CHAPTER 4

INDEPENDENT VARIABLES, DIMENSION INDICATORS AND SUCCESS CRITERIA

4.1 MANIPULATIONS

In the present study, many different values were chosen for the independent variables, and these values were combined in several ways to construct data that captured certain aspects of real data. In order to retain some sight on what is going on, all the values, chosen for the independent variables in the present study (Experiments 1 to 4), and the true pattern matrices, resulting from these choices, will now be summarized. With these choices, the limits of generalizability of the outcomes of the present experiments are defined. In the experiments to follow (Chapters 5 to 8), for each of which different selections of the now presented values are made, the present section will serve as a point of reference. If necessary, the reader may turn to Section 3.1 for the definitions of matrices.

4.1.1 Independent Variables

- The number of variables (m): 12 and 24; the value of the variable m refers to the number of rows in the true pattern matrix $\mathbf{P}_{tk}$ and the number of columns in the matrix $\mathbf{F}_k$;
- The number of factors (q) explaining the variables in the population: 2 and 4; the value of the variable q refers to the number of columns in the true pattern matrix $\mathbf{P}_{tk}$ and the order of the true correlation matrix $\mathbf{\Phi}_{tk}$;
- The correlations between factors ($\mathbf{\Phi}_{tk}$): 0, 0.2 or 0.4. Within each correlation matrix $\mathbf{\Phi}_{tk}$ all the correlations are the same (i.e. 0, 0.2 or 0.4);
- The sample size (n_k): 50, 100 and 150 (and 300, 500 and 1000 in

Experiment 4b); The sample size refers to the number of rows in the matrix $\mathbf{F}_k$;

- The number of samples (groups) taken (p): 2, 3 or 4; The value of p indicates how many samples are constructed with a true pattern matrix $\mathbf{P}_{tk}$, $k=1,...,p$, and a true correlation matrix $\mathbf{\Phi}_{tk}$, $k=1,...,p$;

- The choice of the pattern matrix. In all studies presented here, a limited number of different values was used for the nonzero loadings in the true pattern matrices. The elements of the pattern matrices determine the strength of the factors. For simple patterns with nonoverlapping factors, the squared loading (λ) of each variable is equal to 1 minus the error variance, so the ratio of noise to signal variances for each variable is equal to $(1-\lambda^2)/\lambda^2$, and, in terms of standard deviations, the noise-to-signal ratio (nsr) is defined as

$$nsr \equiv ((1-\lambda^2)/\lambda^2)^{\frac{1}{2}}. \tag{4.1}$$

In a pilot experiment (see Section 4.5) it was found that the most appropriate values for the nsr were .5, 1.5 and 2.5, corresponding with true loadings λ of approximately .89 , .55 and .37, respectively. In the case of overlap, the same nsr's were used for comparability. For the variables, with nonzero loadings on more than one factor, the nsr's do not correspond with the above mentioned loadings.

- Degree of overlap between the factors. Overlap between factors is defined as the situation where a variable has a true loading unequal to zero on two or more factors. In the no overlap situation each variable has only one nonzero true loading. Overlapping factors have been used in Experiment 4c only, where some variables load on two factors in the true pattern matrix instead of one;

- Method of analysis: The four methods used to analyze the constructed data sets were SCA-W, SCA-P, SCA-S and SIFASP-ML, as presented in Chapter 2.

For each condition, described by choices of n, q, $\mathbf{P}_{tk}$ and $\mathbf{\Phi}_{tk}$, ten data sets were analyzed by each method. The maximizations in the methods SCA-W and SCA-S (which require an iterative procedure) were performed six times for each of the ten replications, once starting iterations with a rationally derived matrix for the matrices $\mathbf{W}$ and $\mathbf{S}$, respectively, and

five times starting with a random matrix. Both rational starts are based on the eigendecomposition $\sum_{k=1}^{p} \mathbf{R}_k = \mathbf{K}\Lambda\mathbf{K}'$. The start *weights* matrix is then derived as $\mathbf{W} = \mathbf{K}_q\Lambda_q^{-\frac{1}{2}}$, where $\mathbf{K}_q$ is the $m \times q$ matrix containing the first q columns of matrix $\mathbf{K}$ and Λ_q is the $q \times q$ matrix containing the upper left quadrant of matrix Λ; for the rational *structure* matrix, the first q columns of $\mathbf{K}$ are taken. As random start matrices for both the weights and the structure matrices, an $m \times q$ matrix of random values between plus and minus .5 is taken, which is made columnwise orthonormal by applying Gram–Schmidt orthonormalization. The best of the six solutions was taken. The iterative procedure for the special rotation procedure of the SCA–S solution was performed 21 times for each replication; once starting iterations with a rationally derived rotation matrix $\mathbf{T}$ and 20 times starting with a random rotation matrix $\mathbf{T}$. As the rational start matrix, the varimax rotation matrix, derived when rotating the rowwise normalized structure matrix, was used; as a random start matrix for the rotation, a matrix containing values between plus and minus .5 was derived which was then normalized columnwise.

The convergence criterion for the iterative procedures in the methods SCA–W and SCA–S was determined as follows: When the difference between subsequent function values (for iteration i and i+1) is smaller than .00001 times the function value at iteration i, the program stopped iterating. A justification for this convergence criterion comes from the pilot experiment, presented in Section 4.6.

4.1.2 The true matrices underlying the constructed data

The true pattern matrices were chosen in such a way that they pertained to 2 or 4 factors with non–overlapping clusters of variables (except in Experiment 4c). The true pattern matrices, for 2 and 4 factors, and 12 variables, were rowwise rescaled versions of the matrices

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

respectively. So in the 2 factors situation there were 6 variables per factor and in the 4 factors situation there were 3 variables per factor in the true pattern matrix.

The different true pattern matrices, of which selections were used in each of the different experiments, are given in Appendix A. It can be seen that different combinations of strong and weak factors are used in choosing the true pattern matrices. Only the true pattern matrices in the 12 variables condition are given, for the 2 and 4 factor condition, respectively. In the 24 variables condition, the respective true pattern matrices from the condition of 12 variables were duplicated and stacked one below the other. For instance, the true pattern matrix for 24 variables, analogous to the true pattern matrix $\mathbf{P}_a$ given below for the 12 variables condition, is given as $\tilde{\mathbf{P}}_a = (\mathbf{P}_a' | \mathbf{P}_a')'$. In the presentation of the loadings in Appendix A, only the first two decimals are given, but the loadings used were not rounded.

