

# Three-mode principal component analysis and perfect congruence analysis for sets of covariance matrices

Pieter M. Kroonenberg†

*Department of Education, University of Leiden, PO Box 9507, 2300 RA Leiden, The Netherlands*

Jos M. F. ten Berge

*University of Goningen*

In this paper three-mode principal component analysis and perfect congruence analysis for weights applied to sets of covariance matrices are explained and detailed, and the relationships between the two techniques are explored. It is shown that given several assumptions are made for three-mode principal component analysis close links between the two techniques exist. The methods are illustrated with data pertaining to a theory of self-concept.

## 1. Introduction

Not uncommonly, sets of covariance or correlation matrices are collected in applied research. In such instances, a vital question is whether there exists one underlying set of components for all matrices, and to what extent such a common set of components is shared by all matrices. This issue is often approached in a confirmatory manner via linear structural equations or simultaneous factor analysis (see e.g. Jöreskog, 1971; Jöreskog & Sörbom, 1983; McDonald, 1984). Unless a particular component or factor space has been hypothesized, the use of such techniques reverts to model searches, rather than to pure estimation of parameters. In other words, confirmatory procedures are used for exploration. Even though this may be reasonable in some cases, one might wonder why the problem is not approached directly with exploratory techniques which are far less restrictive. Both exploratory and confirmatory methods have their merits and weaknesses, but we tend to prefer exploratory approaches, certainly in the initial stages of an investigation. There are several exploratory techniques available to address the issue, but here the focus is exclusively on three-mode principal component analysis and perfect congruence analysis. Some alternative procedures are parallel factor analysis (PARAFAC), see

†Requests for reprints.

Harshman (1970), Harshman & Lundy (1984), and several longitudinal procedures developed by Meredith & Tisak (1982), and Tisak & Meredith (in press).

In several applications of three-mode principal component analysis to sets of correlation matrices, results turned out to be very similar to results obtained via perfect congruence analysis for weights. One of the aims of the present paper is to put forward reasons why these similarities could be observed. At the same time, details are worked out for the application and interpretation of three-mode principal component analysis to sets of covariances matrices, which has only been done sketchily and incompletely before. The results presented in this paper are an extension of those by Kroonenberg & ten Berge (1985, 1987), however, the similarities and differences between the techniques were not yet fully understood and developed in the earlier presentations.

Rather than the more general model presented in Tucker (1966), the so-called Tucker2 model will be discussed (Kroonenberg, 1983a; Kroonenberg & de Leeuw, 1980; Tucker, 1972), in which only components are computed for two of the three modes. Perfect congruence analysis for weights—PCW—will not be treated as extensively, because its theory and background have been dealt with adequately in ten Berge (1986b). Previous (incomplete) applications of three-mode PCA to sets of correlation matrices may be found in Kroonenberg (1983a, ch. 11, 1983b), an application of PCW is van Schuur (1984, p. 129ff).

After a short recapitulation of classical PCA, three-mode PCA for sets of covariance matrices is presented as a generalization of classical PCA, followed by a summary statement of PCW. In order to gain some empirical perspective on the two techniques, outcomes are compared using data taken from Marsh & Hocevar (1985) collected to investigate the adequacy of Shavelson's model of self-concept (Shavelson & Bolus, 1982; Shavelson, Hubner & Stanton, 1976). In the second part of the paper, attention will be paid to transforming the component space of the three-mode analysis via orthonormal or non-singular transformations. In particular, it will be investigated why and how the results from the three-mode analysis with these transformations become very similar to those of perfect congruence analysis.

## 2. Classical principal component analysis

In this paper the matrix  $Z$  (and later  $Z_k$ ) with  $l$  subjects as rows and  $m$  variables as columns will be always a matrix of deviation scores scaled in such a way that  $R = Z'Z$  is its covariance matrix, that is factors like  $l^{-1}$  will be ignored. The  $s$  components ( $s \leq m$ ) can be constructed by applying a matrix  $B$  with weights to  $Z$  to obtain the component score matrix  $F$ , that is  $F = ZB$ . The covariances of the variables with the components may be collected in  $S$  ( $m \times s$ ), with

$$S \equiv Z'F = Z'ZB = RB, \quad (1)$$

and the covariances of the components in  $\Phi$  ( $s \times s$ ), with

$$\Phi \equiv F'F = B'Z'ZB = B'RB. \quad (2)$$

The matrix of regression weights  $\mathbf{P}$  ( $s \times m$ ) for 'reconstructing'  $\mathbf{Z}$  from  $\mathbf{F}$  (i.e.  $\hat{\mathbf{Z}} = \mathbf{F}\mathbf{P}'$ ) can be shown to be

$$\mathbf{P} = \mathbf{Z}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1} = \mathbf{S}\mathbf{\Phi}^{-1}. \quad (3)$$

The matrix  $\mathbf{S}$  is called the *structure*,  $\mathbf{P}$  is called the *pattern* (and unfortunately both are often referred to as 'loading' matrix). From  $\hat{\mathbf{Z}} = \mathbf{F}\mathbf{P}'$  and  $\mathbf{F} = \mathbf{Z}\mathbf{B}$ , it follows that the reconstruction of  $\mathbf{Z}$  via the model is achieved via the equation  $\hat{\mathbf{Z}} = \mathbf{Z}\mathbf{B}\mathbf{P}'$ .

For the link between classical PCA and three-mode PCA it is necessary to present some further information on the relationship between  $\mathbf{B}$  and  $\mathbf{P}$ . In particular, it follows from the above development that

$$\mathbf{P} = \mathbf{R}\mathbf{B}(\mathbf{B}'\mathbf{R}\mathbf{B})^{-1}, \quad (4)$$

and thus  $\mathbf{B}'\mathbf{P} = \mathbf{I}$ , which, however, does not imply that  $\mathbf{B} = \mathbf{P}$ . It can be shown that  $\mathbf{B} = \mathbf{P}(\mathbf{P}'\mathbf{P})^{-1}$  if and only if  $\mathbf{B}$  contains (un)rotated eigenvectors of  $\mathbf{R}$ . From this result it can be derived that  $\mathbf{B} = \mathbf{P}$  if and only if  $\mathbf{B}$  is an orthonormal matrix with eigenvectors of  $\mathbf{R}$  or an orthogonal rotation thereof.

An alternative way to derive classical principal component analysis is by specifying the model

$$\mathbf{Z} = \mathbf{F}\mathbf{P}' + \mathbf{E}, \quad (5)$$

where  $\mathbf{Z}$ ,  $\mathbf{F}$ ,  $\mathbf{P}$ , and  $\mathbf{E}$  are the  $l \times m$  data matrix, the  $l \times s$  component scores, the  $m \times s$  component pattern, and the  $l \times m$  residuals respectively. As always there exists a rotational (or fundamental) indeterminacy as  $\tilde{\mathbf{F}} = \mathbf{F}\mathbf{T}$ , and  $\tilde{\mathbf{P}} = \mathbf{P}(\mathbf{T}')^{-1}$  yield the same fit to the data. The  $\mathbf{F}$  and  $\mathbf{P}$  may be found by minimizing the loss function

$$\|\mathbf{Z} - \mathbf{F}\mathbf{P}'\|^2 \quad (6)$$

(see, e.g., Keller, 1962) under the identification constraint that  $\mathbf{F}$  is orthonormal. The above development emphasizes PCA as a decomposition of the data, rather than as a search for linear combinations of  $\mathbf{Z}$ , and three-mode component analysis is generally developed from the latter perspective rather than the former one.

