

Clusterwise HICLAS: A generic modeling strategy to trace similarities and differences in multiblock binary data

T. F. Wilderjans · E. Ceulemans · P. Kuppens

Published online: 16 November 2011
© Psychonomic Society, Inc. 2011

Abstract In many areas of the behavioral sciences, different groups of objects are measured on the same set of binary variables, resulting in coupled binary object \times variable data blocks. Take, as an example, success/failure scores for different samples of testees, with each sample belonging to a different country, regarding a set of test items. When dealing with such data, a key challenge consists of uncovering the differences and similarities between the structural mechanisms that underlie the different blocks. To tackle this challenge for the case of a single data block, one may rely on HICLAS, in which the variables are reduced to a limited set of binary bundles that represent the underlying structural mechanisms, and the objects are given scores for these bundles. In the case of multiple binary data blocks, one may perform HICLAS on each data block separately. However, such an analysis strategy obscures the similarities and, in the case of many data blocks, also the differences between the blocks. To resolve this problem, we proposed the new Clusterwise HICLAS generic modeling strategy. In this strategy, the different data blocks are assumed to form a set of mutually exclusive clusters. For each cluster, different bundles are derived. As such, blocks belonging to the same cluster have the same bundles, whereas blocks of different clusters are modeled with different bundles. Furthermore, we evaluated the performance of Clusterwise HICLAS by means of an extensive simulation study and by applying the

strategy to coupled binary data regarding emotion differentiation and regulation.

Keywords Binary data · Multiblock data · Hierarchical classifications · HICLAS · Multiset data · Simultaneous clusterings of objects and variables · Coupled binary data

About 20 years ago, De Boeck and Rosenberg (1988) proposed the HICLAS model to disclose the structural mechanisms underlying binary object \times variable data. Such data are encountered regularly in many areas of the behavioral sciences (see, e.g., Ceulemans & Van Mechelen, 2008; De Boeck, 2008; Ip, Wang, De Boeck, & Meulders, 2004; Leenen, Van Mechelen, Gelman, & De Knop, 2008; Maris, De Boeck, & Van Mechelen, 1996; Van Mechelen & De Boeck, 1990; Van Mechelen, De Boeck, & Rosenberg, 1995). For example, in the field of psychometrics, such data are obtained when a test consisting of a set of items is administered to a number of persons and the responses to the items are scored as correct or incorrect (see, e.g., De Boeck, 2008; de la Torre, 2011; Wang & Chang, 2011). As a second example, stemming from the field of emotion psychology, a researcher may observe at different time points whether or not an individual experiences a set of emotions (see, e.g., Barrett, Gross, Christensen, & Benvenuto, 2001; Vande Gaer, Ceulemans, Van Mechelen, & Kuppens, *in press*).

In a HICLAS analysis, the variables are reduced to a limited set of binary variables, called “bundles,” that represent the structural mechanisms that underlie the data. In the psychometrics example, the bundles reflect different solution strategies that may be followed to solve the items, whereas in the emotion example, the bundles may represent different emotion types. Moreover, the objects are given binary scores for these bundles, indicating whether or not a

T. F. Wilderjans · E. Ceulemans · P. Kuppens
Katholieke Universiteit Leuven,
Leuven, Belgium

T. F. Wilderjans (✉)
Research Group of Quantitative Psychology and Individual
Differences, Katholieke Universiteit Leuven,
Tiensestraat 102, Box 3713, 3000 Leuven, Belgium
e-mail: tom.wilderjans@psy.kuleuven.be

person has mastered the different solution strategies or denoting the emotion types that are experienced on the measurement occasions. Finally, the binary bundles imply an overlapping clustering of the objects and the variables, called *hierarchical classifications*.

