Behav Res (2013) 45:1011–1023
DOI 10.3758/s13428-013-0329-y

Subspace *K*-means clustering

Marieke E. Timmerman · Eva Ceulemans · Kim De Roover · Karla Van Leeuwen

Published online: 23 March 2013
© Psychonomic Society, Inc. 2013

Abstract To achieve an insightful clustering of multivariate data, we propose subspace *K*-means. Its central idea is to model the centroids and cluster residuals in reduced spaces, which allows for dealing with a wide range of cluster types and yields rich interpretations of the clusters. We review the existing related clustering methods, including deterministic, stochastic, and unsupervised learning approaches. To evaluate subspace *K*-means, we performed a comparative simulation study, in which we manipulated the overlap of subspaces, the between-cluster variance, and the error variance. The study shows that the subspace *K*-means algorithm is sensitive to local minima but that the problem can be reasonably dealt with by using partitions of various cluster procedures as a starting point for the algorithm. Subspace *K*-means performs very well in recovering the true clustering across all conditions considered and appears to be superior to its competitor methods: *K*-means, reduced *K*-means, factorial *K*-means, mixtures of factor analyzers (MFA), and MCLUST. The best competitor method, MFA, showed a performance similar to that of subspace *K*-means in easy conditions but deteriorated in more difficult ones. Using data from a study on parental behavior, we show that subspace *K*-means analysis provides a rich insight into the cluster characteristics, in terms of both the relative positions of the clusters (via the centroids) and the shape of the clusters (via the within-cluster residuals).

M. E. Timmerman (✉)
Heymans Institute for Psychology, Psychometrics & Statistics,
University of Groningen, Grote Kruisstraat 2/1,
9712TS Groningen, The Netherlands
e-mail: m.e.timmerman@rug.nl

E. Ceulemans · K. De Roover
Educational Sciences, K.U. Leuven, Leuven, Belgium

K. Van Leeuwen
Parenting and Special Education, K.U. Leuven, Leuven, Belgium

Keywords Cluster analysis · Cluster recovery · Multivariate data · Reduced *K*-means · *K*-means · Factorial *K*-means · Mixtures of factor analyzers · MCLUST

Introduction

Many different clustering procedures have been proposed for grouping individuals on the basis of observed variables—for instance, the extent to which they display several behaviors. Each cluster is characterized by its centroid, which is usually defined as the mean of the observed variables across the members of the cluster, and by its within-cluster residuals, which are the deviations of the observed scores from the centroid. Two properties determine the attractiveness of a clustering procedure for empirical practice. First, the clustering procedure should properly recover the clusters. That is, when the clusters are well separated in some sense, the procedure should be able to identify them. Second, the clusters should be interpretable, which means that the distinguishing features (e.g., cluster means, covariances of the within-cluster residuals) become apparent in the cluster solution.

Cluster procedures differ widely with respect to those two properties. The differences between procedures generally become more prominent with increasing number of variables, because then the interpretability of the cluster solution and the recovery of the correct partition may become increasingly difficult. In this article, we put forward subspace *K*-means as a useful method for clustering on the basis of multivariate data. Building on the idea both that differences in score level as differences in score covariances may indicate the presence of clusters and that these level and covariance differences may pertain to different variables, subspace *K*-means rests on the key assumption that the cluster centroids and the within-cluster residuals are primarily located in different subspaces of the variables, rather than the full-dimensional space spanned by the observed variables. Because subspace *K*-means accounts

for the clustering, the subspace of the cluster centroids, and, for each cluster, the subspace of the within-cluster residuals, it may handle a wide range of cluster types. Furthermore, interpreting the obtained clusters is facilitated, because the clusters are characterized on the basis of their centroids and their structure of within-cluster residuals.

The remainder of this article is organized as follows. In the following section, we introduce the subspace *K*-means model. Additionally, we review the many clustering models that relate to subspace *K*-means. In the Data Analysis section, we define the loss function, propose an algorithm, and discuss the issue of model selection, including the proposal of a formal model selection criterion. In a simulation study, we evaluate the performance of the subspace *K*-means algorithm and compare it with that of five competitor clustering methods. An empirical example from a study on parental behavior shows the usefulness of subspace *K*-means in empirical practice. We conclude by discussing the strengths and weaknesses of subspace *K*-means.

Subspace *K*-means

A subspace *K*-means analysis groups individuals into mutually exclusive clusters on the basis of their observed multivariate data. Subspace *K*-means simultaneously models the centroids and the within-cluster residuals in subspaces, using a component analysis approach. Specifically, to capture the main differences in level, the location of each centroid is identified via scores on a few between-components (which are weighted sums of the observed variables). To gain insight into the differences within each separate cluster, a few cluster-specific within-cluster components summarize the within-cluster variability in the observed variables. The between-components define the between-subspace, and for each cluster, the within-components the within-subspace per cluster.

As an example, consider Fig. 1a–d, which shows four simulated data sets. Each data set contains the scores of 100 individuals on two observed variables. The individuals can be clustered into two groups. For all data sets, the differences between the two centroids of the clusters can be summarized with one between-component. In contrast, the structure of the within-cluster residuals differs across the four data sets. In data set 1a, the clusters are spherical (i.e., the within-cluster residuals are uncorrelated and have equal variances), implying that one would need two within-components to summarize the data to a reasonable extent. In data sets 1b, 1c, and 1d, using one within-component suffices to summarize the data. Specifically, for data set 1b, the subspace of the within-component is the same for the two clusters and equal to the between-component; for data set 1c, the subspace of the within-component is orthogonal to the between-component. In data set 1d, the subspaces of the within-components differ across the two clusters. Subspace *K*-means can model those types of data sets.

Formally, given that $\mathbf{x}_i$ ($1 \times J$) contains the observed scores on J variables of the ith individual ($i = 1,\ldots, I$), the subspace *K*-means model can be written as

$$\mathbf{x}_i = \mathbf{m} + \sum_{c=1}^{C} u_{ic}\left(\mathbf{fb}^c\ \mathbf{Ab}' + \mathbf{fw}_i^c\ \mathbf{Aw}^{c\prime}\right) + \mathbf{e}_i^c, \qquad (1)$$

where $\mathbf{m}$ ($1 \times J$) denotes a vector containing the variable means across the I individuals; u_{ic} a binary scalar, which specifies to which cluster object i belongs (i.e., $u_{ic} = 1$ if object i belongs to cluster c, and $u_{ic} = 0$ otherwise, with $\sum_{c=1}^{C} u_{ic} = 1$); $\mathbf{fb}^c$ ($1 \times Qb$) a vector with between-component scores of cluster c; $\mathbf{Ab}$ ($J \times Qb$) a between-loading matrix, with Qb the number of between-components; $\mathbf{fw}_i^c$ ($1 \times Qw$) a vector containing within-component scores of the ith individual, with i being a member of cluster c; $\mathbf{Aw}^c$ ($J \times Qw$) the within-loading matrix of cluster c, and Qw the number of within-components; $\mathbf{e}_i^c$ ($1 \times J$) holds the S*K*m residuals of the ith individual. To partly identify the model, the between- and within-loading matrices are constrained to be column-wise orthonormal—that is, $\mathbf{Ab}'\mathbf{Ab} = \mathbf{I}_{Qb}$ and $\mathbf{Aw}^{c\prime}\mathbf{Aw}^c = \mathbf{I}_{Qw}$; furthermore, the component scores are constrained as $\sum_{i=1}^{I}\sum_{c=1}^{C} u_{ic}\mathbf{fb}^c = \mathbf{0}$ and $\sum_{i=1}^{I} \mathbf{fw}_i^c = \mathbf{0}$, with $\mathbf{fw}_i^c = \mathbf{0}$ if $u_{ic} = 0$. Those constraints ensure that $\mathbf{fb}^c\,\mathbf{Ab}'$ provide a model for the centroids of cluster c, and $\mathbf{fw}_i^c\,\mathbf{Aw}^{c\prime}$ a model for the within-cluster residuals of individual i in cluster c. To fit the subspace *K*-means model to observed data, we use a least squares loss function (presented in the Data Analysis section), such that the sum-of-squared S*K*m residuals [i.e., $\sum_{i=1}^{I}\left(\mathbf{e}_i^c\right)^2$] is minimized.

