 3 (which allow for Trait and Method components). Since Model 3 (with correlated Trait compo-

TABLE 6

Percentages of Explained Variance and Numbers of Local Optima for the MBC Data

Model	Constraints	Percentage of Explained Variance	Number of Local Optima
Traits Only	zero correlations	70.8 %	0
	unconstrained	80.0 %	0
Methods Only	zero correlations	38.1 %	3
	unconstrained	48.9 %	0
Traits and Methods	Model 5	84.1 %	0
	Model 4	88.7 %	0
	Model 3	91.4 %	0
	Model 2	91.5 %	2
	Model 1	92.6 %	5

nents, but uncorrelated Method components, and Trait and Method components mutually uncorrelated) is the most parsimonious and fits hardly worse than Model 2, we choose it as the most useful one to describe the data.

Having chosen the model to represent the MBC data, we can study the loadings and the component correlations, and, in addition, we can compute the correlations of the components with the external variables. It may seem that the latter is not straightforward, since we cannot compute component scores, because we do not have the original scores on the variables. However, we can derive these correlations from the correlations of the external scores with the variables. This can be done as follows. The MTMM algorithm is applied to $\mathbf{R}_{Z(12)}$. As has been explained in a previous section, the resulting component scores matrix \mathbf{W} is related to \mathbf{U} by $\mathbf{U} = \mathbf{Q}_Z\mathbf{W}$. It follows that the correlations between *all* variables and the components are given by $\mathbf{Z}'\mathbf{U} = \mathbf{R}'_Z\mathbf{Q}'_Z\mathbf{Q}_Z\mathbf{W} = \mathbf{R}'_Z\mathbf{W}$. Computing the latter matrix is straightforward, and the resulting correlations of *all* variables with the components are given in Table 7, along with the loadings for this model.

The convergent validity is evidenced by the large difference in fit between the "Methods Only" model and the "Traits and Methods" models, as well as by the high loadings of all variables on the associated traits. Discriminant validity is not as clear as with the Lawler data. Clearly, correlations between traits are nonzero, and in particular the "School" component is not well discriminated from the other traits. Further evidence in this direction is obtained from the correlations of the components with the external variables. The School component correlates moderately with all three external variables as one would expect. For the "General" trait, similar results (but to a lesser extent) are found. The traits "English" and "Math" on the other hand are well discriminated, and their interpretation

TABLE 7

Results of Model 3 Fitted to the MBC Data

Loadings							
	Trait Components				Method Components		
	General	School	English	Math	M1	M2	M3
t1m1	0.94				0.08		
t2m1		0.93			0.07		
t3m1			0.73		0.67		
t4m1				0.96	-0.04		
t1m2	0.67					0.71	
t2m2		0.81				0.41	
t3m2			0.96			0.12	
t4m2				0.94		0.24	
t1m3	0.94						-0.07
t2m3		0.74					0.62
t3m3			0.82				0.50
t4m3				0.88			0.35
Correlations of Components Mutually and With External Variables							
	Trait Components				Method Components		
	General	School	English	Math	M1	M2	M3
General	1						
School	.51	1					
English	.24	.51	1				
Math	.20	.53	.00	1			
M1	0	0	0	0	1		
M2	0	0	0	0	0	1	
M3	0	0	0	0	0	0	1
GPA	.07	.52	.13	.39	.05	-.07	.41
Eng-ext	.03	.43	.36	.14	-.04	-.13	.35
Math-ext	.02	.41	.00	.60	.03	.01	.32

is validated externally by the fact that they are correlated highest to the external variables to which they correspond.

Discussion

The above analyses of constructed and empirical data have demonstrated that the CA methods for fitting MTMM models are useful for decomposing a data matrix and for representing the information contained in a data or correlation matrix. It is not claimed that the CA methods are always more useful than CFA methods for fitting the MTMM models under study. However, the CA procedures lead to proper, well-behaved solutions in almost all cases. The only exception is Model 1 fitted to data that are based on large unique variances or nearly equal loadings, as could be expected because of the poor determination of solutions in these cases. CFA, on the other hand, led to unusable solutions very frequently. Surprisingly, this also happened for a relatively large sample size (300). Since in all our simulations the model did hold in the population, improper solutions cannot simply be attributed to misspecification. Hence, respecification may not help in such cases. Apparently, the problem is more serious. Sample sizes as large as 300 are not sufficiently large to guarantee that properties that hold in the population (e.g., common and unique factors uncorrelated) also nearly hold in the sample. In contrast, in many cases where CFA led to improper solutions, CA did give a well-behaved one. This happened in particular for fitting Models 1 and 2. It is well known that CFA performs poorly when fitting Model 1 (e.g., Marsh & Hocevar, 1983; Widaman, 1985). It is therefore surprising that CA did perform well in one condition of fitting Model 1. On the other hand, in cases where CFA led to proper solutions, the CA solution was neither consistently better (in terms of recovery of the underlying structure) nor consistently worse. Thus, the main advantage of the CA methods over the CFA methods is its ability to obtain reasonable estimates when CFA cannot. Although not tested here, this advantage may be particularly profitable in cases where CFA assumptions are grossly violated in the population.

It has been argued (e.g., Jöreskog & Sörbom, 1989, p. 239) that finding an improper solution with CFA should not just be judged as a problem of the CFA method, but could be considered as a meaningful diagnostic as well: Finding an improper solution indicates that the model is misspecified. However, the present simulation study showed that improper solutions do not only arise when models are misspecified. Even in cases where the correct model is specified, improper solutions may arise. For this reason, we view the fact that CA cannot give improper solutions as an asset of the method.

