CHAPTER 1

GENERALIZED COMPONENTS ANALYSIS

1.1 INTRODUCTION

In research in the social sciences, it is often the case that a number of variables (for instance a questionnaire) is measured on two or more different points in time, or at the same time in two or more groups of individuals (differing in some respect). For instance, one may have scores of children at six, eight and ten years of age on a set of intellectual tasks, or one may have scores of comparable groups of individuals from Zimbabwe and Sweden on a personality inventory. One of the first questions that comes to mind when analyzing data of these kinds is whether on the different time points or in the different groups the same factors are found. Are the same intellectual skills present at ages six, eight and ten? Are the same personality traits present in both Zimbabwe and Sweden?

An obvious way to answer these questions is to analyze the groups separately with a factor analysis technique followed by a Varimax rotation and compare the solutions found, using some similarity measure. Factor analysis techniques, such as Common Factor Analysis (CFA) and Principal Components Analysis (PCA) are useful methods to summarize the information, contained in the correlations between a large number of variables and in the scores on a large number of variables, respectively, in an efficient way. Because, in the present study, several generalizations from the method PCA are used, in Section 1.2, the method PCA will be presented.

Performing a Varimax rotation (Kaiser, 1958) on the pattern matrices, found with PCA on each group separately, and comparing the solutions (for instance using a congruence measure) has been the favorite approach of researchers in situations with two or more groups for many

years. The usefulness of this approach will be discussed in Section 1.3.

Another way of comparing the solutions, found with separate PCA's, is to rotate the two (or more) matrices to optimal agreement, using a Generalized Procrustes Rotation (GPR), see for instance Ten Berge (1977; 1986a; 1988) and Ten Berge and Knol (1984), and then applying a congruence measure. With a GPR, it can be investigated to what extent the same factors can be found in two or more groups. In Section 1.4, the GPR technique will be discussed. For both rotation techniques (Varimax and GPR), advantages and disadvantages will be summed up.

As will become clear, the method of using rotation techniques in the search for common factors suffers from a certain shortcoming. However, for the analysis of scores of two or more groups on the same variables, there are methods available that analyze the groups simultaneously, making the above approaches obsolete. A selection of these methods, used in the present study, will briefly be presented in Section 1.5 (for a full description, see Chapter 2), and in Section 1.6 the goal of the present study will be described.

In Section 1.7, reasons for ignoring some of the methods available for simultaneous analysis of two or more groups will be given, and in Section 1.8, one of the methods originally intended for inclusion in the present study will be discarded. In Section 1.9, the comparison of common factor analysis and component analysis for one group will be discussed. Finally, the approach taken in the present study will be presented in Section 1.10.

1.2 PRINCIPAL COMPONENTS ANALYSIS

When a researcher has scores of a group of n individuals on a set of m variables, Principal Components Analysis (PCA) is a popular technique for an exploratory analysis. In PCA, the variables are linearly combined (weighted sums of variables) in such a way that the amount of variance explained by these combinations is maximal. The combinations of variables are called components. In the present study, the term component will be

used for the methods making linear combinations of the variables, and the term factor will be used for the methods making use of hypothetical latent variables.

In the explanation to follow, standardized scores are used. Let the standard scores of n individuals on m variables be collected in an $n \times m$ matrix $\mathbf{Z}$ and the weights for combinations of the m variables to form q components in an $m \times q$ weights matrix $\mathbf{W}$. The component scores matrix $\mathbf{F}$ with the scores of the n individuals on the q components can now be written as $\mathbf{F}=\mathbf{ZW}$. PCA now minimizes the function

$$f(\mathbf{W},\mathbf{P}) = \| \mathbf{Z} - \mathbf{ZWP}' \|^2 = \| \mathbf{Z} - \mathbf{FP}' \|^2, \tag{1.1}$$

where the matrix $\mathbf{P}'$ provides the optimal regression weights for estimating matrix $\mathbf{Z}$ from $\mathbf{ZW}$.

The correlations between the variables are collected in the correlation matrix $\mathbf{R}=n^{-1}\mathbf{Z}'\mathbf{Z}$. Without loss of generality, the matrix $\mathbf{F}$ can be constrained to have standardized and uncorrelated columns. Subject to this constraint, the maximum of function f is found for $\mathbf{W}=\mathbf{K}_q\mathbf{\Lambda}_q^{-\frac{1}{2}}\mathbf{T}$ (cf. Ten Berge, 1993, p. 44), where the matrix $\mathbf{K}_q$ contains the first q columns of matrix $\mathbf{K}$, defined by the eigendecomposition $\mathbf{R}=\mathbf{K\Lambda K}'$, the matrix $\mathbf{\Lambda}_q$ is the upper left submatrix ($q \times q$) of matrix $\mathbf{\Lambda}$, and $\mathbf{T}$ is an arbitrary rotation matrix. The correlations between the variables and the components are collected in the $m \times q$ matrix $\mathbf{S}$, computed as $\mathbf{S}=n^{-1}\mathbf{Z}'\mathbf{F}=n^{-1}\mathbf{Z}'\mathbf{ZW}=\mathbf{RW}$; the correlations between the components are collected in $\mathbf{\Phi}$ ($q \times q$) and calculated as $\mathbf{\Phi}=n^{-1}\mathbf{F}'\mathbf{F}=n^{-1}\mathbf{W}'\mathbf{Z}'\mathbf{ZW}=\mathbf{W}'\mathbf{RW}=\mathbf{W}'\mathbf{S}=\mathbf{S}'\mathbf{W}$. The matrix $\mathbf{S}$ is called the structure matrix. Using the solution for $\mathbf{W}$ and the eigendecomposition of $\mathbf{R}$, any rotated PCA solution for the matrices $\mathbf{S}$ and $\mathbf{\Phi}$ can be written as $\mathbf{S}=\mathbf{K}_q\mathbf{\Lambda}_q^{\frac{1}{2}}\mathbf{T}$ and $\mathbf{\Phi}=\mathbf{T}'\mathbf{T}$ for some matrix $\mathbf{T}$ with $\mathrm{diag}(\mathbf{T}'\mathbf{T})=\mathbf{I}$.