In subspace *K*-means, the maximal numbers of components are the numbers needed to have the centroids and within-cluster residuals, respectively, in full space. For the centroids, the full centroid space is used when the number of between-components Qb is taken equal to $\min(C-1, J)$ (with $C-1$ rather than C because of the sum constraint on $\mathbf{fb}^c$). For the within-part of cluster c, the full within-residual space of cluster c would be used when Qw is taken equal to $\min(n_c - 1, J)$, with n_c the number of individuals in cluster c.

The between-loading and the within-loading matrices have rotational freedom, implying that each of the loading matrices can be rotated without altering the part of the data described by the model, provided that this rotation is compensated for in the accompanying component scores matrix. The rotational freedom can be exploited to facilitate the interpretation.

Besides its interpretational advantages—yielding dimensions that summarize the differences between the centroids (level differences) and the differences between the members

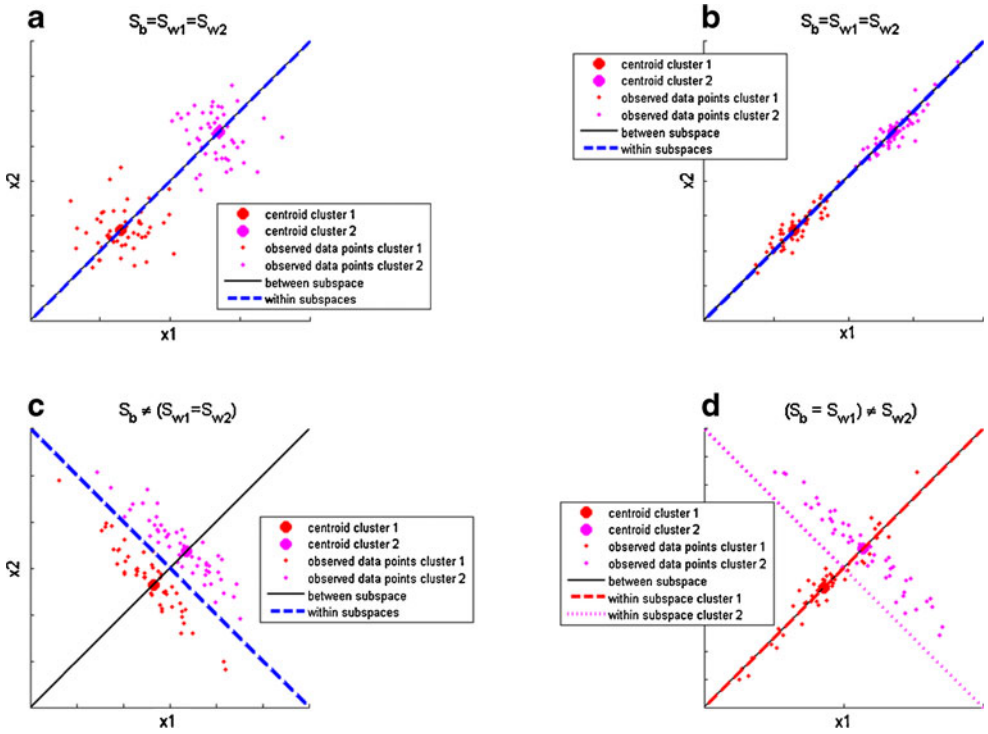


Fig. 1 Four data sets complying with the subspace *K*-means model; see text for details

of each cluster (covariance differences)—subspace *K*-means may be able to recover a partitioning, when *K*-means or alternative subspace clustering models (described in the next section) may fail. A good performance of subspace *K*-means is to be expected when the centroids are located in a subspace with relatively large variance, because subspace *K*-means explicitly models the centroids in a subspace, while maximizing the variance explained, and/or when the within-cluster residuals of cluster c (ε_i^c) have large variances in (a limited) subspace, particularly when those subspaces differ across clusters, because subspace *K*-means explicitly models the within-cluster residuals in a subspace, while maximizing the variance explained.

A word of caution is warranted for a special case—namely, when the centroids and within-models of two (or more) clusters share exactly the same subspace (i.e., $\mathbf{Ab} = \mathbf{Aw}^c = \mathbf{Aw}^{c'}$, for $c \neq c'$)—or a part thereof. Then, the partition of the individuals across the clusters c, c' is nonunique within the shared subspace concerned. That is, for each possible partition, one can construct between-component scores and within-component scores such that the model perfectly describes the data. The empirical implications of this nonuniqueness property will be discussed in the Model Selection section.

Relationships to existing models

Subspace *K*-means is related to a range of deterministic, heuristic, and stochastic approaches that combine clustering and dimension reduction. Previously proposed deterministic methods model either the centroids or the within-cluster residuals in reduced space. In particular, reduced *K*-means (RKM) is a model with only centroids in reduced space. It has been independently proposed by Bock (1987), De Soete and Carroll (1994), and Stute and Zhu (1995) and extended with simplicity constraints for the between-loadings (Vichi & Saporta, 2009). In terms of partition recovery, subspace *K*-means is expected to outperform those models when the within-cluster residuals of cluster c (ε_i^c) have large variances in limited subspaces that vary across clusters and/or when those within-cluster subspaces differ from the between-subspace. The assumption of centroids in reduced space is also found in factorial *K*-means (FKM; Vichi & Kiers, 2001) and factor discriminant *K*-means (Rocci, Gattone, & Vichi, 2011). Those models differ from subspace *K*-means in that they provide a model for the projected data, using a single loading matrix, rather than the observed data itself. Therefore, subspace *K*-means can be expected to outperform those methods when the between- and within-subspaces of the clusters differ considerably. Bock proposed, at a theoretical level only, a model with only within-cluster residuals in reduced space. Subspace *K*-means is the first deterministic model that fully integrates both aspects and subsumes many of the existing deterministic models as special cases.

Stochastic cluster models are well-known as mixture models. In those models, the observed data $\mathbf{x}_i$ are assumed to be a random sample from a mixture of C populations, with

mean $\boldsymbol{\mu}_c$ and (within-cluster) covariance matrix $\boldsymbol{\Sigma}_c$ (c = 1,.., C) (McLachlan & Peel, 2000). For continuous data, it is common to model each population as a multivariate Gaussian. The number of parameters of an unconstrained mixture model grows rapidly with increasing numbers of variables. To resolve the associated identification and estimation problems, one may impose constraints on $\boldsymbol{\mu}_c$ and/or $\boldsymbol{\Sigma}_c$.

We first discuss approaches that leave $\boldsymbol{\mu}_c$ unconstrained and constrain $\boldsymbol{\Sigma}_c$. First, one has the reparametrization approach, in which each $\boldsymbol{\Sigma}_c$ is parameterized in terms of its eigenvalue decomposition; the decompositions of the various clusters are related to each other by imposing equality constraints on certain parameters across clusters (Banfield & Raftery, 1993). As Banfield and Raftery showed, this implies a series of models, generally denoted by MLCUST, that are associated with different optimization criteria. The model for the within part in MLCUST differs from subspace K-means in that subspace K-means implies a low rank approximation and refrains from imposing equality constraints across clusters.

A second type of constraint is to assume that each matrix $\boldsymbol{\Sigma}_c$ (c = 1,.., C) complies with a common factor model. A variety of such mixture models have been proposed. The variant with a separate loading matrix for each cluster is known as mixtures of factor analyzers (MFA; McLachlan & Peel, 2000; McLachlan, Peel, & Bean, 2003). The mixture of probabilistic principal component analyzers (Tipping & Bishop, 1999) is a special case, with constant unique variances across variables and clusters. Furthermore, a sequence of models has been proposed combining a common factor model for $\boldsymbol{\Sigma}_c$ with the reparametrization approach (Bouveyron, Girard, & Schmid, 2007). Finally, Viroli (2010) assumes the common factors for each cluster to be a mixture of factor analyzers themselves. Those models, assuming some common factor model for the within-part, clearly differ from subspace K-means in that the latter uses a low rank approximation for the within-part. This difference boils down to the well-known distinction between common factor analysis and principal component analysis (see, e.g., Jolliffe, 2002).

Mixture models that constrain $\boldsymbol{\mu}_c$ to be in reduced space and leave $\boldsymbol{\Sigma}_c$ unconstrained are known as latent class multidimensional scaling (MDS) models (Böckenholt & Böckenholt, 1991; DeSarbo, Howard, & Jedidi, 1991).They are members of the STUNMIX (stochastic MDS unfolding mixture models) family (Wedel & Desarbo, 1996).