In addition to the above main advantage of CA over CFA, an additional advantage is its suitability for calculating correlations with external variables. As has been illustrated in the second example, such correlations can help assess the external validity of the components. Besides this possibility of computing correlations, the component scores themselves can be used for any other purpose as well.

The CFA methods provide information on the basis of which one can perform tests for determining whether a model fits significantly better than a different model. Such information is not available with the CA methods. Nevertheless, we have demonstrated that it is possible to choose a model, mainly (though not solely) on the basis of model fit. This may seem less rigorous than using significance tests, but then it should be kept in mind that the rigor of significance tests is based on often dubious assumptions (e.g., multinormality) and subjective choices (of significance levels). Moreover, in addition to significance, another important criterion for model selection is whether or not a difference in model fit is considered substantial in the application at hand.

Recently, it has been claimed (e.g., Marsh & Bailey, 1991; Kenny & Kashy, 1992) that a viable alternative to the problematic CFA model is the "correlated uniqueness model".

However, this alternative model also has some disadvantages. For instance, it does not allow for correlations between method factors. This seems, at least in quite a few applications, highly unrealistic. Another problem is that Method factors and Error are confounded, which makes the model complex in terms of interpretation. In fact, this approach was motivated partly by Kenny and Kashy's somewhat disheartening remark concerning CFA that "it then becomes necessary to simplify the complete model" (p. 171). The present paper demonstrates that this is not true: Maintaining the essence of the CFA models is often possible, provided that a different fitting strategy is used.

References

- Algina, J. (1980). A note on identification in the oblique and orthogonal factor analysis models. *Psychometrika*, 45, 393-396.
- Anderson, T. W., & Rubin, H. (1956). Statistical inference in factor analysis. In J. Neymann (Ed.) *Proceedings of the third Berkeley symposium on mathematical statistics and probability* (Vol. 5). Berkeley CA: University of California Press.
- Bagozzi, R. P., & Yi, Y. (1990). Assessing method variance in multitrait-multimethod matrices: The case of self-reported affect and perceptions at work. *Journal of Applied Psychology*, 75, 547-560.
- Bekker, P. A. (1986). A note on the identification of restricted factor loading matrices. *Psychometrika*, 51, 607-611.
- Brannick, M. T., & Spector, P. E. (1990). Estimation problems of the block diagonal model of the multitrait-multimethod matrix. *Applied Psychological Measurement*, 14, 325-340.
- Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology*, 37, 1-21.
- Browne, M. W. (1993). Models for multitrait-multimethod matrices. In R. Steyer, K. F. Wender, & K. F. Widaman (Eds.), *Psychometric Methodology. Proceedings of the 7th European Meeting of the Psychometric Society in Trier* (pp. 61-73). Stuttgart and New York: Gustav Fischer Verlag.
- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56, 81-105.
- Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika*, 31, 33-42.
- 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.
- Grayson, D., & Marsh, H. W. (1994). Identification with deficient rank loading matrices in confirmatory factor analysis: Multitrait-multimethod models. *Psychometrika*, 59, 121-134.
- Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7. A guide to the program and applications*. Chicago: Scientific Software.
- Kenny, D. A. (1979). *Correlation and Causality*. New York: Wiley.
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112, 165-172.
- Kiers, H. A. L., & Takane, Y. (1993). Constrained DEDICOM. *Psychometrika*, 58, 339-355.
- Kruskal, J. B., Harshman, R. A., & Lundy, M. E. (1989). How 3-MFA data can cause degenerate PARAFAC solutions, among other relationships. In R. Coppi & S. Bolasco (Eds.), *Multway Data Analysis* (pp. 115-122). Amsterdam: Elsevier Science Publishers.
- Lawler, E. E. (1967). The multitrait-multimethod approach to measuring managerial job performance. *Journal of Applied Psychology*, 51, 369-381.
- 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., & 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 analyses of MTMM data: The correlated uniqueness model and factorial invariance. *Multivariate Behavioral Research*, 27, 489-507.
- Marsh, H. W., & Hocevar, D. (1983). Confirmatory factor analysis of multitrait-multimethod matrices. *Journal of Educational Measurement*, 20, 231-248.
- Millsap, R. E. (1992). Sufficient conditions for rotational uniqueness in the additive MTMM model. *British Journal of Mathematical and Statistical Psychology*, 45, 125-138.
- Penrose, R. (1956). On best approximate solutions of linear matrix equations. *Proceedings of the Cambridge Philosophical Society*, 52, 17-19.
- Schmitt, N., & Stults, D. M. (1986). Methodology review: Analysis of multitrait-multimethod matrices. *Applied Psychological Measurement*, 10, 1-22.

- Takane, Y., Kiers, H. A. L., & de Leeuw, J. (1995). Component analysis with different sets of constraints on different dimensions. *Psychometrika*, *60*, 259–280.
- Werts, C. E., & Linn, R. L. (1970). Path analysis: Psychological examples. *Psychological Bulletin*, *74*, 193–212.
- Widaman, K. F. (1985). Hierarchically nested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, *9*, 1–26.
- Widaman, K. F. (1990). Bias in pattern loadings represented by common factor analysis and component analysis. *Multivariate Behavioral Research*, *25*, 89–95.
- Widaman, K. F. (1993). Common factor analysis versus principal component analysis: Differential bias in representing model parameters. *Multivariate Behavioral Research*, *28*, 263–311.
- Wothke, W. (1987). Multivariate linear models of the multitrait-multimethod matrix. *Paper presented at the meeting of the American Educational Research Association*. (ERIC ED 283850, TM 870369).
- Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika*, *55*, 255–262.

Manuscript received 6/22/94

Final version received 5/8/95