### 3. Three-mode principal component analysis

In two papers, Tucker (1966, 1972) developed the major characteristics of what he called three-mode factor analysis. In fact, he devised more than one type of model, and more than one method to solve them. In the present paper only the Tucker2 model (using Kroonenberg & de Leeuw's, 1980, terminology), in which components are computed for two of the three modes will be considered. In addition, only those aspects of this model will be taken into account which are relevant for the analysis of sets of covariance matrices.

## 3.1. Common components

Let  $\mathbf{R} = \{\mathbf{R}_1, \dots, \mathbf{R}_n\}$  be a set of  $n$  covariance matrices of order  $m \times m$ , defined on the same set of  $m$  variables. The numbers of subjects in each of the  $n$  samples may be different or the same, but they will not explicitly enter the discussion. They have, of course, influence on the statistical stability of the covariances, but this problem will not be discussed here. For the set of covariance matrices the Tucker2 model is defined as

$$\mathbf{R}_k = \hat{\mathbf{R}}_k + \mathbf{E}_k = \mathbf{P}\mathbf{C}_k\mathbf{P}' + \mathbf{E}_k \quad (k = 1, \dots, n),$$

or

$$\mathbf{R}_k \approx \mathbf{P}\mathbf{C}_k\mathbf{P}', \quad (7)$$

where  $\mathbf{P}$  is an  $m \times s$  matrix of rank  $s$  with component *coefficients*. We use here the somewhat unusual term 'coefficients' to avoid *a priori* interpretations, such as *loadings*, *patterns*, *scores* or *weights*, terms well-defined within classical PCA. The  $\mathbf{C}_k$  are the so-called core slices or core planes, and their elements  $c_{pqk}$  may be referred to as *mutual weights for components* or *mutual weights* for short. Finally the  $\mathbf{E}_k$  represent the matrices with residual covariances for the  $k$ th group, which are not fitted when using  $s$  ( $s \leq m$ ) components. The subjects that have gone into making the covariance matrices may be different in each sample, and even the numbers of subjects in the various samples need not be the same.

There are many models for the data themselves which may underlie the model for covariances in (7). In particular, the data may be modelled as in classical PCA (5), but with one pattern matrix  $\mathbf{P}$  for all  $\mathbf{Z}_k$ , that is

$$\mathbf{Z}_k \approx \mathbf{F}_k\mathbf{P}' \quad (8)$$

so that the covariance matrix becomes

$$\mathbf{Z}_k'\mathbf{Z}_k \approx \mathbf{P}\mathbf{F}_k'\mathbf{F}_k\mathbf{P}' = \mathbf{P}\mathbf{C}_k\mathbf{P}'. \quad (9)$$

Alternatively, in the case that the subjects in each sample are the same, one could assume that the Tucker2 model itself underlies the data, that is

$$\mathbf{Z}_k \approx \mathbf{H}\mathbf{W}_k\mathbf{P}', \quad (10)$$

with  $\mathbf{H}$  an  $l \times t$  matrix of components for the subjects and  $\mathbf{W}_k$  the  $k$ th slice of the core matrix of order  $t \times s$ , so that the covariances become

$$\mathbf{Z}_k'\mathbf{Z}_k \approx \mathbf{P}'\mathbf{W}_k'\mathbf{H}'\mathbf{H}\mathbf{W}_k\mathbf{P}' = \mathbf{P}\mathbf{C}_k\mathbf{P}'. \quad (11)$$

An even more restricted model for the  $\mathbf{Z}_k$  would be Harshman's PARAFAC model (Harshman & Lundy, 1984), that is

$$\mathbf{Z}_k \approx \mathbf{H}\mathbf{D}_k\mathbf{P}', \quad (12)$$



where the diagonal elements of the diagonal  $\mathbf{D}_k$  are the component 'loadings' for the  $k$ th sample, leading to a covariance matrix of the form

$$\mathbf{Z}'_k \mathbf{Z}_k \approx \mathbf{P} \mathbf{D}_k \mathbf{H}' \mathbf{H} \mathbf{D}_k \mathbf{P}' = \mathbf{P} \mathbf{C}_k \mathbf{P}', \quad (13)$$

and other models may be conceived in a similar manner. In other words, depending on a further model for the  $\mathbf{C}_k$ , many different data structures may underlie the model for covariances in (7).

The approach taken in (12) and (13) has been further developed by Harshman & Lundy (1984), and they call this approach 'indirect fitting' of the PARAFAC model. Similarly, (10) and (11) may be called indirect fitting of the Tucker2 model. By further specifying models for the  $\mathbf{C}_k$  one may try to build hierarchies of models with decreasing numbers of parameters, not unlike the PINDIS procedure of Lingoes & Borg (1978). At present, however, only the more general form (8) extending classical PCA is treated, which means that  $\mathbf{P}$  will be interpreted as the common pattern matrix, and the  $\mathbf{C}_k$  as the covariances of the component scores  $\mathbf{F}_k$  on the common components.

In their treatment of three-mode PCA Kroonenberg & de Leeuw (1980; Kroonenberg, 1983a) usually start with an orthonormality assumption for  $\mathbf{P}$  for computational convenience as their computer program TUCKALS2 (Kroonenberg & Brouwer, 1985) is based on the orthonormality of  $\mathbf{P}$ . In the sequel the orthonormal matrix will be referred to as  $\mathbf{G}$ , and  $\mathbf{P}$  will be reserved for pattern (or pattern-like) matrices. Given a solution is found for (7) the orthonormal  $\mathbf{G}$  may be transformed non-singularly without loss of fit provided the inverse transformation is applied to the  $\mathbf{C}_k$ . We will explicitly use such transformations in the sequel. Incidentally, when  $n=1$  and  $s=m$  (7) reduces to the standard eigenvalues-eigenvector decomposition of  $\mathbf{R}_1$  with  $\mathbf{C}_1$  a diagonal matrix. When  $n>1$  the  $\mathbf{C}_k$  are not necessarily diagonal for all  $k$ . When they are, (7) is a special case of Harshman's PARAFAC model (Harshman, 1970; Harshman & Lundy, 1984).

The TUCKALS2 solution for the Tucker2 model (7) is found (cf. Kroonenberg, 1983a, ch. 4; Kroonenberg & de Leeuw, 1980) by minimizing the least squares loss function  $\sum_k \|\mathbf{R}_k - \mathbf{G} \mathbf{C}_k \mathbf{G}'\|^2$ . The loss function is first solved for  $\mathbf{G}$  and then the  $\mathbf{C}_k$  are computed as

$$\mathbf{C}_k = \mathbf{G}' \mathbf{R}_k \mathbf{G} = \mathbf{G}' \mathbf{Z}'_k \mathbf{Z}_k \mathbf{G} \quad (k=1, \dots, n), \quad (14)$$

where the  $\mathbf{Z}_k$  are the column-centred data matrices.