While the loadings in the true pattern matrices almost always can be grouped into nonoverlapping clusters of variables (the only exception is Experiment 4c), the correlations in the true structure matrix can only be grouped into nonoverlapping clusters of variables when the matrix with true correlations between the factors is chosen as the identity matrix, because the true structure matrix is derived as $\mathbf{S}_t = \mathbf{P}_t \mathbf{\Phi}_t$. In all other cases the correlations in the true structure matrix can only be grouped into overlapping clusters of variables.

Six different true correlation matrices $\mathbf{\Phi}_t$, with correlations between the factors, were used. These were

orthogonal: $\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ or $\mathbf{I} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$,

mildly oblique: $\mathbf{\Phi}_{.2} = \begin{pmatrix} 1.0 & 0.2 \\ 0.2 & 1.0 \end{pmatrix}$ or $\mathbf{\Phi}_{.2} = \begin{pmatrix} 1.0 & 0.2 & 0.2 & 0.2 \\ 0.2 & 1.0 & 0.2 & 0.2 \\ 0.2 & 0.2 & 1.0 & 0.2 \\ 0.2 & 0.2 & 0.2 & 1.0 \end{pmatrix}$ and

strongly oblique: $\mathbf{\Phi}_{.4} = \begin{pmatrix} 1.0 & 0.4 \\ 0.4 & 1.0 \end{pmatrix}$ or $\mathbf{\Phi}_{.4} = \begin{pmatrix} 1.0 & 0.4 & 0.4 & 0.4 \\ 0.4 & 1.0 & 0.4 & 0.4 \\ 0.4 & 0.4 & 1.0 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1.0 \end{pmatrix}$,

respectively.

4.2 DIMENSION INDICATORS FOR THE SCA-METHODS

In this section, four measures are presented that were devised in order to determine the number of components to retain, when doing a simultaneous components analysis. In the present study, these measures of fit will be called dimension indicators.

4.2.1 Additional value greater than one (KA1)

An often used rule of thumb says that a component must explain at least as much variance as one variable would, for it to be added to the solution (the Kaiser eigenvalue greater than one rule, K1). Zwick and Velicer (1986) found that in the majority of Monte Carlo studies, analyzing a single group with PCA, this rule overestimated the number of components, but also examples indicating the opposite have been presented (e.g. Cliff, 1988).

In the present study, a variant of this measure was used in determining the number of components to retain, called the "additional value greater than one"-measure (KA1). The variant is based on the

rationale that, to increase the dimensionality of the solution from q–1 to q, the incremental value of the overall percentage of explained variance (denoted as d_q) needs to be larger than the percentage of variance explained by one variable. Accordingly, the number of components indicated by the measure KA1 (denoted as q_{KA1}), is the highest value of q for which the inequality

$$d_q > 100/m \tag{4.2}$$

holds. So, for the KA1–measure, the original 'eigenvalue greater than one'–part from the K1–measure is replaced by 'additional value greater than one'. This is done, because SCA–W and SCA–S do not involve eigenvalues.

4.2.2 Parallel Analysis (PA)

Parallel analysis (Horn, 1965) is a sample–based adaptation of the population–based K1 rule. Horn (1965) noted that at the population level, the eigenvalues of a correlation matrix of uncorrelated variables would all be 1.0. When samples are generated based upon such a matrix, however, the first eigenvalues exceed 1.0 whereas the last eigenvalues are below 1.0. Horn suggested that the eigenvalues of a random sample correlation matrix of m variables should be contrasted with those of the data set in question, based upon the same sample size. Components of the matrix of interest, which corresponds to eigenvalues greater than those of the comparison random matrix, are retained. According to Zwick and Velicer (1986), "this approach integrates the reliability and data–summarizing emphasis of the population–based K1 rule and the effects of sample size." In practice, a number of comparison random matrices is used, and the eigenvalues found are averaged, to attain a more accurate assessment of the random eigenvalue. Glorfeld (1995) proposed an improvement on PA by using a considerably larger number of random correlation matrices to find the complete sampling distribution of eigenvalues for each eigenvalue extracted.

Zwick and Velicer (1986) examined the accuracy of five rules (among which were the K1 rule and PA) in determining the number of components to

retain in PCA. Of the five rules, PA gave by far the most accurate determination of the number of components. For the present study, a generalization of the PA rule to two or more samples was proposed. For a proper generalization, sets of two or more random samples have to be analyzed by each method separately. The incremental amount of explained variance for each dimension corresponds to the eigenvalue for a dimension in the one sample situation. The generalization of Glorfeld's (1995) modification was considered too laborious for the present study.

For each of the three SCA–methods, an analysis of the random comparison matrices was done. To increase the dimensionality of the solution from q–1 to q, the incremental value of overall explained variance (denoted as d_q), explained by the components found with an SCA–method of the data matrices, needs to be higher than the incremental value of overall variance explained with the components found with the same SCA–method in the analysis of the random comparison matrices. The incremental value of overall variance, from the analysis of the comparison matrices, was taken as the mean value found by each method over 50 replications (and denoted by IV_q; Incremental Value found when analyzing 50 (sets of) random comparison matrices). In formula, the number of components indicated by the measure PA (denoted as q_{PA}), is found at the highest value of q for which the inequality

$$d_q > IV_q \tag{4.3}$$

holds.

4.2.3 The Quotient of Additional values (QA)

When doing a simultaneous component analysis, choosing a q–dimensional solution, the amount of variance explained by the q components can be calculated. For the (q+1)–dimensional solution, the amount of variance explained by these q+1 components will be at least as large as the amount of variance explained by the q components.

The rationale behind the QA measure is that it is expected that, when a solution is calculated with one dimension higher than the correct dimension, there will be a drop in the additional amount of explained

variance, and it is expected that increasing the dimensionality further will not give comparatively larger drops in explained variance. Denoting the increase in the total percentage of variance explained when increasing the dimensionality of the solution from q–1 to q as d_q, the quotient of additional values (QA) is defined as the quotient of d_q and d_{q+1}:

$$\mathrm{QA}_q = \frac{d_q}{d_{q+1}}, \tag{4.4a}$$

with the requirement that

$$d_q > 100/m; \tag{4.4b}$$

the value of d_q must be larger than the amount of variance, explained with one variable. If not, the value QA_q is not calculated. Due to (4.4b), the maximum number of components possible equals q_{KA1}, the number of components indicated by KA1.

