Psychometric
Society

THREE-MODE FACTOR ANALYSIS BY MEANS OF CANDECOMP/PARAFAC

ALWIN STEGEMAN AND TAM T.T. LAM

HEYMANS INSTITUTE FOR PSYCHOLOGICAL RESEARCH, UNIVERSITY OF GRONINGEN

A three-mode covariance matrix contains covariances of N observations (e.g., subject scores) on J variables for K different occasions or conditions. We model such an $JK \times JK$ covariance matrix as the sum of a (common) covariance matrix having Candecomp/Parafac form, and a diagonal matrix of unique variances. The Candecomp/Parafac form is a generalization of the two-mode case under the assumption of parallel factors. We estimate the unique variances by Minimum Rank Factor Analysis. The factors can be chosen oblique or orthogonal. Our approach yields a model that is easy to estimate and easy to interpret. Moreover, the unique variances, the factor covariance matrix, and the communalities are guaranteed to be proper, a percentage of explained common variance can be obtained for each variable-condition combination, and the estimated model is rotationally unique under mild conditions. We apply our model to several datasets in the literature, and demonstrate our estimation procedure in a simulation study.

Key words: three-mode factor analysis, multitrait-multimethod, Candecomp, Parafac, minimum rank factor analysis.

1. Introduction

Three-way data refer to data that can be arranged in a three-dimensional array or three-way array. Such data is found in many different contexts. For example: scores on various anxiety scales of a number of individuals in various situations; scores on various competencies of a number of workers by several different assessors; scores on food quality indicators of a number of food products by several different judges; and fMRI brain scan measurements for different areas of the brain over a period of time for different individuals. The three sets of entities associated with such three-way data sets are called the three *modes* of the array.

In this paper, we consider three-way data of N observations of J variables for K different occasions or conditions. We focus on three-mode factor analysis, i.e., a factor model for the $JK \times JK$ covariance matrix containing the covariances of all J variables and K conditions together. In Section 1.1, we introduce the general framework of three-mode factor analysis. In Section 1.2, we discuss existing models and methods for three-mode factor analysis, and introduce our novel model. In Section 1.3, we discuss Minimum Rank Factor Analysis (MRFA) for two-way factor analysis. We use MRFA to estimate the unique variances in our method for three-mode factor analysis.

1.1. Three-Mode Factor Analysis

Let $\mathbf{X}_k$ be an $N \times J$ matrix containing N observations of J variables for occasion k or under condition k, for $k = 1, \ldots, K$. We assume the columns of $\mathbf{X}_k$ have mean zero for all k. That is, offset terms that are constant across observations have been removed for each variable and condition. We suppose that in theory the data $\mathbf{X}_k$ can be written as the sum of a *common part* and a *unique part*: $\mathbf{X}_k = \mathbf{X}_k^{(\mathrm{com})} + \mathbf{E}_k$, for $k = 1, \ldots, K$. The common part $\mathbf{X}_k^{(\mathrm{com})}$ contains the part of each variable under condition k that correlates with other variables in the data. The

unique part $\mathbf{E}_k$ contains the part of each variable under condition k that does not correlate with other variables. The unique part of a variable may contain measurement error as well as a reliable part measuring a trait that is uncorrelated with any other variable. Both $\mathbf{X}_k^{(\text{com})}$ and $\mathbf{E}_k$ have mean-zero columns.

We look for a small number of R factors that best summarizes the common parts: $\mathbf{X}_k^{(\text{com})} \approx \mathbf{F}\mathbf{B}_k^T$, where the factors $\mathbf{F}$ ($N \times R$) are the same for all k, but the loadings $\mathbf{B}_k$ ($J \times R$) may be different. For all $\mathbf{X}_k$ together, and perfect fit, we have the following factor model:

$$\mathbf{X}_{(N \times JK)} = \mathbf{F}\begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix}^T + \mathbf{E}_{(N \times JK)}, \tag{1}$$

where $\mathbf{X}_{(N \times JK)} = [\mathbf{X}_1 \ldots \mathbf{X}_K]$ and $\mathbf{E}_{(N \times JK)} = [\mathbf{E}_1 \ldots \mathbf{E}_K]$. The covariance model corresponding to (1) is

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{11} & \cdots & \mathbf{\Sigma}_{1K} \\ \vdots & \ddots & \vdots \\ \mathbf{\Sigma}_{K1} & \cdots & \mathbf{\Sigma}_{KK} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix} \mathbf{\Phi} \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix}^T + \mathbf{U}, \tag{2}$$

where $\mathbf{\Sigma} = N^{-1}\mathbf{X}_{(N \times JK)}^T\mathbf{X}_{(N \times JK)}$ is the data covariance matrix, $\mathbf{\Sigma}_{kl} = N^{-1}\mathbf{X}_k^T\mathbf{X}_l$ contains the covariances between the J variables for conditions k and l, the factor covariance matrix is $\mathbf{\Phi} = N^{-1}\mathbf{F}^T\mathbf{F}$, and $\mathbf{U} = N^{-1}\mathbf{E}_{(N \times JK)}^T\mathbf{E}_{(N \times JK)}$ is the diagonal matrix of unique variances. Note that $\mathbf{\Sigma}$ and $\mathbf{U}$ have size $JK \times JK$, that $\mathbf{\Sigma}_{kl}$ has size $J \times J$, and that $\mathbf{\Sigma}_{kl} = \mathbf{\Sigma}_{lk}^T$.

The factors $\mathbf{F}$ are usually scaled such that they have variance 1, which makes $\mathbf{\Phi}$ the factor correlation matrix. If the factors are chosen uncorrelated (also called *orthogonal*), then $\mathbf{\Phi} = \mathbf{I}_R$. Otherwise, the factors are called *oblique*. The diagonal entries of $\mathbf{\Sigma} - \mathbf{U}$ are the variances of the common parts of the variables for each condition, and are called *communalities* or common variances. The diagonal entries of $\mathbf{B}_{(\text{all})}\mathbf{\Phi}\mathbf{B}_{(\text{all})}^T$ are called the estimated common variances, where $\mathbf{B}_{(\text{all})} = [\mathbf{B}_1^T \ldots \mathbf{B}_K^T]^T$. The diagonal entries of $\mathbf{U}$ are called the unique variances.

A probabilistic version of the three-mode factor model (1) is

$$\begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_K \end{pmatrix} = \begin{bmatrix} \mathbf{B}_1 \\ \vdots \\ \mathbf{B}_K \end{bmatrix}\mathbf{f} + \begin{pmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_K \end{pmatrix}, \tag{3}$$

where $\mathbf{x}_k$ is the $J \times 1$ random vector corresponding to the J variables under condition k, the $R \times 1$ random vector $\mathbf{f}$ contains the factors, and the $J \times 1$ random vector $\mathbf{e}_k$ corresponds to the unique parts of the variables under condition k. It is assumed that $\mathbf{x}_k$, $\mathbf{f}$, and $\mathbf{e}_k$ have zero expectation for all k, that $\mathbf{f}$ and $\mathbf{e}_k$ are uncorrelated for all k, and that $\mathbf{e}_k$ and $\mathbf{e}_l$ are uncorrelated for $k \neq l$. Under these assumptions, the covariance model corresponding to (3) is equal to (2). The covariances between the variables and factors are given by $\mathbf{B}_{(\text{all})}\mathbf{\Phi}$. The correlations between variables and factors are used to interpret the factors when the fit of the factor model is high.

Some examples of data for which three-mode factor analysis models may be useful and are the following:

- A depression scale of J items filled in by N persons on K time points. The first measurement may be pretreatment, while the consecutive measurements include the effect of the treatment. A three-mode factor model shows the loadings $\mathbf{B}_k$ for each measurement k on the same factors $\mathbf{F}$.
- Multitrait-multimethod data, where J traits are measured for N persons, using K different methods. Here, the loadings $\mathbf{B}_i$ are method-specific. See below for a discussion, and Section 3.1 for an example.

- A belief in a just world scale of J items filled in by N persons, where each item is both formulated as "justice for yourself" and "justice for others." Hence, we have $K = 2$ conditions. See Section 3.2 for a worked example.

So far, we have presented a general form of three-mode factor analysis. Various specific forms for the loadings $\mathbf{B}_k$ have been proposed in the literature. In the next subsection, we will give a short overview, and introduce our own model as an alternative. While existing models may suffer from estimation problems (e.g., convergence problems or nonadmissable solutions), identification problems, interpretation difficulties, or lack of parsimony, our model does not have these shortcomings, as we will see in Section 1.2.