The optimal solution for $\mathbf{P}$ is given as $\mathbf{P}=\mathbf{Z}'\mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}=\mathbf{Z}'\mathbf{F}(n\mathbf{\Phi})^{-1}$ $=n^{-1}\mathbf{Z}'\mathbf{F}\mathbf{\Phi}^{-1}=\mathbf{S}\mathbf{\Phi}^{-1}$. The matrix $\mathbf{P}$ is called the pattern matrix. For any rotated PCA solution, the pattern matrix can be written as $\mathbf{P}=\mathbf{K}_q\mathbf{\Lambda}_q^{\frac{1}{2}}(\mathbf{T}')^{-1}$. From these descriptions, it can be easily seen that, for orthogonal components, the matrices $\mathbf{S}$ and $\mathbf{P}$ are equal. In this situation, this matrix is called the loadings matrix and denoted by $\mathbf{A}$, so $\mathbf{P}=\mathbf{S}=\mathbf{A}$.

From the component scores matrix $\mathbf{F}$ and the pattern matrix $\mathbf{P}$, found by PCA, the original variable scores matrix $\mathbf{Z}$ and correlation matrix $\mathbf{R}$

can be estimated as $\hat{\mathbf{Z}}=\mathbf{FP}'$ and $\hat{\mathbf{R}}=\mathbf{P\Phi P}'$, respectively. The estimate $\hat{\mathbf{Z}}$ is the best estimate of $\mathbf{Z}$, according to the least squares criterion. The fit of the solution can now be described as the amount of variance, explained by the components, found with PCA. This can be obtained as the sum of the variances of the columns of $\hat{\mathbf{Z}}$. Using the above expressions, the explained variance (EV) can be written as a function of matrices $\mathbf{P}$ and $\mathbf{\Phi}$, $\mathbf{S}$ and $\mathbf{\Phi}$ or $\mathbf{W}$ and $\mathbf{R}$, respectively, as

$$\text{EV} = n^{-1}\text{tr}(\hat{\mathbf{Z}}'\hat{\mathbf{Z}}) = \text{tr}(\mathbf{P\Phi P}') = \text{tr}(\mathbf{S\Phi}^{-1}\mathbf{S}') = \text{tr}(\mathbf{W}'\mathbf{R}^2\mathbf{W}(\mathbf{W}'\mathbf{RW})^{-1}). \qquad (1.2)$$

Thus, when analyzing scores of a single group of individuals on a number of variables, PCA produces three matrices, each describing a relation between the scores on the variables and the scores on the components. These are: The components weights matrix $\mathbf{W}$ by which the components are defined as weighted sums of the variables; the structure matrix $\mathbf{S}$ with correlations between variables and components; and the pattern matrix $\mathbf{P}$ with regression weights for the reconstruction of the variables from the components. Each of the matrices $\mathbf{W}$, $\mathbf{S}$ and $\mathbf{P}$ can be used for interpretation of the components, although using the pattern matrix $\mathbf{P}$ is somewhat inconvenient. Because the pattern matrix conveys how the variables depend on the components rather than the other way around, it can only be used indirectly for interpretation of the components (cf. Brogden, 1969; Gorsuch, 1983, p. 207, see Ten Berge, 1986a). Furthermore, Ten Berge (1986a) has shown that the columns of $\mathbf{P}$ are not uniquely defined for each separate component. That is, rotations which leave the position of one component unaltered do affect the column of the matrix $\mathbf{P}$ which corresponds to that component. Still, the pattern matrix is widely used for interpretation of components.

The solutions, found with PCA, are typically rotated to some form of simple pattern, before comparisons over groups are made. The favorite rotation for many years has been the Varimax rotation. The Varimax rotation technique was developed in order to facilitate the interpretation of orthogonal components. Kaiser (1958) suggested maximizing the variance of the squared loadings summed across components, in order to simplify the loadings matrix $\mathbf{A}$ ($=\mathbf{P}=\mathbf{S}$). Specifically, that rotated matrix $\mathbf{AT}$ is sought for which the sum (over columns) of

columnwise variances of squared loadings is maximized. Its goal is to rotate the components in such a way that each component contains large loadings for some variables and small loadings for the other variables. The Varimax rotated loadings matrix thus contains as many near zero and large elements as possible, and as few as possible elements in between. Applying this rotation, therefore, generally increases the interpretability of the resulting components, compared to the unrotated components.