The interpretation of this measure is straightforward. When the value QA_q is not considerably larger than one, adding the $(q+1)^{th}$ component increases the total amount of explained variance almost as much as adding the q^{th} component. Therefore, a value of QA_q close to one indicates there is no reason to choose q components. When, however, the QA_q has a value of, for instance, six, this means that adding the q^{th} component increases the total amount of explained variance six times as much as adding the $(q+1)^{th}$ component. Thus we can say that the profit from adding the $(q+1)^{th}$ component is much less than the profit from adding the q^{th} component, and a choice for no more than q components seems warranted.

In the present study, the QA was used as a comparative measure. That is, the dimension that received the highest QA value was chosen as the correct dimension, so q_{QA} is the value of q for which

$$\mathrm{QA}_q = \max(\mathrm{QA}_k),\ k=1,\ldots,q_{KA1}. \tag{4.5}$$

From the definitions of the measures QA and KA1, it is obvious that the dimension, indicated by the measure QA, will always be lower than or equal to the dimension, indicated by the measure KA1, so

$q_{QA} \leq q_{KA1}$.

For the methods PCA and SCA–P, the amount by which the total amount of explained variance increases, decreases with each component added, because the solutions found with these methods are nested. Because the solutions, found with the methods SCA–W and SCA–S, are not nested, for these methods, this conclusion cannot be drawn. In looking at the outcomes of the experiments in this study, we have seen for the methods SCA–W and SCA–S that it was possible that adding the q^{th} component showed a smaller increase in the total amount of explained variance than adding the $(q+1)^{th}$ component. This means that the measure QA has a minimum possible value of one for the methods PCA–sep and SCA–P, whereas for the methods SCA–W and SCA–S we only know that it can be smaller than one.

4.2.4 The Quotient of Differences in Additional values (QDA)

For the QA–measure, it was mentioned that, when a solution is calculated with one dimension higher than the correct dimension, a drop in the additional amount of explained variance is expected. For the 'Quotient of Differences in Additional values'–measure (QDA), it is argued that, after the significant drop in additional amount of explained variance, further decreases in additional explained variance will be relatively small.

Denoting the increase in the total percentage of variance explained by adding the q^{th} component as d_q, the QDA for component q was defined as the quotient of $d_q - d_{q+1}$ and $d_{q+1} - d_{q+2}$:

$$\mathrm{QDA}_q = \frac{d_q - d_{q+1}}{d_{q+1} - d_{q+2}}, \qquad (4.6a)$$

again, subject to requirement (4.4b): the value of d_q must be larger than the amount of variance, explained with one variable, for the same reason as for the measure QA. The rationale behind the QDA measure can now be thought of as follows. If a drop in additional amount of explained variance occurs when increasing the dimension of the solution with the correct dimension by one, as was argued, the value of d_{q+1} will be small

compared to d_q, so the difference $d_q - d_{q+1}$ will be considerably larger than zero. However, if further decreases in d_q are small when adding another dimension to the solution, the value of d_{q+2} will be close to the value of d_{q+1}, so the difference $d_{q+1} - d_{q+2}$ will be close to zero. Therefore, the solution with the correct dimension is expected to receive the highest value of the QDA measure.

Note from the remarks at the end of Section 4.2.3, that for the methods SCA–W and SCA–S, the value of the measure QDA can become negative. It was found to be possible that the absolute value of this negative value was higher than the highest positive value. To incorporate this in formula (4.6a) the absolute value of the differences in additional values was taken, so for the methods SCA–W and SCA–S, the adjusted formula

$$\mathrm{QDA}_q = \frac{|d_q - d_{q+1}|}{|d_{q+1} - d_{q+2}|} \tag{4.6b}$$

was used for the QDA measure.

In the present study, the dimension that received the highest QDA value was chosen as the correct dimension, so q_{QDA} is the value of q for which

$$\mathrm{QDA}_q = \max(\mathrm{QDA}_k),\ k=1,\ldots,q_{KA1}. \tag{4.7}$$

Again, it is clear that $q_{QDA} \leq q_{KA1}$.

4.2.5 The use of the dimension indicators in the present study

In the experiments to be reported below, interest is not so much in the values of the dimension indicators themselves, but rather in how they compare to the "true dimensionality" underlying the data. Specifically, it was inspected whether the correct dimension was indicated, or over– or underestimation of the correct dimension occurred.

4.3 DIMENSION INDICATORS FOR SIFASP-ML

In the present study, SIFASP–ML models were fitted using the LISREL program (specifically, LISREL 8E, Jöreskog & Sörbom, 1993). The reasons for using LISREL8E rather than LISREL8 are that it can handle larger problems and executes considerably faster than LISREL8.

In the evaluation of the fit of a model, the LISREL program presents several fit indices. In the present study, a selection of these fit indices was used to evaluate the fit of competing models, and thereby to determine the number of factors to retain. Therefore, the fit indices will be called dimension indicators. Besides the sequential chi–square difference test, based on the chi–square value, selected because this is the (mis)fit index mostly used, also a fit index using the error of approximation and a fit index based on cross–validation were selected, because of their promising features.

Before introducing these fit indices, let it be noted that it is not possible to use the various indices of fit, used in factor analysis, in such a way, that all researchers with experience in the field of factor analysis will agree on the number of factors to retain. Let it also be noted that many researchers deem it unwise to base the choice for the number of factors on only one of the fit measures to be described. In the present study, the values of the fit measures were used both separately (Section 4.3.1 to 4.3.4) and in conjunction with one another (Section 4.3.6) to decide on the number of factors to retain. The purpose of utilizing the fit indices in these two different ways was to assess what usage of these fit indices is preferable. The fit measures were evaluated taking into account only the solutions for which no irregularities occurred (Section 4.3.7).

For the introduction of the fit indices, chosen for the present study, some definitions are useful, additional to the definitions given in Section 3.1.1. For reasons of consistency, some of the following notations differ from what is usual in multivariate statistics (e.g., the letter **C** is used to denote the sample covariance matrix, instead of the letter **S** that is normally used, because the letter **S** is already in use to

denote the structure matrix).