1.2. Models for Three-Mode Factor Analysis

An important application of three-mode factor analysis is on multitrait-multimethod data, where the J variables measure one or several personality traits of N individuals, using K different methods. The covariance matrix $\mathbf{\Sigma}$ can be used to study trait validity across the different measurement methods (Campbell and Fiske, 1959). In this context, it has been proposed to assign each factor to a specific trait or to a specific method (e.g., Widaman, 1985). For example, if $J = 3$ variables measure three traits and $K = 2$ methods are used, then $R = 5$ factors may be included with loadings

$$\mathbf{B}_{(\text{all})} = \begin{bmatrix} \mathbf{B}_1 \\ \hline \mathbf{B}_2 \end{bmatrix} = \begin{bmatrix} * & 0 & 0 & * & 0 \\ 0 & * & 0 & * & 0 \\ 0 & 0 & * & * & 0 \\ \hline * & 0 & 0 & 0 & * \\ 0 & * & 0 & 0 & * \\ 0 & 0 & * & 0 & * \end{bmatrix},$$

where $*$ denotes an arbitrary nonzero entry. Hence, the first three factors correspond to the three traits, and the last two factors correspond to the two methods. Additionally, it may be required that the trait factors are uncorrelated with the two method factors. This can be done by constraining the corresponding covariances in $\mathbf{\Phi}$ to zero. In that case, the estimated common variances on the diagonal of $\mathbf{B}_{(\text{all})} \mathbf{\Phi} \mathbf{B}_{(\text{all})}^T$ can be written as the sum of a part due to the trait factors and a part due to the method factors. That is, it holds that

$$\mathbf{B}_{(\text{all})} \mathbf{\Phi} \mathbf{B}_{(\text{all})}^T = \mathbf{B}_{(\text{trait})} \mathbf{\Phi}_1 \mathbf{B}_{(\text{trait})}^T + \mathbf{B}_{(\text{meth})} \mathbf{\Phi}_2 \mathbf{B}_{(\text{meth})}^T, \tag{4}$$

where $\mathbf{B}_{(\text{trait})}$ and $\mathbf{B}_{(\text{meth})}$ contain the columns of $\mathbf{B}_{(\text{all})}$ corresponding to trait and method factors, respectively, and $\mathbf{\Phi}_1$ and $\mathbf{\Phi}_2$ are the covariance matrices of the trait and method factors, respectively.

These type of models are confirmatory factor analysis models and can be estimated by Maximum Likelihood (MLFA) (Jöreskog, 1970, 1971). For a detailed overview of this approach, and the related method of covariance component analysis, we refer to Wothke (1996). Identification results for factor models of this type can be found in Millsap (1992). Some problems may occur when fitting confirmatory factor analysis models for multitrait-multimethod data. Both convergence problems and improper solutions (e.g., where $\mathbf{\Phi}$ or $\mathbf{\Sigma} - \mathbf{U}$ is not a covariance matrix) are well known (Kiers, Takane, & Ten Berge, 1996). To overcome these problems, Kiers et al. (1996) proposed to fit constrained component models on the data instead. Here, the unique parts are treated as errors or noise, and the factor model (1) with constrained $\mathbf{B}_{(\text{all})}$ and $\mathbf{\Phi}$ is fitted directly to the data instead of fitting covariances. However, this alternative method does not overcome the problems in all cases.

Eid (2000) has proposed a class of confirmatory factor models for multitrait-multimethod data based on classical psychometric test theory. In such models, the number of method factors

is one less than the number of methods. As a result, one method is used as a comparison standard. Eid (2000) shows that his models are globally identified. Eid, Nussbeck, Geiser, Cole, Gollwitzer, and Lischetzke (2008) discuss how to choose an appropriate confirmatory factor model for different types of methods in multitrait-multimethod data. Although the approach based on Eid (2000) solves the identification problems and seems to solve the convergence problems of confirmatory factor models, it does not always produce a proper factor covariance matrix. For example, the factor covariance matrix in Table 4 of Eid et al. (2008) has many negative eigenvalues. Also, the fit of the model depends on which method is chosen as comparison standard, a choice which may not be obvious from a substantive point of view.

A different approach to three-mode factor analysis is based on the three-mode component model by Tucker (1966). In this model, each mode of the data has its own components, and their interaction strengths are given by numbers g_{rpq} of the so called *core array*. For our three-way data, suppose we have R components for the N observations, P components for the J variables, and Q components for the K conditions. The Tucker 3 model can then be written in the form of (1) as

$$\mathbf{X}_{(N \times JK)} = \mathbf{FG}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E}_{(N \times JK)},$$ (5)

where $\mathbf{C}$ is an $K \times Q$ matrix containing the Q method components as columns, $\mathbf{B}$ is a $J \times P$ matrix containing the P variable components as columns, $\mathbf{C} \otimes \mathbf{B}$ is the right direct or Kronecker product of $\mathbf{C}$ and $\mathbf{B}$, and $\mathbf{G}$ is the $R \times PQ$ matrix of interaction strengths, with

$$\mathbf{G} = \begin{bmatrix} g_{111} & \cdots & g_{1P1} & & \cdots & & g_{11Q} & \cdots & g_{1PQ} \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ g_{R11} & \cdots & g_{RP1} & & \cdots & & g_{R1Q} & \cdots & g_{RPQ} \end{bmatrix}.$$

It is well known that the three-mode Tucker model is not unique. All of $\mathbf{F}$, $\mathbf{C}$, and $\mathbf{B}$ can be rotated, with inverse transformations applied to the interaction strengths in $\mathbf{G}$, without affecting the model part $\mathbf{FG}(\mathbf{C} \otimes \mathbf{B})^T$. The covariance model corresponding to (5) is

$$\mathbf{\Sigma} = (\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T \Phi \mathbf{G}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{U}.$$ (6)

The associated probabilistic model (3) has been introduced by Bloxom (1968) and is further analyzed by Bentler and Lee (1978, 1979). For simplicity, one may consider $\mathbf{\Psi} = \mathbf{G}^T \Phi \mathbf{G}$ as factor covariance matrix. The case of diagonal $\mathbf{\Psi}$ can be rewritten as $\mathbf{\Sigma} = (\mathbf{CC}^T \otimes \mathbf{BB}^T) + \mathbf{U}$ and is known as the direct product model; see Browne (1984) or Wothke (1996). In this model, $\mathbf{CC}^T$ and $\mathbf{BB}^T$ may be seen as covariance matrices corresponding to methods and variables, respectively. The estimated common variances on the diagonal of $\mathbf{CC}^T \otimes \mathbf{BB}^T$ are obtained by multiplying the corresponding variances in $\mathbf{CC}^T$ (for the method) and $\mathbf{BB}^T$ (for the variable). In this sense, the direct product model is multiplicative, whereas confirmatory factor analysis models are additive; see (4).

Fitting the covariance model (6) can be done by directly fitting the component model (5) to the data. Alternating least squares algorithms minimizing the sum-of-squares of $\mathbf{E}_{(N \times JK)}$ can be found in Kroonenberg and De Leeuw (1980) and Kiers, Kroonenberg, and Ten Berge (1992). Contrary to the confirmatory factor analysis approach, convergence problems do not often occur. One may also fit the covariance model (6) by an algorithm for nonlinear optimization. For example, Bentler and Lee (1978, 1979) propose to use a Gauss–Newton algorithm. For an overview of three-mode component and factor models based on Tucker (1966), we refer to Kroonenberg and Oort (2003). For an accessible introduction to three-mode component analysis, see Kiers and Van Mechelen (2001).

As an alternative to confirmatory factor analysis and the three-mode component model of Tucker (1966), we propose a special case of the latter. Namely, the three-mode component model

known as Candecomp (Carroll & Chang, 1970) or Parafac (Harshman, 1970), which was originally introduced in mathematics (Hitchcock, 1927a, 1927b). The Candecomp/Parafac model is a special case of (5) in which $R = P = Q$ and $g_{rrr} = 1$ and $g_{rpq} = 0$, otherwise. Hence, we have an equal number of components in each mode, and there are no interactions between different component numbers. Analogous to (5), the Candecomp/Parafac model is given by

$$\mathbf{X}_{(N \times JK)} = \mathbf{F}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{E}_{(N \times JK)}, \tag{7}$$

where $\mathbf{C} \odot \mathbf{B} = [\mathbf{c}_1 \otimes \mathbf{b}_1 | \ldots | \mathbf{c}_R \otimes \mathbf{b}_R]$ is the (column-wise) Khatri–Rao product, and contains Kronecker products between the corresponding pairs of columns of $\mathbf{C}$ and $\mathbf{B}$. In the general three-mode factor model (1), the Candecomp/Parafac model corresponds to the case $\mathbf{B}_k = \mathbf{B}\mathbf{C}_k$, where $\mathbf{C}_k$ is the diagonal matrix with row k of $\mathbf{C}$ as its diagonal. In the Candecomp/Parafac model part, the matrices $\mathbf{F}$, $\mathbf{C}$, and $\mathbf{B}$ are unique up to scaling and a simultaneous column permutation under, e.g., the condition (Kruskal, 1977)

$$k_{\mathbf{F}} + k_{\mathbf{B}} + k_{\mathbf{C}} \geq 2R + 2, \tag{8}$$

where $k_{\mathbf{Y}}$ denotes the k-rank of a matrix $\mathbf{Y}$. The latter is defined as the largest number x such that every subset of x columns of $\mathbf{Y}$ is linearly independent. Other, more relaxed uniqueness conditions exist. However, (8) is easy to check and satisfies our current needs.

The covariance model corresponding to (7) is given by

$$\mathbf{\Sigma} = (\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{U}. \tag{9}$$

Also in the common covariance part of (9), matrices $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{\Phi}$ are unique up to scaling and permutation when (8) holds. For orthogonal factors, the covariance matrix of the common part can be written as

$$(\mathbf{C} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{B})^T = \sum_{r=1}^{R} (\mathbf{c}_r \otimes \mathbf{b}_r)(\mathbf{c}_r \otimes \mathbf{b}_r)^T = \sum_{r=1}^{R} (\mathbf{c}_r \mathbf{c}_r^T \otimes \mathbf{b}_r \mathbf{b}_r^T),$$

where $\mathbf{c}_r$ and $\mathbf{b}_r$ denote column r of $\mathbf{C}$ and $\mathbf{B}$, respectively. The estimated common variance for condition k and variable j can be expressed as $\sum_{r=1}^{R} c_{kr}^2 b_{jr}^2$. This shows that the estimated common variances in the Candecomp/Parafac covariance model (9) are obtained by both multiplications of method and variable coefficients, and additions over the number of factors.