It is instructive to look somewhat more closely at the characteristics of  $\mathbf{G}$ . To do this, the expression for the core slices (14) may be substituted into (7), the equation specifying the Tucker2 model for covariance matrices, so that

$$\mathbf{Z}'_k \mathbf{Z}_k = \mathbf{R}_k \approx \hat{\mathbf{R}}_k = \mathbf{G} \mathbf{G}' \mathbf{R}_k \mathbf{G} \mathbf{G}' = \mathbf{G} \mathbf{G}' \mathbf{Z}'_k \mathbf{Z}_k \mathbf{G} \mathbf{G}' = \hat{\mathbf{Z}}'_k \hat{\mathbf{Z}}_k, \quad (15)$$

which implies that  $\hat{\mathbf{Z}}_k = \mathbf{Z}_k \mathbf{G} \mathbf{G}'$  is the estimation equation for reconstructing the data from the model. As the starting point was  $\hat{\mathbf{Z}}_k = \mathbf{F}_k \mathbf{G}'$ , it follows that  $\mathbf{F}_k = \mathbf{Z}_k \mathbf{G}$ , which suggests that the orthonormal  $\mathbf{G}$  may also be interpreted as an orthonormal weight

matrix to derive components by forming linear combinations with the  $Z_k$ . This in turn indicates that the  $C_k$  are the covariance matrices of component scores  $F_k$ .

As Kroonenberg & de Leeuw (1980; Kroonenberg, 1983a, ch. 4) show,  $G$  is the eigenvector matrix of  $Q = \sum_k R_k G G' R_k$ , or  $G\Gamma = QG$  with  $\Gamma$  the diagonal matrix with the decreasing eigenvalues of  $Q$  on its diagonal. This can be rewritten by premultiplying both sides with  $GG'$  and inserting  $G'G (=I)$  twice as

$$\begin{aligned} G\Gamma &= GG'QG = \sum_k GG'R_k G(G'G)G'R_k G(G'G) \\ &= (\sum_k \{GG'R_k GG'\}^2)G = (\sum_k \{\hat{Z}'_k \hat{Z}_k\}^2)G = (\sum_k \hat{R}_k^2)G. \end{aligned} \quad (16)$$

Thus  $G$  is the *eigenvector matrix of the average squared 'reconstructed covariance matrix'*. From (16) it also follows using the expression (14) for the core slices that

$$\Gamma = \sum_k G'R_k GG'R_k G = \sum_k C_k^2, \quad (17)$$

i.e. the eigenvalues associated with  $G$  can be found from the sum of the squared core slices.

After the orthonormal eigenvectors  $G$  and the  $C_k$  have been found, the eigenvector matrix has to be transformed to an appropriate pattern matrix assuming the model (8) underlies the data. The procedure followed here is analogous to classical PCA, that is first the eigenvectors are scaled to obtain 'structure-like' coefficients, and subsequently they are non-singularly (or orthonormally) transformed to some meaningful orientation of the axes, while the inverse transformations are applied to the  $C_k$ . The technical details of the procedure are discussed in the second part of the paper.

#### 4. Perfect congruence analysis for weights

A different exploratory approach to comparing sets of covariance matrices is perfect congruence analysis for weights—PCW—(ten Berge, 1986b). PCW is essentially a method for cross-validating component weights. The method can also be interpreted as a generalization of the multiple group method of factor analysis, see for instance, Nunnally (1978, pp. 398–400), or Gorsuch (1983, pp. 81–89). A parallel procedure may be derived for component loadings (PCL); however, ten Berge (1986b) has shown PCW to be superior.

The procedure is based on the fact that components are linear combinations of variables, and are defined by the weights in those linear combinations. From a first study the weights are derived, and these weights are then used in one or more other studies to determine the values of the subjects on the components in the other study, parallel to the cross-validation procedure in ordinary regression.

As every component is uniquely defined by its component weights the weights can be taken to define the interpretation of the component, as has been argued by Harris (1984, pp. 317–320). It follows that any component from a first (previous) study can be recovered in another (new) study where the same variables have been used. Component weights from the first study can simply be applied to the variables in the other one to define new components with the same interpretation.

More formally, PCW can be explained as follows. Let  $R_k$  denote a  $m \times m$

covariance matrix of  $m$  variables in the  $k$ th sample ( $k=1, \dots, n$ ), and let  $\mathbf{W}_0$  be a  $m \times s$  matrix of component weights or component-score coefficients, defining  $s$  components of the same variables in a reference population. The  $m \times s$  matrix  $\mathbf{W}_k$  of component weights, defining the same components in the  $k$ th sample, can be obtained as

$$\mathbf{W}_k = \mathbf{W}_0(\text{Diag } \mathbf{W}'_0 \mathbf{R}_k \mathbf{W}_0)^{-1/2}. \quad (18)$$

This normalization of the columns of  $\mathbf{W}_0$  guarantees that  $\mathbf{W}_k$  defines standardized components in the  $k$ th sample. The normalization does not affect the 'behaviour' of these components, but merely serves to simplify the presentation. It should be noted that the weight matrices  $\mathbf{W}_0$  and  $\mathbf{W}_k$  are perfectly proportional or congruent columnwise, as measured by Burt's (1948) and Tucker's (1951) congruence coefficient. This is the reason why the analysis has been christened 'perfect congruence analysis for weights'.

From  $\mathbf{R}_k$  and  $\mathbf{W}_k$  it is easy to compute the  $m \times s$  structure (or loading) matrix  $\mathbf{L}_k$ , as  $\mathbf{L}_k = \mathbf{R}_k \mathbf{W}_k$ , and the  $s \times s$  correlation matrix of the components  $\mathbf{F}_k = \mathbf{Z}_k \mathbf{W}_k$

$$\mathbf{\Theta}_k = \mathbf{F}'_k \mathbf{F}_k = \mathbf{W}'_k \mathbf{Z}'_k \mathbf{Z}_k \mathbf{W}_k = \mathbf{W}'_k \mathbf{R}_k \mathbf{W}_k. \quad (19)$$

The  $j$ th column sum of squares of  $\mathbf{L}_k$  conveys how much variance is explained in the  $k$ th sample by a single component which has the same interpretation as component  $j$  from the reference population. Thus the variances are the diagonal entries in

$$\text{Diag}(\mathbf{L}'_k \mathbf{L}_k) = \text{Diag}(\mathbf{W}'_k \mathbf{R}_k^2 \mathbf{W}_k). \quad (20)$$

The above formula implies that the explained variances are determined for each column  $j$  separately. Generally components in the samples will be correlated, even though they are not correlated in the reference population. Therefore, the sum of the explained variances,  $\text{tr}(\mathbf{L}'_k \mathbf{L}_k)$ , is a meaningless quantity, unless the components are orthogonal in the  $k$ th sample, that is unless  $\mathbf{\Theta}_k$  is  $\mathbf{I}$ . Instead, the amount of variance explained by the  $s$  components jointly must be corrected for the correlations between the components in the  $k$ th sample, and they can be computed as

$$\text{tr}(\mathbf{L}'_k \mathbf{L}_k \mathbf{\Theta}_k^{-1}). \quad (21)$$

In addition, the off-diagonal elements of  $\mathbf{\Theta}_k$  can be inspected to assess the correlations between the components in the other study.