HICLAS analysis has been applied in many fields of the behavioral sciences. For example, such analysis has been used (1) to reveal the latent choice requisites that underlie consumer \times product select/not select data (Van Mechelen & Van Damme, 1994); (2) to reveal the implicit taxonomy, in terms of latent syndromes, that underlies psychiatric patient \times symptom presence/absence data (Rosenberg, Van Mechelen, & De Boeck, 1996; Van Mechelen & De Boeck, 1989); (3) to study a person's self-concept (Ashmore, Deaux, & McLaughlin-Volpe, 2004; Hart & Fegley, 1995); (4) to identify individual differences with respect to psychosocial outcomes after surgery (Wilson, Bladin, Saling, & Pattison, 2005); (5) to determine the situation specificity of traits and the restrictiveness of situations for trait-related behavior (ten Berge & de Raad, 2001); (6) to identify forms of social support and how they are related to individual differences in mental health for HIV-positive persons (Reich, Lounsbury, Zaid-Muhammad, & Rapkin, 2010); (7) to gain insight into the psychological mechanisms that govern the decision to pursue mediation in civil disputes (Reich, Kressel, Scanlon, & Weiner, 2007); and (8) to study the inter- and intracategorical structures of semantic categories (Ceulemans & Storms, 2010).

In many cases, however, the same set of variables is scored for more than one set of objects (i.e., different groups of objects). For example, in the psychometrics case, the same test may be administered to different groups of subjects, and, in the emotion example, different persons may be measured at different (not necessarily the same) time points. A challenging question then becomes whether the same or different psychological processes play a role in the different groups of objects. To tackle this question, up to now, researchers have performed a HICLAS analysis on the data of each group separately (see, e.g., Stirratt, Meyer, Ouellette, & Gara, 2008), resulting in as many sets of bundles as groups. By comparing these sets of bundles, similarities and differences between the groups can be identified. However, when the number of groups is relatively large, comparing all of the obtained bundles to each other may become a very time-consuming (and practically infeasible) task.

Therefore, we introduce in this article a new generic modeling strategy, called Clusterwise HICLAS; this strategy encompasses performing a HICLAS analysis on the data of each group of objects separately as a special case. The basic principle behind Clusterwise HICLAS is that the different groups of objects form a limited but

unknown number of mutually exclusive clusters and that the data of the groups that are assigned to the same cluster can be modeled using the same bundles; for groups that belong to another cluster, other bundles are needed. Hence, in Clusterwise HICLAS the groups of objects are clustered, and a separate HICLAS analysis is performed per cluster. This clusterwise principle has already been successfully used in component analysis (De Roover, Ceulemans, & Timmerman, *in press*; De Roover, Ceulemans, Timmerman, & Onghena, 2011; De Roover et al., *in press*) and regression analysis (DeSarbo, Oliver, & Rangaswamy, 1989; Spath, 1982).

The remainder of this article is organized in five main sections: First, after recapitulating HICLAS, we introduce Clusterwise HICLAS. We then propose an algorithm to estimate the parameters of the Clusterwise HICLAS model, as well as proposing a model selection procedure. Next, the optimization and recovery performance of the Clusterwise HICLAS algorithm is evaluated in an extensive simulation study. In the following section, Clusterwise HICLAS analysis is illustrated with an application to emotion data. Finally, we make some concluding remarks.

Model

Data structure

A Clusterwise HICLAS analysis can be performed on all kinds of multivariate hierarchically organized¹ binary data; in this description, “multivariate” denotes that multiple variables are involved, while “hierarchically organized” or “multiblock” implies that the data can be separated into different data blocks (e.g., as can be seen in Fig. 1, blocks representing different groups of subjects or multiple observations of a single subject).² More formally, Clusterwise HICLAS operates on a coupled binary data set $\tilde{\mathbf{D}}$ that consists of N object \times variable binary data blocks \mathbf{D}_i ($I_i \times J$), where the number of observations I_i ($i = 1, \dots, N$) in each data block may vary between data blocks. When concatenating all N data blocks \mathbf{D}_i vertically (i.e., column-wise), a binary “super” data matrix \mathbf{D}^* is obtained (Kiers, 2000).

¹ Hierarchically organized data comprise both multigroup and multilevel data, which only differ with respect to the target of inference (i.e., whether one is only interested in the groups/subjects under study or wants to generalize the results to a population of groups/subjects; see Timmerman, Kiers, Smilde, Ceulemans, & Stouten, 2009).

² Three-way, three-mode data (for an introduction, see Kroonenberg, 2008) may be conceived of as a special case of multivariate multiblock data; in this case, all groups share the same subjects, or all subjects are measured at the same time points.