Our estimation procedure for the Candecomp/Parafac covariance model first uses Minimum Rank Factor Analysis (see Section 1.3) to compute $\mathbf{U}$, which guarantees proper communalities on the diagonal of $\mathbf{\Sigma} - \mathbf{U}$. This is not the case for both confirmatory factor analysis models and the three-mode factor model based on Tucker (1966) (in the way it has been estimated so far). Next, we estimate $\mathbf{C}$, $\mathbf{B}$, and $\mathbf{\Phi}$ by the alternating least squares algorithm of Harshman (1970). We are able to compute the total percentage of explained common variance, as well as for each variable-condition combination separately. Also, the factor covariance matrix $\mathbf{\Phi}$ is guaranteed to be proper (i.e., have nonnegative eigenvalues). Our algorithm does not often suffer from convergence problems. Compared to the Tucker model of covariances (6), we only need to determine the number R of components instead of R, P, and Q. But a special case of Candecomp/Parafac also enables a choice of $R = P$ and $Q = 1$. Also, our Candecomp/Parafac covariance model (9) is unique under mild conditions, while the Tucker-based covariance model (6) is not. Analogous to two-way factor analysis methods, the common part of the covariance model (6) can be rotated (orthogonally or obliquely) to an approximately sparse matrix $\mathbf{G}$ of interaction strengths (Kiers, 1998a, 1998b; Tendeiro, Ten Berge, & Kiers, 2009). Hence, using the covariance model (6) involves a choice of rotation method.

Whether uniqueness in three-way factor analysis is desirable or not is not agreed upon. For example, as an anonymous reviewer stated, in psychology researchers are used to rotating two-way factor solutions and often consider the rotational freedom an advantage from an interpretational point of view. However, as we will illustrate, our unique Candecomp/Parafac covariance

model is also easy to interpret. Still, when the rotational freedom and interactions between different components of the Tucker-based covariance model are desired, it can also be incorporated in our estimation scheme. In this paper, however, we focus on the Candecomp/Parafac covariance model. A discussion on model selection and the comparison between Candecomp/Parafac and the model of Tucker (1966) can be found in Section 5.

Three-mode component models such as Tucker (1966) and Candecomp/Parafac and their four-mode and higher-mode extensions are applied in the social sciences (Kroonenberg, 2008), chemometrics (Smilde, Bro, & Geladi, 2004), independent component analysis (Comon & De Lathauwer, 2010; De Lathauwer, 2010), and data mining in general. For an overview of applications, see Kolda and Bader (2009) or Acar and Yener (2009).

1.3. Minimum Rank Factor Analysis

For later use, we discuss in some detail the Minimum Rank Factor Analysis (MRFA) method for two-mode factor analysis. Here, the covariance model is $\Sigma \approx \mathbf{B}\Phi\mathbf{B}^T + \mathbf{U}$, where Σ is the $J \times J$ data covariance matrix, $\mathbf{B}$ is the $J \times R$ loadings matrix, and $\mathbf{U}$ is the $J \times J$ diagonal matrix of unique variances. MRFA is used to estimate $(\mathbf{B}, \Phi, \mathbf{U})$; see Ten Berge and Kiers (1991). The MRFA algorithm computes the unique variances $\mathbf{U}$ such that $\mathbf{U}$ is nonnegative, $\Sigma - \mathbf{U}$ is a covariance matrix, and the unexplained common variance in $\Sigma - \mathbf{U} \approx \mathbf{B}\Phi\mathbf{B}^T$ is minimized. The matrix $\Sigma - \mathbf{U}$ is a covariance matrix if it is equal to $\mathbf{H}^T\mathbf{H}$ for some matrix $\mathbf{H}$. This is equivalent to $\Sigma - \mathbf{U}$ having nonnegative eigenvalues. When using MINRES or MLFA, it may happen that $\Sigma - \mathbf{U}$ is not a covariance matrix. The best approximation $\mathbf{B}\Phi\mathbf{B}^T$ is obtained from the R largest eigenvalues and associated eigenvectors of $\Sigma - \mathbf{U}$, and the minimum unexplained common variance in $\Sigma - \mathbf{U} \approx \mathbf{B}\Phi\mathbf{B}^T$ is equal to the sum of the $J - R$ smallest eigenvalues of $\Sigma - \mathbf{U}$; see Eckart and Young (1936).

The advantage of MRFA is that we have proper communalities and we can compute the percentage of explained common variance as

$$100 \cdot \frac{\text{trace}(\mathbf{B}\Phi\mathbf{B}^T)}{\text{trace}(\Sigma - \mathbf{U})}, \tag{10}$$

where trace($\cdot$) is defined as the sum of the diagonal entries of a matrix, which is equal to the sum of the eigenvalues of the matrix. The numerator of (10) equals the sum of the estimated common variances. The denominator equals the sum of the communalities (common variances), under the condition that the eigenvalues of $\Sigma - \mathbf{U}$ are nonnegative. To sum up, for a fixed number of R factors, MRFA minimizes the amount of common variance left unexplained under the constraint of proper communalities. A detailed comparison between MRFA and other factor analysis methods can be found in Sočan (2003).

Finally, we give an outline of the remaining part of the paper. In Section 2, we present our estimation procedure for the Candecomp/Parafac covariance model (9). In Section 3, we apply our Candecomp/Parafac method to several datasets in the literature, and compare the results with other methods. In Section 4, we assess the performance of our estimation procedure in a simulation study. Section 5 contains a discussion of our findings.

2. Three-Mode Factor Analysis by Means of Candecomp/Parafac

Here, we present our estimation procedure for the Candecomp/Parafac covariance model. After having computed the data covariance matrix Σ, the steps of our estimation procedure for the Candecomp/Parafac covariance model (9) are as follows:

1. Use the MRFA algorithm of Ten Berge and Kiers (1991) to estimate $\mathbf{U}$. This implies that $\mathbf{U}$ is nonnegative, $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, and the trace of $\mathbf{\Sigma} - \mathbf{L} - \mathbf{U}$ is minimal, where $\mathbf{L}$ is a best rank-R approximation of $\mathbf{\Sigma} - \mathbf{U}$.

2. Compute the eigendecomposition $\mathbf{\Sigma} - \mathbf{U} = \mathbf{VSV}^T$, with $\mathbf{V}$ having orthonormal columns, and $\mathbf{S}$ the diagonal matrix containing the eigenvalues in decreasing order. This is also the singular value decomposition of $\mathbf{\Sigma} - \mathbf{U}$. Let $\mathbf{P} = \mathbf{VS}^{1/2}$, which implies $\mathbf{\Sigma} - \mathbf{U} = \mathbf{PP}^T$. If $\mathbf{P}_R$ contains the first R columns of $\mathbf{P}$, then $\mathbf{L} = \mathbf{P}_R\mathbf{P}_R^T$ is a best rank-R approximation of $\mathbf{\Sigma} - \mathbf{U}$; see Eckart and Young (1936). In the next step, we approximate $\mathbf{P}$ by a Candecomp/Parafac decomposition with R components.

3. Fit the Candecomp/Parafac model as $\mathbf{P} \approx (\mathbf{C} \odot \mathbf{B})\mathbf{T}^T$ by using the alternating least squares algorithm of Harshman (1970). It is recommended to do, e.g., ten runs of the algorithm with random starting values, and to keep the best run as Candecomp/Parafac solution. Matrix $\mathbf{P}$ is a matricized $K \times J \times JK$ array, and $\mathbf{T}$ is a $JK \times R$ matrix. The columns of $\mathbf{T}$ are scaled such that they have length 1. For orthogonal factors, restrict the columns of $\mathbf{T}$ to be orthogonal. We obtain $\mathbf{\Sigma} - \mathbf{U} \approx (\mathbf{C} \odot \mathbf{B})\mathbf{T}^T\mathbf{T}(\mathbf{C} \odot \mathbf{B})^T$. We evaluate the Candecomp/Parafac fit as

$$100 - 100 \cdot \frac{\mathrm{ssq}(\mathbf{P} - (\mathbf{C} \odot \mathbf{B})\mathbf{T}^T)}{\mathrm{ssq}(\mathbf{P})}, \tag{11}$$

which is the percentage of the sum-of-squares of $\mathbf{P}$ that is fitted by $(\mathbf{C} \odot \mathbf{B})\mathbf{T}^T$. In the alternating least squares algorithm, $(\mathbf{C} \odot \mathbf{B})\mathbf{T}^T$ is the regression of $\mathbf{P}$ on $(\mathbf{C} \odot \mathbf{B})$. Since the regression and the residual are orthogonal, it follows that (11) is equal to $100 \cdot \mathrm{ssq}((\mathbf{C} \odot \mathbf{B})\mathbf{T}^T)/\mathrm{ssq}(\mathbf{P})$.

4. For oblique factors, let $\mathbf{\Phi} = \mathbf{T}^T\mathbf{T}$. For orthogonal factors, we have $\mathbf{\Phi} = \mathbf{T}^T\mathbf{T} = \mathbf{I}_R$.