In presenting PCA, so far only the situation when one has only a single group of scores has been discussed. What happens when the scores of two or more groups of respondents on the same variables need to be analyzed? The most straightforward approach would be to perform the procedure presented above on each group separately and see whether the same components are found. Thus, a PCA would be performed on each group separately (called PCA–sep), followed by a Varimax rotation on each of the solutions. The usefulness of this procedure is examined in the next Section.

1.3 COMPARING VARIMAX ROTATED COMPONENTS

In order to compare the interpretations of corresponding components in the different groups, Tucker's congruence coefficient (Tucker, 1951) can be calculated for the corresponding columns of the loading matrices. This is a measure for the proportionality of columns. When the congruence of the components found with the two separate analyses is higher than .85, one can conclude that the components will get the same interpretations in both groups (see Haven and Ten Berge, 1977). However, when the congruence of the components found with the two analyses is lower than .85, no conclusion can be drawn. Although a congruence lower than .85 does imply the components *found* in the groups are essentially different, it is still possible that the same components are indeed present in both groups. For instance, a different rotation technique might result in components with a higher congruence.

1.4 APPLYING THE GENERALIZED PROCRUSTES ROTATION

To see if indeed the same components are present in different solutions, Generalized Procrustes Rotation (GPR) can be used. GPR rotates p matrices ($p \geq 2$) toward a best least-squares fit. Specifically, if $\mathbf{A}_1,...,\mathbf{A}_p$ is a set of p matrices of order $m \times q$ ($m \geq q$), GPR finds orthonormal matrices $\mathbf{T}_k$ ($k=1,...,p$) for which the function

$$f(\mathbf{T}_1,...,\mathbf{T}_p) = \sum_{k=1}^{p} \sum_{l=1}^{p} \| \mathbf{A}_k\mathbf{T}_k - \mathbf{A}_l\mathbf{T}_l \|^2 \tag{1.3}$$

is minimized.

If after GPR all congruence coefficients between corresponding columns of $\mathbf{A}_k\mathbf{T}_k$ and $\mathbf{A}_l\mathbf{T}_l$ are high, it can be concluded that the same components are present in the solutions. However, if a GPR fails to yield (nearly) equivalent components, it is still possible that common components do exist. This can be argued in two ways. Firstly, all rotational techniques are, in their search for common components, limited to already found component solutions. It is conceivable that, when a larger number of components is taken, the same components suddenly *do* appear in both groups, because they were merely ordered differently (in terms of the amount of variance they explain) in both groups. Secondly, GPR can only use orthogonal rotations. It is possible that, by using oblique rotations of the solutions, one does find the same components. However, allowing for totally unconstrained oblique rotations in GPR would lead to degenerate solutions where all columns of a matrix are transformed into the same vector. Hence, as soon as one is willing to employ totally unconstrained oblique transformations, it no longer makes sense to use GPR.

A more general problem with GPR, and this again goes for all rotational methods, is that when one is comparing components found, some measure for comparison has to be used. Mostly this will be the congruence criterion (for a description of the congruence criterion, see Section 3.4.1). The problem that is always present, when using such a measure, is for what values of the congruence criterion two components must be judged equal or unequal. The threshold value that has been advocated for the congruence criterion by Haven and Ten Berge (1977) is .85, but what does

one do when the congruence value found falls just below this value, say .84? A researcher, facing this value, could very well be inclined to judge the components as being "almost" equal. Hence, with the use of this measure, no strong conclusions can be drawn.

It was mentioned that when, after GPR, congruences are high, it can be concluded that the same components are present in the solution. However, because GPR attempts to rotate matrices towards one another, some caution is advised. Horn and Knapp (1973) have found that, with *random* target matrices, GPR succeeded in rotating solutions from a factor analysis to such a position that 84% of the hypothesized 'salient' (but randomly chosen) loadings were recaptured from the solution. Thus, GPR might indicate that common components exist in two groups, when in fact there are no (strong) common components.

For these reasons, the GPR does not seem to be the ideal approach. Thus, a different approach seems necessary, that, for one thing, also looks outside the range of rotations of the factor solutions. Two types of techniques for this will be presented in the next section.

1.5 METHODS FOR SIMULTANEOUS ANALYSIS IN TWO OR MORE GROUPS, USED IN THE PRESENT RESEARCH

In the methods presented above, first a q–dimensional solution was derived for each group separately, using a PCA on each group separately, followed by some rotational method, and secondly these solutions were compared. A way to avoid the problems accompanying this approach, is to simultaneously analyze the groups, with a method of analysis that puts an equality constraint on one of the matrices that can be used for interpretation of the components (**W**, **P** and **S**). With such a method it is no longer necessary to decide whether or not components with equal interpretations can be found in the groups, because equality is enforced by the method of analysis. For each of the methods coming up, one of the matrices **W**, **P** and **S** is used for the interpretation of the components.

To see how easy it is to find equal components, let the component

interpretation for a group be based on the weights matrix, found with PCA in one group. To find the same components in a different group, one can simply use this weights matrix for combining the variables in this other group (provided the same variables are used in both groups), thus by definition creating components with exactly the same interpretation in both groups. So the question whether or not it is possible to find the same components across groups is a trivial one: It is always possible to obtain the same components in all groups, by simply defining them in the same way in all groups (see Ten Berge, 1986b), although they need not account for much variance in all groups. A more meaningful question is thus how much of the variance is explained by such components in each group. This is directly related to the aim of PCA: The aim of PCA should *not* be thought of as to find *the* components present in a particular data set (because every conceivable component is present in a data set), but to find a (small) number of components that represent the data as well as possible, by accounting for the maximum amount of variance in the data. We are now in a position to introduce the methods for simultaneous analysis of two or more groups, used in the present study.