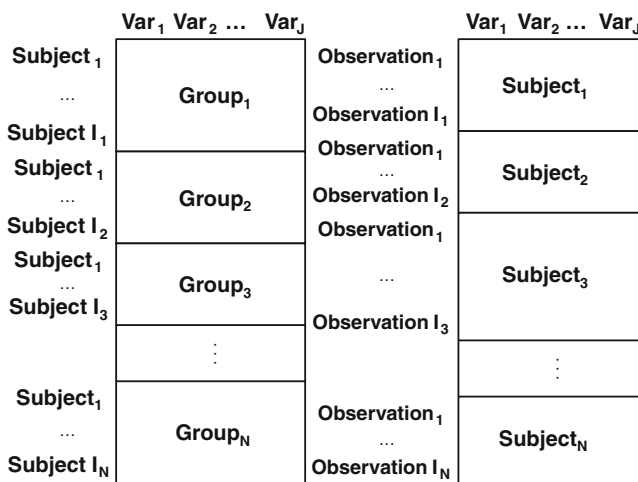


Fig. 1 Graphical representation of a multivariate hierarchically organized (multiblock) data set, consisting of N data blocks, J variables, and a varying number of objects I_i for each data block ($i = 1, \dots, N$). (Left) Groups of subjects measured on the same variables. (Right) Multiple observations of different subjects on the same variables

Hierarchical classes analysis (HICLAS) for one binary data block

In a HICLAS analysis of a single $I \times J$ object \times variable binary data matrix \mathbf{D} , an $I \times J$ binary model matrix \mathbf{M} is fitted to \mathbf{D} . The binary model matrix \mathbf{M} is decomposed into an $I \times P$ binary object bundle matrix \mathbf{A} and a $J \times P$ binary variable bundle matrix \mathbf{B} as follows:

$$\mathbf{M} = \mathbf{A} \otimes \mathbf{B}', \tag{1}$$

or, in terms of the model entries,

$$m_{ij} = \bigoplus_{p=1}^P a_{ip}b_{jp}, \tag{2}$$

where \otimes denotes a Boolean matrix product, \bigoplus indicates a Boolean sum (i.e., $1 \oplus 1 = 1$), and m_{ij} , a_{ip} , and b_{jp} are the entries of \mathbf{M} , \mathbf{A} , and \mathbf{B} , respectively. The P columns of \mathbf{A} and \mathbf{B} define a set of P binary variables, called “bundles”.

To illustrate the HICLAS model, we will make use of the hypothetical testee \times item model matrix \mathbf{M}_1 in Table 1, which contains the scores of six testees on a test with three items. The values of 1 in \mathbf{M}_1 define a “solves” relation between the testees and the items, denoting which testee solves which item. The bundle matrices \mathbf{A}^1 and \mathbf{B}^1 of a HICLAS model with two bundles for \mathbf{M}_1 are presented in Tables 2 and 3, respectively. In our example, the bundles represent solution strategies that may be followed to solve the items, with \mathbf{B}^1 indicating for each item the possible strategies that may be followed to solve the item in question. Matrix \mathbf{A}^1 then contains the scores of each testee on these bundles, denoting which solution strategies the testee in question masters. For this example, the HICLAS decomposition rule in Eqs. 1 and 2 implies that a testee solves an item when he or she masters at least one solution strategy that may be used to solve the item in question. For example, the third testee Te_3^1 solves the first item It_1 , because Te_3^1 masters the second solution strategy Str_2 (see \mathbf{A}^1), and this solution strategy can be used to solve It_1 (see \mathbf{B}^1).

An extra feature of the HICLAS model, in comparison to related models such as Boolean factor analysis (Mickey, Mundle, & Engelman, 1983), is that for both the testees and the items, a quasi-order relation “ \leq ” is defined: When Q^{Te_i} denotes the set of items that testee Te_i answers correctly, then $Te_i \leq Te_{ij}$ iff $Q^{Te_i} \subseteq Q^{Te_{ij}}$, which implies that all items solved by Te_i are also solved by Te_{ij} . For the items, a similar quasi-order relation is defined and represented by the HICLAS model, with Q^{It_j} denoting the set of testees who solve It_j . In the HICLAS model, the quasi-order relations among the testees and the items are represented by means of subset–superset relations among their bundle patterns. For example, in \mathbf{M}_1 , $Te_4^1 \leq Te_6^1$ because $Q^{Te_4^1} \subseteq Q^{Te_6^1}$. Therefore, in \mathbf{A}^1 , the bundle pattern for Te_4^1 is a subset of the bundle pattern for Te_6^1 . Also, in \mathbf{M}_1 , $It_3 \leq It_2$, and therefore, the bundle pattern for It_3 is a subset of the bundle pattern for It_2 in \mathbf{B}^1 . Note that the quasi-order relations among the testees and the items imply a partitioning of the testees and the items into classes (i.e., testees/items with an identical bundle pattern) that are