Finally, some mixture models constrain both $\boldsymbol{\Sigma}_c$ and $\boldsymbol{\mu}_c$. Specifically, these models impose a single common factor model on the full data matrix, in which the factor scores follow a mixture distribution. Thus, the factors cover both the cluster means and within-cluster (co)variances. Hence, unlike subspace K-means, such models do not provide direct insight into the relevance of the different factors for describing between-cluster differences (cluster centroids) or within-cluster differences (within-cluster residuals). Mixture models of this type include models with equal unique variances across clusters and variables (Sanguinetti, 2008; Yoshida, Higuchi, Imoto, & Miyano, 2006). The models that do not impose an equality constraint on the unique variances differ in whether (Galimberti, Montanari, & Viroli, 2008) or not (Baek, McLachlan, & Flack, 2010) they constrain the mean factor scores (i.e., mean across clusters) to zero.

Subspace clustering approaches have also been proposed for unsupervised learning, in which clusters are sought in high-dimensional spaces. This high-dimensionality (for example, in gene expression data, one may have over 20,000 genes to analyze) presents specific challenges to finding meaningful clusters because, in most cases, many of those dimensions are irrelevant for the clustering. For an excellent review of the different strategies and algorithms, we refer to Parsons, Haque, and Liu (2004), and for a recent search strategy to Ilies and Wilhelm (2010). The key difference between the subspace clustering approaches for unsupervised learning and the deterministic and stochastic models for clustering in reduced space is that the latter, unlike the former, involve the estimation of an explicit model, using a particular optimization function. A notable exception is K-subspace clustering (Wang, Ding, & Li, 2009), which boils down to a special case of subspace K-means, with centroids in full space and the within-models for the clusters in maximally two-dimensional space (i.e., $Qw \leq 2$).

Finally, feature selection approaches offer an alternative to subspace clustering in that they search for clusters in the full space, thereby differently weighting the variables involved. Many weighted K-means variants (e.g., Chan, Ching, Ng, & Huang, 2004; Cordeiro de Amorim & Mirkin, 2012; Huang, Ng, Hongqiang & Zichen, 2005; Jing, Ng, & Huang, 2007) have been proposed. Those methods all have an equal weighting for the centroids and within-cluster residuals, which makes them essentially different from subspace K-means.

To summarize, most mixture models involving dimension reduction impose constraints on $\boldsymbol{\Sigma}_c$. Models that apply factor analysis on $\boldsymbol{\mu}_c$ as well do not distinguish between-cluster factors and within-cluster factors, as subspace K-means does. In the class of deterministic models, dimension reduction takes place at either the between-cluster or the within-cluster level. Therefore, subspace K-means is the first model that explicitly summarizes between-cluster and within-cluster differences.

Data analysis

Loss function

To fit the subspace K-means model to observed data $\mathbf{X}$ ($I \times J$), we propose to minimize the following least

squares loss function, given specific values of C, Qb, and Qw:

$$L(\mathbf{m}, \mathbf{U}, \mathbf{Fb}, \mathbf{Ab}, \mathbf{Fw}^c, \mathbf{Aw}^c) = \sum_{i=1}^{I} \left\| (\mathbf{x}_i - \mathbf{m}) - \sum_{c=1}^{C} u_{ic} \left(\mathbf{fb}^c \, \mathbf{Ab}' + \mathbf{fw}_i^c \, \mathbf{Aw}^{c\prime} \right) \right\|^2, \quad (2)$$

where $\mathbf{U}$ ($I \times C$) is the binary partition matrix and $\mathbf{Fw}^c = \left[\mathbf{fw}_1^{c\prime} | \ldots | \mathbf{fw}_I^{c\prime} \right]'$ ($I \times Qw$) is the within-component matrix of cluster c, under the constraints that $\sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic} \, \mathbf{fb}^c = \mathbf{0}$ and $\sum_{i=1}^{I} \mathbf{fw}^c{}_i = \mathbf{0}$, with $\mathbf{fw}^c{}_i = \mathbf{0}$ if u_{ic}=0. This implies that the model parameters collected in $\mathbf{m}$, $\mathbf{U}$, $\mathbf{fb}^c$, $\mathbf{Ab}$, $\mathbf{Fw}^c$, $\mathbf{Aw}^c$ are estimated such that the sum-of-squared SKm residuals [i.e., $\sum_{i=1}^{I} \left(\mathbf{e}_i^c \right)^2$] is minimized. The zero constraints on the sum of the between- and within-component scores [i.e., $\sum_{i=1}^{I} \sum_{c=1}^{C} u_{ic} \, \mathbf{fb}^c = \mathbf{0}$ and $\sum_{i=1}^{I} \mathbf{fw^c_i} = \mathbf{0}$, with $\mathbf{fw}^c{}_i = \mathbf{0}$ if u_{ic} =0] ensure that the between- and within-parts are orthogonal and, hence, can be updated separately (see Timmerman, 2006, for a proof, in MLCA context).

To express the model fit, we consider the percentage of sum of squares accounted for by the model, Fit for short, which can be computed as

$$\text{Fit} = \frac{\|\mathbf{X}\|^2 - \text{L}}{\|\mathbf{X}\|^2} \times 100, \quad (3)$$

where L denotes the loss function value (see Eq. 2).

Algorithm

We propose to use a fast relocation algorithm to fit the subspace K-means model to data. As is generally the case for relocation algorithms (see Steinley, 2003), the subspace K-means algorithm may end in a local minimum, but a feasible alternative that is guaranteed to end in a global minimum is lacking. Specifically, to obtain for a given (C, Qb, Qw) a subspace K-means solution that minimizes the loss function (2), we use an alternating least squares procedure that consists of five steps, of which Steps 3 and 4 are similar to the ones taken in MLCA (Timmerman, 2006).

1. *Compute the variable mean vector* $\mathbf{m}$ as $\mathbf{m} = I^{-1} \, \mathbf{1}'\mathbf{X}$.
2. *Initialize the partition matrix* $\mathbf{U}$ by one of the following starting methods (see also Ceulemans, Van Mechelen, & Leenen, 2007):
   i. Random: randomly assign each of the I individuals to one of the C clusters, where each cluster has equal probability of being assigned to. If a cluster is empty, repeat this procedure until all clusters contain at least one individual.
   ii. Partition from alternative cluster analysis: Perform an alternative cluster analysis (e.g., K-means, RKM) on $\mathbf{X}$, and use the estimated partition matrix $\mathbf{U}_{Alt}$ as $\mathbf{U}$.
   iii. Perturbed partition from alternative cluster analysis: Reallocate 10 % of the individuals, randomly drawn, in $\mathbf{U}_{Alt}$ to a randomly selected different cluster. Use the resulting matrix as $\mathbf{U}$.
3. *Estimate the between-part*: Compute $\mathbf{B}= \mathbf{UC}$, with the cth row of $\mathbf{C}$ as $\mathbf{c}_c = \frac{1}{1c} \sum_{i=1}^{I} u_{ic}(\mathbf{x}_i - \mathbf{m})$, c = 1,…, C. Perform a singular value decomposition on $\mathbf{B}$—that is, $\mathbf{B} = \mathbf{PDQ}'$, with $\mathbf{P}'\mathbf{P} = \mathbf{I}$, $\mathbf{Q}'\mathbf{Q} = \mathbf{QQ}' = \mathbf{I}$, and $\mathbf{D}$ a diagonal matrix with diagonal elements in descending order. Update $\mathbf{Fb}$ and $\mathbf{Ab}$ as $\mathbf{Fb} = \mathbf{P}_{Qb}\mathbf{D}_{Qb}$ and $\mathbf{Ab} = \mathbf{Q}_{Qb}$, with $\mathbf{P}_{Qb}$ (resp. $\mathbf{Q}_{Qb}$) a matrix consisting of the first Qb columns of $\mathbf{P}$ (resp. $\mathbf{Q}$), and $\mathbf{D}_{Qb}$ a matrix containing the first Qb columns and rows of $\mathbf{D}$; $\mathbf{Fb}$ ($I \times Qb$) is the between-component scores matrix, with row i equal to $\mathbf{fb}^c$ if i belongs to c (c = 1,…, C).
4. *Estimate the within-part*: For each cluster c, compute the within-cluster residual matrix $\mathbf{W}^c$ ($I \times$ J) with the ith row of $\mathbf{W}^c : \mathbf{w}_i^c = u_{ic} \, (\mathbf{x}_i - \mathbf{m} - \mathbf{fb}^c\mathbf{Ab}')$, i = 1,…, I. Perform a singular value decomposition on $\mathbf{W}^c$—that is, $\mathbf{W}^c = \mathbf{PDQ}'$. Estimate $\mathbf{Fw}^c$ as $\mathbf{Fw}^c = \mathbf{P}_{Qw} \, \mathbf{D}_{Qw}$ and $\mathbf{Aw} = \mathbf{Q}_{Qw}$.
5. *Reestimate the partition matrix* $\mathbf{U}$: For each individual i, the fit to each of the C (c = 1,…, C) clusters is assessed as $L_{ic} = \left\| \mathbf{x}_i - \mathbf{m} - \mathbf{fb}^c\mathbf{Ab}' - \widetilde{\mathbf{fw}}_i^c \mathbf{Aw}^{c\prime} \right\|^2$, with $\widetilde{\mathbf{fw}}_i^c$ computed as $(\mathbf{x}_i - \mathbf{m} - \mathbf{fb}^c \, \mathbf{Ab}')\mathbf{Aw}^c$ (i.e., the optimal estimate in least-squares sense). Subsequently, individual i is assigned to the cluster for which L_{ic} is minimal.

