PSYCHOMETRIKA—VOL. 78, NO. 4, 648–668
OCTOBER 2013
DOI: 10.1007/S11336-013-9318-4

MODELING DIFFERENCES IN THE DIMENSIONALITY OF MULTIBLOCK DATA BY MEANS OF CLUSTERWISE SIMULTANEOUS COMPONENT ANALYSIS

KIM DE ROOVER AND EVA CEULEMANS

KU LEUVEN

MARIEKE E. TIMMERMAN

UNIVERSITY OF GRONINGEN

JOHN B. NEZLEK

COLLEGE OF WILLIAM & MARY

UNIVERSITY OF SOCIAL SCIENCES AND HUMANITIES, FACULTY IN POZNAŃ

PATRICK ONGHENA

KU LEUVEN

Given multivariate multiblock data (e.g., subjects nested in groups are measured on multiple variables), one may be interested in the nature and number of dimensions that underlie the variables, and in differences in dimensional structure across data blocks. To this end, clusterwise simultaneous component analysis (SCA) was proposed which simultaneously clusters blocks with a similar structure and performs an SCA per cluster. However, the number of components was restricted to be the same across clusters, which is often unrealistic. In this paper, this restriction is removed. The resulting challenges with respect to model estimation and selection are resolved.

Key words: multigroup data, multilevel data, principal component analysis, simultaneous component analysis, clustering, dimensionality.

1. Introduction

When researchers measure the same variables for a number of subjects who are nested in groups (e.g., students of different classes, inhabitants of different countries), the obtained data have a hierarchical structure. The same holds when a set of variables (e.g., a number of emotions) is measured multiple times for some subjects, where the measurement occasions are nested within subjects. Such hierarchical data can also be called "multiblock" data as each group in the first example and each subject in the second example corresponds to a "block" of data.

By their nature, multiblock data raise questions about the underlying dimensional structure of each data block and the extent to which this structure differs across the data blocks. For instance, the well-known Big Five model (Goldberg, 1990) in personality psychology states that individual differences in dispositional characteristics can be described using five dimensions or "factors." Although there is considerable support for this five factor model, some argue that this may not hold for all cultures in that the number and/or the nature of the dimensions may differ across cultures (e.g., Diaz-Loving, 1998). As another example, there appear to be differences

Requests for reprints should be sent to Kim De Roover, Methodology of Educational Sciences Research Unit, Faculty of Psychology and Educational Sciences, KU Leuven, Andreas Vesaliusstraat 2, 3000 Leuven, Belgium. E-mail: Kim.DeRoover@ppw.kuleuven.be

among individuals in how strongly emotions covary across time, with some people showing stronger covariation than others (Barrett, 1998).

To explore the dimensional structure of multiblock data, component analysis may be used. In component analysis, the observed variables are reduced to a few components that summarize the data as adequately as possible. For instance, analyzing the scores of a sample of persons on a set of personality-related variables with standard principal component analysis (PCA; Jolliffe, 1986; Meredith & Millsap, 1985; Pearson, 1901), one obtains a component score matrix **F**, containing the scores of the persons on the components, and a loading matrix **B**, indicating the extent to which the scores on the personality variables can be summarized by the respective components. In other words, the loadings express the dimensional structure of the variables.

For analyzing multiblock data, De Roover and colleagues recently presented Clusterwise SCA-ECP (where "SCA-ECP" refers to simultaneous component analysis with Equal Cross-Product constraints on the data blocks within a cluster; De Roover, Ceulemans, & Timmerman, 2012a; De Roover, Ceulemans, Timmerman, Vansteelandt, Stouten & Onghena, 2012c), which captures the most important structural differences and similarities between the data blocks by clustering the data blocks according to their dimensional structure. Specifically, data blocks that are assigned to the same cluster are modeled using the same loadings, while the loadings across clusters may, and usually will, differ. Clusterwise SCA-ECP is a generic modeling strategy, as it includes standard (i.e., non-clusterwise) SCA-ECP (Kiers, 1990; Kiers & ten Berge, 1994; Timmerman & Kiers, 2003) and PCA on each of the data blocks separately as special cases, that is, when the number of clusters is set to one or to the number of data blocks, respectively. The applications in De Roover et al. (2012c)—disentangling two groups of eating disordered patients and describing differences between a number of experimental conditions—illustrate that Clusterwise SCA is a very useful modeling technique that can help explain differences and similarities in dimensional structures.

Nonetheless, Clusterwise SCA-ECP has an important drawback in that the number of components is restricted to be equal across the clusters, which is often not realistic. For instance, when subjects report on their emotions at randomly selected measurement occasions, studies on emotional granularity (Barrett, 1998; Tugade, Fredrickson, & Barret, 2004) suggest that the number of underlying dimensions, and thus the number of components, can differ between subjects, with a higher number of components corresponding to a higher emotional granularity level. Therefore, in this paper, we present a more general version of Clusterwise SCA-ECP in which the number of components may vary over the clusters. This generalization is not as straightforward as it may seem, as serious challenges arise at the level of model estimation and model selection. These challenges are discussed and resolved in this paper. More specifically, we propose a model estimation procedure in which data blocks are only added to a cluster with a relatively high number of components when the increase in fit outweighs the increase in the number of model parameters to be estimated. Subsequently, to select among the huge set of possible solutions, which differ with respect to the number of clusters and the number of components per cluster, four model selection procedures are presented and evaluated.

The remainder of the paper consists of four sections. Throughout these sections, we use a real psychological data set for illustration purposes. In the Model section, we recapitulate the required data structure and the recommended preprocessing and discuss the generalized Clusterwise SCA-ECP model. In the Model Estimation section, we present a model estimation procedure and we evaluate it in a simulation study. In the Model Selection section, we present several procedures for model selection, of which the performance is compared in a simulation study. Next, we compare the performance of the original Clusterwise SCA-ECP method with that of the generalized version in terms of cluster recovery. Finally, in the Discussion, we end with some directions for future research.

2. Model

2.1. Data Structure and Preprocessing

Multiblock data consist of I data blocks $\mathbf{X}_i\,(N_i \times J)$ containing scores on J variables, where the number of observations N_i $(i = 1, \ldots, I)$ may differ across data blocks. The vertical concatenation of the I data blocks gives an $N \times J$ data matrix $\mathbf{X}$, where $N = \sum_{i=1}^{I} N_i$. To apply Clusterwise SCA-ECP, it is required that each N_i exceeds the number of components to be fitted, and for the sake of stable model estimates, N_i is preferably larger than J for each data block.

Note that "multiblock data" is often used in a more general way in that the data blocks under study may be coupled in the row mode (observations, e.g., persons, situations, measurement occasions, etc.) or the column mode (variables, e.g., items, symptoms, etc.) (see, e.g., Van Mechelen & Smilde, 2010). In this paper, we only consider data that are coupled in the column mode, as Clusterwise SCA-ECP is developed for multiple data blocks containing the same variables. In this case, a further distinction can be made between "multigroup" and "multilevel" data, based on whether the data blocks are considered fixed or random, respectively (Timmerman, Kiers, Smilde, Ceulemans, & Stouten, 2009). For instance, if the data blocks correspond to subjects, the data blocks are fixed when one is only interested in the subjects in the study, and random when one wants to generalize the conclusions toward a larger population of subjects. As Clusterwise SCA-ECP is a deterministic method (i.e., no distributional assumptions are made with respect to the component scores), it is applicable to both multigroup and multilevel data.

Clusterwise SCA-ECP was designed for modeling the dimensional or correlational structure of the data blocks. Therefore, to make sure that between-block differences in correlational structure are not confounded with between-block differences in means or variances of the variables, we standardize each variable per data block $\mathbf{X}_i$.

We illustrate the application of this technique using a real data set consisting of daily measures of self-consciousness. Initially, Feningstein, Scheier, and Buss (1975) distinguished two kinds of self-consciousness: private and public. Public self-consciousness is being conscious of one's appearance to others, whereas private self-consciousness is the process of examining one's thoughts, feelings, and behaviors.

Although this distinction was useful, a body of research accumulated that suggested the existence of what came to be called the "self-focused (or self-absorption) paradox." Sometimes, greater private self-consciousness was found to be associated with adaptive behaviors or outcomes, whereas in other instances it was associated with maladaptive outcomes (Trapnell & Campbell, 1999). Partially in an attempt to resolve this paradox, Trapnell and Campbell posited and provided evidence in support of a model that broke private self-consciousness into two, unrelated components, rumination and reflection, which Trapnell and Campbell (1999) described as "neurotic" and "intellectual self-attentiveness," respectively. Implicit in this distinction is that rumination (neurotic) self-focused thinking is the more maladaptive of the pair, whereas reflection (intellectual self-attentiveness) is the more adaptive.

In this paper, we focus on the within-subject variability of public self-consciousness, rumination, and reflection, using data collected by Nezlek (2005). The 204 participants in the study provided measures of public self-consciousness, rumination, and reflection each day for 10 to 20 days. Specifically, participants logged onto a website and answered, among others, nine questions, three for each of the three types of self-consciousness: public self-consciousness ("How much today did you . . ." (Publ1) "think about what other people thought of you," (Publ2) "worry about making a good impression," and (Publ3) "think about your physical appearance [clothes, grooming, etc.]?"), rumination ("Today . . ." (Rum1) "how much time did you spend "ruminating" or dwelling on things that happened to you for a long time afterward?" (Rum2) "I played back over in my mind how I acted in a past situation." (Rum3) "how much time did you spend rethinking things that are over and done with?"), and reflection ((Refl1) "How much today did you think about the nature and meaning of things?" (Refl2) "How much did you care today for introspective or self-reflective thinking?" (Refl3) "Today, how much did you think about your attitudes and feelings?"). Responses were made on 7-point scales with endpoints of "not at all" to "very much." For a more detailed discussion of the methods used to collect these data, see Nezlek (2005, 2012). Across all participants, there were 2,796 valid observations (days of data).