Table 1 Hypothetical testee (stemming from four different groups) \times item model matrices \mathbf{M}_1 , \mathbf{M}_2 , \mathbf{M}_3 , and \mathbf{M}_4

	\mathbf{M}_1			\mathbf{M}_2			\mathbf{M}_3			\mathbf{M}_4					
	It_1	It_2	It_3	It_1	It_2	It_3	It_1	It_2	It_3	It_1	It_2	It_3			
Te_1^1	0	1	1	Te_1^2	0	1	0	Te_1^3	1	0	0	Te_1^4	1	1	1
Te_2^1	1	1	1	Te_2^2	0	1	0	Te_2^3	1	0	0	Te_2^4	1	1	0
Te_3^1	1	1	0	Te_3^2	1	0	0	Te_3^3	0	0	0	Te_3^4	0	1	1
Te_4^1	1	1	0	Te_4^2	0	0	0	Te_4^3	1	0	0	Te_4^4	0	1	1
Te_5^1	0	0	0					Te_5^3	0	1	0				
Te_6^1	1	1	1												

Table 2 Testee bundle matrices $A^1, A^2, A^3,$ and A^4 for the Clusterwise HICLAS model with two bundles (and two clusters) for the model matrices $M_1, M_2, M_3,$ and M_4 in Table 1

	A^1		A^2		A^3		A^4				
	Str ₁	Str ₂	Str ₁	Str ₂	Str ₁	Str ₂	Str ₁	Str ₂			
Te ₁ ¹	1	0	Te ₁ ²	1	0	Te ₁ ³	0	1	Te ₁ ⁴	1	1
Te ₂ ¹	1	1	Te ₂ ²	1	0	Te ₂ ³	0	1	Te ₂ ⁴	0	1
Te ₃ ¹	0	1	Te ₃ ²	0	1	Te ₃ ³	0	0	Te ₃ ⁴	1	0
Te ₄ ¹	0	1	Te ₄ ²	0	0	Te ₄ ³	0	1	Te ₄ ⁴	1	0
Te ₅ ¹	0	0				Te ₅ ³	1	0			
Te ₆ ¹	1	1									

hierarchically ordered (for more information and for applications that make use of the quasi-order relations, see De Boeck & Rosenberg, 1988).

Regarding the uniqueness of the HICLAS decomposition, the following sufficient condition has been proven: If all bundle-specific classes (i.e., classes of testees/items that belong to one bundle only) of a HICLAS decomposition with P bundles of an $I \times J$ binary array M are nonempty, this decomposition is unique upon a permutation of the bundles (Ceulemans & Van Mechelen, 2003; Van Mechelen et al., 1995).

The Clusterwise HICLAS model

In order to trace the similarities and differences between the mechanisms that underlie the different data blocks D_i ($i = 1, \dots, N$), Clusterwise HICLAS partitions the N data blocks (i.e., groups of objects/testees) into K mutually exclusive and nonempty clusters. Data blocks that belong to the same cluster are assumed to be governed by the same underlying processes, whereas different mechanisms play a role for data blocks belonging to different clusters. To gain insight into these processes, in each cluster a HICLAS analysis is performed using P bundles, with P being assumed to be the same across clusters. Formally, the $I_i \times J$ model matrices M_i ($i = 1, \dots, N$) are decomposed by the following rule:

$$M_i = \sum_{k=1}^K p_{ik} (A^i \otimes (B^k)')$$
(3)

Table 3 Item bundle matrices B^1 and B^2 and the partition matrix P for the Clusterwise HICLAS model with two bundles and two clusters for the model matrices $M_1, M_2, M_3,$ and M_4 in Table 1