   Steps 3 to 5 are repeated until convergence—for example, until the decrease of the loss function value L (Eq. 2) for the current iteration is smaller than 1e – 6. To reduce the probability that one ends in a local minimum, we strongly recommend the use of multiple and various types of initializations. For example, in an empirical subspace K-means analysis, we use as starts the solutions of K-means, FKM, RKM, MFA, and MCLUST, 500 perturbed versions of each of them, and 500 random starts.
6. *Select the best-fitting solution*: Select from the solutions resulting from the various initializations, the solution with the highest Fit, and take the associated estimates as the subspace K-means solution.

Model selection

To estimate a subspace *K*-means model, the numbers of clusters *C*, between-components *Qb*, and within-components *Qw* have to be specified. Ideally, both substantive considerations and formal criteria would play a role in deciding which (*C*, *Qb*, *Qw*) values are reasonable for a given data set. The substantive considerations for deciding upon reasonable (*C*, *Qb*, *Qw*) values should comprise the theoretical background of the data and the interpretability of the resulting models. To what extent those considerations guide in solving the selection problem depends on the nature of the empirical data analysis problem at hand—that is, whether it is more exploratory or confirmatory in nature. For instance, one may have available knowledge on the number of clusters. In empirical practice, the numbers of between- and within-components will almost always have to be estimated.

As is indicated in the Subspace *K*-Means section, the partition of two (or more) clusters is nonunique when the centroids and within-models of two (or more) clusters share exactly the same subspace. This means that in empirical practice, the partitioning may be arbitrary if the estimated within-subspaces and the between-subspace overlap considerably. Therefore, it is important to examine the amount of overlap among the obtained between- and within-subspaces. Overlap can be assessed by computing, for each pair of loading matrices, the congruence coefficient (Tucker, 1951) between each pair of columns of the loading matrices concerned, after orthogonal Procrustes rotation (to account for arbitrary differences in axes position). In general, the probability of overlap and, hence, of an arbitrary partitioning becomes higher with increasing dimensionalities of the modeled between- and within-subspaces. Because it is commonly desirable to have sufficient separation among cluster centroids, one would generally favor a reduction in the size of within-subspaces, rather than of the between-subspace. In the most extreme case—namely, when considerable overlap between within- and between-subspaces still exists with a single within-component—it will be necessary to use zero within-components.

A formal criterion: the CHull procedure

In addition to substantive considerations, a formal criterion is useful to guide the choice of the *C*, *Qb*, and *Qw* values. In case one has theory-based hypotheses about these values, the application of a formal criterion may also help to prevent overinterpreting a model. As a formal criterion, we propose to use the CHull procedure (Ceulemans & Kiers, 2006). CHull has proved to be a powerful tool for solving a wide variety of model selection problems (Ceulemans & Kiers, 2006, 2009; Ceulemans, Timmerman, & Kiers, 2011; Lorenzo-Seva, Timmerman, & Kiers, 2011; Schepers, Ceulemans, & Van Mechelen, 2008), among which is MFA (Bulteel, Wilderjans, Tuerlinckx, & Ceulemans, in press). CHull is a numerical procedure that aims at selecting a model with an optimal balance between some goodness-of-fit measure and some complexity measure. Software for applying CHull is available (Wilderjans, Ceulemans, & Meers, in press).

To apply the CHull procedure in subspace *K*-means analysis, the fit has to be estimated for a sensible range of models—that is, with different values of *C*, *Qb*, and *Qw*. The maximal number of clusters to consider can be based on theoretical knowledge and/or related to the number of observations involved (e.g., about *I*/30). The maximal number of between-components equals min($C - 1$, *J*) (see the Subspace *K*-Means section). As was explained above, the maximal number of within-components should be kept low and can be chosen as a function of the number of observed variables (e.g., about *J*/4) or, presumably seldom, on the basis of expectations about the within-cluster structures. As fit, we use the Fit (see Eq. 3). As a complexity measure, we take the sum of the numbers of components—that is, $sum = Qb + CQw$—analogous to the most successful strategy in two-mode clustering (Schepers et al., 2008).

Simulation study: sensitivity to local minima and goodness of recovery

Questions and hypotheses

To evaluate the absolute and relative performance of the subspace *K*-means algorithm, we performed a simulation study. In this study, we assessed the sensitivity of subspace *K*-means to local minima and the quality of the obtained partition. We also compared the quality of the subspace *K*-means partition with that of five other clustering methods. Specifically, we selected five methods that model similar aspects as subspace *K*-means but also differ considerably from each other in that some are deterministic and some stochastic and in that some have the centroids in reduced space and others the within-residuals. The competitor methods were the deterministic *K*-means, RKM (e.g., De Soete & Carroll, 1994), and FKM (Vichi & Kiers, 2001) methods and the stochastics MCLUST (Banfield & Raftery, 1993) and MFA (McLachlan & Peel, 2000). RKM and FKM model the centroids in reduced space, and MCLUST and MFA the within-residuals, while *K*-means models in full space.

Moreover, we also inspected whether the performance of the six clustering methods considered depend on the following three data characteristics: (1) congruence, (2) between-variance, and (3) error. We selected those factors because they are theoretically interesting and, for the latter two, their influence on clustering algorithms is known. Factor 1

(congruence) pertains to the amount of overlap between the within-cluster subspaces (i.e., $\mathbf{fw}_i^c\mathbf{Aw}^{c\prime}$). This overlap is operationalized in terms of the congruence (Tucker, 1951) between the within-loading matrices. We expect the performance of subspace *K*-means to deteriorate when the within-loading matrices are more congruent. Factor 2 is the percentage of between-variance relative to the total of the between- and within-model variances (i.e., excluding the variance of the S*K*m residuals $\mathbf{e}_i^c$, because this variance is manipulated separately). To manipulate the between-variance factor, we varied the distances between the between-component scores ($\mathbf{fb}^c$, $c = 1,\ldots, C$). If the within-cluster residuals (i.e., $\varepsilon_i^c = \mathbf{fw}_i^c\mathbf{Aw}^{c\prime} + \mathbf{e}_i^c$) of the different clusters are (partly) in the same subspace, decreasing the distances between centroids and, thus, decreasing the percentage of between-variance results in increasing overlap between clusters. Therefore, we hypothesize the recovery performance to deteriorate with decreasing percentages of between-variance, with larger effects in case of higher congruence. For factor 3, the percentage of error variance, we expect that the higher the amount of error, the worse the recovery (cf. Brusco & Cradit, 2005).

To keep the study feasible, other properties of the simulated data were not varied. Specifically, we fixed the number of individuals, clusters, components, and variables, because those factors appeared to affect the performance of *K*-means, RKM, and/or FKM to a relatively minor extent only (e.g., Steinley, 2006; Timmerman, Ceulemans, Kiers, & Vichi, 2010).