Since we were not able to construct an algorithm for the simultaneous estimation of $\mathbf{U}$ and $\mathbf{C}, \mathbf{B}, \mathbf{\Phi}$ under the restriction that $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix, we instead estimate $\mathbf{U}$ and $\mathbf{C}, \mathbf{B}, \mathbf{\Phi}$ sequentially. First, we estimate $\mathbf{U}$ by MRFA based on a rank-R factor model for $\mathbf{\Sigma} - \mathbf{U}$. For fitting $(\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B})^T$ to $\mathbf{\Sigma} - \mathbf{U}$, we have considered two options. Either we approximate the best rank-R approximation of $\mathbf{\Sigma} - \mathbf{U}$ by the Candecomp/Parafac covariance part as $\mathbf{P}_R \approx (\mathbf{C} \odot \mathbf{B})\mathbf{T}_R^T$, where $\mathbf{T}_R$ is an $R \times R$ matrix, or we use steps 2 and 3 above, in which the Candecomp/Parafac covariance part is fitted on the full $\mathbf{P}$. Using the approximation of $\mathbf{P}_R$ has as advantage that it is consistent with the estimation of $\mathbf{U}$, which assumes a rank-R factor model. Moreover, we observed that the fit percentage of $\mathbf{P}_R \approx (\mathbf{C} \odot \mathbf{B})\mathbf{T}_R^T$ is usually above 99 % for datasets in the literature, which reduces the need for a simultaneous estimation procedure. However, when using the approximation of $\mathbf{P}_R$ it cannot be guaranteed that the estimated common variance is at most equal to the communality for each combination of variable and condition. Therefore, we have chosen to use the approximation of $\mathbf{P}$ instead, for which this is guaranteed (see further below). Conceptually, this introduces an inconsistency between the two estimation steps. However, the performance of both variants of the estimation procedure is very similar in our simulation study.

In the obtained matrices $\mathbf{C}$ and $\mathbf{B}$, there is still a scaling indeterminacy for each pair of columns (i.e., $(\lambda\mathbf{c}_r) \otimes (\lambda^{-1}\mathbf{b}_r) = \mathbf{c}_r \otimes \mathbf{b}_r$). This can be fixed by rescaling the columns of $\mathbf{B}$ to length 1. Note that step 4 guarantees a factor covariance matrix $\mathbf{\Phi}$ that is proper, i.e., has nonnegative eigenvalues.

The Candecomp/Parafac matrices $\mathbf{C}, \mathbf{B}, \mathbf{T}$ are unique up to permutation and scaling under mild conditions, e.g., (8). When both $(\mathbf{C} \odot \mathbf{B})$ and $\mathbf{T}$ have rank R, then also the Candecomp/Parafac covariance model $(\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B})^T$ is unique up to permutation and scaling. Note that $(\mathbf{C} \odot \mathbf{B})$ must have rank R for Candecomp/Parafac uniqueness (Liu & Sidiropoulos, 2001; Stegeman & Sidiropoulos, 2007). When desired, the Tucker-based covariance model

can replace our Candecomp/Parafac covariance model in step 3 above. In that case, $\mathbf{P} \approx (\mathbf{C} \otimes \mathbf{B})\mathbf{G}^T \mathbf{T}^T$ is estimated. However, this is outside the scope of this paper.

For orthogonal factors, it has been proven that a best-fitting Candecomp/Parafac model $\mathbf{P} \approx (\mathbf{C} \odot \mathbf{B})\mathbf{T}^T$ always exists (Krijnen, Dijkstra, & Stegeman, 2008). For oblique factors, this is not necessarily true, and one may encounter convergence problems and uninterpretable solutions (Harshman & Lundy, 1984; Kruskal, Harshman, & Lundy, 1989; Paatero, 2000; Harshman, 2004; Stegeman, 2006, 2007, 2008, 2009; Krijnen et al., 2008; De Silva & Lim, 2008). However, this problem has not occurred when analyzing reallife datasets such as those in Section 3. There are no general results stating conditions on a three-way array under which a best-fitting Candecomp/Parafac model exists. This is still an open problem (De Silva & Lim, 2008; Stegeman, 2006, 2007, 2008).

As goodness-of-fit measure we wish to use the percentage of explained common variance. The percentage of explained common variance in our Candecomp/Parafac covariance model is given by (11), which can be written as

$$100 \cdot \frac{\text{trace}((\mathbf{C} \odot \mathbf{B})\mathbf{T}^T \mathbf{T}(\mathbf{C} \odot \mathbf{B})^T)}{\text{trace}(\mathbf{\Sigma} - \mathbf{U})}. \tag{12}$$

If the factors are chosen orthogonal, then we may also obtain a percentage of explained common variance due to each factor. Namely, $\mathbf{T}^T \mathbf{T} = \mathbf{I}_R$ in (12), and we can define the explained common variance due to factor r as the sum-of-squares of column r of $(\mathbf{C} \odot \mathbf{B})$. This column is $\mathbf{c}_r \otimes \mathbf{b}_r$ and has sum-of-squares equal to $(\mathbf{c}_r^T \mathbf{c}_r)(\mathbf{b}_r^T \mathbf{b}_r)$. Hence, we have

$$100 \cdot \frac{\text{trace}((\mathbf{C} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{B})^T)}{\text{trace}(\mathbf{\Sigma} - \mathbf{U})} = \sum_{r=1}^{R} \left(100 \cdot \frac{(\mathbf{c}_r^T \mathbf{c}_r)(\mathbf{b}_r^T \mathbf{b}_r)}{\text{trace}(\mathbf{\Sigma} - \mathbf{U})} \right), \tag{13}$$

where the summands in the right-hand side express the percentage of explained common variance due to each factor.

To obtain a percentage of explained common variance for each variable-condition combination separately, we proceed as follows. Contrary to the two-mode case of MRFA, it may happen that some diagonal entry of $(\mathbf{C} \odot \mathbf{B})\mathbf{T}^T \mathbf{T}(\mathbf{C} \odot \mathbf{B})^T$ is larger than the corresponding communality on the diagonal of $\mathbf{\Sigma} - \mathbf{U}$. Because of this, we formulate the explained common variance per variable-condition combination analogous to (11) rather than to (12). A particular variable-condition corresponds to a row of $\mathbf{P} - (\mathbf{C} \odot \mathbf{B})\mathbf{T}^T$. Let row m of this matrix be denoted as $\mathbf{q}_m^T$. Then we define the corresponding percentage of explained common variance as

$$100 - 100 \cdot \frac{\text{ssq}(\mathbf{q}_m^T)}{(\mathbf{\Sigma} - \mathbf{U})_{mm}},$$

where $(\mathbf{\Sigma} - \mathbf{U})_{mm}$ is the corresponding communality.

In the Candecomp/Parafac covariance model (9) we have R components in the individuals mode, in the condition mode, and in the variable mode. Using a constrained Candecomp/Parafac model, it is possible to have different numbers of components in different modes. For example, instead of the Candecomp/Parafac model $(\mathbf{C}, \mathbf{B}, \mathbf{T})$ with $R = 2$, we can have $(\mathbf{c}\mathbf{\Psi}, \mathbf{B}, \mathbf{T})$, with $\mathbf{\Psi} = [1\ 1]$ being fixed. Hence, we have one component $\mathbf{c}$ in the condition mode, which interacts with both components in the individuals and variables modes. This type of Candecomp/Parafac model is known as Paralind (Bro, Harshman, Sidiropoulos, & Lundy, 2009) or Confac (De Almeida, Favier, & Mota, 2008a, 2008b), and can be fitted by an alternating least squares algorithm. Uniqueness conditions for Paralind models are proven in Stegeman and de Almeida (2009) and Stegeman and Lam (2012).The general Paralind model has form $(\mathbf{C}\mathbf{\Psi}, \mathbf{B}\mathbf{\Phi}, \mathbf{T}\mathbf{\Omega})$, where $\mathbf{C}, \mathbf{B}, \mathbf{T}$ have linearly independent columns, and $\mathbf{\Psi}, \mathbf{\Phi}, \mathbf{\Omega}$ are fixed constraint matrices.

TABLE 1.
Correlations of personality variables ambition (Am), attractiveness (At), leadership (Le), and extraversion (Ex), measured by self-report and peer-report.

Variable	Self-report				Peer-report			
	Am	At	Le	Ex	Am	At	Le	Ex
Am	1	0.223	0.337	0.223	0.402	0.035	0.160	0.093
At	0.223	1	0.418	0.290	0.070	0.442	0.196	0.180
Le	0.337	0.418	1	0.693	0.226	0.251	0.603	0.451
Ex	0.223	0.290	0.693	1	0.210	0.219	0.639	0.645
Am	0.402	0.070	0.226	0.210	1	0.233	0.379	0.269
At	0.035	0.442	0.251	0.219	0.233	1	0.314	0.283
Le	0.160	0.196	0.603	0.639	0.379	0.314	1	0.582
Ex	0.093	0.180	0.451	0.645	0.269	0.283	0.582	1

Note: Data taken from Bentler and Lee (1978).