With our Clusterwise SCA-ECP analysis, we intended to answer the following questions: (a) Are the items that are intended to measure the same construct (rumination, reflection, and public self-consciousness) strongly correlated? (b) Are there any within-subject relationships between these three constructs; for example, are the items measuring rumination and reflection strongly correlated implying that they reflect one underlying dimension which can be labeled private self-consciousness? (c) Do these within-subject relationships differ across subjects; for example, does it hold that rumination and reflection are strongly correlated for some subjects, whereas they are not for others?

2.2. *Clusterwise SCA-ECP Model*

Clusterwise SCA-ECP models differences and similarities in dimensional structure by simultaneously clustering the data blocks and fitting an SCA-ECP model per cluster. Stated differently, data blocks with a similar dimensional structure will be assigned to the same cluster and thus modeled by the same loading matrix, while data blocks with a different structure will be assigned to different clusters.

For ease of presentation, we will first recapitulate SCA-ECP (Kiers & ten Berge, 1994; Timmerman & Kiers, 2003), whose model equation reads as follows:

$$\mathbf{X}_i = \mathbf{F}_i \mathbf{B}' + \mathbf{E}_i, \tag{1}$$

where $\mathbf{F}_i\,(N_i \times Q)$ and $\mathbf{E}_i\,(N_i \times J)$ denote the component score matrix and the residual matrix of the ith data block, and $\mathbf{B}$ $(J \times Q)$ denotes the loading matrix. The $\mathbf{F}_i$ matrices are constrained to have Equal Cross-Products (ECP), that is, $N_i^{-1}\mathbf{F}_i'\mathbf{F}_i = \mathbf{\Phi}$. To partly identify the solution (without loss of fit), the variances of the component scores per block (i.e., of the columns of $\mathbf{F}_i$) are fixed at 1. Because $\mathbf{X}_i$ is standardized, $\mathbf{F}_i$ is centered, implying that the matrix $\mathbf{\Phi}$ is the correlation matrix of the component scores.

It is noteworthy that the components of an SCA-ECP solution have rotational freedom. Specifically, to obtain solutions that are easier to interpret, the loading matrix $\mathbf{B}$ can be multiplied by any rotation matrix, provided that such a transformation is compensated for in the component score matrices $\mathbf{F}_i$ $(i = 1, \dots, I)$ (for more details, see De Roover et al., 2012c).

The Clusterwise SCA-ECP model is expressed by the following equation:

$$\mathbf{X}_i = \sum_{k=1}^{K} p_{ik} \mathbf{F}_i^{(k)} \mathbf{B}^{(k)\prime} + \mathbf{E}_i, \tag{2}$$

where K is the number of clusters, p_{ik} is an entry of the binary partition matrix $\mathbf{P}$ $(I \times K)$, which equals 1 when data block i is assigned to cluster k and 0 otherwise, $\mathbf{F}_i^{(k)}$ $(N_i \times Q^{(k)})$ denotes the component score matrix of data block i when assigned to cluster k, $Q^{(k)}$ is the number of components for cluster k, and $\mathbf{B}^{(k)}$ $(J \times Q^{(k)})$ denotes the loading matrix of cluster k. The components have rotational freedom per cluster. Note that De Roover et al. (2012c) imposed the restriction that $Q^{(k)} = Q$ for all clusters. For an overview of the relations between Clusterwise SCA-ECP and existing models, we refer the reader to De Roover et al. (2012c).

To illustrate the model, we will present the Clusterwise SCA-ECP solution with two clusters for the self-consciousness data, using two components for the first cluster (105 subjects) and one

TABLE 1.
Normalized Varimax rotated loadings for the two clusters of the self-consciousness data. Loadings greater than ±0.30 are highlighted in bold face. 'Rum' is rumination, 'Refl' is reflection and 'Publ' is public self-consciousness.

	Cluster 1		Cluster 2
	Private self-consciousness	Public self-consciousness	Self-consciousness
Rum1	**0.76**	0.01	**0.65**
Rum2	**0.78**	0.09	**0.73**
Rum3	**0.76**	0.07	**0.71**
Refl1	**0.69**	0.06	**0.73**
Refl2	**0.58**	−0.01	**0.56**
Refl3	**0.68**	0.05	**0.69**
Publ1	0.19	**0.79**	**0.73**
Publ2	0.05	**0.81**	**0.67**
Publ3	−0.06	**0.74**	**0.57**

component for the second cluster (99 subjects). In the fourth section, we will discuss why we selected this solution.

The cluster loading matrices of this solution, rotated orthogonally according to the Normalized Varimax criterion (Kaiser, 1958), are presented in Table 1. In the first cluster, the two components correspond to private self-consciousness and public self-consciousness, given the high loadings of the rumination and reflection items on the first component and the high loadings of the public self-consciousness items on the second. In the second cluster, all items load high on the obtained component; therefore, it is labeled "self-consciousness." We conclude that for all subjects, more rumination on a certain day co-occurs with more self-reflection on that same day as these constructs are not recovered as separate components. In addition, the subjects differ with respect to the relation between daily private and public self-consciousness: Whereas for the subjects in Cluster 2, a higher level of daily private self-consciousness is also associated with a higher level of daily public self-consciousness; these two constructs vary independently for the subjects in Cluster 1.

3. Model Estimation

In this section, we will first describe the original Clusterwise SCA-ECP algorithm (De Roover et al., 2012c), that is, for estimating a model where the number of components is the same in each cluster. Second, we discuss why this algorithm is not appropriate for fitting a Clusterwise SCA-ECP model in which the number of components may differ across clusters, and how it may be adapted. Third, the performance of the original and adapted algorithm is evaluated and compared in a simulation study.

3.1. Procedure

3.1.1. Clusterwise SCA-ECP with $Q^{(k)} = Q$ for All Clusters To fit Clusterwise SCA-ECP solutions with an equal number of components Q across the clusters, De Roover et al. (2012c) propose to minimize the following objective function, given specific values of K and Q:

$$SSE = \sum_{i=1}^{I} \sum_{k=1}^{K} p_{ik} \left\| \mathbf{X}_i - \mathbf{F}_i^{(k)} \mathbf{B}^{(k)\prime} \right\|^2. \tag{3}$$

Note that because $\mathbf{X}_i$ is centered, minimizing Equation (3) is equivalent to maximizing the percentage of variance accounted for (VAF %), which can be calculated as follows:

$$\text{VAF}\,\% = \frac{\|\mathbf{X}\|^2 - SSE}{\|\mathbf{X}\|^2} \times 100. \tag{4}$$

To this end, an alternating least squares (ALS) algorithm[1] is used, consisting of the following steps:

1. *Randomly initialize partition matrix* $\mathbf{P}$: The partition matrix $\mathbf{P}$ $(I \times K)$ contains the binary cluster memberships p_{ik} (Equation (2)). Randomly assign the I data blocks to one of the K clusters, where each cluster has an equal probability of being assigned to and empty clusters are not allowed.
2. *Update the SCA-ECP model for each cluster*: Estimate the $\mathbf{F}_i^{(k)}$ and $\mathbf{B}^{(k)}$ matrices for each cluster by performing a rationally started SCA-ECP analysis (Timmerman & Kiers, 2003) on the $\mathbf{X}_i$ data blocks assigned to the cluster. Specifically, the loading matrix $\mathbf{B}^{(k)}$ is rationally initialized, based on the singular value decomposition (svd) of the vertical concatenation of the data blocks within cluster k, denoted by $\mathbf{X}^{(k)}$. Next, $\mathbf{F}^{(k)}$ and $\mathbf{B}^{(k)}$ are iteratively reestimated, where $\mathbf{F}^{(k)}$ is the vertical concatenation of the $\mathbf{F}_i^{(k)}$ matrices for the data blocks assigned to cluster k. $\mathbf{F}^{(k)}$ is (re-)estimated by performing an svd for each data block $\mathbf{X}_i$ that belongs to cluster k: $\mathbf{X}_i\mathbf{B}^{(k)}$ is decomposed into $\mathbf{U}_i$, $\mathbf{S}_i$, and $\mathbf{V}_i$ with $\mathbf{X}_i\mathbf{B}^{(k)} = \mathbf{U}_i\mathbf{S}_i\mathbf{V}_i'$ and a least squares estimate of $\mathbf{F}_i^{(k)}$ is then given by $\mathbf{F}_i^{(k)} = \sqrt{N_i}\mathbf{U}_i\mathbf{V}_i'$ (ten Berge, 1993). $\mathbf{B}^{(k)}$ is updated by $\mathbf{B}^{(k)} = ((\mathbf{F}^{(k)\prime}\mathbf{F}^{(k)})^{-1}\mathbf{F}^{(k)\prime}\mathbf{X}^{(k)})'$.
3. *Update the partition matrix* $\mathbf{P}$: Each cluster membership is updated by quantifying the extent to which data block i fits in each cluster using a block- and cluster-specific partition criterion:

$$SSE_i^{(k)} = \left\| \mathbf{X}_i - \mathbf{F}_i^{(k)}\mathbf{B}^{(k)\prime} \right\|^2 \tag{5}$$

   and assigning it to the cluster k for which $SSE_i^{(k)}$ is minimal. To this end, $\mathbf{F}_i^{(k)}$ in Equation (5) is computed by means of the svd-step described in Step 2. When one or more clusters are empty after this procedure, the data blocks with the lowest $SSE_i^{(k)}$-values are moved to the empty clusters.
4. Steps 2 and 3 are repeated until the partition $\mathbf{P}$ no longer changes.