Firstly, there are three types of Simultaneous Components Analysis (SCA) methods. With these methods, correlated components are found that explain a maximum amount of variance over the groups of individuals, under the constraint that either the components weights matrix, the pattern matrix, or the structure matrix is kept equal (or at least columnwise proportional) over all groups (see Levin, 1966; Millsap and Meredith, 1988; and Kiers and Ten Berge, 1989; 1994a, respectively), thus ensuring that components with the same interpretation (based on a common weights matrix **W**, a common pattern matrix **P** or a common structure matrix **S**, respectively) are found. These methods are called SCA-W, SCA-P and SCA-S, respectively.

Secondly, there are methods that impose common factors in statistical models for covariance or correlation matrices. These models are fitted using the maximum likelihood principle and will be called ML-methods. In the present study, one ML-method was investigated, namely Simultaneous Factor Analysis in Several Populations (SIFASP-ML, Jöreskog,

1971). Another ML–method, called Partial Common Principal Components Analysis (PCPC, Flury, 1987), originally intended for inclusion in the present study, was discarded from the study in an early stage. This method and reasons for discarding it are presented in Section 1.7.

The methods mentioned above are not the only available methods. In cases where a researcher wants to analyze scores of two or more groups of individuals on the same variables, several other methods for simultaneous factor analysis methods can be used. This study is limited to the four methods, presented above. These methods have been selected because of their promising features. For a detailed description of the methods see Chapter 2.

1.6 GOALS OF THE PRESENT STUDY

When analyzing scores from different groups of respondents on the same variables by a simultaneous components or factor analysis, researchers are faced with a variety of available methods, as we have seen above. Which method is chosen by a researcher seems to be a matter of taste. A choice between the methods on scientific grounds appears to be impossible, because there has been almost no research, systematically comparing these methods. In an attempt to fill this void, in the present study, the usefulness of a selection of the available methods of analysis is examined in a 'comparative commodities research', by applying these methods on constructed data of which the factor solutions are known. In this study, these known solutions are defined as the desired outcomes for the four methods. The main question in the present study is: "To what extent are the methods of analysis doing what they are supposed to do, under different circumstances? The circumstances are manipulated through simulation. The goal of this comparative study is therefore to verify which method is to be preferred under what circumstances and to come up with directions for selecting an appropriate method of simultaneous factor analysis, depending on the circumstances. It is, for instance, conceivable that one (or several) of the methods outperforms some of the

others when the number of variables is small or when the factors are only faintly present in the data. On the other hand, some of the methods can be especially suited for detecting differences between populations while the other methods are not.

The above methods are not the only methods available for the analysis of scores of two or more groups on a set of variables. A number of alternative methods, not dealt with in the present study, will be described in the next section, along with the reasons for discarding them.

1.7 OTHER METHODS FOR THE SIMULTANEOUS ANALYSIS IN TWO OR MORE GROUPS: IDIOSCAL, PARAFAC2 AND INDSCAL

Besides the techniques for simultaneous components or factor analysis, presented in Section 1.5, several other techniques, that might be considered for inclusion in the present study, have been suggested over the last decades. These techniques, called IDIOSCAL, PARAFAC2 and INDSCAL, respectively, will now be presented in short. After that, it will be discussed why these techniques were not included in the present study.

The three methods, presented next, can be viewed as three-way Multidimensional scaling (MDS) methods. Therefore, in the next section, the theory behind MDS will be briefly introduced. At this point, it must be noted that MDS is a technique usually applied to matrices (of order $n \times n$, where n is the number of individuals) containing scores of observer k ($k=1,\ldots,p$), and describing the (dis)similarity between the n individuals, based on the scores of each person on m variables. Because the methods presented in Section 1.5 all focus on the analysis of matrices containing covariances or correlations between m variables, obtained for p ($p \geq 2$) groups of individuals, MDS and the three methods of analysis for MDS will be presented in a form directly applicable to this kind of data.

1.7.1 Multidimensional Scaling

Many situations arise in the social and behavioral sciences in which several covariance matrices for the same variables are available for two or more groups. One would generally like to account for all of these data matrices in a single comprehensive analysis, based on a plausible model. Multidimensional scaling (MDS) is a useful mathematical tool for this, because it can use covariance measures as input, does not require a priori knowledge of the attributes of the variables to be scaled and provides a space that reveals dimensions (components) relevant to the groups. That is, MDS procedures represent variables judged experimentally similar to one another as points close to each other in a resultant spatial map.

As said, the most typical application of MDS, in the present context, arises when covariance matrices are available from two or more groups of individuals. However, instead of groups of individuals, the different matrices may also be associated with other types of data sources, such as different experimental conditions, tasks, or occasions. The term 'three–way MDS' is used to characterize the models and methods, used in these situations. The input for any three–way method consists of a three–way array $\mathbf{C}$ that can be thought of as comprising a set of p (≥ 2) $m \times m$ symmetric slices $\mathbf{C}_1,...,\mathbf{C}_p$, for instance, covariance matrices.