In many applied studies  $\mathbf{W}_0$  is not known or has not been reported. Typically, only component loadings are reported. However, this need not be a problem. If the components from a reference population are truncated and/or rotated principal components with  $m \times r$  pattern matrix  $\mathbf{V}_0$ , then  $\mathbf{W}_0$  can explicitly be obtained as (see ten Berge, 1986a, p. 36, Lemma 1)

$$\mathbf{W}_0 = \mathbf{V}_0(\mathbf{V}'_0 \mathbf{V}_0)^{-1}. \quad (22)$$

### 5. Example

To illustrate the use of three-mode PCA and PCW for sets of covariance matrices, we will reanalyse data from Marsh & Hocevar (1985; see also Marsh, Barnes, Cairns & Tidman, 1984). These authors presented confirmatory simultaneous factor analyses of self-concept ratings on 28 subscales of the Self-Description Questionnaire (Marsh, Relich & Smith, 1983) by pre-adolescent children from Grades 2, 3, 4, and 5 in Sydney, Australia, using models with and without higher order factors. Their major aim was to examine the invariance of an *a priori* factor structure based on the aspects that were used to construct the questionnaire, both of the first- and higher-order factors. The (exploratory) aim of the present analysis is to investigate the dimensionality of the questionnaire, to what extent one common (compromise) structure is shared by all grades, and to what extent differences between grades with respect to this common structure exist. By using the procedures described above, it is assumed that a common identical pattern exists, and that its values need to be estimated. In a sense these assumptions resemble Marsh & Hocevar's Model 11 most. In those cases where *a priori* structure is not known or cannot be trusted the analysis could be a first step for a more confirmatory approach.

In Table 1 the three-component solution for the Tucker2 model (7) both in its principal component orientation, and after a direct oblimin transformation computed via BMDP4M (Dixon, 1981); for further details see below. Note, by the way, that Marsh & Hocevar's factor correlation matrix for Grade 5 (p. 570; the only one reproduced) also suggests an approximate three-dimensional solution (eigenvalues: 10.5, 4.1, 3.1, 1.8, 1.3, 0.7, and 0.5). The first principal component reflects the overall positive covariances between the 28 subscales, with somewhat lower values for the 'Relations with parents' and the 'Physical ability' variables. This suggests that in a simultaneous factor analysis at some (higher-) order a general factor might be hypothesized. The structure in the subscales is easily discernable from a plot of the second against the third component (Fig. 1). All subscales group according to the design of the questionnaire, and the figure suggests that Marsh & Hocevar's oblique seven-factor solution in which each factor has exclusively loadings for one aspect of the Self-Description Questionnaire is a reasonable proposition.

If there is interest in specifying second-order factors in a simultaneous factor analysis (as was the case with Marsh & Hocevar) the obliquely rotated solution (right-hand part of Table 1) shows that either two second-order factors could be hypothesized (i.e. 'Non-academic aspects' and 'Academic aspects'), or three second-order factors ('Non-academic aspects', 'Mathematics', and 'Reading'). This can also be seen from Fig. 1 by looking at the correlations of the subscales, to be deduced from their angles in the figure and the near equal coefficients on the first component.

In order not to overburden the table, percentages explained variances were omitted for the three-mode PCA solution. The results shown in Table 2 indicate that all grades have comparable overall explained variances, and that in Grades 2 and 3 the first principal component carries more weight than in Grades 4 and 5, and the reverse for the last two principal components ('Non-academic versus Academic', and 'Reading versus Mathematics (and Physical ability)').

Turning to the oblique components there hardly seem to be any *systematic*



**Table 1.** Unrotated and rotated component coefficients (TUCKALS2 solution)

Aspects	Subscale	Principal component			Direct oblimin		
		1	2	3	1	2	3
Physical ability	1	0.7	0.5	-0.5	<b>1.0</b>	0.2	-0.3
	2	0.8	0.5	-0.6	<b>1.0</b>	0.3	-0.4
	3	<b>1.0</b>	0.5	-0.7	<b>1.1</b>	0.5	-0.4
	4	<b>1.0</b>	0.5	-0.5	<b>1.1</b>	0.3	-0.3
Physical appearance	5	<b>1.3</b>	<b>1.0</b>	0.1	<b>1.6</b>	-0.3	0.4
	6	<b>1.4</b>	<b>1.1</b>	0.4	<b>1.7</b>	-0.5	0.7
	7	<b>1.5</b>	<b>1.3</b>	-0.0	<b>2.0</b>	-0.3	0.3
	8	<b>1.5</b>	<b>1.0</b>	0.0	<b>1.7</b>	-0.1	0.3
Relations with peers	9	0.8	0.6	-0.0	<b>1.1</b>	-0.1	0.1
	10	<b>1.2</b>	0.7	-0.2	<b>1.4</b>	0.1	0.0
	11	<b>1.3</b>	0.8	-0.0	<b>1.5</b>	-0.0	0.3
	12	<b>1.4</b>	0.8	-0.2	<b>1.5</b>	0.1	0.2
Relations with parents	13	0.5	0.1	0.2	0.3	0.1	0.3
	14	0.6	-0.1	0.3	0.2	0.2	0.5
	15	<b>1.0</b>	0.1	0.4	0.5	0.2	0.7
	16	0.7	0.1	0.3	0.4	0.1	0.5
Reading	17	0.9	-0.4	<b>1.1</b>	-0.1	0.1	<b>1.5</b>
	18	<b>1.0</b>	-0.3	<b>1.1</b>	0.1	0.0	<b>1.6</b>
	19	<b>1.3</b>	-0.3	<b>1.1</b>	0.2	0.2	<b>1.5</b>
	20	<b>1.2</b>	-0.5	<b>1.0</b>	-0.0	0.4	<b>1.5</b>
Mathematics	21	<b>1.5</b>	-0.9	-0.9	0.1	<b>2.0</b>	-0.2
	22	<b>1.5</b>	-0.9	-0.6	0.1	<b>1.8</b>	0.1
	23	<b>1.5</b>	-1.1	-0.7	-0.1	<b>2.0</b>	0.1
	24	<b>1.6</b>	-1.0	-0.8	0.1	<b>2.0</b>	-0.0
General school aspects	25	<b>1.2</b>	-0.5	0.2	0.1	0.9	0.7
	26	<b>1.3</b>	-0.4	-0.1	0.4	<b>1.0</b>	0.4
	27	<b>1.4</b>	-1.1	0.3	-0.3	<b>1.4</b>	0.9
	28	<b>1.4</b>	-0.8	0.2	-0.0	<b>1.2</b>	0.8
Variance explained		41.1	15.0	8.8	25.1	21.9	14.0

Note. For the transformation of the orthonormal components (23) was used with  $D = \text{Diag } \hat{C}$  (with values 41.2, 15.0, and 8.8, respectively), and  $K = I$  for the left-hand part, and the oblimin transformation for the right-hand part.

differences between the (explained) variances between grades related to age. As indicated above, the explained variances need not and do not add up to the total explained variances, because they are computed on a per component basis, and not from a partitioning of the total explained variance. There are obvious differences between explained variances between grades, such as lower explained variances of the Non-academic component in Grades 1 and 4, than in Grades 2 and 3, and a somewhat higher explained variance of 'Reading' in Grade 2 than in the other grades, but these are not related to age. Grades 2 and 3 generally show higher component correlations than Grades 4 and 5 with a notable exception of the correlation between the 'Non-academic' component and the 'Reading' one. This pattern concurs with analyses of Marsh & Hocevar (1985, p. 575), although we should mention that the conclusions depend very much on the particular rotation one examines as the