$\boldsymbol{\Sigma}_o$ = the covariance matrix of the population from which a sample is drawn. It is assumed that the sample covariance matrix $\mathbf{C}$ converges in probability to the population covariance matrix $\boldsymbol{\Sigma}_o$ with increasing sample size n;

$\boldsymbol{\theta}_j$ = the vector of parameters specifying model M_j;

$\tilde{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}}_j)$ represents the best possible fit of the postulated model M_j to the population covariance matrix $\boldsymbol{\Sigma}_o$ in terms of the discrepancy function F_{ML};

$\hat{\boldsymbol{\Sigma}}_j = \boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j)$ represents the best fit of the postulated model M_j to the sample covariance matrix $\mathbf{C}$ in terms of the discrepancy function F_{ML};

$\hat{F}_j = F_{ML}(\mathbf{C};\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}}_j))$ is the minimum value of the discrepancy function F_{ML}, based on the best fit $\hat{\boldsymbol{\Sigma}}_j$ of the postulated model M_j to the sample covariance matrix $\mathbf{C}$;

k = subscript indicating the k^{th} group;

t_j = the number of parameters estimated in model M_j;

d_j = the degrees of freedom of model M_j;

m = the number of observed variables;

p = the number of groups;

$N = \sum_{k=1}^{p} n_k$.

In the literature, for the multiple group situation (p>1), the fit measures encountered were all based on χ^2. In fact, three different definitions of the chi-square measure were given, by Bentler (1989), Bollen (1989), and Jöreskog and Sörbom (1993), respectively. The only difference between the three definitions lies in the weights used for the minimized function values for each group, that is, the three authors used the weights $(n_k-1)/(N-p)$, $(n_k-1)/N$, and n_k/N, respectively. Although these differences are almost negligible, because the calculation of the chi-square value is based on asymptotic theory, it was checked which of the three weights was used in the LISREL program. It turned out to be Bentler's definition (see Section 4.3.1).

For the other fit indices used in the present study, both the

definitions of the fit indices for the single group situation and generalizations to the multiple group situation will be discussed. For the fit indices RMSEA and ECVI (Sections 4.3.3 and 4.3.4), these generalizations for the multiple group situation were derived from the values presented for these fit measures by the LISREL program.

4.3.1 The Chi-square statistic

To assess the fit of a model, many researchers will first look at the chi-square value. Let $\boldsymbol{\theta}_o$ contain the values of the parameters $\boldsymbol{\theta}$ that minimize the discrepancy function $F_o = \sum_{k=1}^{p}\left((n_k-1)/(N-p)\right) F_{ML}(\boldsymbol{\Sigma}_{ko};\boldsymbol{\Sigma}_k(\boldsymbol{\theta}))$, describing the difference between $\boldsymbol{\Sigma}_{ko}$ and $\boldsymbol{\Sigma}_k(\boldsymbol{\theta})$, $k=1,...,p$. Then the model is true when $\boldsymbol{\Sigma}_{ko} = \boldsymbol{\Sigma}_k(\boldsymbol{\theta}_o)$ for $k=1,...,p$, in other words when $F_{ML}(\boldsymbol{\Sigma}_{ko};\boldsymbol{\Sigma}_k(\boldsymbol{\theta}_o)) = 0$. However, in practice only a sample covariance matrix $\mathbf{C}_k$ based on a sample of size n_k is available. In that case, $\hat{\boldsymbol{\theta}}$ contains the values of the parameters $\boldsymbol{\theta}$ minimizing $\hat{F}_j = \sum_{k=1}^{p}\left((n_k-1)/(N-p)\right) F_{ML}(\mathbf{C}_k;\boldsymbol{\Sigma}_k(\boldsymbol{\theta}))$, $k=1,...,p$. To test whether a model M_j is valid, under the assumption that the model is true in the population, the chi-square statistic

$$\chi_j^2 = (N-p)\hat{F}_j = \sum_{k=1}^{p}(n_k-1)\, F_{ML}(\mathbf{C}_k,\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)) \qquad (4.8)$$

can be used (cf. Bentler, 1989, p. 222–223), where F_{ML} is the discrepancy function for ML-estimation

$$F_{ML}(\mathbf{C}_k,\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)) = \log|\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)| + \mathrm{tr}(\mathbf{C}_k\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)^{-1}) - \log|\mathbf{C}_k| - m, \qquad (4.9)$$

and $\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}_j)$ is the estimate of the population covariance matrix using model parameters $\hat{\boldsymbol{\theta}}_j$. The degrees of freedom are given as

$$df = \tfrac{1}{2}p(m)(m+1) - t_j.$$

For the maximum likelihood estimation method with discrepancy function $F_{ML}(\mathbf{C}_k;\boldsymbol{\Sigma}_k(\hat{\boldsymbol{\theta}}))$, the chi-square measure is distributed asymptotically as a central chi-square distribution under the assumption of multinormality of the observed variables, if a covariance matrix is analyzed and under the null hypothesis that the model is exactly true. If a correlation matrix

is analyzed with ML, the χ^2–distribution applies when the model is scale–invariant and diag($\hat{\mathbf{\Sigma}}$)=diag($\mathbf{R}$) (Jöreskog & Sörbom, 1993, p. 32). In the present study, the models used were not scale–invariant. Therefore, the chi–square measure was only used as a descriptive measure for the badness of fit of a model, rather than as a measure for testing whether a model was valid, as will be described next.

4.3.2 The Sequential Chi–square Difference Test (SCDT)

In the present study, the sequential chi–square difference test (SCDT) (see, for instance, Steiger, Shapiro & Browne, 1985) was used to ascertain the "best" value for the number of common factors to retain. In this procedure, two nested models are compared by treating the difference of their chi–square test statistics as a chi–square statistic with degrees of freedom equal to the difference between the degrees of freedom for the individual chi–squares. Here, the basic idea is to interpret the difference chi–square as an "improvement index". If the difference chi–square is significant (using α=.05), it indicates that a significant improvement in the fit of the model has been produced by increasing the number of free parameters, i.e. adding one (or more) factors. The SCDT is applied repeatedly to compare all the nested models (from 1 to 6 factors), adding one factor at a time.

In the present study, in many instances only a subset of the models (with one to six factors) was present, and the SCDT could therefore not always be done. This problem was handled as follows. If the model with one factor did not fit (so no chi–square value was present), the model with two factors was taken as the first model and it was inspected whether this model had acceptable fit ($P(\chi^2_{df}<x)>.05$). If the model with two factors did not fit either, the model with three factors was inspected for acceptable fit, and so on. From the first model that fitted on, the SCDT was used. When a next model that did not fit was encountered, of course, again no SCDT could be done. In such a case, the first subsequent model with a larger number of factors was taken and the difference in chi–square values between the two models was tested for

significance.