The three–way MDS method IDIOSCAL is presented in Section 1.7.2. The three–way MDS methods PARAFAC2 and INDSCAL, presented in Section 1.7.3 and 1.7.4, respectively, can be considered as models imposing restrictions on the model IDIOSCAL. The three models presented can be viewed as differing essentially in how they decompose the covariances for group k, k=1,...,p.

1.7.2 IDIOSCAL

The most general model in this class of three–way MDS models has been called (by Carroll & Chang, 1972) the IDIOSCAL (*I*ndividual *DI*fferences *I*n *O*rientation *SCAL*ing) model. In matrix notation, this model

can be expressed as

$$\mathbf{C}_k = \mathbf{A}\mathbf{H}_k\mathbf{A}' + \mathbf{E}_k, \tag{1.4}$$

$k=1,...,p$, where $\mathbf{H}_k$ is a $q \times q$ symmetric positive semi-definite (p.s.d) matrix, $\mathbf{A}$ is an arbitrary $m \times q$ $(m \geq q)$ matrix, and $\mathbf{E}_k$ is a matrix containing the part not explained by the model. IDIOSCAL minimizes the function

$$\sigma_1(\mathbf{A},\mathbf{H}_1,...,\mathbf{H}_p) = \sum_{k=1}^{p} \| \mathbf{C}_k - \mathbf{A}\mathbf{H}_k\mathbf{A}' \|^2. \tag{1.5}$$

Kiers, Cléroux and Ten Berge (1994) proposed an algorithm that minimizes the IDIOSCAL function and converges monotonically, for p.s.d. matrices $\mathbf{C}_k$, $k=1,...,p$.

Ten Berge, Bekker and Kiers (1994) considered the TUCKALS2 (T2) algorithm applied to the IDIOSCAL problem, as first proposed by Kroonenberg and De Leeuw (1980). Ten Berge et al. proved that, in case TUCKALS2 is applied to p.s.d. matrices, the TUCKALS2 algorithm does converge to a proper IDIOSCAL solution.

Note that the solution for the matrix $\mathbf{A}$ for the IDIOSCAL problem is not unique, although all solutions span the same column spaces.

1.7.3 PARAFAC2

In order to introduce PARAFAC2 properly, first the method from which it was derived, PARAFAC1 (*PARA*llel *FAC*tors), will be presented. PARAFAC1 is an interesting generalization of principal components analysis for the situation where the same variables have been measured on the same observation units (individuals, groups) a number of times (Harshman, 1970; Harshman & Lundy, 1984). One of the main features of PARAFAC1 is that it gives factors that (under mild conditions) are unique up to scaling and permutation. However, the standard PARAFAC1 method cannot be used for analyzing data from different groups. For an exploratory analysis of such cross-sectional data, taking into account that the same variables were measured in all groups, Harshman (1972) proposed to fit the PARAFAC2 model to the covariance matrices obtained in the different groups (see also Carroll & Wish, 1974, p. 94-96; Harshman & Lundy, 1984, p. 187).

If $\mathbf{C}_k$ denotes the $m \times m$ covariance matrix at occasion k, then the PARAFAC2 model can be described as

$$\mathbf{C}_k = \mathbf{A}\mathbf{D}_k\mathbf{H}\mathbf{D}_k\mathbf{A}' + \mathbf{E}_k, \tag{1.6}$$

$k=1,...,p$, where $\mathbf{A}$ denotes an $m \times q$ matrix of 'loadings' for the m variables on a set of q 'latent factors', $\mathbf{D}_k$ denotes a $q \times q$ diagonal matrix containing 'saliences' for the different factors for $\mathbf{C}_k$, $\mathbf{H}$ denotes a $q \times q$ symmetric (and usually p.s.d) matrix with relations between the factors, and $\mathbf{E}_k$ is a matrix with error terms. For details on the interpretation of the various parameter sets, the reader is referred to Harshman (1972), Carroll and Wish (1974, pp. 94–96), Dunn and Harshman (1982), or Kroonenberg (1983, pp. 53–56), among others.

Compared with the IDIOSCAL model, Harshman's (1972) PARAFAC2 model is the specific case of IDIOSCAL in which the matrix $\mathbf{H}_k$ is constrained to be of the form

$$\mathbf{H}_k = \mathbf{D}_k\mathbf{H}\mathbf{D}_k, \tag{1.7}$$

where $\mathbf{D}_k$ is a diagonal matrix.

The PARAFAC2 model is fitted to a set of covariance matrices in the least squares sense by minimizing

$$\sigma_2(\mathbf{A},\mathbf{H},\mathbf{D}_1,...,\mathbf{D}_p) = \sum_{k=1}^{p} \| \mathbf{C}_k - \mathbf{A}\mathbf{D}_k\mathbf{H}\mathbf{D}_k\mathbf{A}' \|^2. \tag{1.8}$$

Kiers (1993) proposed an alternating least squares algorithm for PARAFAC2, using an adjusted version of the three-way DEDICOM (DEDICOM3) algorithm, alternately updating the matrices $\mathbf{H}$, $\mathbf{A}$, and $\mathbf{D}_1,...,\mathbf{D}_p$. Kiers found the PARAFAC2 algorithm to show good convergence properties, always explaining at least 99% of variance in errorless data constructed with $\mathbf{H}$ p.s.d. and according to (1.6), although the procedure converged very slowly when the function value in (1.8) approached zero. Kiers also demonstrated the well known nonuniqueness of the PARAFAC2 solutions in case only two covariance matrices are analyzed.