Assuming that the convergence proofs for k-means (e.g., Selim & Ismail, 1984) can be generalized to the Clusterwise SCA-ECP problem, convergence of the above procedure—which we will call ALS$_{SSE}$—is guaranteed when the optimal partition is unique. The optimal partition is not unique when the correlation structure underlying the different clusters is identical (e.g., because the imposed number of clusters is too high). In empirical practice this will almost never occur as the correlation structure of a data block is always partly driven by random error. But, as with all ALS algorithms, the algorithm may converge to a local minimum. To increase the probability of obtaining the partition that corresponds to the global minimum, it is advised to use a number of different random starts (e.g., 25) and retain the best-fitting solution (i.e., with the lowest SSE) as the final solution.

3.1.2. Clusterwise SCA-ECP with $Q^{(k)}$ Varying Across the Clusters

It might be tempting to also use the ALS$_{SSE}$ approach described above to fit Clusterwise SCA-ECP solutions with the number of components $Q^{(k)}$ varying across clusters. However, in the case where $Q^{(k)}$ varies,

[1] This algorithm is implemented in an easy-to-use software program that can be downloaded at http://ppw.kuleuven.be/okp/software/MBCA/ (De Roover et al., 2012a).

using ALS$_{SSE}$ may imply that the majority of the data blocks are assigned to the cluster(s) with the highest $Q^{(k)}$-value(s), since such solutions happen to have the lowest SSE-value. Specifically, this phenomenon may occur when the clusters are relatively difficult to distinguish in terms of their correlational structure, which can be due to a high congruence between the cluster loading matrices and/or to a high amount of error variance. For instance, when we use the ALS$_{SSE}$ procedure to estimate a Clusterwise SCA-ECP solution for the self-consciousness data with two clusters using one and two as the $Q^{(k)}$-values for the respective clusters, the vast majority of the subjects (i.e., 181 of the 204 subjects) are assigned to the cluster with two components.

To solve this problem, we propose to use an alternative objective function. Specifically, inspired by the penalty approach that is successfully used in, for instance, regression analysis and simultaneous component analysis to circumvent multicollinearity problems or to enforce sparse or simple structure models (Hoerl, 1962; Tibshirani, 1996; Van Deun, Wilderjans, van den Berg, Antoniadis, & Van Mechelen, 2011), we will add a penalty that is higher, respectively lower, when a data block is assigned to a cluster with a higher, respectively lower, number of components. This will be achieved by using the well-known Akaike information criterion (AIC; Akaike, 1974) as the objective function in the above ALS algorithm; this approach will be referred to as the ALS$_{AIC}$ approach.

The AIC reads as follows (Akaike, 1974):

$$AIC = -2\,\text{loglik}(\mathbf{X}|\mathbf{M}) + 2fp, \tag{6}$$

where loglik($\mathbf{X}|\mathbf{M}$) refers to the log-likelihood of data $\mathbf{X}$ given model $\mathbf{M}$ and fp denotes the number of free parameters to be estimated. To define loglik($\mathbf{X}|\mathbf{M}$), we need to stochastically extend the Clusterwise SCA-ECP model by assuming the residuals $e_{n_i j}$ to be independent and identically distributed as $e_{n_i j} \sim N(0, \sigma^2)$. In that case, it follows that

$$\text{loglik}(\mathbf{X}|\mathbf{M}) = \log\left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{NJ}{2}} \exp\left(-\frac{SSE}{2\sigma^2}\right)\right] = -\frac{NJ}{2}\log\left(2\pi\sigma^2\right) - \frac{1}{2\sigma^2}SSE. \tag{7}$$

Inserting $\hat{\sigma}^2 = \frac{SSE}{NJ}$ as a post-hoc estimator of the error variance σ^2 (Wilderjans, Ceulemans, Van Mechelen, & van den Berg, 2011) yields

$$\begin{aligned}\text{loglik}(\mathbf{X}|\mathbf{M}) &= -\frac{NJ}{2} - \frac{NJ}{2}\log(2\pi) + \frac{NJ}{2}\log(NJ) - \frac{NJ}{2}\log(SSE) \\ &= -\frac{NJ}{2}\left[1 + \log(2\pi) - \log(NJ) + \log(SSE)\right],\end{aligned} \tag{8}$$

where the first three terms are invariant across solutions and thus can be discarded when the AIC is minimized.

Following Ceulemans, Timmerman, and Kiers (2011), the number of free parameters fp of a Clusterwise SCA-ECP model can be calculated as follows:

$$\begin{aligned}fp &= I + \sum_{k=1}^{K} fp^{(k)} \\ &= I + \sum_{k=1}^{K}\left(N^{(k)}Q^{(k)} + JQ^{(k)} - \left(Q^{(k)}\right)^2 - \left(I^{(k)} - 1\right)Q^{(k)} - \left(I^{(k)} - 1\right)\frac{Q^{(k)}(Q^{(k)} - 1)}{2}\right),\end{aligned} \tag{9}$$

where I reflects the number of cluster memberships (when $K > 1$) and $fp^{(k)}$ is the number of free parameters within each cluster k, with the first two terms of $fp^{(k)}$ indicating the number of cluster-specific component scores (i.e., $N^{(k)} = \sum_{i=1}^{I} p_{ik} N_i$), and loadings. The other terms of $fp^{(k)}$ correct for the rotational freedom and the restrictions on the variances and the correlations of the component scores respectively, where $I^{(k)}$ indicates the number of data blocks in cluster k. When estimating a model given specific values of K and $Q^{(k)}$, the first, third, and fourth terms of *fp* are invariant, and thus can be discarded when minimizing the function. Therefore, the *AIC* objective function[2] boils down to:

$$AIC = NJ\log(SSE) + 2\sum_{k=1}^{K}\left(N^{(k)}Q^{(k)} - \left(I^{(k)}-1\right)Q^{(k)} - \left(I^{(k)}-1\right)\frac{Q^{(k)}(Q^{(k)}-1)}{2}\right). \tag{10}$$

The *AIC*-based partition criterion can be derived accordingly (see Appendix):

$$AIC_i^{(k)} = N_i J\log\left(SSE_i^{(k)}\right) + 2N_i Q^{(k)}, \tag{11}$$

by noting that, when updating p_{ik}, the number of free parameters is equal to the number of component scores to be estimated for data block i when assigned to cluster k. This partition criterion ensures that data blocks are only assigned to a cluster with more components when the corresponding increase in fit outweighs the increase in complexity (i.e., the number of component scores to be estimated). We will refer to the algorithm using the *AIC* objective function and its corresponding partition criterion as the ALS_{AIC} procedure.[3] This procedure consists of the same steps as described above for the $Q^{(k)} = Q$ case,[4] where the *AIC* objective function directly influences the estimation of the partition in Step 3 (through the *AIC*-based partition criterion) on the one hand and the choice of the best and thus final solution from the multistart procedure on the other hand. Note that in the case where $Q^{(k)} = Q$, the number of free parameters is not influenced by the cluster memberships of the data blocks, and thus the solutions retained by means of the ALS_{SSE} and ALS_{AIC} procedures are identical.

When we use the ALS_{AIC} procedure to estimate a Clusterwise SCA-ECP solution for the self-consciousness data with two clusters using one and two as the $Q^{(k)}$-values for the respective clusters, 105 of the 204 subjects are assigned to the cluster with two components and the remaining 99 subjects to the cluster with one component.

3.2. *Simulation Study to Compare the* ALS_{SSE} *and* ALS_{AIC} *Approaches*

3.2.1. Problem To evaluate and compare the performance of the ALS_{SSE} and ALS_{AIC} approaches, we performed a simulation study. In particular, we assessed the recovery of the clustering.[5] We hypothesize that ALS_{SSE} performs worse than ALS_{AIC}, in that ALS_{SSE} will tend to

[2] It was confirmed for the simulation study reported below that multiplying the second term of the loss function (and partition criterion) with two—like in the *AIC*—gives an optimal cluster recovery for 99.6 % of the simulated data sets, as opposed to using another factor. In particular, multiplying *fp* with $\log(N)$—like in the Bayesian information criterion (*BIC;* Schwarz, 1978)—appeared to lead to a too high penalty, in that too few data blocks were assigned to the higher-dimensional clusters.

[3] The adapted procedure will be added to the above mentioned software program in the near future and the updated program will be made available at http://ppw.kuleuven.be/okp/software/MBCA/.

[4] In Step 2 of the ALS_{AIC} procedure, the estimation of the SCA-ECP model per cluster is also based on the least squares estimates for the $\mathbf{F}_i^{(k)}$ and $\mathbf{B}^{(k)}$ matrices described by Timmerman and Kiers (2003), which implies that this step minimizes the *SSE* objective function. This is equivalent to minimizing the *AIC* objective function, because the number of free parameters is fixed within Step 2 and the minimal *SSE* corresponds to the minimal $\log(SSE)$.