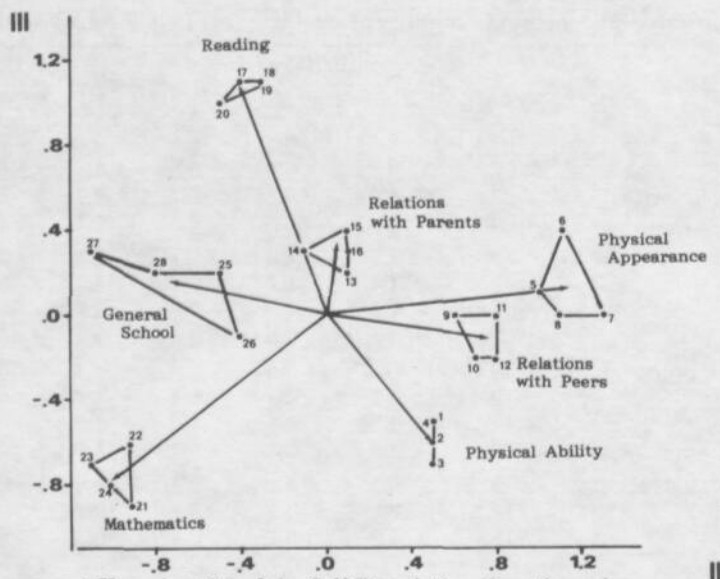


Figure 1. Component II against III of the Self-Description Questionnaire.

comparison of the correlations between the principal components and those between the oblique components shows (see Table 2, bottom part).

The results from our three-mode and perfect congruence analyses show that the three-dimensional configuration captures most of the information in the covariance matrices, and that before rotation in the higher grades other components gain more importance than in the lower grades, but that the situation is far more complex for the rotated components. In the oblique solution there is a clear tendency for lower component correlations in the higher grades, but this is not the case in the principal component solution. Taking the two representations together, they both point towards increasing component differentiation with age; in the principal component orientation this follows from the explained variances, and in the oblique orientation from the component correlations (see also Marsh & Hocevar, 1985, p. 576). A more detailed analysis of these data is possible using three-mode methods, for instance by considering the differences between the 'reconstructed' or fitted variances and the observed variances, but this would take us too far afield.

## 6. Transforming components from three-mode analyses

As pointed out in the previous discussion, and as becomes clear from perusal of especially Table 2, certain systematic patterns can be observed between the outcomes from PCW and three-mode PCA. In this second part of the paper, further details are provided on the way transformations can be used in three-mode PCA, and in particular what kind of choices can and have to be made, if the technique is to be used on sets of covariance matrices, while assuming that the extension of classical PCA as given in (8) is the model for the data. And, secondly, the similarities between three-mode PCA and PCW are theoretically explored.

**Table 2.** Explained variances and correlations of principal and oblique components from perfect congruence analysis (P) and three-mode PCA (T2)

Grade	Tot. var.	Explained variance								Percentage Explained variance PCW			
		1		2		3		Tot.		1	2	3	Tot
		P	(T2)	P	(T2)	P	(T2)	P	(T2)				
<i>Explained variances of principal components<sup>a</sup></i>													
2	137	42	(42)	18	(17)	9	(8)	64	(62)	0.31	0.13	0.07	0.47
3	123	49	(58)	12	(8)	7	(5)	68	(65)	0.40	0.09	0.06	0.55
4	113	40	(37)	21	(24)	11	(13)	67	(66)	0.35	0.18	0.10	0.59
5	122	38	(35)	18	(21)	12	(16)	68	(66)	0.31	0.15	0.10	0.56
Av.	124	42	(43)	17	(18)	10	(10)	67	(65)	0.34	0.14	0.08	0.54
<i>Explained variances of direct oblimin components<sup>a</sup></i>													
2	137	26	(21)	37	(39)	26	(26)	64	(62)	0.19	0.27	0.19	0.47
3	123	41	(42)	36	(35)	30	(26)	68	(65)	0.33	0.29	0.24	0.55
4	113	37	(47)	23	(21)	17	(14)	67	(66)	0.33	0.20	0.15	0.59
5	122	27	(24)	29	(31)	22	(27)	68	(66)	0.22	0.24	0.18	0.56
Av.	124	33	(34)	31	(31)	24	(23)	67	(65)	0.27	0.25	0.19	0.54
Grade	Principal components <sup>b</sup>			Rotated components <sup>b</sup>									
	(1,2)	(1,3)	(2,3)	(1,2)	(1,3)	(2,3)							
<i>Correlations between components (PCW and TUCKALS2)</i>													
2	0.26	0.02	-0.25	0.44	0.14	0.50							
3	0.05	-0.03	0.13	0.52	0.53	0.45							
4	-0.28	0.09	0.12	0.28	0.27	0.03							
5	0.09	-0.05	0.01	0.25	0.18	0.21							
Overall	0.00	0.00	0.00	0.37	0.28	0.31							

<sup>a</sup>The explained variance of the components in each of the grades was assessed with equations (20) and (33) for both the principal component and the oblique orientation of the axes, while the total explained variances were computed using (21) and (31), respectively.

<sup>b</sup>(*i, j*) is correlation between components *i* and *j*.

The basic solution of three-mode PCA as given in (14) consists of the orthonormal eigenvector matrix  $\mathbf{G}$  and the matrices with mutual weights  $\mathbf{C}_k$  ( $k=1, \dots, n$ ). Assuming (8) underlies the data the eigenvector matrix  $\mathbf{G}$  has to be transformed to an appropriate pattern matrix. The procedure followed here is analogous to classical PCA, that is, first the eigenvectors are scaled with a scaling matrix  $\mathbf{D}$ , to obtain 'structure-like' coefficients, and subsequently they are non-singularly (or orthonormally) transformed to some meaningful orientation of the axes, while the inverse transformations are applied to the  $\mathbf{C}_k$ .

Considering the eigenvector-eigenvalue equation of  $\mathbf{G}$  (16), one obvious choice for the scaling matrix  $\mathbf{D}$  for  $\mathbf{G}$  would be  $\mathbf{D}=\mathbf{\Gamma}$ , i.e. based on the diagonal of the sum of the squared core slices. One disadvantage of such a choice is, however, that numerically this  $\mathbf{\Gamma}$  is associated with the squared covariances, as  $\mathbf{G}$  is the eigenvector matrix for the average squared reconstructed covariance matrix. In classical PCA, however, the eigenvalues which go into  $\mathbf{D}$  are derived from the covariance matrices

themselves. For comparison with separate analyses on each of the covariance matrices, this choice of  $\mathbf{D}$  might, therefore, not be very fortuitous.

Another choice for scaling the coefficients is the diagonal of the average core slice  $\mathbf{D} = \text{Diag } \bar{\mathbf{C}} = \text{Diag } (n^{-1} \Sigma \mathbf{C}_k)$ . In this way the diagonal of  $\mathbf{D}$  is the diagonal of the average covariance matrix of the component scores. Such a choice is more closely related to the strategy to model the data as in classical PCA, see equation (8), than is the choice of  $\mathbf{\Gamma}$  from (17). The non-singular (or orthonormal) transformation matrix will be denoted by  $\mathbf{K}$ , and it will be assumed that any reasonable strategy to find  $\mathbf{K}$  in classical PCA may be applied in three-mode PCA as well.