Ten Berge and Kiers (1996) have shown that PARAFAC2 is usually unique with four matrices, in the rank two case. With three matrices, the PARAFAC2 solution is unique when an additional assumption is introduced. If, for instance, the diagonal matrices $\mathbf{D}_k$, $k=1,2,3$, are constrained to

be nonnegative, three matrices are enough to have uniqueness in the rank two case of PARAFAC2. See also Harshman and Lundy (1996) for uniqueness proofs of PARAFAC2 for ranks larger than 2.

1.7.4 INDSCAL

INDSCAL (*IN*dividual *D*ifferences *SCAL*ing) is probably the most widely used three–way MDS method. INDSCAL decomposes the covariance matrix $\mathbf{C}_k$ as

$$\mathbf{C}_k = \mathbf{A}\mathbf{D}_k\mathbf{A}' + \mathbf{E}_k, \tag{1.9}$$

k=1,...,p, with $\mathbf{D}_1,...,\mathbf{D}_p$ diagonal and nonnegative (Carroll & Chang, 1970, p. 284; Carroll, 1972, p. 111; Ten Berge, 1993, p. 64). In the present situation, INDSCAL is the special case of IDIOSCAL in which the matrix $\mathbf{H}_k$ is diagonal and nonnegative.

The INDSCAL model is based on the premise that for all groups the same matrix $\mathbf{A}$ with loadings of the variables on q dimensions is used, but that for each group the idiosyncratic weights attached to these dimensions differ. These weights are the diagonal elements of $\mathbf{D}_1,...,\mathbf{D}_p$.

The problem how to fit the INDSCAL model in the least squares sense amounts to minimizing

$$\sigma_3(\mathbf{A},\mathbf{D}_1,...,\mathbf{D}_p) = \sum_{k=1}^{p} \| \mathbf{C}_k - \mathbf{A}\mathbf{D}_k\mathbf{A}' \|^2. \tag{1.10}$$

This problem has not been solved directly. The most popular approach, due to Carroll and Chang (1970), is based on a technique called splitting (De Leeuw & Heiser, 1982). That is, the two appearances of $\mathbf{A}$ in (1.10) are represented by different matrices, $\mathbf{A}$ and $\mathbf{B}$. These are optimized independently, along with $\mathbf{D}_1,...,\mathbf{D}_p$, by means of the "CANDECOMP" algorithm which minimizes $\sum_{k=1}^{p} \| \mathbf{C}_k - \mathbf{A}\mathbf{D}_k\mathbf{B}' \|^2$. After convergence of CANDECOMP, it is hoped that the $\mathbf{A}$ and $\mathbf{B}$ obtained happen to be equal. In practice, this always appears to be the case. However, Ten Berge and Kiers (1991) have shown that essentially different matrices $\mathbf{A}$ and $\mathbf{B}$ can be found at stationary points of the CANDECOMP function, when Gramian matrices $\mathbf{C}_k$ are analyzed. They did not arrive at a proof for the equality of $\mathbf{A}$ and $\mathbf{B}$ at the global minimum of the CANDECOMP function, nor did they encounter any counterexamples.

Also, one must hope that the solutions obtained for $\mathbf{D}_1,...,\mathbf{D}_p$ have nonnegative diagonal elements throughout. This is not always the case in practice. It is possible to avoid negative saliences by imposing nonnegativity constraints (for instance see Carroll, Pruzansky & Kruskal, 1980; Ten Berge, Kiers & Krijnen, 1993).

A special property of INDSCAL is that its solutions are unique up to mild conditions (Harshman, 1972; see also Ten Berge & Kiers, 1991).

1.7.5 Reasons for not selecting the MDS methods

The three methods, presented above, were discarded from inclusion in the present study on account of practical limitations. The methods do not minimize the residual variance of the data in one way or another. Instead of presenting a model for the data itself, covariance matrices (or correlation matrices) are fitted. This kind of indirect fitting does not tie in nicely with SCA.

1.8 DISCARDING THE PARTIAL COMMON PRINCIPAL COMPONENTS METHOD

As said, the ML-method, called Partial Common Principal Components Analysis (PCPC, Flury, 1987), originally intended for inclusion in the present study, was discarded from the study in an early stage. Next, a short overview of this method will be given, together with the reasons for discarding it.

1.8.1 Partial Common Principal Components (PCPC)

Flury (1984) conceived of PCA as a method of diagonalizing a covariance matrix. Accordingly, he has developed a generalization of PCA, aimed at diagonalizing a set of $m \times m$ covariance matrices $\mathbf{\Sigma}_k$ (using his notation) of p samples simultaneously, called the Common Principal Components model (CPC). In CPC, it is assumed that the p population

covariance matrices have the same m eigenvectors, that is, the same orthonormal matrix diagonalizes all matrices $\mathbf{\Sigma}_k$ simultaneously. For the CPC model, Flury derived a normal theory maximum likelihood estimator.