[5] We also assessed the sensitivity to local minima and the recovery of the within-cluster component structures. A sufficiently low sensitivity to local minima was established for both procedures (i.e., 5.17 % and 0.29 % local minima over all conditions for ALS_{SSE} and ALS_{AIC}, respectively) and the recovery of the cluster loading matrices was found to be really good (i.e., mean congruence coefficient of 0.9968 ($SD = 0.02$) between estimated and simulated loadings across all conditions) for the ALS_{AIC} procedure. Note that previous studies on Clusterwise SCA (De Roover et al., 2012c; De Roover, Ceulemans, Timmerman, & Onghena, 2012b) have already indicated that the within-cluster component loadings are recovered very well in cases where the data blocks are clustered correctly.

assign too many data blocks to the clusters with a higher number of components. The recovery performance is evaluated in light of six manipulated factors: (1) the number of data blocks, (2) the number of observations per data block, (3) the number of underlying clusters K and components $Q^{(k)}$, (4) the cluster size, (5) the amount of error on the data, and (6) the structure of the cluster loading matrices. The first five factors are often varied in simulation studies to evaluate clustering algorithms (see, e.g., Brusco & Cradit, 2001, 2005; Hands & Everitt, 1987; Milligan, Soon, &, Sokol, 1983; Timmerman, Ceulemans, Kiers, & Vichi, 2010; Steinley, 2003), and also in the original Clusterwise SCA-ECP simulation study (De Roover et al., 2012c). With respect to these factors, we expect that Clusterwise SCA-ECP will perform better when more information is available (i.e., more data blocks and/or more observations per data block; Brusco & Cradit, 2005; De Roover et al., 2012c; Hands & Everitt, 1987), in cases of less clusters and less within-cluster components (Brusco & Cradit, 2005; De Roover et al., 2012c; Milligan et al., 1983; Timmerman et al., 2010), when the clusters are of equal size (Brusco & Cradit, 2001; Milligan et al., 1983; Steinley, 2003), and when the data contain less error (Brusco & Cradit, 2005; De Roover et al., 2012c). Factor 6 was included because it is empirically relevant and theoretically interesting to evaluate the effect of different kinds of relations between the cluster loading matrices. We conjecture that it will be harder to distinguish the clusters when the cluster loading matrices are strongly related.

3.2.2. Design and Procedure The number of variables J was fixed at 12 and the six factors mentioned above were varied in a complete factorial design:

1. the *number of data blocks I* at 2 levels: 20, 40;
2. the *number of observations per data block* N_i at 4 levels: $N_i \sim$ U[15; 20], $N_i \sim$ U[30; 70], $N_i \sim$ U[80; 120], $N_i \sim$ U[20; 120], with U indicating a discrete uniform distribution between the given numbers and where the fourth level was added to evaluate the effect of large differences in block size;
3. the *number of underlying clusters and components* (i.e., K and $Q^{(k)}$-values) at 6 levels: [2 1], [4 2], [2 1 2], [4 2 4], [2 1 4 2], [4 2 4 2], where K equals the length of the vector and the $Q^{(k)}$-values are the elements of the vector;
4. the *cluster size*, at 3 levels (see Milligan et al., 1983): equal (equal number of data blocks in each cluster); unequal with minority (10 % of the data blocks in one cluster and the remaining data blocks distributed equally over the other clusters); unequal with majority (60 % of the data blocks in one cluster and the remaining data blocks distributed equally over the other clusters); note that the minority and majority cluster is chosen randomly from the available clusters, implying that they are expected to have one, two and four component(s) in 13/72, 1/2 and 23/72 of the cases, respectively;
5. the *error level e*, which is the expected proportion of error variance in the data blocks $\mathbf{X}_i$ at 2 levels: 0.20, 0.40;
6. the *structure of the cluster loading matrices* at 2 levels: random structure, simple structure.

With respect to factor 6, the loading matrices with random structure were obtained by sampling the loadings uniformly between -1 and 1. In the simple structure conditions, all variables have a high loading on one of the components and a zero loading on the other components. Moreover, the subgroups of variables constituting the components are assembled in such a way that the components of the less complex cluster(s) can be obtained by a pairwise merge of two components of the preceding more complex cluster. We selected this type of loading matrices because we expect that it would be relatively difficult to disentangle the associated less and more complex clusters from each other. For instance, in the case where the $Q^{(k)}$-vector equals [4 2 4 2], the simple structure cluster loading matrices were constructed as follows:

$$\mathbf{B}^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{B}^{(2)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix},$$

$$\mathbf{B}^{(3)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \qquad \mathbf{B}^{(4)} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

where, for example, merging the first two (respectively last two) components of $\mathbf{B}^{(1)}$ gives the first (respectively last) component of $\mathbf{B}^{(2)}$. A similar relationship exists between $\mathbf{B}^{(3)}$ and $\mathbf{B}^{(4)}$. To quantify the degree of relatedness or similarity among the cluster loading matrices, a mean RV-coefficient ("RV_{mean}") was calculated for each data set:

$$RV_{mean} = \frac{\sum_{k_1=1}^{K}\sum_{k_2=k_1+1}^{K} RV(\mathbf{B}^{(k_1)}, \mathbf{B}^{(k_2)})}{K(K-1)/2}$$

$$\text{with } RV\left(\mathbf{B}^{(k_1)}, \mathbf{B}^{(k_2)}\right) = \frac{\|\mathbf{B}^{(k_1)\prime}\mathbf{B}^{(k_2)}\|^2}{\sqrt{\|\mathbf{B}^{(k_1)\prime}\mathbf{B}^{(k_1)}\|^2 \times \|\mathbf{B}^{(k_2)\prime}\mathbf{B}^{(k_2)}\|^2}}. \tag{12}$$

The RV-coefficient (Robert & Escoufier, 1976) is a rotation-independent correlation between two matrices, which allows for the number of columns to differ between the matrices and which takes values between 0 and 1. In Equation (12), the RV-coefficient is computed for each pair of true cluster loading matrices and then averaged over all cluster pairs. On average, RV_{mean} amounts to 0.17 ($SD = 0.09$) and 0.67 ($SD = 0.04$) for the random and simple structure loadings, respectively,[6] which indicates that the random loadings matrices are very different among clusters while the simple structure loading matrices are moderately related as intended by the manipulation.

For each cell of the factorial design, 50 data matrices $\mathbf{X}$ were generated, consisting of I $\mathbf{X}_i$ data blocks. Specifically, the partition matrix $\mathbf{P}$ was obtained by randomly assigning the correct number of data blocks (i.e., given the cluster size factor) to each of the clusters. The component score matrices $\mathbf{F}_i^{(k)}$ as well as the error matrices $\mathbf{E}_i$ were generated by randomly sampling entries from a standard normal distribution. Subsequently, the error matrices $\mathbf{E}_i$ and the cluster loading matrices $\mathbf{B}^{(k)}$ were rescaled to obtain data that contain an expected proportion e of error

[6] The mean values for the modified RV-coefficient (Smilde, Kiers, Bijlsma, Rubingh, & van Erk, 2009), are 0.02 ($SD = 0.09$) and 0.59 ($SD = 0.08$), respectively.

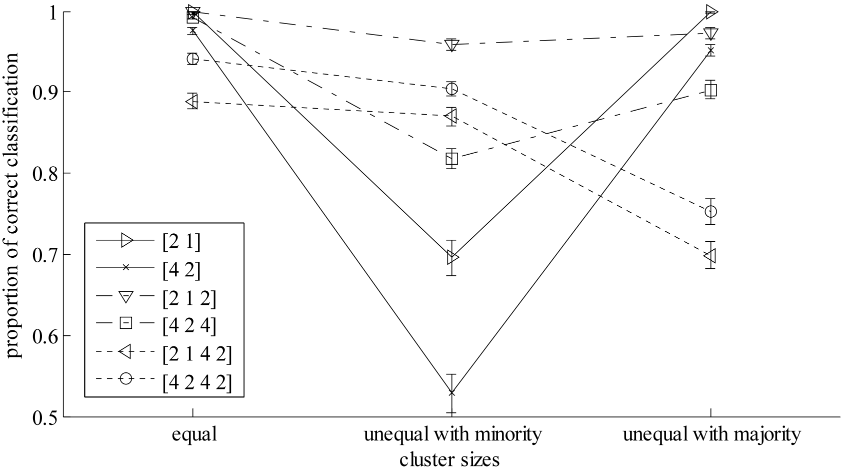

FIGURE 1.
Mean values and associated 95 % confidence intervals of the proportion of correct classification as a function of the number of clusters and components and the cluster sizes for the ALS_{SSE} procedure.

variance. Finally, the resulting $\mathbf{X}_i$ matrices were standardized per variable, and were vertically concatenated into the matrix $\mathbf{X}$.

In total, 2 (number of data blocks) × 4 (number of observations per data block) × 6 (number of clusters and components) × 3 (cluster size) × 2 (error level) × 2 (structure of cluster loading matrices) × 50 (replicates) = 28,800 simulated data matrices were generated. Each data matrix $\mathbf{X}$ was analyzed with the ALS_{SSE} and ALS_{AIC} procedures, each time using the correct K and $Q^{(k)}$-values and starting from 25 different random partition matrices.