4.3.3 Root Mean Square Error of Approximation (RMSEA)

Steiger's (1990; see also Steiger & Lind, 1980) Root Mean Square Error of Approximation (RMSEA) is a measure for the error of approximation per degree of freedom for the model. For the derivation of this measure, it is assumed that the postulated model is not valid, so the null hypothesis is false. If a postulated model is not valid, so if we have $\boldsymbol{\Sigma}_o \neq \boldsymbol{\Sigma}(\boldsymbol{\theta}_o)$ or $F_o > 0$, the statistic $\hat{F}_j$ in (4.8) asymptotically has a non–central chi–square distribution with d_j degrees of freedom and non–centrality parameter λ, defined as $\lambda = (n-1)F_o$. For the single group situation, the RMSEA is defined as

$$\text{RMSEA} = (F_o/d_j)^{\frac{1}{2}}, \tag{4.10}$$

where $F_o = F(\boldsymbol{\Sigma}_o, \tilde{\boldsymbol{\Sigma}}_j)$ can be any fit function. The sample discrepancy function value $\hat{F}_j$ is an appreciably biased estimator of F_o, the discrepancy due to approximation, which is therefore not suited as an estimate of F_o. Because $(n-1)\hat{F}_j$ has, according to the assumption made above, an asymptotic noncentral chi–square distribution with noncentrality parameter $\lambda=(n-1)F_o$ and d_j degrees of freedom, the expected value of $\hat{F}_j$ may be approximated as

$$\mathcal{E}(\hat{F}_j) \approx F_o + d_j/(n-1). \tag{4.11}$$

Consequently, a less biased point estimator of the discrepancy due to approximation F_o, would be given by $\hat{F}_o = \hat{F}_j - d_j/(n-1)$ (cf. McDonald, 1989). This statistic can have negative values, therefore

$$\max\left(\hat{F}_j - d_j/(n-1),\ 0\right)$$

was used as a point estimate for the discrepancy due to approximation. A point estimate of the RMSEA, for model M_j, is given by

$$\text{RMSEA}_j = (\hat{F}_o/d_j)^{\frac{1}{2}} = \left(\max\left(\hat{F}_j/d_j - 1/(n-1),\ 0\right)\right)^{\frac{1}{2}}, \tag{4.12}$$

(Browne & Cudeck, 1992). The generalization of the point estimate of the RMSEA for model M_j to two or more groups is given by

$$\text{RMSEA}_j = (\hat{F}_o/d_j)^{\frac{1}{2}} = \left(\max\left(\hat{F}_j/d_j - 1/(N-p),\ 0\right)\right)^{\frac{1}{2}}. \qquad (4.13)$$

Browne and Cudeck (1992) concluded on the basis of their practical experience that "a value of the RMSEA_j of about 0.05 or less would indicate a *close fit* of the model in relation to the degrees of freedom. This value is based on subjective judgment. It cannot be regarded as infallible or correct, but is more reasonable than the requirement of exact fit with the RMSEA = 0.0. We are also of the opinion that a value of about 0.08 or less for the RMSEA_j would indicate a reasonable error of approximation and would not want to employ a model with a RMSEA_j greater than 0.1" (p. 239). So, according to Browne and Cudeck, there lies a grey area between the values 0.05 and 0.1.

For the present study, the model (number of factors) with the lowest RMSEA_j, which had to be lower than 0.05, was taken as the model indicated by the RMSEA_j measure. If none of the models had an RMSEA_j lower than 0.05, it was concluded that none of the models could be chosen on the basis of the RMSEA measure. If competing models had the same lowest RMSEA value (i.e. zero), the model with the lowest number of factors was seen as the one indicated by the measure RMSEA.

In order to test the quality of the solution with the correct dimension, as found by SIFASP–ML, the test for close fit, proposed by Browne and Cudeck (1992) was used. Browne and Cudeck (1992, pp. 240–241) find the point hypothesis of exact fit

H_o: RMSEA = 0.0,

of little practical use, because this null hypothesis is invariably false in practical situations. Therefore, they proposed a test for close fit, adopted in the present study, testing the null hypothesis

H_o: RMSEA $\leq$ 0.05

against the alternative

H_1: RMSEA > 0.05.

To test this hypothesis, the test statistic $(N-p)\hat{F}_j$, having a non–central chi–square distribution, is used. Let $G((N-p)\hat{F}_j|\lambda,d_j)$ be the cumulative distribution function of the noncentral chi–square distribution with

noncentrality parameter $\lambda = (N-p)F_o = (N-p)d_j(\text{RMSEA})^2$ and d_j degrees of freedom. The probability that the stochastic variable $(N-p)\underline{\hat{F}}_j$ has a value larger than the observed value $(N-p)\hat{F}_j$, is given by

$$P\left((N-p)\underline{\hat{F}}_j > (N-p)\hat{F}_j\right) = 1 - P\left((N-p)\underline{\hat{F}}_j \leq (N-p)\hat{F}_j\right)$$

$$= 1 - G\left((N-p)\hat{F}_j | (N-p)d_j(0.05)^2, d_j\right). \qquad (4.14)$$

The null hypothesis is rejected, when the p–value is less than a prespecified 'subjective' level. In the present study, for the model with the correct number of factors, the null hypothesis was tested with α=0.05.

4.3.4 Expected Cross-Validation Index (ECVI)

Cudeck and Browne (1983) suggested that the entire approach for evaluating models should be modified: "Selection methods should de–emphasize the practice of developing a model in a single sample of data, and instead should be concerned with identifying models which will perform optimally in future samples. One reasonable method for this purpose is cross–validation" (o.c., pp. 150–151). Therefore, Browne and Cudeck (1989) proposed to use an expected value of the cross–validation index (CVI) for model selection. For this index, two samples with scores on the same variables are required. One of the samples is (arbitrarily) used as a calibration sample and the other sample is used as a validation sample, denoted by $\mathbf{C}_c$ and $\mathbf{C}_v$, respectively. The CVI is defined by Cudeck and Browne (1983, p. 151) as

$$\text{CVI} = F(\mathbf{C}_v;\hat{\mathbf{\Sigma}}_c),$$

where the estimate $\hat{\mathbf{\Sigma}}_c$ is based on the calibration sample. The smaller the value of the CVI, the more a researcher can trust that his model will hold in independent, new samples, drawn from the same population.