As a modification of the CPC model, Flury (1987) assumed that only q out of m eigenvectors are common to all matrices $\mathbf{\Sigma}_k$, while the remaining $m-q$ eigenvectors are specific to each group. This is called the Partial Common Principal Components model (PCPC). Flury (1987) defined the PCPC model as

$$\mathbf{\Sigma}_k = \mathbf{B}^{(k)}\mathbf{\Delta}_k\mathbf{B}^{(k)\prime} \quad (k=1,\ldots,p), \tag{1.11}$$

with $\mathbf{B}^{(k)} = (\mathbf{B}_1:\mathbf{B}_2^{(k)})$, and $\mathbf{\Delta}_k$ diagonal, where all $\mathbf{B}^{(k)}$ are orthonormal $m\times m$ matrices, the common part $\mathbf{B}_1$ has dimension $m\times q$ ($q \leq m-2$), and $\mathbf{B}_2^{(k)}$ has dimension $m\times(m-q)$, $k=1,\ldots,p$. It is the common part $\mathbf{B}_1$ that is currently of interest, because the columns of $\mathbf{B}_1$ define the factors common for all groups.

For the PCPC model, the common likelihood function is given as

$$L(\mathbf{\Sigma}_1,\ldots,\mathbf{\Sigma}_p) = c \times \Pi_{k=1}^{p} \exp(\mathrm{tr}(-\tfrac{1}{2}N_k\mathbf{\Sigma}_k^{-1}\mathbf{C}_k))|\mathbf{\Sigma}_k|^{-\frac{1}{2}N_k}, \tag{1.12}$$

in which c is a constant and N_k is n_k-1. Equivalent to maximizing the likelihood function is the minimization of

$$g(\mathbf{\Sigma}_1,\ldots,\mathbf{\Sigma}_p) = -2 \log L(\mathbf{\Sigma}_1,\ldots,\mathbf{\Sigma}_p) + 2 \log c =$$

$$\sum_{k=1}^{p} N_k(\log|\mathbf{\Sigma}_k|+\mathrm{tr}\mathbf{\Sigma}_k^{-1}\mathbf{C}_k). \tag{1.13}$$

1.8.2 Reasons for discarding PCPC

Flury (1987) states that "Solving the likelihood equations for PCPC is extremely laborious and no simple or elegant method has yet been found to accomplish this. [...] However, an approximate solution can easily be obtained, provided that maximum likelihood estimates for the ordinary common principal component model are available. The approximation is based on the observation that, if the partial model holds, the q common components are estimated almost correctly in the ordinary common principal component model, irrespective of the other components that may

be specific in each population (p. 61)." This means that, because in the present study, the point of interest lies in the q common components, the solution found with PCPC is nothing else than a selection of q components from the solution, found with the CPC model. Which q components must be chosen from the m–dimensional solution of the CPC model, however, is not univocally clear.

Two of the possibilities, suggested by Flury (personal communication, 1994), are (1) one can have some idea about a component one expects to find: (2) one can look at the correlations between the components found in the different groups. When two components correlate high in one group and not at all in another, you can not consider these components to be common for both groups and therefore both must be abandoned. However, these guidelines depend to a large extent on subjective decisions.

Ten Berge (1995) has criticized the criterion of diagonalization, used in PCPC (see also Flury, 1988, pp. 205–210). Ten Berge states that "It is by no means granted that components which (partially) diagonalize the covariance or correlation matrices reflect the most important sources of variation in the data." In a small example a clear demonstration of this is given. Let

$$\mathbf{R}_1 = \begin{pmatrix} 1 & .8 & .6 & 0 \\ .8 & 1 & .7 & 0 \\ .6 & .7 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{R}_2 = \begin{pmatrix} 1 & .6 & .75 & 0 \\ .6 & 1 & .85 & 0 \\ .75 & .85 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

and suppose a solution with q=1 is desired. The PCPC model attains the best diagonalizing solution when $\mathbf{B}$=(−.0002 −.0002 .0004 1.0000)'. Clearly, the resulting common component is connected to the fourth variable, rather than with the first three variables. Thus, it fails to reflect the major source of variation underlying the data. This demonstrates a serious drawback of the criterion of diagonalizability.

For the reasons cited above it was decided to discard PCPC from the present study.

1.9 FACTOR ANALYSIS VERSUS COMPONENTS ANALYSIS IN THE ONE GROUP SITUATION

In the present study, generalizations of both components analysis and factor analysis are studied. Factor analysis and principal components analysis use different models for a rather similar analysis of the relations among a number of variables, both describing these relations by means of loadings of the variables on mediating new variables (factors and components, respectively), as well as by correlations with and among these new variables. Considering the common purpose, many researchers have addressed the issue whether one should use a components analysis or a factor analysis, when analyzing the data for one group. Unfortunately, the issue has not been solved. In the present section, some results of the discussion and some of the findings of research concerning factor and component analyses of a single group of scores will be summarized, mainly on the basis of Velicer and Jackson's (1990) review article. It was stated by Velicer and Jackson (1990) that the choice between the methods was not obvious because they serve a similar purpose and share many characteristics, but it can also be said that the choice between the methods is not obvious because they serve different purposes and differ on certain characteristics (Widaman, 1993, p.308).

Velicer and Jackson (1990) discussed a number of theoretical and practical issues. The practical issues included: the degree of numeric similarity between solutions from the two methods, some common rules for the number of factors to be retained, problems with improper solutions (only for factor analysis), and comparisons in the computational efficiency. The theoretical issues included: the factor indeterminacy issue, the differences between exploratory and confirmatory procedures, and the issue of latent versus manifest variables.