Thus with either of the above choices the orthonormal coefficient matrix  $\mathbf{G}$  will be scaled and transformed as

$$\mathbf{P}^* = \mathbf{G} \mathbf{D}^{1/2} \mathbf{K}. \quad (23)$$

As mentioned above, within the model such transformations can only be made without loss of fit by applying the inverse transformations to the core slices, that is

$$\mathbf{C}_k^* = \mathbf{K}^{-1} \mathbf{D}^{-1/2} \mathbf{C}_k \mathbf{D}^{-1/2} (\mathbf{K}')^{-1}, \quad (24)$$

such that the basic model equation (7) remains unchanged. The  $\mathbf{C}_k^*$  are now the covariance matrices of the transformed component scores

$$\mathbf{F}_k^* = \mathbf{F}_k \mathbf{D}^{-1/2} (\mathbf{K}')^{-1}. \quad (25)$$

From (25) it follows that the correlations between the components are

$$\Phi_k^* = \mathbf{D}_k^*{}^{-1/2} \mathbf{C}_k^* \mathbf{D}_k^*{}^{-1/2} \quad (26)$$

with  $\mathbf{D}_k^* = \text{Diag } \mathbf{C}_k^*$ , that is the diagonal of  $\mathbf{C}_k^*$ . As is common in classical PCA one may scale the component scores per sample to unit length by adjusting the  $\mathbf{F}_k^*$  with  $\mathbf{D}_k^*$ , that is  $\tilde{\mathbf{F}}_k = \mathbf{F}_k^* \mathbf{D}_k^*{}^{-1/2}$ .

The change from orthonormal eigenvectors  $\mathbf{G}$  to scaled and transformed patterns  $\mathbf{P}^*$  has an interesting consequence for the equivalent of (14), i.e. the expression of the  $\mathbf{C}_k$  in terms of  $\mathbf{G}$  and  $\mathbf{R}_k$ ; substituting (14) into (24) leads to

$$\begin{aligned} \mathbf{C}_k^* &= \mathbf{K}^{-1} \mathbf{D}^{-1/2} \mathbf{C}_k \mathbf{D}^{-1/2} (\mathbf{K}')^{-1} \\ &= \mathbf{K}^{-1} \mathbf{D}^{-1/2} \mathbf{G}' \mathbf{R}_k \mathbf{G} \mathbf{D}^{-1/2} (\mathbf{K}')^{-1} \cong \mathbf{P}^*{}' \mathbf{R}_k \mathbf{P}^*. \end{aligned} \quad (27)$$

Equation (27) implies that the transformation of  $\mathbf{G}$  to  $\mathbf{P}^*$  and the associated transformation of  $\mathbf{C}_k$  to  $\mathbf{C}_k^*$  affect the decomposition of the core slices. The simple decomposition of  $\mathbf{C}_k$  into  $\mathbf{G}$  and  $\mathbf{R}_k$  does *not* carry over to a similar expression for the transformed quantities (except when  $\mathbf{D}$  is the identity matrix and  $\mathbf{K}$  is orthonormal). In other words, (14) has no direct equivalent for the transformed covariances. However, with  $\mathbf{B}^* \equiv \mathbf{G} \mathbf{D}^{-1/2} (\mathbf{K}')^{-1}$ , (27) may be written [using (25)]



$$\mathbf{F}_k^* \mathbf{F}_k^{*\prime} = \mathbf{C}_k^* = \mathbf{B}^{*\prime} \mathbf{R}_k \mathbf{B}^* = \mathbf{B}^{*\prime} \mathbf{Z}'_k \mathbf{Z}_k \mathbf{B}^*, \quad (28)$$

thus allowing the interpretation of  $\mathbf{B}^*$  as weights to derive the transformed components from the original data. With the definitions of  $\mathbf{P}^*$  and  $\mathbf{B}^*$  (15) becomes

$$\mathbf{R}_k \approx \hat{\mathbf{R}}_k = \mathbf{P}^* \mathbf{B}^{*\prime} \mathbf{R}_k \mathbf{B}^* \mathbf{P}^{*\prime}, \quad (29)$$

and the reconstruction of the data with the model, that is  $\hat{\mathbf{Z}}_k = \mathbf{Z}_k \mathbf{B}^* \mathbf{P}^{*\prime}$ , parallels the equation in classical PCA,  $\hat{\mathbf{Z}} = \mathbf{Z} \mathbf{B} \mathbf{P}'$ .

Summing up, the results from a three-mode PCA on a set of covariance matrices may be interpreted in terms of classical PCA assuming that the data are modelled by (8), the generalization of classical PCA. This gives as an interpretation that  $\mathbf{P}^*$  is the common pattern matrix for all samples,  $\mathbf{B}^*$  the corresponding weight matrix, the  $\mathbf{F}_k^*$  are the component scores of each sample  $k$  (which can only be computed if the raw data, and not only their covariances are available), the  $\mathbf{C}_k^*$  are the covariances, and  $\Phi_k^*$  the correlations of the component scores for each sample.

### 6.1. Explained variance

Given the solution via the Tucker2 model applied to a set of covariance matrices, the *covariances* between the components in each sample are known, but still lacking is the total amount of *variance explained* by the model in each sample, and the *variance explained* by each component in each sample.

The total variance of a sample follows from  $\text{tr}(\mathbf{Z}'_k \mathbf{Z}_k)$ , and the explained variance from  $\text{tr}(\hat{\mathbf{Z}}'_k \hat{\mathbf{Z}}_k)$ . The latter expression becomes, using the reconstruction formula for  $\hat{\mathbf{Z}}_k$  (29)

$$\begin{aligned} \text{tr}(\hat{\mathbf{Z}}'_k \hat{\mathbf{Z}}_k) &= \text{tr}(\mathbf{P}^* \mathbf{B}^{*\prime} \mathbf{Z}'_k \mathbf{Z}_k \mathbf{B}^* \mathbf{P}^{*\prime}) \\ &= \text{tr}(\mathbf{G} \mathbf{G}' \mathbf{R}_k \mathbf{G} \mathbf{G}') = \text{tr} \mathbf{C}_k. \end{aligned} \quad (30)$$

Thus the total explained variance is independent of the transformations applied to  $\mathbf{G}$ . This is, of course, not surprising as the patterns, weights, and component covariances were defined in such a way to leave the overall solution unchanged. The percentage explained variance is

$$100 \times (\text{tr} \mathbf{C}_k) / (\text{tr} \mathbf{R}_k). \quad (31)$$

The per component explained variance may be derived in a way analogous to that in classical PCA from the column sums of squares of the structure matrices  $\mathbf{S}_k$ , with

$$\mathbf{S}_k \equiv \mathbf{Z}'_k \mathbf{F}_k = \mathbf{Z}'_k \mathbf{Z}_k \mathbf{B}^* = \mathbf{R}_k \mathbf{B}^*. \quad (32)$$

Thus the explained variances per component are the diagonal entries in

$$\mathbf{\Gamma}_k = \text{Diag}(\mathbf{S}'_k \mathbf{S}_k) = \text{Diag}(\mathbf{B}^{*\prime} \mathbf{R}_k^2 \mathbf{B}^*), \quad (33)$$

and the percentages explained variance are the diagonal elements of

$$100 \times \Gamma_k(\text{Diag } \mathbf{R}_k)^{-1}. \quad (34)$$

The unsatisfying aspect of the per component explained variances as computed in (33) is that they are not represented directly in the model, in contrast with the other matrices and quantities discussed so far. On the other hand this is also the case in classical PCA.