Design and procedure

The number of individuals was fixed at $I = 200$, which is within the range of 50–300 that is commonly found in clustering simulations (Steinley, 2003). Noting that the number of clusters is often taken between two and eight in simulations (Steinley & Brusco, 2011a, b), we used four clusters. The number of observed variables was set at 15, and the numbers of between- and within-components at two. Each between-loading and within-loading matrix was an orthogonal matrix. Moreover, the between-loading matrix was generated orthogonal to the within-loading matrices, to ensure that the centroids and within-model would be in different subspaces. The within-component scores $\left(\mathbf{fw}_i^c\right)$ and the S*K*m residuals $\left(\mathbf{e}_i^c\right)$ were drawn from multivariate normal distributions with mean zero and diagonal covariance matrices, with, respectively, variances equal to one, and dependent on the error (factor 3). We did not use alternative distributions, like a uniform or triangular distribution, since for *K*-means, this had a relatively small impact on partition recovery (Steinley, 2006) and we expect this to hold for subspace *K*-means as well.

The three experimental factors were varied in a following full factorial design.

1. Congruence: Minimal, maximal. We selected the two most extreme conditions, with congruences (Tucker, 1951) between the columns of the within-loading matrices equal to zero and one, respectively. This was achieved by taking the within-loading matrices ($\mathbf{Aw}^c$, $c = 1,\ldots, C$) either columnwise orthogonal (minimal—i.e., congruence zero) or completely equal (maximal—i.e., congruence one). The elements of these within-loading matrices were taken as the orthonormal basis of a matrix with elements from a uniform [0,1] distribution. Note that in the condition with maximal congruence, the data were in fact simulated according to a constrained version of subspace *K*-means (i.e., with $\mathbf{Aw}^c = \mathbf{Aw}$, for all $c = 1,\ldots, C$).
2. Between-variance: 50 %, 25 %, 15 %. This percentage was defined as the expected percentage of variance of the between-component scores (i.e., of **Fb**) relative to the total variance of the between- and within-component scores (i.e., of **Fb** and $\mathbf{Fw}^c$, $c = 1,\ldots, C$ jointly). In other words, this percentage expresses how much of the modeled variance is due to between-cluster differences. To this end, we manipulated the joint overlap between the four clusters in the two-dimensional space using the OCLUS procedure (Steinley & Henson, 2005). The between-variance percentages of 50, 25, and 15 correspond to proportions of joint overlap (in the two-dimensional space) of .25, .48, and .60, respectively.
3. Error: 10 %, 30 %. This is the expected percentage of variance of the S*K*m residuals $\left(\mathbf{e}_i^c\right)$ relative to the total variance of the observed data.

For each cell of the factorial design, we generated 20 simulated data matrices according to Eq. (1), yielding $3 \times 2 \times 2 \times 20 = 240$ simulated data matrices in total. Each data matrix was analyzed with the *K*-means algorithm (using the built-in *K*-means function in MATLAB R2010a), FKM and RKM (using functions written for MATLAB R2010a), MFA (using the EMFAC software package; McLachlan et al., 2003) and MCLUST (McLachlan et al., 2003), and the subspace *K*-means algorithm (using functions written for MATLAB R2010a). For each model, we used the correct model specification—that is, the correct number of clusters in all models, the correct number of between-components for FKM, RKM, and subspace *K*-means, and the correct numbers of within-factors for MFA and within-components for subspace *K*-means. With regard to the within-cluster covariance structure, we let MCLUST decide which structure is the most appropriate for a specific data set (which is the default option in the MCLUST package). *K*-means was run using 1,000 random starts. For MFA, we used 501 starts: the best *K*-means solution and 500 random starts. To obtain RKM and FKM

solutions, we ran the associated algorithms 1,001 times: once from the best *K*-means partition, 500 random runs, and 500 perturbed *K*-means runs. The Subspace *K*-means algorithm was run using 3,005 starts of 12 different types—namely, the *K*-means, FKM, RKM, MFA and MCLUST (5) starts, 500 perturbed versions of each of them ($5 \times 500 = 2{,}500$), and 500 random starts. From these 3,005 solutions, we retained the best one (Step 6 of the algorithm). Moreover, we ran the algorithm also once from the true partition, to assess the sensitivity to local minima.

Results

Sensitivity to local minima To evaluate the sensitivity of the subspace *K*-means algorithm to local minima, the loss function value of the best of the 3,005 obtained solutions should be compared with that of the global minimum. Since the latter is unknown, we resort to a comparison with a proxy. We took the proxy as the minimal loss function value across all runs of the subspace *K*-means algorithm (i.e., including the true started runs).

First, we evaluated whether the best-fitting solution had a loss function value higher than the proxy, which implies that this solution is a local minimum for sure. This occurred in 12 (5.0 %) of all 240 data sets, and only in the conditions with maximal congruence, 30 % error, and 25 % or 15 % between-variance.

To assess the sensitivity to local minima further, Table 1 presents, per cell of the design, the percentage of data sets for which the 11 types of starts yielded a solution with a loss value equal to the proxy. It can be concluded that these percentages depend on the type of start and the condition under study. For all but the FKM-based types of starts, the percentages decrease with increasing congruence, increasing error, and decreasing between-variance. In contrast, the success rates of the FKM-based starts improve with decreasing Between variance. Comparing the different types of starts across all conditions learns that the perturbed MFA and perturbed MCLUST starts perform best (both yield for 73 % of the data sets solutions that are equal to the proxy), followed by perturbed RKM (66 %) and perturbed *K*-means (65 %). FKM-based starts and random starts perform well in specific conditions (e.g., minimal congruence and large within variance).

Overall, the results show that the subspace *K*-means algorithm appears sensitive to local minima, where no type of start performs reasonably well across all conditions considered. To reduce the probability to end in a local minimum, we recommend using starts of different types—that is, the *K*-means, FKM, RKM, MFA and MCLUST partition, and perturbed versions of those starts (e.g., 500 starts each) and random starts (e.g., 500).

Recovery of partition We studied how well subspace *K*-means, *K*-means, FKM, RKM, MFA, and MCLUST succeeded in recovering the correct partition. We assessed the cluster membership recovery by computing the Adjusted Rand Index (ARI; Hubert & Arabie, 1985) between the true and estimated partition matrices. The ARI has a maximal value of 1 in case of perfect agreement; a value of 0 indicates that the partition matrices do not correspond more than expected by chance.

In Table 2, we present the means (and standard errors) of the ARI, per cluster method, both per cell in the experimental design and across all conditions. Comparing for each condition, the performances across cluster methods show that subspace *K*-means and MFA perform generally very well, and rather similarly. Only in the maximal congruence conditions do substantial differences occur. That is, subspace *K*-means clearly outperforms MFA in the "more difficult" conditions—that is, with decreasing between-variance and increasing error variance. Furthermore, the condition with an error variance of 30 % and a small between-variance (15 %) poses a real challenge for all methods involved. RKM, MCLUST, and *K*-means perform well (e.g., ARI > 0.95) in a very limited set of conditions—namely, with between-variance equal to 50 %. FKM performs badly in all conditions.

To summarize the findings on recovery, subspace *K*-means performs very well across all conditions investigated. MFA appears second best, with a very good performance in case the between- and within-subspaces are all orthogonal toward each other, but deteriorates in cases in which the within subspaces are equal for all clusters.

Illustrative application: parenting types and dimensions

To illustrate the merits of subspace *K*-means, we reanalyzed data from a study on parenting. Research in this field typically uses either a variable-centered approach or a person-centered approach. In the variable-centered approach, one typically uses component or factor analysis to obtain a limited number of parenting dimensions that account for most of the interindividual differences in frequency and/or relevance of a broad set of parenting behaviors. Well-known examples of such parenting dimensions are support, behavioral control, and psychological control (Galambos, Barker, & Almeida, 2003). In contrast, in the person-centered approach, one is interested in distinguishing (mutually exclusive) parenting types, each referring to particular combinations of parenting behavior. An example is the distinction between authoritarian, authoritative, permissive, and indifferent parents (Maccoby & Martin, 1983). To obtain such parenting types, one often uses cluster analysis.