3. Application to Datasets in the Literature

Here, we apply our three-mode factor model (9) to datasets in the literature. We use the estimation procedure outlined in Section 2. In Section 3.1, we consider a multitrait-multimethod (MTMM) dataset with four personality traits and two methods. Using a three-mode factor model (6) based on the Tucker (1966) three-mode component model, this dataset was analyzed by Bentler and Lee (1978). We show that our Candecomp/Parafac approach can incorporate the model of Bentler and Lee (1978), but has lower unique variances and a proper covariance matrix $\Sigma - U$. The latter is not the case in Bentler and Lee (1978). In Section 3.2, we consider a belief in a just world scale with eight items, where each item is asked from two perspectives: a just world for oneself and a just world for others. The data is part of a recent study among online American respondents.

3.1. MTMM Data from Bentler and Lee (1978)

In this dataset, $J = 4$ personality traits are measured with $K = 2$ methods. The traits are ambition, attractiveness, leadership, and extraversion. The methods are self-report and peer-report. The number of individuals in the dataset is $N = 72$. The correlation matrix Σ reported in Bentler and Lee (1978) is given in Table 1 above.

In Bentler and Lee (1978), the three-mode factor model (6) is used with $Q = 1$ method component, $P = 2$ components for the variables, and $R = 2$ orthogonal factors. The covariance model (6) is simplified to $G^T \Phi G = I_2$. As estimation procedure for (6), Bentler and Lee (1978) use a Gauss–Newton algorithm. Their results are:

$$\mathbf{C} = \begin{bmatrix} 1 \\ 0.85 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0.74 & 0 \\ 0.30 & 0.36 \\ 0.41 & 0.71 \\ 0.27 & 0.80 \end{bmatrix}, \tag{14}$$

and diag($\mathbf{U}$) = (0.63 0.71 0.52 0.47 0.69 0.74 0.58 0.63). The nonuniqueness of $\mathbf{C}$ and $\mathbf{B}$ has been fixed by setting the first entry of $\mathbf{C}$ equal to one, and rotating $\mathbf{B}$ such that its entry (1,2) is zero. The eigenvalues of $\Sigma - U$ are: $-0.31, -0.29, -0.27, -0.13, 0.25, 0.43, 0.52, 2.83$. The negative eigenvalues make tr($\Sigma - U$) useless as a measure of total common variance. The model has 8 parameters (not counting the unique variances) and the sum-of-squares of $\Sigma - U - (\mathbf{C} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{B})^T$ equals 0.67.

TABLE 2.
Percentages of estimated common variances (ECV %) and communalities for the Candecomp/Parafac covariance model
(9) with $R = 2$ factors fitted to the MTMM data from Bentler and Lee (1978).

Variable	ECV % oblique	ECV % orthogonal	ECV % Paralind	Communalities
Am, self	98	97	89	1.00
At, self	36	35	36	0.63
Le, self	81	81	83	0.81
Ex, self	87	88	89	0.81
Am, peer	55	58	63	0.46
At, peer	37	36	35	0.54
Le, peer	88	86	82	0.70
Ex, peer	84	82	78	0.63

The interpretation of (14) could be as follows. The same factors underly the variables for self-report and for peer-report. In the peer-report condition, the factors are a bit less pronounced (weight 0.85 opposed to weight 1 for self-report). The first factor can be interpreted as ambition (variable 1), and the second factor as leadership and extraversion (variables 3 and 4). The two factors are uncorrelated. Variable 2 (attractiveness) is not a good indicator of either of the two factor dimensions.

Next, we estimate the Candecomp/Parafac covariance model (9) with $R = 2$ oblique factors, using the estimation procedure in Section 2. That is, we fit $\mathbf{P} \approx (\mathbf{C} \odot \mathbf{B})\mathbf{T}^T$ with $\mathbf{P}$ such that $\mathbf{\Sigma} - \mathbf{U} = \mathbf{PP}^T$. The results are:

$$\mathbf{C} = \begin{bmatrix} 1.56 & 1.29 \\ 1.37 & 0.56 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0.39 & 0.90 \\ 0.36 & 0.25 \\ 0.62 & 0.32 \\ 0.58 & 0.15 \end{bmatrix}, \qquad \mathbf{\Phi} = \begin{bmatrix} 1 & -0.53 \\ -0.53 & 1 \end{bmatrix}, \qquad (15)$$

and diag($\mathbf{U}$) = (0 0.37 0.19 0.19 0.54 0.46 0.30 0.37). Here, $\mathbf{B}$ is rescaled to have columns of length 1. The unique variance of zero is a boundary solution. This may also occur for other models and estimation methods; see, e.g., Bentler and Lee (1979). For the Candecomp/Parafac decomposition ($\mathbf{C}, \mathbf{B}, \mathbf{T}$) we have $k_{\mathbf{C}} = k_{\mathbf{B}} = k_{\mathbf{T}} = 2$. Hence, by condition (8) it is unique up to permutation and scaling. The eigenvalues of $\mathbf{\Sigma} - \mathbf{U}$ are all nonnegative: 0, 0, 0, 0.18, 0.50, 0.73, 1.02, 3.14. The percentage of explained common variance is 74.56. The percentages for each trait-method combination are given in Table 2. Note that the unique variances in $\mathbf{U}$ are much lower than in the solution of Bentler and Lee (1978). The model has 11 parameters and the sum-of-squares of $\mathbf{\Sigma} - \mathbf{U} - (\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B})^T$ equals 0.84. This is larger than for the solution of Bentler and Lee (1978), but we estimate under the restriction that $\mathbf{\Sigma} - \mathbf{U}$ is a covariance matrix.

For the interpretation of (15) we compute $(\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}$; see Table 3. The first factor can be interpreted as leadership and extraversion (variables 3 and 4), and is similar to the second factor of the solution (14). The second factor can be interpreted as ambition, but only for self-report. This factor represents a difference in ambition measured by self-report and ambition measured by peer-report. This information cannot be obtained from (14) where both factors have the same weight for each condition (i.e., matrix $\mathbf{C}$ has only one column). The first factor in (14) is also interpreted as ambition, but there is no clear difference between methods. The larger weights in $\mathbf{C}$ for self-report in (14) also occur in $\mathbf{C}$ in (15), but are much more pronounced for the second factor. Note that in (15), the unique variance of ambition measured by self-report equals zero. The variance of this variable-method combination is almost completely explained in the covariance matrix $(\mathbf{C} \odot \mathbf{B})\mathbf{\Phi}(\mathbf{C} \odot \mathbf{B})^T$; see Table 2. The columns of $\mathbf{B}$ in (15) are somewhat similar to those of $\mathbf{B}$ in (14), except they are in reversed order, and the zero entry after rotation is not present in (15). The factors correlate -0.53 as opposed to being uncorrelated in (14).

TABLE 3.
Estimated values of $(\mathbf{C} \odot \mathbf{B})\boldsymbol{\Phi}$ for the Candecomp/Parafac covariance model (9) with $R = 2$ factors fitted to the MTMM data from Bentler and Lee (1978).

Variable	Oblique		Orthogonal		Paralind	
Am, self	−0.00	**0.84**	0.26	**0.97**	**0.85**	0
At, self	0.40	−0.02	0.45	0.15	0.24	0.45
Le, self	**0.75**	−0.09	**0.80**	0.16	0.40	**0.79**
Ex, self	**0.80**	−0.28	**0.80**	−0.01	0.26	**0.83**
Am, peer	0.26	0.23	0.25	0.40	**0.67**	0
At, peer	0.42	−0.12	0.43	0.06	0.18	0.35
Le, peer	**0.75**	−0.26	**0.76**	0.07	0.31	**0.62**
Ex, peer	**0.75**	−0.33	**0.76**	−0.00	0.20	**0.65**

Note: Numbers larger than 0.6 are in boldfont.

We also estimate the Candecomp/Parafac covariance model (9) with $R = 2$ orthogonal factors ($\boldsymbol{\Phi} = \mathbf{I}_2$). The results are as follows:

$$\mathbf{C} = \begin{bmatrix} 1.25 & 1.00 \\ 1.19 & 0.41 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0.21 & 0.97 \\ 0.36 & 0.15 \\ 0.64 & 0.17 \\ 0.64 & -0.01 \end{bmatrix}, \tag{16}$$

where $\mathbf{B}$ has columns of length 1. Again, the Candecomp/Parafac decomposition $(\mathbf{C}, \mathbf{B}, \mathbf{T})$ is unique up to permutation and scaling by condition (8), since $k_{\mathbf{C}} = k_{\mathbf{B}} = k_{\mathbf{T}} = 2$. The percentage of explained common variance equals 74.16, of which 53.4 % is due to factor one and 20.8 % is due to factor two; see (13). The model has 10 parameters and the sum-of-squares of $\boldsymbol{\Sigma} - \mathbf{U} - (\mathbf{C} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{B})^T$ equals 0.95. This is larger than for estimation with two oblique factors, since we use the restriction of orthogonal factors here.

We use the values of $(\mathbf{C} \odot \mathbf{B})$ to interpret (16); see Table 3. As (15), the two factors are clearly related to leadership and extraversion (factor one) and ambition (factor two). Matrix $\mathbf{C}$ is similar to (15), but now the difference in weights for self-report and peer-report pertains almost only to factor two. With $\boldsymbol{\Phi} = \mathbf{I}_2$, one only has to use $\mathbf{C}$ and $\mathbf{B}$ to interpret the factors. Compared to the solution (14) of Bentler and Lee (1978), which also features two orthogonal factors, solution (16) yields a unique and proper solution, is easier to interpret, and contains more information.