When only a single sample is available, however, this sample has to be divided at random into two subsamples (a calibration and a validation sample), in order to calculate a CVI. To avoid this, Browne and Cudeck (1989) devised an estimator for an unconditional expected

cross–validation index (ECVI), i.e. the expected value over calibration and validation samples:

$$\text{ECVI} = \mathcal{E}_c\mathcal{E}_v(\text{CVI}) \approx \text{F}(\boldsymbol{\Sigma}_o;\tilde{\boldsymbol{\Sigma}}_j) + (m^*+t)/(n-1), \tag{4.15}$$

where $m^* = \frac{1}{2}m(m+1)$. The point–estimate for the ECVI, under the postulated model M_j, is given by

$$\text{ECVI}_j = \text{F}(\mathbf{C};\hat{\boldsymbol{\Sigma}}_j) + 2t_j/(n-1). \tag{4.16}$$

The smaller the estimate of the ECVI, the better. Note that the ECVI_j depends on the number of parameters estimated and on the sample size. With the addition of extra parameters (increase in t_j), the function value $\text{F}(\mathbf{C};\hat{\boldsymbol{\Sigma}}_j)$ decreases. The ratio $2t_j/(n-1)$ in the formula for the ECVI makes sure that adding more parameters will not automatically lead to a lower value for the ECVI_j. The ratio $2t_j/(n-1)$ is therefore called the penalty function of the ECVI_j. When looking for the number of factors to retain, the number of factors with the lowest ECVI_j can be taken (if a minimum is reached for a number of factors). Note that it is possible that the value of the ECVI_j keeps on decreasing with each factor added. Browne and Cudeck (1992, p. 247) also noted that by inspection of the 90% confidence intervals (not given here) it is often not clear what the exact number of factors should be. According to Browne and Cudeck (1992, p. 247), the purpose of the ECVI, therefore, is merely "to give a rough indication of the number of factors that can subsequently be modified by subjective judgment." Furthermore, the value of the ECVI_j for a certain model or number of factors in itself is not important, because it can only be interpreted in comparison with values for the ECVI_j, obtained for other models.

In comparing the ECVI_j with Akaike's Information Criterion (AIC, not presented), it appears that, when ML–methods are used, there exists a linear relationship between the two (a rescaled AIC equals the ECVI_j, Browne & Cudeck,, 1992, p. 243; Jöreskog & Sörbom, 1993), so both indices lead to the same order of models. Therefore, we did not use the AIC in addition to the ECVI_j in the present study. Next, the motivation for selecting the ECVI_j will be given.

Oud, Haughton and Jansen (1996) conducted a simulation study,

evaluating fifteen criteria for model selection in LISREL, among which the fit index AIC (= $ECVI_j$). They found that in the presence of overspecified, underspecified and correctly specified analysis models, the fit index AIC showed one of the highest frequencies of selection of the correctly specified model. For this reason, the fit index $ECVI_j$ was used in the present study.

For the multiple group case, the $ECVI_j$ is calculated by LISREL as

$$ECVI_j = \hat{F}_j + 2t_j/(N-p). \tag{4.17}$$

In the present study, the number of factors for which the $ECVI_j$ measure reached its minimum value was taken as the number of factors indicated by the $ECVI_j$ measure.

4.3.5 Check of fit indices for the model with the correct dimension

In addition to the use of the fit indices as dimension indicators, as presented above, it was inspected whether the model with the correct dimension satisfied the requirement of acceptable fit. For this, it was inspected whether the p–value of the chi–square measure was larger than .05, indicating reasonable fit, and it was inspected whether the p–value for the RMSEA measure was smaller than .05, which also indicates reasonable fit.

4.3.6 The use of the fit indices as dimension indicators in the present study

In the present study, the dimension indicators were used in a comparative way (compare Section 4.2.5). It was inspected whether the correct dimension is indicated, or over– or underestimation of the correct dimension occurred.

As a test of the conjunctive use of the fit indices, a combination of the fit indices was chosen, named 'COMBI'. The dimension indicated by this measure is the dimension indicated by most of the three dimension indicators, as specified in previous sections. In the case that three different dimensions are indicated by the three dimension indicators,

according to this rule, the lowest dimension is preferred.

4.3.7 Checks on the solutions of LISREL 8

Even when given an appropriate headstart (see Sections 2.3.2 and 2.6.2), the solution found with the program LISREL–ML may contain Heywood cases, show nonconvergence and/or have otherwise improper solutions (such as non–Gramian matrices Φ_k), which make the solution suspicious, because parameters should not have values that are mathematically absurd, such as negative variances.

For the solutions, found with the LISREL program, it was inspected whether any irregularities occurred. For this, the warnings and error messages in the LISREL output were used. The irregularities can be classified into two main categories:

1.) The solution did not converge. In this case no final solution was found with the LISREL program. Nonconvergence could be attributed to: a.) The program was unable to start iterations because the fitted covariance matrix was not positive definite. In this case the Two Stage Least Squares analysis (the default start) produced nonadmissible starting values; b.) The solution, when checked after the first 20 iterations, was found to be nonadmissible. The LISREL program has a built–in check on admissibility of the solution, that checks whether the matrix Λ has full column rank and no rows of only zeros, and whether the matrices Φ and Θ are positive definite; c.) The program run reached the maximum number of iterations allowed by the program; and d.) Serious numerical problems were encountered during the program run. When this happens, the program is unable to continue iterations;
2.) The converged solution was nonadmissible. The matrix theta–delta (Θ_δ) and/or the matrix Φ, arrived at when the program had converged, were not positive definite.

In the present study, all irregular solutions were left out of the analysis. For the dimension indication, however, it was inspected what the effect was of including solutions that had converged but contained

nonadmissible values, because it was suspected that in the case that either of the matrices Φ or Θ_δ was not positive definite, this could be due to a very small divergence of parameter values beyond acceptable values.

4.4 SUCCESS CRITERIA

4.4.1 The Recovery Rate (RR)

To assess to what degree the columns of a matrix **A**, found with one of the methods of analysis, lead to the same interpretation as the columns of the appropriate true matrix **B**, Tucker's congruence measure has been adopted (see the definition of the CM in Section 3.4.1). Tucker's congruence measure is also used in the definition of an important overall success criterion called the Recovery Rate (RR). The RR gives the percentage of factors retrieved by each method. For this, for each method the choice has to be made which matrix is to be used for interpreting the components or factors, that is, the pattern matrix or the structure matrix. For the methods SCA-W, SCA-P and SIFASP-ML, the pattern matrix was the matrix for which recovery of the components or factors was measured, and for the method SCA-S this was the structure matrix. The choice for these matrices comes from the fact that these are the matrices made simple in the rotation procedure for each method. However, in Experiment 1, for each method, the RR was also calculated for the other matrix (pattern or structure), and in Experiment 3, with SCA-S also the pattern matrix was used. Now the calculation of the RR will be presented.