In their discussion of the strengths and limitations of variable- and person-centered approaches, Mandara (2003)

Table 1 Percentages of the data sets with a loss value equal to the proxy, per type of start, per cell in the experimental design (based on 20 replicates per cell)

Congruence	Minimal						Maximal					
Between variance (%)	50		25		15		50		25		15	
Error (%)	10	30	10	30	10	30	10	30	10	30	10	30
Random	0	0	85	80	85	65	0	0	90	0	65	5
Pert. *K*-means	100	100	85	75	85	65	100	60	65	0	50	0
K-means	100	100	55	30	60	5	100	10	0	0	0	0
Pert. RKM	100	100	85	80	80	40	100	95	55	0	60	0
RKM	100	100	75	45	45	5	100	45	0	0	0	0
Pert. FKM	0	0	95	75	80	65	0	0	55	0	25	5
FKM	0	0	10	0	5	5	0	0	5	0	0	0
Pert. MFA	100	95	85	80	85	85	70	45	75	75	45	35
MFA	95	85	85	55	85	20	70	20	75	0	20	0
Pert. MCLUST	100	100	85	80	85	80	100	80	80	5	65	15
MCLUST	100	100	60	30	80	40	90	35	35	0	0	00

Pert., perturbed; RKM, reduced *K*-means; FKM, factorial *K*-means; MFA, mixtures of factor analyzers

and Bergman and Andersson (2010) pointed out that a strictly variable- or person-centered approach narrows the possibilities of understanding a complex reality and, therefore, hampers progress in research on child and family psychology. Since subspace *K*-means combines both approaches (i.e., clustering of parents and reduction of parenting behaviors to broader parenting dimensions), we expect that it will provide more fine-grained information on parenting than could be obtained with either factor or cluster analysis alone, which will be useful in understanding child functioning.

The data that we will reanalyze, are taken from Van Leeuwen, Mervielde, Braet, and Bosmans (2004). These authors asked both parents of 600 children (47 % boys, 53 % girls) 7 to 15 years of age (M = 10.9, SD = 1.8), who were recruited on the basis of a stratified random sampling procedure (for a detailed discussion, see Van Leeuwen et al., 2004), to independently complete a series

Table 2 Mean (standard error) of the ARI for each cluster method, per cell in the experimental design (based on 20 replicates per cell), and across all conditions

Loading congruence	Minimal						Maximal						Across all conditions
Between variance (%)	50		25		15		50		25		15		
Error (%)	10	30	10	30	10	30	10	30	10	30	10	30	
Subspace *K*-means	***1.00***	***1.00***	***1.00***	***.97***	***1.00***	**.93**	***1.00***	***.99***	*.98*	**.85**	*.96*	**.42**	**.93**
	(.00)	(.00)	(.00)	(.00)	(.00)	(.01)	(.00)	(.00)	(.02)	(.01)	(.02)	(.02)	(.01)
K-means	***1.00***	*.98*	.47	.42	.25	.22	***1.00***	.76	.03	.04	.01	.01	.43
	(.00)	(.01)	(.01)	(.01)	(.02)	(.01)	(.00)	(.05)	(.01)	(.01)	(.00)	(.00)	(.03)
RKM	***1.00***	*.99*	.83	.60	.18	.13	***1.00***	***.99***	.01	.01	.01	.00	.48
	(.00)	(.00)	(.05)	(.04)	(.02)	(.01)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.03)
FKM	.00	.01	.00	.01	.01	.01	.01	.01	.01	.01	.00	.01	.01
	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)	(.00)
MFA	*.99*	*.95*	***1.00***	***.97***	**1.00**	**.93**	.90	.81	***1.00***	.77	***.99***	.37	.89
	(.00)	(.03)	(.00)	(.00)	(.00)	(.01)	(.04)	(.04)	(.00)	(.02)	(.00)	(.02)	(.01)
MCLUST	***1.00***	*.98*	.78	.28	.79	.09	*.95*	.84	.29	.07	.07	.04	.52
	(.00)	(.00)	(.04)	(.03)	(.05)	(.04)	(.04)	(.05)	(.07)	(.02)	(.02)	(.00)	(.03)

Best cluster method(s) per condition and across conditions are printed in bold face; ARI values > 0.95 are printed in italic. RKM, reduced *K*-means; FKM, factorial *K*-means; MFA, mixtures of factor analyzers

of questionnaires, covering parenting behavior, child problem behavior, and child and parent personality. In our analysis, we focus on the mother ratings of the Parental Behavior Scale (PBS; Van Leeuwen & Vermulst, 2004), a questionnaire designed to assess parenting behavior on the basis of social learning theory (Patterson, Reid, & Dishion, 1992). These ratings indicate how frequently the mothers display a number of parenting behaviors toward the target child, on a 5-point Likert scale ranging from *never* to *always*. The parenting behaviors are assigned to nine subscales (with item examples in brackets): positive parental behavior ("I make time to listen to my child, when he/she wants to tell me something"), autonomy ("I teach my child that he/she is responsible for his/her own behavior"), rules ("I teach my child to obey rules"), monitoring ("I keep track of the friends my child is seeing"), discipline ("When my child has done something wrong, I punish him/her by taking away something nice [for instance, the child can't watch TV, . . .]"), harsh punishment ("I slap my child when he/she has done something wrong"), ignoring unwanted behavior ("When my child does something that is not allowed, I only talk to him/her again when he/she behaves better"), inconsistent discipline ("When I have punished my child, it happens that I let my child out of the punishment early"), and material rewarding ("I give my child money or a small present when he/she has done something that I am happy about"). Evidence for an adequate internal consistency of these PBS subscales has been provided by Van Leeuwen and Vermulst.

For our analysis, we standardized the 600 mothers × 9 PBS scales data matrix and discarded the 9 mothers for whom some PBS scores were missing. Since we did not have strong theory-based hypotheses on the number of clusters and/or dimensions to expect, we applied the CHull procedure, considering subspace *K*-means models with *C* varying from 2 to 5, *Qb* varying from 1 to 4, and *Qw* varying from 1 to 4. For each subspace *K*-means analysis, we used 3,005 starts (i.e., of *K*-means, SKM, RKM, MFA, MCLUST, 500 perturbed versions of each of them, and 500 random starts). The CHull procedure indicates the ($C = 2$, $Qb = 1$, $Qw = 4$) solution, since this solution is located at the elbow in the higher boundary of the convex hull plot.

The selected solution distinguishes two parenting types. To detect possible nonuniqueness of the partitioning, we first assessed the degree of overlap between the two within-cluster spaces. The congruence coefficients computed between the four columns of the within-loading matrices (after Procrustes rotation) ranged from 0.39 to 0.59, indicating moderate degrees of overlap at maximum, suggesting that no uniqueness problems are to be expected.

To interpret the two parenting types, we examine the loadings of the nine parenting scales on the single between-component, which are given in Table 3. This between-component mainly shows a contrast between rules and harsh punishment, having highly positive and highly negative loadings on this between-component, which implies that the two parenting types mainly differ as to what extent they lay down rules versus punish harshly. The between-component scores amount to 0.57 and −0.44, and we labeled the two obtained types as *authoritative parenting* and *authoritarian parenting*, respectively.

The interpretation of the two parenting types was validated by comparing the two types with respect to stress in the parent–child relationship. The amount of stress was assessed with the Dutch version of the Parenting Stress Index (de Brock, Vermulst, Gerris, & Abidin, 1992), including a parent domain score, referring to stress evoked by characteristics of the parent (e.g., feelings of incompetence). The authoritarian parenting mothers experience significantly (taking $\alpha = .05$) more parent domain stress than do the authoritative parenting mothers, $t\,(549.3) = -2.44$ $p = .02$, as would be expected.

The within-loadings per cluster offer further insight into the parenting differences between mothers who belong to the same parenting type. To facilitate the interpretation, the within-loading matrices were Varimax rotated. As can be seen in Table 3, the broader dimensions summarizing the differences among mothers within the same parenting type do differ across the types. We first discuss the interpretation of the within-component per parenting type.

For the authoritative parenting type, the first component mainly includes positive parental behavior, monitoring (as a way of showing interest in the child), and material rewarding. Positive parental behavior consists (among others) of items referring to social rewarding (e.g., compliment the child), and therefore, it is not illogical that it is correlated with material rewarding. We label this component *support*. The second component is labeled *autonomy*, since it mainly refers to the autonomy subscale. Autonomy can be viewed as being related to support. The third component refers to inconsistent discipline and ignoring of unwanted behavior; therefore, it can be identified as lack of behavioral control. The fourth component shows a contrast between discipline and inconsistent discipline and can be considered as behavioral control—that is, "parental behaviors that attempt to control or manage children's behavior" (Barber, 1996, p. 3296). In short, the authoritative parenting type includes two components that are related to support, whereas the two other components are related to behavioral control, thus reflecting the assumed balance between support and demandingness in authoritative parenting (Darling & Steinberg, 1993).