Finally, we estimate a solution for the Candecomp/Parafac covariance model with two orthogonal factors and $\mathbf{C} = [\mathbf{c} \ \mathbf{c}]$. This is a Paralind model, as explained at the end of Section 2. We have $(\mathbf{C} \odot \mathbf{B}) = (\mathbf{c} \otimes \mathbf{B})$, which implies that we actually have the same model as in Bentler and Lee (1978). Since the solution is not unique, we rotate $\mathbf{B}$ such that its (1,2) entry is zero, and rescale $\mathbf{C}$ such that its first entry equals 1. In this way, we obtain a solution that can be compared directly to (14). The result is

$$\mathbf{c} = \begin{bmatrix} 1 \\ 0.78 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0.85 & 0 \\ 0.24 & 0.45 \\ 0.40 & 0.79 \\ 0.26 & 0.83 \end{bmatrix},$$

which looks a lot like (14). However, the differences are that in our solution $\boldsymbol{\Sigma} - \mathbf{U}$ is a covariance matrix, and the unique variances in $\mathbf{U}$ are much smaller. The percentage of explained common variance equals 72.76. The model has 8 parameters and the sum-of-squares of $\boldsymbol{\Sigma} - \mathbf{U} - (\mathbf{c} \otimes \mathbf{B})(\mathbf{c} \otimes \mathbf{B})^T$ equals 1.09. This is larger than for the model with two orthogonal factors, since we use the restriction $\mathbf{C} = [\mathbf{c} \ \mathbf{c}]$.

3.2. Belief in a Just World Items of Lipkus et al. (1996)

To measure people's belief in a just world, Lipkus, Dalbert, and Siegler (1996) argue that it is important to distinguish between seeing the world as just or unjust for oneself and seeing the world as just or unjust for others. They administered a belief in a just world scale containing eight items, where each item is formulated both as "for yourself" and as "for others." The items for yourself are listed below. The items for others are the same, except that the words "me" and "I" are replaced by "people" and "they" (and the sentence is grammatically corrected).

1. I feel that I get what I am entitled to have.
2. I feel that my efforts are noticed and rewarded.
3. I feel that people treat me fairly in life.
4. I feel that I earn the rewards and punishments I get.
5. I feel that when I meet with misfortune, I have brought it upon myself.
6. I feel that I get what I deserve.
7. I feel that people treat me with the respect I deserve.
8. I feel that the world treats me fairly.

Each item is answered on a 7-point scale, where 1 represents "totally disagree" and 7 represents "totally agree." As part of a larger study, the items of Lipkus et al. (1996) were administered in September 2012 in an online survey containing 246 paid American respondents aged 18 to 82. After deleting persons with missing data, $N = 236$ persons are kept. Note that we have $J = 8$ items and $K = 2$ conditions (yourself/others). The 16×16 correlation matrix Σ is given in the Appendix.

Next, we estimate the Candecomp/Parafac covariance model (9), using the estimation procedure in Section 2. We use orthogonal factors to ease the interpretation of the solution. The solution with $R = 2$ factors yields one general factor and one "for yourself" factor, and has 74.51 % explained common variance. In the solution with $R = 3$, the factors have a more interesting interpretation. The results for $R = 3$ are:

$$\mathbf{C} = \begin{bmatrix} 1.78 & 1.32 & -0.68 \\ 1.35 & 1.62 & 1.13 \end{bmatrix}, \qquad \mathbf{B} = \begin{bmatrix} 0.30 & 0.35 & 0.16 \\ 0.44 & 0.24 & 0.29 \\ 0.45 & 0.17 & 0.48 \\ 0.28 & 0.44 & 0.25 \\ -0.02 & 0.56 & 0.18 \\ 0.28 & 0.46 & 0.30 \\ 0.42 & 0.16 & 0.49 \\ 0.42 & 0.23 & 0.50 \end{bmatrix},$$

(17)

$$\text{diag}(\mathbf{U}) = (0.34 \quad 0.13 \quad 0.16 \quad 0.19 \quad 0.29 \quad 0.15 \quad 0.21 \quad 0.11 \quad 0.21 \quad 0.25 \quad 0.15 \quad 0.23$$
$$0 \quad 0.11 \quad 0.19 \quad 0.11).$$

Again, $\mathbf{B}$ is rescaled to have columns of length 1. The Candecomp/Parafac decomposition $(\mathbf{C}, \mathbf{B}, \mathbf{T})$ is unique up to permutation and scaling by condition (8), since $k_\mathbf{C} = 2$ and $k_\mathbf{B} = k_\mathbf{T} = 3$. The eigenvalues of $\Sigma - \mathbf{U}$ are all nonnegative: 0, 0, 0, 0.01, 0.02, 0.08, 0.11, 0.17, 0.18, 0.27, 0.28, 0.33, 0.53, 1.24, 2.07, 7.88. The percentage of explained common variance is 84.54, where 38.0 % is due to factor 1, 33.3 % is due to factor 2, and 13.2 % is due to factor 3. Note that the fit of the model is quite good, with high explained common variance and low unique variances in $\mathbf{U}$. The sum-of-squares of $\Sigma - \mathbf{U} - (\mathbf{C} \odot \mathbf{B})(\mathbf{C} \odot \mathbf{B})^T$ equals only 1.06.

For the interpretation of the solution (17), we compute $(\mathbf{C} \odot \mathbf{B})$, which can be found in Table 4. Factor 1 is a mixture of especially items 2, 3, 7, 8, and has somewhat smaller loadings for especially items 1, 4, 6. The loadings are higher for the self-condition than for others. Factor 2 is a mixture of items 1, 4, 5, 6, and has higher loadings for others than for the self-condition.

TABLE 4.
Estimated values of $(\mathbf{C} \odot \mathbf{B})$ for the Candecomp/Parafac covariance model (9) with $R = 3$ orthogonal factors fitted to the belief in a just world data.

Item	Factor 1	Factor 2	Factor 3	Item	Factor 1	Factor 2	Factor 3
1, self	**0.53**	0.47	−0.11	1, others	0.40	**0.57**	0.18
2, self	**0.79**	0.31	−0.19	2, others	**0.60**	0.38	0.32
3, self	**0.80**	0.23	−0.32	3, others	**0.61**	0.28	**0.54**
4, self	**0.50**	**0.58**	−0.17	4, others	0.38	**0.71**	0.28
5, self	−0.04	**0.73**	−0.12	5, others	−0.03	**0.90**	0.21
6, self	0.49	**0.60**	−0.20	6, others	0.37	**0.74**	0.34
7, self	**0.76**	0.21	−0.33	7, others	**0.57**	0.25	**0.55**
8, self	**0.75**	0.31	−0.33	8, others	**0.57**	0.37	**0.56**

Note: Numbers larger than 0.5 are in boldfont.

One could say that the items of factor 1 pertain to a more implicit or passive idea of justice (e.g., how you are treated), while the items of factor 2 pertain to a more explicit or active idea of justice (e.g., what you get). The third factor represents a contrast between the two conditions. Its loadings are largest for items 3, 7, and 8 for the others condition. The presence of this contrast factor is in line with the substantive arguments in Lipkus et al. (1996) for distinguishing the two conditions (yourself/others) in belief in a just world scales.

4. Simulation Study

Here, we assess the performance of the estimation procedure outlined in Section 2 to retrieve underlying factors in simulated three-mode data. For given $\mathbf{C}, \mathbf{B}, \mathbf{\Phi}, \mathbf{U}$, we create random data $\mathbf{X}_{(N \times JK)}$ with population correlation matrix $\mathbf{\Sigma}$ in (9). Next, we apply the estimation procedure of Section 2 on the sample covariance matrix of $\mathbf{X}_{(N \times JK)}$, and check whether the underlying factors are retrieved. Below, we explain the details of our simulations.

We consider the cases of $R = 2$ and $R = 3$ factors. The true matrices $\mathbf{C}, \mathbf{B}$, and $\mathbf{\Phi}$ are the following. For $R = 2$, we use

$$\mathbf{C}_1 = \begin{bmatrix} 1.00 & 0.80 \\ 0.80 & 1.20 \end{bmatrix}, \qquad \mathbf{\Phi}_1 = \begin{bmatrix} 1 & -0.40 \\ -0.40 & 1 \end{bmatrix},$$

and

$$\mathbf{B}_1 = \begin{bmatrix} 0.80 & 0.10 \\ 0.10 & 0.83 \\ 0.83 & 0.10 \\ 0.10 & 0.79 \\ 0.83 & 0.10 \\ 0.10 & 0.82 \end{bmatrix}, \qquad \mathbf{B}_2 = \begin{bmatrix} 0.65 & 0.10 \\ 0.10 & 0.63 \\ 0.61 & 0.10 \\ 0.10 & 0.70 \\ 0.64 & 0.10 \\ 0.10 & 0.62 \end{bmatrix}.$$

Hence, we use one true matrix $\mathbf{C}_1$, and two true matrices $\mathbf{B}_1$ and $\mathbf{B}_2$. In the latter matrices, each item loads high on one factor only, and each factor has high loadings from three items. Matrix $\mathbf{B}_1$ has higher loadings than matrix $\mathbf{B}_2$. We consider both orthogonal factors ($\mathbf{\Phi} = \mathbf{I}_2$) and oblique factors ($\mathbf{\Phi} = \mathbf{\Phi}_1$). For $R = 3$, we take the true $\mathbf{A}$ and $\mathbf{B}$ from the solution (17) for the belief in a just world scales. The three factors are orthogonal.