First, the columns of the pattern matrices found (for SCA-W, SCA-P and SIFASP-ML) or the structure matrices found (for SCA-S) were permuted and reflected to find the maximum CM (as described in Section 3.4.1) for the p matrices, found by a method, and the p true matrices. In the p matrices, compared with p true matrices, the same ordering of columns was maintained. Next, from the rows of congruences, thus found for each matrix, the number of values above .85 (the threshold obtained by Haven &

Ten Berge, see Section 3.4.1) was calculated. From the number of times a component or factor is retrieved (had a congruence value with the true factor above .85), the Recovery Rate (RR) is calculated as

$$RR = 100\% \times \frac{\text{the number of retrieved factors}}{\text{the total number of factors}}.$$

4.4.2 The Difference in Factor Correlations (DFC)

In all experiments, the correlations between the true factors were specified in the data construction. In order to investigate whether the methods of analysis, used in the present study, retrieved these correlations correctly or showed a bias, all off-diagonal elements of the true matrix Φ used were subtracted from the values of the off-diagonal elements of the matrix Φ, found with each method. For each method, the matrix Φ was calculated from the weights matrix **W** (with columns permuted and/or reflected to maximize the CM as described in Section 4.4.1) and the correlation matrix **R**. The mean of the absolute values of these differences was called the Difference in Factor Correlations measure (DFC).

4.4.3 Additional criteria for the SCA-methods: Amounts of explained variance.

As explained in Section 3.3, the simulated data used in the present study can be divided into two main categories: Data coming from one population (one-population data) and data coming from two or more (different) populations (two-or-more-populations data). In the present study, we are interested in whether or not it is possible to find an indication in the results from the SCA-methods to distinguish the data from these two categories. For this, a discriminant analysis was done in Experiment 1, with as predictors the amount of variance explained by each of the SCA-methods and the method PCA-sep, and the variables sample size, number of variables and number of factors. The number of misclassifications will be presented, together with the discriminant

function.

Imagine a situation where the strengths of factors differ across samples, while the interpretation of these factors remains the same, for instance, in a longitudinal study where some factors grow weaker over time and others grow stronger. This situation illustrates that the interpretation of a factor (discussed in Section 4.4.1) and the amount of variance it explains are entirely different and unrelated concepts, which require separate treatment. Factors can be correctly recovered, while, for instance, the strengths of weak factors are gravely overestimated, compared to strong factors. To investigate this, in Experiment 3 (Chapter 7), besides looking at the RR of factors, also the *relative* strength of components found with the SCA-methods was compared with the *relative* strength of the true factors. A measure for assessing this is discussed in Section 7.3.2.

4.5 PILOT EXPERIMENT TO DETERMINE THE PREFERABLE RANGE OF NOISE TO SIGNAL RATIOS FOR CONSTRUCTED DATA, USING THE SCA-METHODS

One of the most important preliminary questions of this study was: "At what noise-to-signal ratio do the methods considered start to yield unrealistic results?" A pilot study was conducted to answer this question. This pilot study is presented in the present section.

4.5.1 Data

In all data sets there were two groups (p=2), 12 variables (=m), 2 or 4 factors (=q), and sample sizes of n=50, n=100 or n=150. The noise-to-signal ratio's (*nsr's*) were chosen as .5, 1, 1.5, 2, 2.5 and 3. These were only varied over data sets, not over the two groups within a data set nor over factors, so we had one-population data. Each *nsr* value corresponded with a true loading (rounded to two decimals) in the pattern matrices and with a reliability (i.e., squared loading) of the

constructed variables (also rounded to two decimals) as follows (see Section 4.1.1):

nsr value	0.5	1.0	1.5	2.0	2.5	3.0
true loading	.89	.71	.55	.45	.37	.32
reliability	.80	.50	.31	.20	.14	.10

In the populations we have factors with non–overlapping clusters of variables. The true correlation matrices $\mathbf{\Phi}_{tk}$ with correlations between the factors were taken as the identity matrix. So the true matrices, for 2 and 4 factors, were $\mathbf{P}_a$, $\mathbf{P}_b$ and $\mathbf{P}_c$, given in Appendix A, and $\mathbf{P}_x$, $\mathbf{P}_y$, $\mathbf{P}_z$, defined analogously to $\mathbf{P}_a$, $\mathbf{P}_b$ and $\mathbf{P}_c$, but with nonzero loadings of .71, .45 and .32, respectively. In total, there were (2×3×6=) 36 different conditions.

In each condition, the simulation of the data was replicated five times, using different starting values for the random generation of data. The presented results of the analyses with the methods under study are the mean values for these five replications. The methods of analysis used were SCA–W, SCA–P, SCA–S and (only for comparison) PCA–sep.

4.5.2 Measures

There were two measures used to answer the question asked at the beginning of this section. Firstly, it was inspected how often the correct dimension was indicated by each of the SCA–methods in the different conditions. Because this experiment was conducted in an early stage of the study, only the measure QA (which was considered the most promising measure at the time) was used. In all conditions, for each data set the solution was calculated for 2, 3 and 4 components drawn, and from these values three quotients of additional values were calculated. Using only these three values of the QA–measure had the shortcoming that, besides indicating the correct number of components, it could only overestimate the number of components when q=2 and only underestimate the number of components when q=4. However, it still served the purpose of

the present experiment, in that we were able to see at what level the results started getting unrealistic.

Secondly, it was inspected how often the components received the correct interpretation in the different conditions: "At what error level do the SCA-methods fail to retrieve the true factors?" To answer this question, the congruence measure was adopted (instead of the RR measure presented in Section 4.4.1 just a mean congruence value was used), because this experiment was conducted in an early stage of the study, at which time the congruence measure was considered the ideal measure for recovery of factors. For each condition it was inspected, for each SCA-method separately, whether the CM for the pattern and the structure matrices for the solution with the correct dimension was above or below .85.

4.5.3 Results

The first question to be answered is: "For what noise-to-signal ratio does the measure QA fail to indicate the number of true factors for the SCA-methods (and PCA)?" Here, the measure QA was used because it was, at that point in time, deemed the most reasonable and feasible dimension indicator. The results for the QA measure are given in Table 4.1, aggregated over the four methods of analysis, resulting in a maximum of 20 possible failures for each cell (5 replications for each of the 4 methods).

From Table 4.1 it can be seen that, when large noise-to-signal ratios are used (nsr=3), the number of components, indicated by the measure QA, is often overestimated for q=2, n=50 and often underestimated for q=4, for all group sizes. On the basis of this result a choice of nsr=3 for the noise-to-signal ratio (giving true pattern matrix $\mathbf{P}_z$) seems too high.