For the authoritarian parenting type, the first component mainly refers to harsh punishment and ignoring of unwanted behavior. The latter scale in the PBS has items such as "When my child does something that is not allowed, I give him/her an angry look and pretend he/she is not there." In

Table 3 Subspace *K*-means solution of the parenting data: Between- and within-loadings of two clusters

	Between- loadings	Within-loadings Cluster 1 (n = 261) (Authoritative parenting)				Within-loadings Cluster 2 (n = 330) (Authoritarian parenting)			
	Rules vs. harsh punishment	Support	Autonomy	Lack of beha-vioral control	Beha-vioral control	Harsh discipline	Positive parental behavior / Autonomy	Proactive beha-vioral control	Reactive beha-vioral control
Harsh punishment	**−.70**	−0.07	0.01	0.02	0.00	**0.79**	0.03	0.05	0.03
Rules	**.58**	0.12	0.19	−0.01	0.01	−0.02	0.33	**0.62**	0.10
Monitoring	.11	**0.64**	0.21	0.06	0.14	−0.01	−0.22	**0.68**	−0.10
Positive parental behavior	.35	**0.48**	0.09	−0.11	−0.05	−0.16	**0.62**	0.14	0.21
Material rewarding	−.07	**0.51**	−0.08	0.32	−0.10	−0.04	0.13	−0.26	**0.72**
Discipline	.06	−0.15	0.07	0.25	**0.78**	0.28	−0.01	0.17	**0.40**
Inconsistent discipline	.12	−0.17	0.10	**0.39**	**−0.59**	−0.12	−0.43	0.17	**0.50**
Ignoring unwanted behavior	.10	−0.06	−0.03	**0.82**	0.06	**0.49**	−0.01	−0.07	−0.01
Autonomy	.01	−0.14	**0.94**	−0.01	−0.03	0.08	**0.52**	−0.06	−0.10

Loadings >|0.40| are indicated in boldface

Barber's view (1996), this can be seen as an example of "love withdrawal" (withdrawing love or attention if a family member did not do what the other expected), which is one of the components of psychological control ("control attempts that intrude into the psychological and emotional development of the child"; Barber, 1996, p. 3296). Therefore, we label this component as *harsh* discipline. The second component mainly pertains to positive parental behavior and autonomy. Those subscales refer to particular types of support, which are not necessarily related in general among parents, and therefore we label this component as *positive parental behavior/autonomy*. Components 3 and 4 are both related to behavioral control. One component can be labeled as *proactive behavioral control*, with parenting behaviors that prevent unwanted behavior, such as setting rules and monitoring. The other component refers to reactive behavioral control, including parenting behaviors related to discipline and material rewarding, which are actions that directly follow desired or undesired child behavior in order to readjust the child's behavior.

Comparing the two parenting types shows that the parental behaviors correlate in a different way. It is particularly striking that among the authoritative parenting mothers, levels of monitoring, positive parental behavior, and material rewarding are substantially (positively) correlated. In contrast, among the authoritarian parents, levels of monitoring appear to be substantially (positively) correlated with rules only. As was already concluded from different analyses of the same data (Van Leeuwen & Vermulst, 2004), and based on previous research (Jacob, Moser, Windle, Loeber, & Stouthamer-Loeber, 2000), monitoring appears to have a twofold meaning. For the authoritative parenting mothers, it can be viewed as an expression of care and attention of parents toward their children, whereas for the authoritarian parents, it is an expression of control to prevent undesired behavior.

As a second example, consider discipline and inconsistent discipline, which have relatively large within-loadings for both parenting types. For the authoritative parenting mothers, those loadings are in opposite sign, which implies that, for mothers of this type, discipline is negatively correlated with inconsistent discipline. In contrast, for the authoritarian parenting mothers, the sign is equal, implying a positive correlation between discipline and inconsistent discipline. Thus, authoritative parenting mothers who discipline relatively much appear to be more consistent in their discipline behavior than the authoritarian parenting mothers. Authoritative parenting mothers who discipline relatively little appear to be relatively inconsistent. A possible reason is that the undesired behavior of the child concerned is not that serious, or because the mother is not inclined to discipline at all and only threatens punishment.

Discussion

Behavioral research questions may pertain to the identification of subgroups of individuals with similar characteristics in multivariate data. In this article, we proposed subspace *K*-means as an alternative to existing cluster procedures. Indeed, the results of our simulation study indicate that

subspace *K*-means outperforms the competitor cluster procedures examined, which covered a broad range of (subspace) clustering methods available. Unlike any other existing procedure, subspace *K*-means models both the centroids and within-cluster residuals in subspaces, which may facilitate and enrich the interpretation. Considering the centroids in a subspace rather than the full space has the advantage that fewer scores are to be compared between clusters (i.e., the scores on between-components, instead of scores on separate variables) and that the relevance of the different variables in separating the (centroids of the) clusters becomes clear. Furthermore, subspace *K*-means may provide insight into the most eminent differences between individuals belonging to the same cluster. As such, subspace *K*-means might show that different within-residual subspaces are important for different clusters. To facilitate the use of subspace *K*-means in empirical practice, we made available subspace *K*-means software at http://www.rug.nl/research/heymans-institute/research/psychometrics_and_statistics/software/.

To enlarge the application range of subspace *K*-means, it could be useful to extend the model. We see different potentially useful options that may improve the recovery and/or may provide additional interpretational advantages. First, one could consider different numbers of within-components across clusters, rather than a fixed number for all clusters. Second, one could impose equality constraints across subspaces of the modeled within-cluster residuals. These constraints can take different forms, ranging from full equality across all clusters to subsets of clusters to partly overlapping subspaces. Although those extensions are relatively easy to define from a mathematical point of view, they are not so easy to work with in empirical practice, since they involve intricate model estimation and model selection problems.

Author Note This research 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 K.U. Leuven (GOA/10/02).

References

Baek, J., McLachlan, G. J., & Flack, L. K. (2010). Mixtures of factor analyzers with common factor loadings: Applications to the clustering and visualization of high-dimensional data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32*(7), 1298–1309.

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and Non-Gaussian clustering. *Biometrics, 49*(3), 803–821.

Barber, B. K. (1996). Parental psychological control: Revisiting a neglected construct. *Child Development, 67*(6), 3296–3319. doi:10.1111/j.1467-8624.1996.tb01915.x

Bergman, L. R., & Andersson, H. (2010). The person and the variable in developmental psychology. *Journal of Psychology, 218,* 155–165.

Bock, H. H. (1987). On the interface between cluster analysis, principal component analysis, and multidimensional scaling. In H. Bozdogan & A. K. Gupta (Eds.), *Multivariate statistical modeling and data analysis* (pp. 17–34). Dordrecht: D. Reidel Publishing Company.

Böckenholt, U., & Böckenholt, I. (1991). Constrained latent class analysis: Simultaneous classification and scaling of discrete choice data. *Psychometrika, 56*(4), 699–716. doi:10.1007/BF02294500

Bouveyron, C., Girard, S., & Schmid, C. (2007). High-dimensional data clustering. *Computational Statistics & Data Analysis, 52*(1), 502–519. doi:10.1016/j.csda.2007.02.009

Brusco, M. J., & Cradit, D. J. (2005). ConPar: A method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses. *Journal of Mathematical Psychology, 49*(2), 142–154. doi:10.1016/j.jmp.2004.11.004

Bulteel, K., Wilderjans, T. F., Tuerlinckx, F., & Ceulemans, E. (in press). CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers. *Behavior Research Methods.*

Ceulemans, E., & Kiers, H. A. L. (2006). Selecting among three-mode principal component models of different types and complexities: A numerical convex hull based method. *British Journal of Mathematical and Statistical Psychology, 59*(1), 133–150. doi:10.1348/000711005X64817

Ceulemans, E., & Kiers, H. A. L. (2009). Discriminating between strong and weak structures in three-mode principal component analysis. *British Journal of Mathematical and Statistical Psychology, 62*(3), 601–620. doi:10.1348/000711008X369474

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. doi:10.1016/j.chemolab.2010.08.001

Ceulemans, E., Van Mechelen, I., & Leenen, I. (2007). The local minima problem in hierarchical classes analysis: An evaluation of a simulated annealing algorithm and various multistart procedures. *Psychometrika, 72*(3), 377–391. doi:10.1007/s11336-007-9000-9

Chan, E. Y., Ching, W. K., Ng, M. K., & Huang, J. Z. (2004). An optimization algorithm for clustering using weighted dissimilarity measures. *Pattern Recognition, 37*(5), 943–952. doi:10.1016/j.patcog.2003.11.003

Cordeiro de Amorim, R., & Mirkin, B. (2012). Minkowski metric, feature weighting and anomalous cluster initializing in K-Means clustering. *Pattern Recognition, 45*(3), 1061–1075. doi:10.1016/j.patcog.2011.08.012

Darling, N., & Steinberg, L. (1993). Parenting style as context: An integrative model. *Psychological Bulletin, 113*(3), 487–496. doi:10.1037/0033-2909.113.3.487

de Brock, A. J. L. L., Vermulst, A. A., Gerris, J. R. M., & Abidin, R. R. (1992). *NOSI. Nijmeegse Ouderlijke Stress Index. Handleiding experimentele versie*. Lisse: Swets & Zeitlinger.