After the true $\mathbf{C}$ and $\mathbf{B}$ are chosen, the unique variances on the diagonal of $\mathbf{U}$ are determined such that the population correlation matrix $\mathbf{\Sigma}$ in (9) has ones on the diagonal. For the sample size N, we consider $N = 100$ and $N = 500$. The data are generated as

$$\mathbf{X}_{(N \times JK)} = \mathbf{Z}_{(N \times JK)} \mathbf{\Sigma}^{1/2}, \tag{18}$$

TABLE 5.
Congruence coeffcients (mean and standard deviation, rounded to two decimals) for the simulated cases with $R = 2$ factors and sample size $N = 100$.

C	B	Φ	Comm.	Estimation	Congr. coeff. C		Congr. coeff. B		Div. comp.
C_1	B_1	I_2	0.63	Orthogonal	1.00	1.00	0.99	0.99	–
					(0.00)	(0.00)	(0.01)	(0.01)	
C_1	B_1	I_2	0.63	Oblique	1.00	1.00	0.99	0.99	0
					(0.00)	(0.00)	(0.01)	(0.01)	
C_1	B_2	I_2	0.39	Orthogonal	1.00	1.00	0.97	0.98	–
					(0.00)	(0.00)	(0.04)	(0.03)	
C_1	B_2	I_2	0.39	Oblique	1.00	1.00	0.96	0.97	0
					(0.00)	(0.00)	(0.07)	(0.04)	
C_1	B_1	Φ_1	0.57	Orthogonal	1.00	1.00	0.97	0.97	–
					(0.00)	(0.00)	(0.03)	(0.02)	
C_1	B_1	Φ_1	0.57	Oblique	1.00	1.00	0.98	0.98	2
					(0.00)	(0.00)	(0.03)	(0.02)	
C_1	B_2	Φ_1	0.35	Orthogonal	1.00	1.00	0.95	0.96	–
					(0.00)	(0.00)	(0.05)	(0.04)	
C_1	B_2	Φ_1	0.35	Oblique	1.00	1.00	0.94	0.96	2
					(0.00)	(0.00)	(0.07)	(0.05)	

TABLE 6.
Congruence coeffcients (mean and standard deviation, rounded to two decimals) for true C and B in (17), $R = 3$ orthogonal factors and sample size $N = 100$. Average communality is 0.69.

Estimation	Congr. coeff. C			Congr. coeff. B			Div. comp.
Orthogonal	1.00	1.00	0.98	0.99	0.98	0.99	–
	(0.00)	(0.00)	(0.03)	(0.02)	(0.03)	(0.01)	
Oblique	1.00	1.00	0.98	0.98	0.94	0.99	1
	(0.01)	(0.00)	(0.04)	(0.02)	(0.07)	(0.01)	

where $\mathbf{Z}_{(N \times JK)}$ has random entries from the standard normal distribution, and $\boldsymbol{\Sigma}$ is the population correlation matrix. For each choice of true model and sample size, we generate 100 datasets (18), and fit the Candecomp/Parafac covariance model (9) to the sample covariance matrix.

To assess factor retrieval, we compare the true values of the loadings in matrices $\mathbf{C}$ and $\mathbf{B}$ to their estimates. For this, we use the congruence coefficient of the true value and estimate of each column of $\mathbf{C}$ and $\mathbf{B}$. For two vectors $\mathbf{h}_1$ and $\mathbf{h}_2$, the congruence coefficient is given by

$$\frac{\mathbf{h}_1^T \mathbf{h}_2}{\sqrt{\mathbf{h}_1^T \mathbf{h}_1} \sqrt{\mathbf{h}_2^T \mathbf{h}_2}}.$$

For two-mode factor analysis, congruence coefficients of the columns of two loading matrices are used as a measure of factor similarity. Absolute values of 0.85 to 0.94 correspond to fair similarity, while an absolute value higher than 0.95 implies equal factors (Lorenzo-Seva & Ten Berge, 2006). Note that by using congruence coefficients, we do not need to deal with the scaling indeterminacy of the columns of $\mathbf{C}$ and $\mathbf{B}$.

In Tables 5 and 6, we report the mean and standard deviation of the congruence coefficients of the columns of $\mathbf{C}$ and $\mathbf{B}$ for each case. Also, the average communality is given for each case. In each case, we estimate the model using both orthogonal factors and oblique factors. We only report the results for sample size $N = 100$. For $N = 500$, the congruence coefficients are slightly larger and have slightly smaller standard deviation.

For $R = 2$, the recovery of the true loadings is very good in general. For larger communalities, the recovery is clearly better. Also, using orthogonal factors in the estimation improves the recovery when the true model has orthogonal factors. Note that using orthogonal factors in the estimation while the true model has oblique factors still results in relatively high congruence coefficients. For $R = 3$, the recovery results are also very good.

When oblique factors are used in the estimation, we also report the number of cases diverging components are encountered in the Candecomp/Parafac algorithm. Two components are labeled diverging when their congruence coefficient is smaller than -0.90; see, e.g., Stegeman (2012). The occurrence of diverging components often indicates nonexistence of a best-fitting Candecomp/Parafac model. As can be seen, there were only few such cases. For sample size $N = 500$, no cases of diverging components were encountered.

5. Discussion

In this paper, we have proposed and demonstrated a method for three-mode factor analysis using MRFA to estimate unique variances $\mathbf{U}$ and Candecomp/Parafac to estimate the covariance matrix of the common part. By using MRFA, the matrix $\mathbf{\Sigma} - \mathbf{U}$ is guaranteed to be a covariance matrix. This makes it possible to compute the percentage of explained common variance. For other methods of (two-mode or three-mode) factor analysis, this is often not possible. Also, our factor correlation matrix $\mathbf{\Phi}$ is guaranteed to be a covariance matrix.

By using the Candecomp/Parafac covariance model, the factors and weight matrices are unique up to permutation and scaling under mild conditions. This is not the case for covariance models based upon the three-mode model by Tucker (1966). Also, identifiability is often hard to prove for confirmatory factor analysis models. Besides yielding a unique solution, the Candecomp/Parafac covariance model is parsimonious and easy to interpret.

The simulation study shows that our relatively simple estimation procedure performs very well in retrieving underlying factors when the data is randomly sampled from a normal distribution with a covariance matrix $\mathbf{\Sigma}$ satisfying the Candecomp/Parafac covariance model.

To select the number of factors in our Candecomp/Parafac covariance model, one may use the criterion of increase in explained common variance. When adding an additional factor increases the percentage of explained common variance only little, one may decide not to include an additional factor in the analysis. However, one should also consider the added value of the additional factor in terms of interpretation.

Although we have focused on the Candecomp/Parafac covariance model, our approach can be generalized to any model for the common covariance part. That is, after estimating $\mathbf{U}$ by MRFA, any model can be fit to $\mathbf{\Sigma} - \mathbf{U}$. As mentioned in Sections 1.2 and 2, it may be convenient to apply a Tucker-based covariance model instead of Candecomp/Parafac. To determine whether a Candecomp/Parafac or Tucker-type model is appropriate, and which number(s) of factors should be used, Ceulemans and Kiers (2006) have proposed a method based on selecting the model with the highest fit for an acceptable number of free parameters. Naturally, also the size of the interaction terms in a Tucker-type model, the desirability of rotational nonuniqueness, and ease of interpretation should be taken into account when selecting an appropriate model.

Acknowledgements

Appendix: Belief in a Just World Data

Item	For yourself								For others							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
1	1	0.56	0.46	0.51	0.25	0.53	0.42	0.52	0.61	0.50	0.40	0.46	0.29	0.52	0.39	0.41
2	0.56	1	0.75	0.63	0.24	0.55	0.68	0.72	0.40	0.50	0.49	0.37	0.12	0.38	0.35	0.38
3	0.46	0.75	1	0.60	0.19	0.51	0.77	0.80	0.33	0.44	0.37	0.32	0.05	0.31	0.29	0.27
4	0.51	0.63	0.60	1	0.47	0.67	0.52	0.61	0.50	0.51	0.41	0.61	0.33	0.54	0.34	0.37
5	0.25	0.24	0.19	0.47	1	0.56	0.16	0.30	0.35	0.30	0.28	0.47	0.65	0.41	0.22	0.27
6	0.53	0.55	0.51	0.67	0.56	1	0.51	0.65	0.47	0.48	0.43	0.55	0.40	0.61	0.35	0.39
7	0.42	0.68	0.77	0.52	0.16	0.51	1	0.73	0.35	0.43	0.35	0.31	0.04	0.34	0.34	0.31
8	0.52	0.72	0.80	0.61	0.30	0.65	0.73	1	0.37	0.49	0.43	0.40	0.18	0.46	0.36	0.42
1	0.61	0.50	0.40	0.46	0.29	0.52	0.39	0.41	1	0.62	0.46	0.62	0.52	0.66	0.50	0.54
2	0.50	0.50	0.49	0.37	0.12	0.38	0.35	0.38	0.62	1	0.67	0.61	0.34	0.65	0.57	0.57
3	0.40	0.49	0.37	0.32	0.05	0.31	0.29	0.27	0.46	0.67	1	0.55	0.35	0.59	0.74	0.73
4	0.46	0.37	0.32	0.61	0.47	0.55	0.31	0.40	0.62	0.61	0.55	1	0.62	0.77	0.51	0.56
5	0.29	0.12	0.05	0.33	0.65	0.40	0.04	0.18	0.52	0.34	0.35	0.62	1	0.62	0.36	0.40
6	0.52	0.38	0.31	0.54	0.41	0.61	0.34	0.46	0.66	0.65	0.59	0.77	0.62	1	0.57	0.65
7	0.39	0.35	0.29	0.34	0.22	0.35	0.34	0.36	0.50	0.57	0.74	0.51	0.36	0.57	1	0.77
8	0.41	0.38	0.27	0.37	0.27	0.39	0.36	0.42	0.54	0.57	0.73	0.56	0.40	0.65	0.77	1

Note: Data provided by K. Stroebe, Department of Social Psychology, University of Groningen.