3.2.3. Results To quantify how well the clustering of the data blocks is recovered, we will use the proportion of correctly classified data blocks. When calculating this proportion, we took into account the number of components in a cluster. That is, clusters that contain the same blocks as the true clusters, but that are modeled with a too low or too high number of components, are considered incorrect. The overall mean proportion of correct classification is 0.88 ($SD = 0.27$) and 1.00 ($SD = 0.02$) when using the ALS_{SSE} and ALS_{AIC} approach, respectively. This implies that the ALS_{AIC} approach is clearly superior, as hypothesized.

To examine which conditions induce classification errors when using the ALS_{SSE} approach, an analysis of variance was performed with the proportion of correct classification as the dependent variable and the six manipulated factors and their interactions as independent variables. The main effects of the manipulated factors were all significant ($p < 0.001$) and correspond to the expectations formulated in the Problem section above. With respect to factor 2, the results indicate an effect of the expected mean block size but not of the differences in block size. Subsequently, to examine which main and interaction effects have a large effect size, we computed for each effect the partial eta-squared statistic (Cohen, 1973), which indicates the proportion of explained variance of the effect concerned that is not explained by the other effects. Only discussing effects larger than 0.05, the strongest main effect was that of the amount of error ($\hat{\eta}^2_{partial} = 0.11$), where the proportion of correct classification was 0.95 ($SD = 0.20$) and 0.81 ($SD = 0.31$) for error variances of 0.20 and 0.40, respectively. Moreover, we found important effects of the cluster size ($\hat{\eta}^2_{partial} = 0.10$) and the number of clusters and components ($\hat{\eta}^2_{partial} = 0.07$): the proportion of correct classification is lower in case of unequal cluster sizes with majority and lowest for the unequal cluster sizes with minority (Figure 1) and the proportion is also lower when two or four clusters are estimated and when four components are estimated in at least one of the clusters.

The interaction effect between these two factors ($\hat{\eta}^2_{partial} = 0.18$), implies that the latter effect becomes more pronounced when clusters are not of equal size (see Figure 1).

To test the hypothesis that with the ALS$_{SSE}$ approach too many data blocks are assigned to more complex cluster(s), we further examine the misclassifications that result from using this approach. Specifically, we computed the ratio of the number of data blocks that were assigned to a too simple cluster to the total number of misclassified data blocks, for the 5,842 data sets for which misclassifications were made. According to our hypothesis we expect this ratio to be low, which would indicate that most classification errors imply that data blocks are assigned to more or equally complex clusters. This was confirmed by our results in that the mean value of the ratio (i.e., proportion misclassified towards simpler clusters) across the 5,842 data sets amounts to 0.08 ($SD = 0.11$). To examine which factors have a large effect on this ratio, an unbalanced analysis of variance was performed with the ratio as the dependent variable and the six manipulated factors and their interactions as independent variables. A main effect of the structure of the cluster loading matrices is found ($\hat{\eta}^2_{partial} = 0.06$), where the ratio equals 0.18 ($SD = 0.09$) and 0.04 ($SD = 0.09$) for the random and simple structure cluster loading matrices, respectively. Further inspection revealed that in the case of random loadings some of the misclassifications concern relocations of blocks into less complex clusters, while in the simple structure conditions the classification errors mostly imply that data blocks are assigned to more or equally complex clusters. This effect can be explained by the fact that for 1,795 data sets, which mainly (i.e., 1,777 of the 1,795) belong to the random loadings conditions, two correctly assembled clusters of data blocks are swapped and thus modeled with an incorrect number of components. Moreover, for 3,319 of the 3,904 data sets with simple structure loading matrices, data blocks are exclusively misclassified to a more complex cluster of which the dimensions can be merged pairwise to obtain the correct dimensional structure underlying the data block.

The ALS$_{AIC}$ approach yielded classification errors for only 257 data sets so no analysis of variance was performed. The majority (i.e., 240) of these errors are made in conditions with an error variance of 0.40 and/or unequal cluster sizes. Specifically, 174 of the errors imply that only one data block is misclassified, often in a less complex cluster, and for the remaining 83 data sets the proportion of data blocks that are assigned to a too simple cluster amounted to 0.45 ($SD = 0.20$). Therefore, we conclude that the tendency of assigning too many data blocks to the more complex clusters is indeed mitigated by using the *AIC*-based objective function. Because of its clear superiority over the ALS$_{SSE}$ procedure, we will use the ALS$_{AIC}$ procedure in the remainder of the paper.

4. Model Selection

When applying Clusterwise SCA-ECP, the most appropriate number of clusters K, K^{best}, is often unknown, as well as the best number of components $Q^{(k)}$ within each cluster, $Q^{(k),\text{best}}$. To tackle the resulting model selection problem, one may estimate Clusterwise SCA-ECP solutions using 1 to $K^{\max}$ clusters and 1 to $Q^{\max}$ components, where $K^{\max}$ and $Q^{\max}$ are larger values than can be reasonably expected for the data at hand. This implies that, given specific $K^{\max}$ and $Q^{\max}$-values, the number of models among which one has to choose equals:

$$S = \sum_{K=1}^{K^{\max}} \frac{(Q^{\max} + K - 1)!}{((Q^{\max} - 1)!K!)}. \tag{13}$$

This number rapidly becomes very large: For example, if $K^{\max}$ and $Q^{\max}$ equal six, 923 different solutions are obtained. To compare, in the case where $Q^{(k)} = Q$ one has to choose among only 36 models (i.e., $K^{\max} \times Q^{\max}$). In this section, we will first recapitulate the stepwise model selection procedure that De Roover et al. (2012a) proposed for the $Q^{(k)} = Q$ case, which showed good performance in a large simulation study (De Roover et al., 2012a, 2012b). Subsequently, we expand this stepwise procedure to accommodate different $Q^{(k)}$-values for the clusters. Additionally, we discuss a number of alternative model selection techniques, which select among all possible solutions simultaneously. Finally, the performance of the different model selection techniques is evaluated in a simulation study.

4.1. Procedure

4.1.1. Clusterwise SCA-ECP with $Q^{(k)} = Q$ for All Clusters Given the $K^{\max} \times Q^{\max}$ different solutions for the same data, De Roover et al. (2012a) propose to evaluate the balance between fit and complexity on the basis of a generalization of the scree test (Cattell, 1966), and to retain the model with the best balance. Specifically, these authors present a stepwise procedure in which one first selects the best number of clusters, K^{best}, and subsequently the best number of components, Q^{best}. To determine K^{best}, scree ratios $sr_{(K|Q)}$ are calculated for each value of K, given different Q-values:

$$sr_{(K|Q)} = \frac{\text{VAF}_{K|Q} - \text{VAF}_{K-1|Q}}{\text{VAF}_{K+1|Q} - \text{VAF}_{K|Q}}, \tag{14}$$

where $\text{VAF}_{K|Q}$ indicates the VAF % of the solution with K clusters and Q components; these scree ratios indicate the extent to which the increase in fit with additional clusters levels off. Subsequently, K^{best} is chosen as the K-value with the highest average scree ratio across the different Q-values. In the second step, similar scree ratios are calculated for each Q-value, given K^{best}:

$$sr_{(Q|K^{\text{best}})} = \frac{\text{VAF}_{Q|K^{\text{best}}} - \text{VAF}_{Q-1|K^{\text{best}}}}{\text{VAF}_{Q+1|K^{\text{best}}} - \text{VAF}_{Q|K^{\text{best}}}}. \tag{15}$$

The best number of components Q^{best} is the number of components Q for which the above scree ratio is maximal.

To illustrate this stepwise procedure we apply it to the self-consciousness data set. First, we analyzed the data using six as $K^{\max}$- and $Q^{\max}$-values. In Figure 2, the VAF % is plotted against Q, for each K-value. The corresponding scree ratios to determine K^{best} and Q^{best} are presented in Table 2. According to these scree ratios, K^{best} is two as the average scree ratio in the upper part of Table 2 is highest for two clusters. This corresponds to the fact that in Figure 2 an increase in VAF % of about 2.5 % is observed when going from one to two clusters, while adding more clusters gives markedly smaller increases in fit. The lower part of Table 2 shows that Q^{best} equals two, as the corresponding scree ratio is the highest. Indeed, the scree curve for two clusters in Figure 2 displays a mild elbow at two components.