### 6.2. Comparisons with separate PCA solutions

Given a simultaneous solution for the components and the core slices, it is often desirable to compare the joint solution for the  $\mathbf{R}_k$  with the separate principal component solutions for each of the  $\mathbf{R}_k$ . The latter have the form

$$\mathbf{R}_k \approx \mathbf{G}_k \mathbf{\Omega}_k \mathbf{G}_k' \quad (k = 1, \dots, n), \quad (35)$$

where the right-hand side is the truncated eigenvector-eigenvalue decomposition of  $\mathbf{R}_k$ . Again it will be assumed that only  $s \leq m$  components have been derived. The diagonal elements of  $\mathbf{\Omega}_k$  provide the variances accounted for by each component in  $\mathbf{G}_k$ . The most important aspect for comparison is the total amount of variance which is necessary to compare how well the simultaneous solution has succeeded in explaining the variance in the  $k$ th group (with the same number of components) compared to that of the separate analysis. In other words  $\text{tr}(\mathbf{G}_k \mathbf{\Omega}_k \mathbf{G}_k')$  is to be compared with  $\text{tr}(\mathbf{G} \mathbf{C}_k \mathbf{G}') = \text{tr } \mathbf{C}_k$ . Large discrepancies indicate that the joint components have not much in common with the first principal components of the separate solution in terms of variance accounted for. Comparing per component explained variance is less meaningful as in general  $\mathbf{G}_k$  will not be equal to  $\mathbf{G}$ , even after (non-singular) transformations.

Further information on the quality of the Tucker2 solution in each group can be found by inspecting the elements of  $\mathbf{E}_k$  in equation (7), which indicate the residual covariances, that is the differences between the observed covariances and those estimated by the model.

## 7. Relationships

Under the condition that the weight matrix from the three-mode PCA is taken as a starting point for PCW, several comparisons may be made to show the similarities and differences between the techniques. This choice of weight matrix is only one of the possible starting points for PCW. For instance, the weight matrix derived from the average covariance matrix and the weight matrix from an entirely different study would be other alternatives. In the latter two cases, no close formal correspondence to three-mode PCA is to be expected.

As  $\mathbf{B}^*$  is the component weight matrix from three-mode PCA, it will be taken as starting point for the comparison, that is  $\mathbf{W}_0 = \mathbf{B}^*$ , i.e. the common component space

is taken as the reference solution. Then, for the  $k$ th sample the weight matrix associated with the reference components,  $\mathbf{W}_k$ , becomes using equation (18)

$$\begin{aligned}\mathbf{W}_k &= \mathbf{B}^*(\text{Diag } \mathbf{B}^* \mathbf{R}_k \mathbf{B}^*)^{-1/2} \\ &= \mathbf{B}^*(\text{Diag } \mathbf{C}_k^*)^{-1/2} = \mathbf{B}^* \mathbf{D}_k^{*-1/2}.\end{aligned}\quad (36)$$

### 7.1. Component correlations

In PCW the correlations between the transformed components using the weights  $\mathbf{B}^*$  from the three-mode PCA in the  $k$ th sample were found to be  $\Theta_k = \mathbf{W}_k' \mathbf{R}_k \mathbf{W}_k$  [see equation (19)]. Substituting  $\mathbf{W}_k$  from (36), using the expressions derived in (28) for  $\mathbf{C}_k^*$ , and those for  $\Phi_k^*$  in (26) respectively, leads to

$$\begin{aligned}\Theta_k &= \mathbf{W}_k' \mathbf{R}_k \mathbf{W}_k = \mathbf{D}_k^{*-1/2} \mathbf{B}^* \mathbf{R}_k \mathbf{B}^* \mathbf{D}_k^{*-1/2} \\ &= \mathbf{D}_k^{*-1/2} \mathbf{C}_k^* \mathbf{D}_k^{*-1/2} = \Phi_k^*.\end{aligned}\quad (37)$$

Thus, it may be concluded that the transformations applied within the three-mode model to transform the orthonormal  $\mathbf{G}$  into a pattern-like matrix  $\mathbf{P}^*$  with corresponding weights  $\mathbf{B}^*$ , have the effect of producing the same component correlations as those found in PCW provided the weights  $\mathbf{B}^*$  from three-mode PCA are used as reference solution in PCW.

### 7.2. Explained variances per component

The explained variances of the separate components within PCW for the  $k$ th sample were given in (20) as the diagonal elements of  $\mathbf{L}_k' \mathbf{L}_k$ . Substituting in (20) the results for  $\mathbf{W}_k$  from (36), and comparing the subsequent expression with that for the explained variance in three-mode PCA leads to

$$\begin{aligned}\text{Diag}(\mathbf{L}_k' \mathbf{L}_k) &= \text{Diag}(\mathbf{W}_k' \mathbf{R}_k^2 \mathbf{W}_k) \\ &= \text{Diag}(\mathbf{D}_k^{*-1/2} \mathbf{B}^* \mathbf{R}_k^2 \mathbf{B}^* \mathbf{D}_k^{*-1/2}) = \mathbf{D}_k^{*-1} \Gamma_k.\end{aligned}\quad (38)$$

This shows that the explained variances in the TUCKALS2 solution are closely related to but different from those in PCW. In particular, PCW makes use of more sample information via the  $\mathbf{D}_k^{*-1}$ , than three-mode PCA.

## 7.3. Total explained variance

Finally, the total explained variance for the  $k$ th sample is in PCW

$$\begin{aligned} \text{tr}(\mathbf{L}'_k \mathbf{L}_k \mathbf{\Theta}_k^{-1}) &= \text{tr}(\mathbf{W}'_k \mathbf{R}_k^2 \mathbf{W}_k (\mathbf{W}'_k \mathbf{R}_k \mathbf{W}_k)^{-1}) \\ &= \text{tr}(\mathbf{G}' \mathbf{R}_k^2 \mathbf{G} (\mathbf{G}' \mathbf{R}_k \mathbf{G})^{-1}) \geq \text{tr}(\mathbf{G}' \mathbf{R}_k \mathbf{G} (\mathbf{G}' \mathbf{G})^{-1}) \\ &= \text{tr}(\mathbf{G}' \mathbf{R}_k \mathbf{G}) = \text{tr} \mathbf{C}_k, \end{aligned} \quad (39)$$

using Corollary 1 from ten Berge (1986b, p. 56) for the inequality, equation (36) for the  $\mathbf{W}_k$ , the definition of  $\mathbf{B}^*$  and the fundamental equation (14) for the  $\mathbf{C}_k$ . It should be noted that (39) holds for every orthonormal  $\mathbf{G}$ . The conclusion from (39) is that the total explained variance in PCW is always greater than or equal to that of three-mode PCA. This can be explained as follows. In PCW, the variance explained is maximized in each sample by choosing  $\mathbf{P}_k$  such that  $\|\mathbf{Z}_k - \mathbf{F}_k \mathbf{P}'_k\|^2$  is minimal, for fixed  $\mathbf{F}_k = \mathbf{Z}_k \mathbf{G}$ . In TUCKALS2, on the other hand, the residual variance is merely  $\|\mathbf{Z}_k - \mathbf{F}_k \mathbf{G}'\|^2$ , cf. (8). Replacing the optimal  $\mathbf{P}_k$  matrices by the constant  $\mathbf{G}$  entails higher residual variances. Thus follows that the explained variance in PCW will exceed that of TUCKALS2.