Next, the CM of the pattern and the structure matrices, found with the SCA-methods, and the true pattern and true structure matrices, respectively, for the combinations of true pattern matrices, were

Table 4.1: *The times the QA measure indicated an incorrect number of components, summed over the different methods of analysis*[*)]

		noise–to–signal ratio					
n	q	.5	1.0	1.5	2.0	2.5	3.0
50	2	0	0	0	1	6	9
100	2	0	0	0	0	0	3
150	2	0	0	0	0	0	1
50	4	0	0	10	17	17	15
100	4	0	0	0	4	17	18
150	4	0	0	0	5	14	16

*) *read table* 1 *as follows*: '0' *means the measure QA indicated the correct number of components in all* 20 *replications*; '4' *means the measure QA overestimated* (*when* $q=2$) *or underestimated* (*when* $q=4$) *the number of components* 4 *times.*

inspected. In Table 4.2 it is indicated by a "+" that the mean CM was over .85 and by a "–" that it was smaller than .85. Because the three SCA–methods all behaved approximately the same on this measure, no distinction was made between the SCA–methods. Also, there was approximately no difference in the results found for the pattern matrix and the structure matrix. Therefore, the results in Table 4.2 go for both the structure and the pattern matrix. The only exception to this was SCA–P, that attained a mean value of the CM just above .85 for the pattern matrix in the condition n=50, q=2 and nsr=3.0. Both other methods attained a CM for the pattern and the structure matrices smaller than .85, as did SCA–P with the structure matrix, explaining the minus in this condition. In Table 4.2, the conditions for which the measure CM was above and below .85 are given.

From Table 4.2, it can be seen that high noise–to–signal ratios (nsr=3), small sample sizes (n=50) and a high number of factors (q=4 as opposed to q=2), facilitate the failure of the SCA–methods of retrieving the true factors.

Table 4.2: *The conditions for which the mean value of the CM was above (+) or below (−) .85.*

		noise-to-signal ratio					
n	*q*	.5	1.0	1.5	2.0	2.5	3.0
50	2	+	+	+	+	+	−
100	2	+	+	+	+	+	+
150	2	+	+	+	+	+	+
50	4	+	+	+	−	−	−
100	4	+	+	+	+	+	−
150	4	+	+	+	+	+	−

4.5.4 Conclusion about meaningful noise levels

From the results found with the two different measures, it was concluded that, in order to span the complete range of useful and meaningful noise-to-signal ratios, the values 0.5, 1.5 and 2.5 would be used in the bulk of this study, corresponding to true loadings of .89, .55 and .37, respectively. This explains the decision (Section 4.1.1) to settle for noise-to-signal ratios of .5, 1.5 and 2.5.

4.6 PILOT EXPERIMENT TO DETERMINE THE PREFERABLE CONVERGENCE VALUE FOR THE ITERATIVE PROCEDURES IN THE METHODS SCA-W AND SCA-S

In order to assess what value to use for the convergence criterion in the iterative procedures in the methods SCA-W and SCA-S, a pilot experiment was done, using three different values for the convergence criterion. The minimum value of the convergence criterion was sought, for which the iterative procedures arrived at the global optimum most of the time (when multiple starts were used) within a reasonable amount of time.

4.6.1 Data

In all data sets there were four groups (p=4), 18 variables (=m), 3 factors (=q), and sample sizes of n=50, n=100 or n=150. Only one noise–to–signal ratio was used: nsr=1, resulting in a true loading of .71 in the true pattern matrix, so we have one–population data. The true correlation matrix $\mathbf{\Phi}_{tk}$ with correlations between the factors was taken as the identity matrix for all groups. So the true pattern matrix is given as

$$\mathbf{P}_t = .71 \times \begin{pmatrix} \mathbf{1}_6 & 0_6 & 0_6 \\ 0_6 & \mathbf{1}_6 & 0_6 \\ 0_6 & 0_6 & \mathbf{1}_6 \end{pmatrix},$$

where $\mathbf{1}_6$ is a 6×1 vector of ones and $\mathbf{0}_6$ a 6×1 vector of zeros.

The values, chosen for the convergence criterion were .0001, .00001 and .000001. In total there were (3×3=) nine conditions. In each condition the simulation of the data was replicated three times, using different starting values for the random generation of data. For each of the three replications, both SCA–W and SCA–S were run eleven times, one time starting the iterations with a rationally derived weights matrix or structure matrix for the methods SCA–W and SCA–S, respectively, and ten times starting iterations with a randomly generated matrix. In every condition the same ten random starts were used. It was assumed that the best of the eleven solutions was the global optimum. Because in this pilot experiment, it was only of interest what the effect of convergence values used was in finding the global optimum, the solutions, found with the methods SCA–W and SCA–S, were not rotated.

4.6.2 Measures

In all conditions, the two, three and four components solutions were calculated, each solution for each method consisting of eleven function values (i.e., explained variances). The highest of these eleven values was taken as the global optimum, and the other ten values were compared with this global optimum. When the difference of one of these ten values

with the global optimum was smaller than .01, the iterative procedure was considered to have arrived at the global optimum for that particular instance.

4.6.3 Results

The main question was: "For what value of the convergence criterion is the global optimum reached for most of the starts used in the iterative procedure." When summed over sample size, method and (rational and random) starts, giving a total number of starts of (2×11×3×3×3=) 594, the percentage of times the global optimum is reached by the iterative procedures (from now on called success rates) for the three convergence values .0001, .00001 and .000001, were 55%, 83% and 97%, respectively. The random starts appear to be a little more sensitive to the choice of convergence value than the rational starts, with success rates of 78%, 91% and 93% for the rational starts for the values .0001, .00001 and .000001, respectively, and success rates of 52%, 83% and 97%, for the random starts, respectively.

Sample size has almost no effect. The number of components drawn, however, does have a large effect on whether the global optimum is reached. When the correct number of components is drawn (three), the global optimum is reached in all cases, for all three values of the convergence criterion. However, the convergence criterion does affect the success of the iterative procedures when two or four components are drawn.

4.6.4 Conclusion about meaningful convergence criteria

On the basis of the results, it was concluded a convergence value of .00001 was most appropriate for the present study (see Section 4.1.1, final paragraph) . The value .0001 showed a too high a chance to not reach the global optimum, and using the value .000001, although giving higher success rates than when using the value .00001, appeared to be very time consuming, especially for the method SCA–S. It can also be seen

that using eleven starts is wasteful, given that more than 80% of the starts tend to lead to the global optimum. Hence the decision (Section 4.1.1) to use five random and one rational start seems to be adequate.