4.1.2. Clusterwise SCA-ECP with $Q^{(k)}$ Varying Across the Clusters

Stepwise Procedure We propose to expand the above described stepwise Clusterwise SCA-ECP model selection procedure as follows:

1. *Obtain K^{best} and Q^{best} (and the corresponding partition)*: see above.
2. *Determine the best number of components $Q^{(k),\text{best}}$ for each cluster k*: Perform a scree test per cluster k, using the partition found in Step 1. Specifically, $Q^{(k),\text{best}}$ is set to the Q-value that maximizes the following scree ratio:

$$sr_{(Q)} = \frac{\text{VAF}_Q^{(k)} - \text{VAF}_{Q-1}^{(k)}}{\text{VAF}_{Q+1}^{(k)} - \text{VAF}_Q^{(k)}}, \tag{16}$$

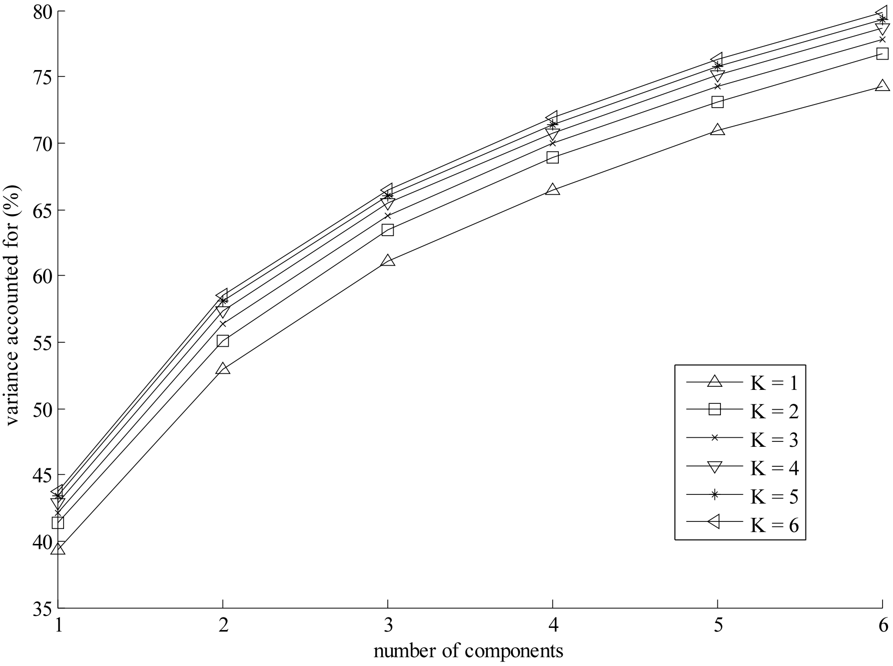

FIGURE 2.
Percentage of explained variance for Clusterwise SCA-ECP solutions for the self-consciousness data, with the number of clusters K varying from 1 to 6, and the number of components Q fixed over clusters and varying from 1 to 6. The solutions with K equal to 1 are equivalent to standard SCA-ECP solutions.

TABLE 2.
Scree ratios for the numbers of clusters K given the fixed numbers of components Q and averaged over the numbers of components (top part of table), and for the fixed numbers of components Q given two clusters (bottom part of table), for the self-consciousness data. The maximal scree ratio in each column is highlighted in bold face.

	1 comp	2 comp	3 comp	4 comp	5 comp	6 comp	average
2 clusters	**2.42**	**1.75**	**2.28**	**2.37**	**1.95**	**2.16**	**2.15**
3 clusters	1.09	1.32	1.16	1.33	1.28	1.32	1.25
4 clusters	1.40	1.34	1.67	1.42	1.58	1.29	1.45
5 clusters	1.55	1.48	1.34	0.96	0.95	1.26	1.25

	2 clusters
2 components	**1.65**
3 components	1.54
4 components	1.29
5 components	1.18

where $\text{VAF}_Q^{(k)}$ is the VAF % of the SCA-ECP solution with Q components for the data blocks in cluster k.

3. *Estimate the Clusterwise SCA-ECP model with the selected complexity*: Run the Clusterwise SCA-ECP algorithm, using the ALS_{AIC} approach with the selected K^{best} and $Q^{(k),\text{best}}$-values, to estimate the corresponding optimal partition and within-cluster models. Apart from (e.g.) 25 random starts, use one rational start, by taking the partition that resulted from Step 1 as the initial partition.

TABLE 3.
Scree ratios for the numbers of components $Q^{(k)}$ for Cluster 1 (left) and Cluster 2 (right), for the self-consciousness data. The maximal scree ratio in each column is highlighted in bold face.

	Cluster 1	Cluster 2
1 component	1.27	**3.77**
2 components	**2.43**	1.26
3 components	1.46	1.29
4 components	1.20	1.34
5 components	1.24	1.18

4. *Check convergence*: Repeat Step 2 using the partition that results from Step 3, to evaluate whether the possible changes in the partition affect the selected $Q^{(k)}s$. If this is the case, repeat Steps 3 and 4.

The scree ratios in Equation (16) (and Equation (15)) cannot be calculated for Q equal to one, unless we specify a VAF$_0$-value (i.e., for the solution with zero components). In this paper, we will evaluate two alternative values: 0 and $100/J$. Using a value of 0 was proposed in the DIFFIT scree test (Timmerman & Kiers, 2000). Although this value may be intuitively appealing, a disadvantage is that it is very low and invariant over data sets; therefore, in some cases it might lead to a maximal scree ratio for one-component solutions when the true number of components is higher. The value of $100/J$ was inspired by considering that it would not make sense to apply component analysis to data in which the observed variables are all uncorrelated, as in such a case no components can be extracted that summarize multiple variables (i.e., each variable would constitute a different component), and realizing that the VAF % by one of the J observed variables would equal $100/J$.

Note that using the (negative) log-likelihood rather than the VAF %, which improves the correspondence among the model estimation and model selection criteria, led to almost identical model selection results in our simulation study. This can be explained by the fact that (1) the logarithmic transformation of the *SSE*-values (see Equation (8)) closely resembles a linear transformation for large *SSE*-values (say, larger than 3,000) and (2) the value of a scree ratio is insensitive to linear transformations. Indeed, using the log-likelihood gives a different model selection result for only 23 data sets, the majority of which are situated in the conditions with only 15 to 20 observations per data block, implying smaller *SSE*-values.

To illustrate this procedure for the self-consciousness data, we first note that in the first step the solution with two clusters and two components is retained (see above). Next, the best number of components is determined for each cluster separately. The corresponding scree ratios are presented in Table 3. For the first cluster, the best number of components remains two, but for the second cluster, one component seems to be enough to adequately summarize the data. In Figure 3, the VAF % within each cluster are plotted against the numbers of components in that cluster; note that we used $100/J$ as VAF % for a solution with 0 components. We can see that the elbow for Cluster 1 corresponds to two components while for Cluster 2 more variance is explained by one component already and the increase in fit indeed seems to level off after one component. The resulting solution was discussed in the Model section.

Simultaneous Procedures To simultaneously select among all S (Equation (13)) possible Clusterwise SCA-ECP solutions—which, thus, all have to be estimated—, one may consider (a) the well-known *AIC* (Akaike, 1974), which selects the solution for which the *AIC*-value (Equation (6)) is the lowest, (b) the equally popular *BIC* (Schwarz, 1978), which retains the solution for which the *BIC*-value, that is, $BIC = -2\,\text{loglik}(\mathbf{X}|\mathbf{M}) + \log(N)fp$, is minimal, and

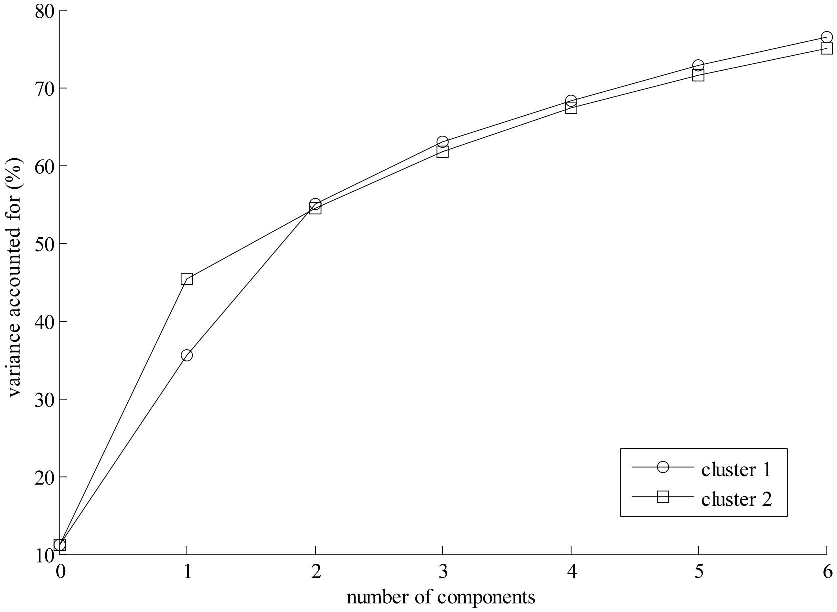

FIGURE 3.
Percentage of explained variance for SCA-ECP solutions within Cluster 1 and Cluster 2 for the self-consciousness data, with the number of components varying from 1 to 6.

(c) the CHULL procedure (Ceulemans & Van Mechelen, 2005), which generalizes the scree test (Cattell, 1966) to multidimensional model selection problems (e.g., three-mode component analysis, Ceulemans & Kiers, 2006, 2009; multilevel component analysis, Ceulemans et al., 2011). Specifically, the CHULL procedure balances fit and complexity, by comparing the VAF % (Equation (4)) and the number of free parameters *fp* (Equation (9)) of the obtained solutions.

4.2. *Simulation Study to Evaluate the Model Selection*

4.2.1. Problem This simulation study aims to evaluate the overall performance of the stepwise model selection procedure that we proposed above, and to compare it to some existing simultaneous model selection procedures: *AIC*, *BIC*, and CHULL. Analyzing a subset of the data sets from the simulation study carried out to examine model estimation, we will investigate the overall frequency of correct model selection as well as the effect of the six manipulated factors.