The equality sign in equation (39) holds if and only if the columns of  $\mathbf{G}$  are equal to the eigenvectors of  $\mathbf{R}_k$ , or a rotation thereof. This means that the total explained variance in PCW for the  $k$ th sample is an upper bound for the total explained variance derived by three-mode PCA, if it is defined by (30). If the equality sign holds in (39), the explained variances per component of the two techniques become equal as well. Of course, with  $\mathbf{G}$  as the (truncated) eigenvector matrix for each  $\mathbf{R}_k$  simultaneously, the problem is no longer really three mode. For perfect equality of interpretation of the TUCKALS2 results in terms of PCW, one could use (21) and (38), rather than (30) and (33), however, the latter are more 'natural' choices with the Tucker2 framework.

## 8. Discussion and conclusion

Three-mode principal component analysis and perfect congruence analysis for weights can be fruitfully applied to the study of sets of covariance matrices. The results of three-mode PCA can be interpreted in terms of classical PCA, and can be made to resemble PCW by postulating a component model with an identical pattern matrix for each subset of observations. By computing the explained variances and the component correlations the relevance of the common compromise component space derived by three-mode PCA can be assessed for each sample by either technique, be it that the explained variances of the components and their overall explained variance differ if one wants to use 'natural' definitions in the three-mode PCA case. Furthermore, for each sample, the overall explained variance of PCW is an upper bound for that in three-mode PCA. In several practical examples, the structures in the  $n$  samples were very similar. This implies near fulfilment of the equality condition of (39). Moreover, in those cases the component solution from three-mode



PCA and the component solution from the average covariance matrix (which was routinely used in PCW) are virtually identical as well. Therefore very similar results for both techniques were observed, even though they started from rather different positions and philosophies. This seemingly curious concurrence of results, was what led to the above investigation. The results presented in this paper are an extension of previous work by Kroonenberg & ten Berge (1985, 1987), however, the similarities between the two techniques were not yet fully understood and developed in the earlier presentations. Evaluating the algebraic developments, it seems that in most cases three-mode PCA is most useful in deriving a common solution, while PCW is more effective and simpler in evaluating separate samples given a reference solution.

In the example presented, the initially surprising similarity of PCW and three-mode PCA was demonstrated. Furthermore, the power of the combination of the two techniques was elucidated as well. Most of the features necessary to describe the structure of the Self-Description Questionnaire could be found in a low (three-) dimensional component solution, in contrast with the seven-factor solution employed by Marsh & Hocevar. However, it is clear that confirming a number of hypothesized oblique factors plus several higher-order factors without regard to the dimensionality of the solution, is an entirely different aim for analysis than describing the structure of a data set in a low-dimensional space.

#### Acknowledgements

This paper is a revision of a paper presented at the 4th European Meeting of the Psychometric Society, Cambridge, UK, 2-5 July 1985. Thanks are due to Henk Kiers for his comments on an earlier version of this manuscript.

#### References

- Burt, C. (1948). The factorial study of temperament traits. *British Journal of Psychology, Statistical Section*, **1**, 178-203.
- Dixon, W. J. (Ed) (1981). *BMDP Statistical Software 1981*. Berkeley, CA: University of California Press.
- Gorsuch, R. L. (1983). *Factor Analysis*, 2nd ed. Hillsdale, NJ: Erlbaum.
- Harris, R. J. (1984). *A Primer of Multivariate Statistics*, 2nd ed. New York: Academic Press.
- Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an 'explanatory' multi-mode factor analysis. *UCLA Working Papers in Phonetics*, **16**, 1-84. (Reprinted by Xerox University Microfilms, Ann Arbor, MI: Order no. 10,085.)
- Harshman, R. A. & Lundy, M. E. (1984). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder Jr., J. A. Hattie & R. P. McDonald (Eds), *Research Methods in Multimode Data Analysis*. New York: Praeger.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, **36**, 409-426.
- Jöreskog, K. G. & Sörbom, D. (1983). *LISREL V: Analysis of Linear Structural Relationships by the Method of Maximum Likelihood*. Chicago, IL: International Educational Services.
- Keller, J. B. (1962). Factorization of matrices by least-squares. *Biometrika*, **49**, 239-242.
- Kroonenberg, P. M. (1983a). *Three-mode Principal Component Analysis: Theory and Applications*. Leiden, The Netherlands: DSWO Press.
- Kroonenberg, P. M. (1983b). Correlational structure of the subtests of the Snijders-Oomen Non-verbal Intelligence Scale. *Kwantitatieve Methoden*, **4**(11), 40-51.
- Kroonenberg, P. M. & Brouwer, P. (1985). User's Guide to TUCKALS2, Version 4.0. (WEP Reeks, WR 85-12-RP). University of Leiden, Department of Education.

- 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. & ten Berge, J. M. F. (1985). Cross-validation of components from several correlation matrices with perfect congruence analysis and three-mode principal component analysis. Paper presented at the 4th European Meeting of the Psychometric Society, Cambridge, UK, July.
- Kroonenberg, P. M. & ten Berge, J. M. F. (1987). Cross-validation of the WISC-R factorial structure using three-mode principal component analysis and perfect congruence analysis. *Applied Psychological Measurement*, **11**, 195-210.
- Lingoes, J. C. & Borg, I. (1978). A direct approach to individual differences scaling using increasingly complex transformations. *Psychometrika*, **43**, 491-520.
- Marsh, H. W., Barnes, J., Cairns, L. & Tidman, M. (1984). The Self Description Questionnaire (SDQ): Age effects in the structure and level of self-concept for preadolescent children. *Journal of Educational Psychology*, **76**, 940-956.
- Marsh, H. W. & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First- and higher order factor models and their invariance across groups. *Psychological Bulletin*, **97**, 562-582.
- Marsh, H. W., Relich, J. D. & Smith, I. D. (1983). Self-concept: The construct validity of interpretations based on the SDQ. *Journal of Personality and Social Psychology*, **45**, 173-187.
- McDonald, R. P. (1984). The invariant factors model. In H. G. Law, C. W. Snyder Jr., J. A. Hattie & R. P. McDonald (Eds), *Research Methods for Multimode Data Analysis*. New York: Praeger.
- Meredith, W. & Tisak, J. (1982). Canonical analysis of longitudinal and repeated measures data with stationary weights. *Psychometrika*, **47**, 47-67.
- Nunnally, J. C. (1978). *Psychometric Theory*. New York: McGraw-Hill.
- Shavelson, R. J. & Bolus, R. (1982). Self-concept: The interplay of theory and methods. *Journal of Educational Psychology*, **74**, 3-17.
- Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Validation of construct interpretations. *Review of Educational Research*, **46**, 407-441.
- ten Berge, J. M. F. (1986a). Some relationships between descriptive comparisons of components from different studies. *Multivariate Behavioral Research*, **21**, 29-40.
- ten Berge, J. M. F. (1986b). Rotation to perfect congruence and the cross-validation of component weights across populations. *Multivariate Behavioral Research*, **21**, 41-64; 262-266.
- Tisak, J. & Meredith, W. (in press). Exploratory longitudinal factor analysis in multiple populations. *Psychometrika*.
- Tucker, L. R. (1951). A method for synthesis of factor analysis studies. Personnel Research Section Reports No. 984. Washington DC: Department of the Army.
- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika*, **31**, 279-311.
- Tucker, L. R. (1972). Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, **37**, 3-27.
- van Schuur, H. (1984). Structure in political beliefs. A new model for stochastic unfolding with applications to European party activists. Doctoral dissertation, University of Groningen.

Received 20 July 1988; revised version received 26 October 1988