References

Acar, E., & Yener, B. (2009). Unsupervised multiway data analysis: a literature survey. *IEEE Transactions on Knowledge and Data Engineering*, *21*, 1–15.

Bentler, P.M., & Lee, S.-Y. (1978). Statistical aspects of a three-mode factor analysis model. *Psychometrika*, *43*, 343–352.

Bentler, P.M., & Lee, S.-Y. (1979). A statistical development of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology*, *32*, 87–104.

Bloxom, B. (1968). A note on invariance in three-mode factor analysis. *Psychometrika*, *33*, 347–350.

Bro, R., Harshman, R.A., Sidiropoulos, N.D., & Lundy, M.E. (2009). Modeling multi-way data with linearly dependent loadings. *Journal of Chemometrics*, *23*, 324–340.

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

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

Carroll, J.D., & Chang, J.J. (1970). Analysis of individual differences in multidimensional scaling via an *n*-way generalization of Eckart–Young decomposition. *Psychometrika*, *35*, 283–319.

Ceulemans, E., & Kiers, H.A.L. (2006). Selecting among three-mode principal component models of different types and complexities: a numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology*, *59*, 133–150.

Comon, P., & De Lathauwer, L. (2010). Algebraic identification of under-determined mixtures. In P. Comon & C. Jutten (Eds.), *Handbook of blind source separation: independent component analysis and applications* (pp. 325–366). San Diego: Academic Press.

De Almeida, A.L.F., Favier, G., & Mota, J.C.M. (2008a). Constrained tensor modeling approach to blind multiple-antenna CDMA schemes. *IEEE Transactions on Signal Processing*, *56*, 2417–2428.

De Almeida, A.L.F., Favier, G., & Mota, J.C.M. (2008b). A constrained factor decomposition with application to MIMO antenna systems. *IEEE Transactions on Signal Processing*, *56*, 2429–2442.

De Lathauwer, L. (2010). Algebraic methods after prewhitening. In P. Comon & C. Jutten (Eds.), *Handbook of blind source separation: independent component analysis and applications* (pp. 155–178). San Diego: Academic Press.

De Silva, V., & Lim, L.-H. (2008). Tensor rank and the ill-posedness of the best low-rank approximation problem. *SIAM Journal on Matrix Analysis and Applications*, *30*, 1084–1127.

Eckart, C., & Young, G. (1936). The approximation of one matrix by another of lower rank. *Psychometrika*, *1*, 211–218.

Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*, 241–261.

Eid, M., Nussbeck, F.W., Geiser, C., Cole, D.A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: different models for different types of methods. *Psychological Methods*, *13*, 230–253.

Harshman, R.A. (1970). Foundations of the Parafac procedure: models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, *16*, 1–84.

Harshman, R.A., & Lundy, M.E. (1984). Data preprocessing and the extended Parafac model. In H.G. Law, C.W. Snyder Jr., J.A. Hattie, & R.P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 216–284). New York: Praeger.

Harshman, R.A. (2004). The problem and nature of degenerate solutions or decompositions of 3-way arrays. In *Talk at the tensor decompositions workshop*, Palo Alto, USA.

Hitchcock, F.L. (1927a). The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics*, *6*, 164–189.

Hitchcock, F.L. (1927b). Multiple invariants and generalized rank of a *p*-way matrix or tensor. *Journal of Mathematics and Physics*, *7*, 39–70.

Jöreskog, K.G. (1970). A general method for analysis of covariance structures. *Biometrika*, *57*, 239–251.

Jöreskog, K.G. (1971). Statistical analysis of sets of congeneric tests. *Psychometrika*, *36*, 109–133.

Kiers, H.A.L. (1998a). Three-way SIMPLIMAX for oblique rotation of the three-mode factor analysis core to simple structure. *Computational Statistics and Data Analysis*, *28*, 307–324.

Kiers, H.A.L. (1998b). Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis. *Journal of Classification*, *15*, 245–263.

Kiers, H.A.L., Kroonenberg, P.M., & Ten Berge, J.M.F. (1992). An efficient algorithm for TUCKALS3 on data with large number of observation units. *Psychometrika*, *57*, 415–422.

Kiers, H.A.L., Takane, Y., & Ten Berge, J.M.F. (1996). The analysis of multitrait-multimethod matrices via constrained component analysis. *Psychometrika*, *61*, 601–628.

Kiers, H.A.L., & Van Mechelen, I. (2001). Three-way component analysis: principles and illustrative application. *Psychological Methods*, *6*, 84–110.

Kolda, T.G., & Bader, B.W. (2009). Tensor decompositions and applications. *SIAM Review*, *51*, 455–500.

Krijnen, W.P., Dijkstra, T.K., & Stegeman, A. (2008). On the non-existence of optimal solutions and the occurrence of "degeneracy" in the Candecomp/Parafac model. *Psychometrika*, *73*, 431–439.

Kroonenberg, P.M., & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika*, *45*, 69–97.

Kroonenberg, P.M., & Oort, F.J. (2003). Three-mode analysis of multimode covariance matrices. *British Journal of Mathematical and Statistical Psychology*, *56*, 305–335.

Kroonenberg, P.M. (2008). *Wiley series in probability and statistics*. *Applied multiway data analysis*. Hoboken: Wiley.

Kruskal, J.B. (1977). Three-way arrays: rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and Its Applications*, *18*, 95–138.

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.), *Multiway data analysis* (pp. 115–121). Amsterdam: North-Holland.

Lipkus, I.M., Dalbert, C., & Siegler, I.C. (1996). The importance of distinguishing the belief in a just world for self versus for others: implications for psychological well-being. *Personality and Social Psychology Bulletin*, *22*, 666–677.

Liu, X., & Sidiropoulos, N.D. (2001). Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays. *IEEE Transactions on Signal Processing*, *49*, 2074–2086.

Lorenzo-Seva, U., & Ten Berge, J.M.F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *2*, 57–64.

Millsap, R.E. (1992). Sufficient conditions for rotational uniqueness in the additive MTMM model. *British Journal of Mathematical and Statistical Psychology*, *45*, 125–138.

Paatero, P. (2000). Construction and analysis of degenerate Parafac models. *Journal of Chemometrics*, *14*, 285–299.

Smilde, A., Bro, R., & Geladi, P. (2004). *Multi-way analysis: applications in the chemical sciences*. Chichester: Wiley.

Sočan, G. (2003). *The incremental value of minimum rank factor analysis*. Ph.D. Thesis, Department of Psychometrics & Statistics, University of Groningen, Groningen, The Netherlands.

Stegeman, A. (2006). Degeneracy in Candecomp/Parafac explained for $p \times p \times 2$ arrays of rank $p + 1$ or higher. *Psychometrika*, *71*, 483–501.

Stegeman, A. (2007). Degeneracy in Candecomp/Parafac explained for several three-sliced arrays with a two-valued typical rank. *Psychometrika*, *72*, 601–619.

Stegeman, A. (2008). Low-rank approximation of generic $p \times q \times 2$ arrays and diverging components in the Candecomp/Parafac model. *SIAM Journal on Matrix Analysis and Applications*, *30*, 988–1007.

Stegeman, A. (2009). Using the simultaneous generalized Schur decomposition as a Candecomp/Parafac algorithm for ill-conditioned data. *Journal of Chemometrics*, *23*, 385–392.

Stegeman, A. (2012). Candecomp/Parafac—from diverging components to a decomposition in block terms. *SIAM Journal on Matrix Analysis and Applications*, *33*, 291–316.

Stegeman, A., & Sidiropoulos, N.D. (2007). On Kruskal's uniqueness condition for the Candecomp/Parafac decomposition. *Linear Algebra and Its Applications*, *420*, 540–552.

Stegeman, A., & de Almeida, A.L.F. (2009). Uniqueness conditions for constrained three-way factor decompositions with linearly dependent loadings. *SIAM Journal on Matrix Analysis and Applications*, *31*, 1469–1490.

Stegeman, A., & Lam, T.T.T. (2012). Improved uniqueness conditions for canonical tensor decompositions with linearly dependent loadings. *SIAM Journal on Matrix Analysis and Applications*, *33*, 1250–1271.

Ten Berge, J.M.F., & Kiers, H.A.L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*, 309–315.

Tendeiro, J.N., Ten Berge, J.M.F., & Kiers, H.A.L. (2009). Simplicity transformations for three-way arrays with symmetric slices, and applications to Tucker-3 models with sparse core arrays. *Linear Algebra and Its Applications*, *430*, 924–940.

Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, *31*, 279–311.

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

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