4.2.2. Design To keep the total computation time within reasonable limits, we used the first five replications in each cell of the model estimation study. Thus, we are now dealing with 2 (number of data blocks) × 4 (number of observations per data block) × 6 (number of clusters and components) × 3 (cluster size) × 2 (error level) × 2 (structure of cluster loading matrices) × 5 (replicates) = 2,880 simulated data matrices. To each data set, we applied the four model selection procedures under consideration, setting the $K^{\max}$- and $Q^{\max}$-values to six, implying that we had to select among 923 possible solutions. When computing *AIC*, *BIC* and CHULL, the number of free parameters *fp* (Equation (9)) is used as complexity measure. Like Ceulemans et al. (2011), we slightly adjusted *fp* to account for redundancy in the component scores of large data blocks. Specifically, the number of observations in a cluster is computed as $N^{(k)} = \sum_{i=1}^{I} p_{ik} \min(N_i, J \log(N_i))$. Because of the expected difficulty of the simultaneous model selection (i.e., choosing among 923 solutions), we retain the three best solutions for each criterion. The stepwise procedure was executed as described above and only the best solution was selected.

4.2.3. Results The stepwise procedure (with VAF_0 equal to $100/J$) selects the correct Clusterwise SCA-ECP model—correct number of clusters K as well as correct number of components $Q^{(k)}$ for each cluster k—for 87 % of the cases (2,500 out of 2,880). The simultaneous model selection by means of *AIC* displays a very strong tendency to overestimate K as well as $Q^{(k)}$, and never selects the correct model. *BIC* performs better in that the correct model is among the three selected models for 67 % of the data sets. Specifically, the best model according to *BIC* is the correct one in 54 % of the cases, while the second best and third best are correct in 9 % and 4 % of the cases, respectively. When using CHULL, the correct model is one of the three retained models in 78 % of the cases, with the correct one being the best model for 38 % of the simulated data sets, the second best for 31 % and the third best for 9 %. We conclude that the stepwise procedure performs best; note that this also holds when focusing on specific levels of the factors. Moreover, the stepwise procedure required a mean computation time of about five minutes only, while the simultaneous model selection methods require a mean computation time of about two hours per data set (with a 3.33 GHz processor).

Since the stepwise method is the best in terms of performance and time efficiency, we will take its performance under scrutiny to see when it can go wrong (i.e., which data characteristics play a role) and what happens in those cases. The majority of the model selection mistakes (324 out of the 380) correspond to an underestimation of the number of clusters in the simple structure conditions or an underestimation of $Q^{(k)}$ for at least one cluster in the conditions with random loadings and with four components underlying one or more of the clusters. The underestimations of the number of clusters are due to the structural relations between less and more complex simple structure clusters (see description of manipulated factor 6). The $Q^{(k)}$ underestimations may be explained by the fact that each component accounts for about the same VAF % in the simple structure case, while in the random loadings case the VAF % may differ strongly across the components. Consequently, in the latter case, it is more difficult to distinguish components that are explaining less variance from the error, and even more so when $Q^{(k)}$ is higher.

We also compared the effect of using 0 instead of $100/J$ as the VAF_0-value when computing the scree ratio for one-component solutions. While using $100/J$ yielded the results described in the previous paragraphs, using zero implied 11 additional model selection mistakes. More specifically, for these 11 data sets—all in the conditions with random loadings—one component was wrongly selected as $Q^{(k),\text{best}}$ for one or more of the clusters.

5. Performance of Clusterwise SCA-ECP with Varying $Q^{(k)}$ and Clusterwise SCA-ECP with $Q^{(k)} = Q$ for the Simulated Data

Based on a simulation study reported in De Roover et al. (2012c) on the effects of overextraction (i.e., using too many components), one might hypothesize that Clusterwise SCA-ECP with $Q^{(k)} = Q$ will adequately recover the clustering of the data blocks when the number of components Q is sufficiently high for all clusters, raising doubts about the necessity of Clusterwise SCA-ECP with varying $Q^{(k)}$. Indeed, if Clusterwise SCA-ECP with $Q^{(k)} = Q$ would correctly reveal the underlying clustering when $Q^{(k)}$ actually varies across clusters, one could simply determine the most appropriate number of components $Q^{(k)}$ per cluster and the corresponding loadings and component scores in a post-processing step. To put the doubts about the added value of Clusterwise SCA-ECP with varying $Q^{(k)}$ to rest, we will demonstrate its superior cluster recovery by re-analyzing the 2,880 data sets from the model selection simulation with the two Clusterwise SCA-ECP approaches. As the underlying $Q^{(k)}$s are unknown in empirical practice, we will apply the stepwise model selection procedure described above and compare the clustering of the selected model with $Q^{(k)} = Q$ (i.e., clustering after Step 1 of the procedure) to that of the selected model with varying $Q^{(k)}$ (i.e., clustering at the end of the procedure). Note that this implies that the number of clusters K^{best} will always be identical for both selected models.

Computing the proportion of correctly classified data blocks (as defined in Section 3.2.3), reveals that using $Q^{(k),\text{best}} = Q^{\text{best}}$ yields an incorrect clustering (i.e., proportion of correct classification < 1) for 215 out of the 2,880 data sets, while the clustering of the selected model with a varying $Q^{(k),\text{best}}$ is incorrect for only 23 data sets. Thus, the cluster recovery by means of Clusterwise SCA-ECP with $Q^{(k)} = Q$ is relatively good, but clearly inferior to the recovery by means of the new Clusterwise SCA-ECP model with a varying number of components across clusters.

6. Discussion

The key idea behind the original Clusterwise SCA-ECP method—capturing differences and similarities in the dimensional structure of a number of data blocks—is very useful for behavioral research. The method had an important drawback, however, in that the number of dimensions was restricted to be the same across the obtained clusters of data blocks. As this restriction is often unrealistic and inappropriate, it was removed in the current paper.

To improve the model estimates of the Clusterwise SCA-ECP model with possibly varying numbers of components, we made use of a stochastic extension of the Clusterwise SCA-ECP model. In this model, the residuals are assumed to be independently, identically and normally distributed. The robustness of the model against violations of this assumption was not examined in the current paper. Previous work by Wilderjans et al. (2011) on the influence of between-block differences in error variance on the performance of a stochastically extended SCA, indicated that the performance is only hampered when large differences in error variance are combined with large differences in the size of the data blocks. Since the influence of block size differences is eliminated by the autoscaling, we conjecture that Clusterwise SCA-ECP can withstand between-block differences in error variance. However, no research has been conducted concerning the robustness of (Clusterwise) SCA models with respect to differences in the block-specific error variance of the different variables, non-normality of the residuals or dependencies between the residuals and, thus, it would be useful to do so in the future.

Additionally, it would be interesting to provide some uncertainty information about the obtained parameter estimates. However, as the stochastic extension in this paper only implies distributional assumptions about the residuals, we cannot use classic inferential procedures. Therefore, it might be useful to develop a bootstrap procedure to assess the stability of the clustering of a Clusterwise SCA-ECP model on the one hand and to construct confidence intervals around the loadings within the clusters (thus, given a certain partition) on the other hand. Whereas for the cluster stability the work of Hofmans, Ceulemans, Steinley, and Van Mechelen (2012) might be relevant, for the confidence intervals around the loadings a similar procedure as the one described by Timmerman et al. (2009) could be applied.

Even though a number of generally applicable model selection procedures exists to select the best out of a large number of models (i.e., *AIC*, *BIC*, CHULL), the stepwise procedure that we specifically developed for the generalized Clusterwise SCA-ECP model clearly performs better. We see two possible reasons for this difference in performance. First, whereas *AIC* and *BIC* do not differentiate between the often large set of possible solutions, the stepwise procedure first conducts an initial screening among a subset of the solutions (i.e., the solutions that impose an equal number of dimensions in each cluster) to obtain a good guess about the complexity of the best solution. Subsequently, this guess is fine-tuned in the remaining steps of the stepwise procedure. Second, complexity is operationalized in different ways in the stepwise and simultaneous model selection procedures. The stepwise procedure is based on the number of clusters and components as complexity measures, which do not change as a function of the data size (the number of observations, data blocks, and variables). In contrast, the simultaneous procedures all use the number of free parameters, which strongly depends on the size of the data. Further research into this matter may provide better insight into the mechanisms behind different model selection procedures, which would be highly useful for statistical practice.

Finally, our findings about the usefulness of penalized model estimation as well as about the characteristics of different model selection procedures might also be useful for other models combining some form of clustering with a dimension reduction within each cluster. For instance, mixtures of factor analyzers (MFA; McLachlan & Peel, 2000; Yung, 1997) is a widely used approach that combines mixture modeling and factor analysis for single-block data (i.e., data that are not hierarchically structured). Following the line of reasoning that was given for Clusterwise SCA-ECP, applying MFA with the number of factors varying across the mixture components might be interesting for some data sets. Indeed, if our self-consciousness data would have been cross-sectional in that each person had rated how much they ruminate, reflect and are aware of their public appearance on average, MFA could be applied to distinguish between a number of person types on the basis of their mean level as well as the correlation structure. Herewith, it makes sense to assume that rumination, reflection, and public self-consciousness are more strongly correlated in some person types, implying that the number of factors should be allowed to differ over the mixtures. Such a generalization of MFA might cause the same model estimation problems as reported in this paper, and thus MFA might also benefit from using the number of free parameters as a penalty during model estimation. Also, up to now, the best MFA model was selected using *BIC*. It may be interesting to compare the performance of *BIC* and a stepwise procedure similar to the one presented in this paper.