De Soete, G., & Carroll, J. D. (1994). K-means clustering in a low-dimensional Euclidean space. In E. Diday, Y. Lechevallier, M. Schader, P. Bertrand, & B. Burtschy (Eds.), *New approaches in classification and data analysis* (pp. 212–219). Heidelberg: Springer.

DeSarbo, W., Howard, D., & Jedidi, K. (1991). Multiclus: A new method for simultaneously performing multidimensional scaling and cluster analysis. *Psychometrika, 56*(1), 121–136. doi:10.1007/BF02294590

Galambos, N. L., Barker, E. T., & Almeida, D. M. (2003). Parents do matter: Trajectories of change in externalizing and internalizing problems in early adolescence. *Child Development, 74*(2), 578–594. doi:10.1111/1467-8624.7402017

Galimberti, G., Montanari, A., & Viroli, C. (2008). Latent classes of objects and variable selection. 373–383.

Huang, J. Z., Ng, M. K., Hongqiang, R., & Zichen, L. (2005). Automated variable weighting in k-means type clustering. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 27*(5), 657–668.

Hubert, L. J., & Arabie, P. (1985). Comparing partitions. *Journal of Classification, 2,* 193–218.

Ilies, I., & Wilhelm, A. (2010). Projection-based partitioning for large, high-dimensional datasets. *Journal of Computational and Graphical Statistics, 19*(4), 474–492. doi:10.1198/jcgs.2010.194in

Jacob, T., Moser, R. P., Windle, M., Loeber, R., & Stouthamer-Loeber, M. (2000). A new measure of parenting practices involving pre-adolescent and adolescent-aged children. *Behavior Modification, 24*(5), 611–634. doi:10.1177/0145445500245001

Jing, L., Ng, M. K., & Huang, J. Z. (2007). An entropy weighting k-Means algorithm for subspace clustering of high-dimensional sparse data. *Knowledge and Data Engineering, IEEE Transactions on, 19*(8), 1026–1041.

Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer-Verlag.

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. L. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research, 46*(2), 340–364. doi:10.1080/00273171.2011.564527

Maccoby, E. E., & Martin, J. A. (1983). Socialization in the context of family: parent–child interaction. In P. H. Mussen (Ed.), *Handbook of child psychology. Vol. IV. Socialization, personality and social development.* (pp. 1–101)

Mandara, J. (2003). The typological approach in child and family psychology: A review of theory, methods, and research. *Clinical Child and Family Psychology Review, 6*(2), 129–146. doi:10.1023/A:1023734627624

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

McLachlan, G. J., Peel, D., & Bean, R. W. (2003). Modelling high-dimensional data by mixtures of factor analyzers. *Computational Statistics & Data Analysis, 41,* 379–388. doi:10.1016/S0167-9473(02)00183-4

Parsons, L., Haque, E., & Liu, H. (2004). Subspace clustering for high dimensional data: A review. *SIGKDD Explorations, 6*(1), 90–105. doi:10.1145/1007730.1007731

Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *A social learning approach. IV. Antisocial boys*. Eugene: Castalia.

Rocci, R., Gattone, S., & Vichi, M. (2011). A New dimension reduction method: Factor discriminant K-means. *Journal of Classification, 28*. doi:10.1007/s00357-011-9085-9

Sanguinetti, G. (2008). Dimensionality reduction of clustered data sets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on, 30*(3), 535–540.

Schepers, J., Ceulemans, E., & Van Mechelen, I. (2008). Selecting among multi-mode partitioning models of different complexities: A comparison of four model selection criteria. *Journal of Classification, 25*(1), 67–85. doi:10.1007/s00357-008-9005-9

Steinley, D. (2003). Local optima in K-Means clustering: What you don't know may hurt you. *Psychological Methods, 8*(3), 294–304. doi:10.1037/1082-989X.8.3.294

Steinley, D. (2006). Profiling local Optima in K-means clustering: Developing a diagnostic technique. *Psychological Methods, 11*(2), 178–192. doi:10.1037/1082-989X.11.2.178

Steinley, D., & Brusco, M. J. (2011a). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods, 16*(1), 63–79. doi:10.1037/a0022673

Steinley, D., & Brusco, M. J. (2011b). K-means clustering and mixture model clustering: Reply to McLachlan (2011) and Vermunt (2011). *Psychological Methods, 16*(1), 89–92. doi:10.1037/a0022679

Steinley, D., & Henson, R. (2005). OCLUS: An analytic method for generating clusters with known overlap. *Journal of Classification, 22,* 221–250.

Stute, W., & Zhu, L. X. (1995). Asymptotics of k-Means clustering based on projection pursuit. *Sankhyā: The Indian Journal of Statistics, Series A, 57*(3), 462–471.

Timmerman, M. E. (2006). Multilevel component analysis. *British Journal of Mathematical and Statistical Psychology, 59*(2), 301–320. doi:10.1348/000711005X67599

Timmerman, M. E., Ceulemans, E., Kiers, H. A. L., & Vichi, M. (2010). Factorial and reduced K-means reconsidered. *Computational Statistics & Data Analysis, 54*(7), 1858–1871. doi:10.1016/j.csda.2010.02.009

Tipping, M. E., & Bishop, C. M. (1999). *Mixtures of Probabilistic Principal Component Analyzers* - MIT Press. doi:10.1162/089976699300016728

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* No. 984). Washington, DC: Department of the Army.

Van Leeuwen, K. G., Mervielde, I., Braet, C., & Bosmans, G. (2004). Child personality and parental behavior as moderators of problem behavior: Variable- and person-centered approaches. *Developmental Psychology, 40*(6), 1028–1046. doi:10.1037/0012-1649.40.6.1028

Van Leeuwen, K. G., & Vermulst, A. A. (2004). Some psychometric properties of the Ghent parental behavior scale. *European Journal of Psychological Assessment, 20*(4), 283–298. doi:10.1027/1015-5759.20.4.283

Vichi, M., & Kiers, H. A. L. (2001). Factorial k-means analysis for two-way data. *Computational Statistics and Data Analysis, 37,* 49–64.

Vichi, M., & Saporta, G. (2009). Clustering and disjoint principal component analysis. *Computational Statistics & Data Analysis, 53*(8), 3194–3208. doi:10.1016/j.csda.2008.05.028

Viroli, C. (2010). Dimensionally reduced model-based clustering through mixtures of factor mixture analyzers. *Journal of Classification, 27*(3), 363–388. doi:10.1007/s00357-010-9063-7

Wang, D., Ding, C., & Li, T. (2009). K-Subspace clustering. *Machine Learning and Knowledge Discovery in Databases, 5782,* 506–521. doi:10.1007/978-3-642-04174-7_33

Wedel, M., & Desarbo, W. S. (1996). An exponential-family multidimensional scaling mixture methodology. *Journal of Business & Economic Statistics, 14*(4), 447–459.

Wilderjans, T. F., Ceulemans, E., & Meers, K. (in press). CHull: A generic convex hull based model selection method. *Multivariate Behavioral Research,*

Yoshida, R., Higuchi, T., Imoto, S., & Miyano, S. (2006). ArrayCluster: An analytic tool for clustering, data visualization and module finder on gene expression profiles. *Bioinformatics, 12,* 1538–1539.