With respect to the similarity issue, they state that "when the same number of components or factors are extracted, the results from different types of component or factor analysis typically yield highly similar results. Discrepancies are rarely, if ever, of any practical importance in subsequent interpretations. [...] Arguments that the slightly larger

loadings are *better* or *worse* are not likely to convince anyone."

In their discussion of some common rules for the number of factors to be retained, Velicer and Jackson's final conclusion is that "alternative, more accurate procedures, should replace the common, but problematic procedures."

In factor analysis, it is frequently the case that solutions are improper (for instance, because uniqueness estimates are zero or negative, making the solution, which they are part of, uninterpretable). This is often the case in so–called boundary conditions. These are conditions, such as the presence of only low saturation variables, small sample size, and low variable to factor ratio. The occurrence of improper solutions is usually treated as an annoyance, but according to Velicer and Jackson, could turn into an advantage for factor analysis compared to component analysis, when treated as a diagnostic, signaling that the present model is incorrect for the presented data set.

Concerning the computational issue, Velicer and Jackson state that "in general, component analysis will involve less computing processor time" and "in addition, can be performed on larger variable sets."

Coming to the theoretical issues next, Velicer and Jackson address the factor indeterminacy problem: "Factor indeterminacy is the inability to determine uniquely the common and unique factor variables of the common factor model from the uniquely defined observed variables because the number of observed variables is smaller than the number of common and unique factors" (Mulaik & McDonald, 1978, p.177). Extensive reviews of this problem are provided by Steiger (1979), Steiger and Schönemann (1978) and McDonald and Mulaik (1979). The implications of the factor indeterminacy issue are still being debated, although most writers simply ignore the issue.

The latent versus manifest variables issue concerns the fact that component analysis, being a manifest variable procedure that makes weighted linear composites of the observed variables, is limited to the space of the observed variables, while factor analysis, being a latent variable procedure, is not. Latent variables are estimates of unobserved, underlying, error free variables that (with the addition of sampling

error) account for the observed variables. The major assumed advantage of a latent variable model is the perceived added degree of generalization to other, unsampled variables. However, the existing empirical evidence does not support this claim. In fact, Velicer and Jackson (1990, p. 19) report that the slight differences found favor component analysis.

Some authors, according to Velicer and Jackson, "have made the distinction between component analysis as an exploratory procedure and factor analysis as a confirmatory procedure. The basis of this distinction is unclear, because almost identical operations are employed when either technique is employed in an exploratory manner." They prefer exploratory analytic approaches except for those cases where a well defined theory exists or where there are two (or more) competing legitimate theories, because exploratory approaches avoid confirmation bias, do not force a theory–oriented approach prematurely, and represent a conservative strategy.

The major conclusion of the Velicer and Jackson article is that there is little basis to prefer either component or factor analysis, because although there are some differences, for practical purposes, the choice of method will not greatly affect empirical results or substantive conclusions.

Widaman (1993) compared common factor analysis and principal component analysis, using error–free population data to ensure that sampling variability would neither obscure results nor be available as an alternative explanation for trends noted. He found, among other things, when using component analysis, nonzero pattern loadings were larger than the population loadings and the estimates of component intercorrelations were consistently lower than in the population. Common factor analysis, however, reproduced pattern loadings and interfactor correlations accurately. Widaman further showed that component analysis and common factor analysis define different parameters in the population, and therefore estimate different parameters in samples from that population, as was also shown by Snook and Gorsuch (1989).

Because, in the one group situation, there is no agreement on which method is the best, we have no expectations about the relative

performance of the generalizations of these methods to the two or more groups situation. Therefore, no expectation as to which is the best method for the two or more groups situation is given.

1.10 THE ORGANIZATION OF THE PRESENT STUDY

In Chapter 2, the four methods under investigation (SCA–W, SCA–P, SCA–S and SIFASP–ML) will be presented. For each method, the rotational methods used for simplification of the solution will be presented and its choice motivated. At the end of Chapter 2, some comparisons of the methods on theoretical grounds will be discussed.

In Chapter 3, the method of data construction used in the present study, will be presented. Two data categories will be defined, one for samples from one population and one for samples from two (or more) populations, and a measure for quantifying overall differences between populations will be presented.

In Chapter 4, the manipulations of the independent variables in this study will be described shortly, and measures used for dimension indication for the SCA–methods and SIFASP–ML will be described. Next, the success criteria, used in the present study, are introduced. At the end of Chapter 4, two pilot experiments will be presented, in which the necessary and useful range of amount of error to be added to scores is explored and a suitable convergence criterion for iterative procedures, used in two of the methods under investigation (SCA–W and SCA–S), is determined.

In Chapter 5, the main experiment comparing the three SCA–methods is presented. Findings in this experiment were used as a basis for comparison in following experiments.

In Chapter 6, the SCA–methods are being compared with the method SIFASP–ML. Findings in this experiment were used as a basis for comparison in following experiments.

In Chapter 7, the SCA–methods are compared in an experiment simulating a longitudinal process, in which some components grow weaker

over time, while others grow stronger.

In Chapter 8, the performance of the SCA–methods and the method SIFASP–ML is investigated in three special conditions: analyzing four groups simultaneously, large samples, and overlapping factors.

In Chapter 9, final conclusions are drawn and some guidelines are presented for how to use the most successful method(s) from the present study in practical research situations.