Acknowledgements

The research reported in this paper was partially supported by the fund for Scientific Research-Flanders (Belgium), Project No. G.0477.09 awarded to Eva Ceulemans, Marieke Timmerman, and Patrick Onghena and by the Research Council of KU Leuven (GOA/2010/02).

Appendix: Derivation of an *AIC*-Based Partition Criterion

Conditional upon a specific Clusterwise SCA-ECP model $\mathbf{M}$, the log-likelihood of data block $\mathbf{X}_i$ when assigned to cluster k (and thus modeled by $\mathbf{M}_i^{(k)}$) amounts to

$$\begin{aligned}\operatorname{loglik}\big(\mathbf{X}_i|\mathbf{M}_i^{(k)}\big) &= \log\left[\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{N_iJ}{2}}\exp\left(-\frac{SSE_i^{(k)}}{2\sigma^2}\right)\right] \\ &= -\frac{N_iJ}{2}\log\big(2\pi\sigma^2\big) - \frac{1}{2\sigma^2}SSE_i^{(k)},\end{aligned} \tag{A.1}$$

which is the block-specific counterpart of Equation (7), given $SSE_i^{(k)}$ as defined in Equation (5). When inserting $\hat{\sigma}^2 = \frac{SSE_i^{(k)}}{N_iJ}$ as a post-hoc estimator of the error variance σ^2 (Wilderjans et al., 2011), the log-likelihood can be rewritten as

$$\operatorname{loglik}\big(\mathbf{X}_i|\mathbf{M}_i^{(k)}\big) = -\frac{N_iJ}{2}\big[1 + \log(2\pi) - \log(N_iJ) + \log\big(SSE_i^{(k)}\big)\big], \tag{A.2}$$

where the first three terms are not influenced by the cluster assignment and can thus be discarded. The number of free parameters for data block i, when it is tentatively assigned to cluster k, is denoted by $fp_i^{(k)}$ and can be computed as follows:

$$fp_i^{(k)} = N_i Q^{(k)}. \tag{A.3}$$

It corresponds to the size of the component score matrix $\mathbf{F}_i^{(k)}$ that is computed to evaluate the fit of data block i in cluster k. When combining Equations (A.2) (omitting the invariant terms) and (A.3) as in the *AIC* (Akaike, 1974), we obtain the *AIC*-based partition criterion in Equation (11).

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*, 716–723.

Barrett, L.F. (1998). Discrete emotions or dimensions? The role of valence focus and arousal focus. *Cognition and Emotion*, *12*, 579–599.

Brusco, M.J., & Cradit, J.D. (2001). A variable selection heuristic for K-means clustering. *Psychometrika*, *66*, 249–270.

Brusco, M.J., & Cradit, J.D. (2005). ConPar: a method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses. *Journal of Mathematical Psychology*, *49*, 142–154.

Cattell, R.B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276.

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

Ceulemans, E., & Kiers, H.A.L. (2009). Discriminating between strong and weak structures in three-mode principal component analysis. *British Journal of Mathematical & Statistical Psychology*, *62*, 601–620.

Ceulemans, E., Timmerman, M.E., & Kiers, H.A.L. (2011). The CHULL procedure for selecting among multilevel component solutions. *Chemometrics and Intelligent Laboratory Systems*, *106*, 12–20.

Ceulemans, E., & Van Mechelen, I. (2005). Hierarchical classes models for three-way three-mode binary data: interrelations and model selection. *Psychometrika*, *70*, 461–480.

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, *33*, 107–112.

De Roover, K., Ceulemans, E., & Timmerman, M.E. (2012a). How to perform multiblock component analysis in practice. *Behavior Research Methods*, *44*, 41–56.

De Roover, K., Ceulemans, E., Timmerman, M.E., & Onghena, P. (2012b). A clusterwise simultaneous component method for capturing within-cluster differences in component variances and correlations. *British Journal of Mathematical & Statistical Psychology*. doi:10.1111/j.2044-8317.2012.02040.x. Advance online publication.

De Roover, K., Ceulemans, E., Timmerman, M.E., Vansteelandt, K., Stouten, J., & Onghena, P. (2012c). Clusterwise simultaneous component analysis for the analysis of structural differences in multivariate multiblock data. *Psychological Methods*, *17*, 100–119.

Diaz-Loving, R. (1998). Contributions of Mexican ethnopsychology to the resolution of the etic-emic dilemma in personality. *Journal of Cross-Cultural Psychology*, *29*, 104–118.

Feningstein, A., Scheier, M.F., & Buss, A. (1975). Public and private self-consciousness. *Journal of Consulting and Clinical Psychology*, *43*, 522–527.

Goldberg, L.R. (1990). An alternative “description of personality”: the Big-Five factor structure. *Journal of Personality and Social Psychology*, *59*, 1216–1229.

Hands, S., & Everitt, B. (1987). A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. *Multivariate Behavioral Research*, *22*, 235–243.

Hoerl, A.E. (1962). Application of ridge analysis to regression problems. *Chemical Engineering Progress*, *58*, 54–59.

Hofmans, J., Ceulemans, E., Steinley, D., & Van Mechelen, I. (2012). On the added value of bootstrap analysis for K-means clustering. Manuscript conditionally accepted.

Jolliffe, I.T. (1986). *Principal component analysis*. New York: Springer.

Kaiser, H.F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*, 187–200.

Kiers, H.A.L. (1990). *SCA. A program for simultaneous components analysis of variables measured in two or more populations*. Groningen: iec *Pro*GAMMA.

Kiers, H.A.L., & ten Berge, J.M.F. (1994). Hierarchical relations between methods for Simultaneous Components Analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical & Statistical Psychology*, *47*, 109–126.

McLachlan, G.J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.

Meredith, W., & Millsap, R.E. (1985). On component analyses. *Psychometrika*, *50*, 495–507.

Milligan, G.W., Soon, S.C., & Sokol, L.M. (1983). The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *5*, 40–47.

Nezlek, J.B. (2005). Distinguishing affective and non-affective reactions to daily events. *Journal of Personality*, *73*, 1539–1568.

Nezlek, J.B. (2012). Diary methods for social and personality psychology. In J.B. Nezlek (Ed.), *The SAGE library in social and personality psychology methods*. London: Sage Publications.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, *2*, 559–572.

Robert, P., & Escoufier, Y. (1976). A unifying tool for linear multivariate statistical methods: the RV-coefficient. *Applied Statistics*, *25*, 257–265.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, *6*, 461–464.

Selim, S.Z., & Ismail, M.A. (1984). K-means-type algorithms: a generalized convergence theorem and characterization of local optimality. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *6*, 81–87.

Smilde, A.K., Kiers, H.A.L., Bijlsma, S., Rubingh, C.M., & van Erk, M.J. (2009). Matrix correlations for high-dimensional data: the modified RV-coefficient. *Bioinformatics*, *25*, 401–405.

Steinley, D. (2003). Local optima in K-means clustering: what you don't know may hurt you. *Psychological Methods*, *8*, 294–304.

ten Berge, J.M.F. (1993). *Least squares optimization in multivariate analysis*. Leiden: DSWO Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B (Methodological)*, *58*, 267–288.

Timmerman, M.E., Ceulemans, E., Kiers, H.A.L., & Vichi, M. (2010). Factorial and reduced K-means reconsidered. *Computational Statistics & Data Analysis*, *54*, 1858–1871.

Timmerman, M.E., & Kiers, H.A.L. (2000). Three-mode principal component analysis: choosing the numbers of components and sensitivity to local optima. *British Journal of Mathematical & Statistical Psychology*, *53*, 1–16.

Timmerman, M.E., & Kiers, H.A.L. (2003). Four simultaneous component models of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, *86*, 105–122.

Timmerman, M.E., Kiers, H.A.L., Smilde, A.K., Ceulemans, E., & Stouten, J. (2009). Bootstrap confidence intervals in multi-level simultaneous component analysis. *British Journal of Mathematical & Statistical Psychology*, *62*, 299–318.

Trapnell, P.D., & Campbell, J.D. (1999). Private self-consciousness and the five factor model of personality: distinguishing rumination from reflection. *Journal of Personality and Social Psychology*, *76*, 284–304.

Tugade, M.M., Fredrickson, B.L., & Barrett, L.F. (2004). Psychological resilience and positive emotional granularity: examining the benefits of positive emotions on coping and health. *Journal of Personality*, *72*, 1161–1190.

Van Deun, K., Wilderjans, T.F., van den Berg, R.A., Antoniadis, A., & Van Mechelen, I. (2011). A flexible framework for sparse simultaneous component based data integration. *BMC Bioinformatics*, *12*, 448.

Van Mechelen, I., & Smilde, A.K. (2010). A generic linked-mode decomposition model for data fusion. *Chemometrics and Intelligent Laboratory Systems*, *104*, 83–94. doi:10.1016/j.chemolab.2010.04.012.

Wilderjans, T.F., Ceulemans, E., Van Mechelen, I., & van den Berg, R.A. (2011). Simultaneous analysis of coupled data matrices subject to different amounts of noise. *British Journal of Mathematical & Statistical Psychology*, *64*, 277–290.

Yung, Y.F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, *62*, 297–330.

Manuscript Received: 17 JAN 2012
Final Version Received: 4 JUL 2012
Published Online Date: 25 JAN 2013