	B^1		B^2		P			
	Str ₁	Str ₂	Str ₁	Str ₂	Cl ₁	Cl ₂		
It ₁	0	1	It ₁	0	1	Group ₁	1	0
It ₂	1	1	It ₂	1	0	Group ₂	0	1
It ₃	1	0	It ₃	0	0	Group ₃	0	1
						Group ₄	1	0

where K is the number of clusters, p_{ik} are the entries of the binary partition matrix P ($N \times K$) that indicates whether data block i belongs to cluster k ($p_{ik} = 1$) or not ($p_{ik} = 0$), A^i ($I_i \times P$) is the object bundle matrix for block i ($i = 1, \dots, N$), and B^k ($J \times P$) is the variable bundle matrix for cluster k ($k = 1, \dots, K$). The cluster-specific variable bundle matrices B^k are indexed by k , because they are shared by all data blocks that belong to cluster k .

To illustrate the Clusterwise HICLAS decomposition rule in Eq. 3, we will use the hypothetical testee \times item model matrices $M_1, M_2, M_3,$ and M_4 , representing data from four different groups (i.e., $N = 4$), in Table 1. Tables 2 and 3 show the testee bundle matrices $A^1, A^2, A^3,$ and A^4 , the cluster-specific item bundle matrices B^1 and B^2 , and the partition matrix P of a Clusterwise HICLAS model with two bundles and two clusters for $M_1, M_2, M_3,$ and M_4 . For example, in M_4 , the third testee of the fourth group Te_3^4 solves the third item It_3 , because Te_3^4 masters the first solution strategy Str_1 (see A^4 in Table 2), which allows for solving It_3 (see B^1 in Table 3) and the fourth group of testees Group₄ is assigned to the first cluster Cl_1 (see P in Table 3). However, in M_3 , Te_5^3 does not answer It_1 correctly, because the solution strategy that Te_5^3 masters (see A^3) does not match the strategy that allows for solving It_1 (see B^2), since Group₃ is assigned to Cl_2 .

In the Clusterwise HICLAS model, a quasi-order relation is defined on the testees in each M_i . For example, in Table 1, one can see that $Te_3^4 \leq Te_1^4$ in M_4 , because $Q^{Te_3^4} \subseteq Q^{Te_1^4}$, with $Q^{Te_3^4}$ denoting the items that Te_3^4 solves. Also, for each cluster of groups separately, a quasi-order relation is defined on the items, with Q^{It_j} now indicating which testees in the groups (i.e., data blocks) that belong to the cluster in question solve It_j . For example, for the second cluster, $It_3 \leq It_1$ because $Q^{It_3} \subseteq Q^{It_1}$ in M_2 and M_3 (see Table 1), which pertains to the groups that belong to the second cluster (see P in Table 3), with $Q^{It_1} = \{Te_3^2, Te_1^3, Te_2^3, Te_4^3\}$. These quasi-order relations are represented by the Clusterwise HICLAS model in the same way as by the HICLAS model (i.e., by subset–superset relations among the bundle patterns). For example, $Te_3^4 \leq Te_1^4$ (see above), and therefore, in Table 2, the bundle pattern for Te_3^4 is a

subset of the bundle pattern for Te_1^4 . Moreover, $It_3 \leq It_1$ (see above), resulting in the bundle pattern of It_3 in \mathbf{B}^2 (Table 3) being a subset of the bundle pattern of It_1 (for more information, see Wilderjans, Ceulemans, & Van Mechelen, 2008, in press). Note that a Clusterwise HICLAS solution can be interpreted on the basis of the obtained object and (cluster-specific) variable bundles and/or of the implied quasi-orders for the objects and the variables (see Van Mechelen et al., 1995).

When considering the Clusterwise HICLAS decomposition rule in Eq. 3, it can be seen that the Clusterwise HICLAS model is a generic modeling strategy to disclose similarities and differences between coupled binary data blocks. If K equals N , which implies that each data block forms a separate cluster, Clusterwise HICLAS boils down to performing a separate HICLAS analysis on each data block. A second special case is obtained when K equals 1, implying that all data blocks are assigned to the same cluster. In this case, a Clusterwise HICLAS analysis reduces to performing a single HICLAS analysis on \mathbf{D}^* (Kiers, 2000); this case, which can be conceived of as the hierarchical-classes counterpart of simultaneous component analysis (see Kiers & ten Berge, 1989, 1994; Millsap & Meredith, 1988; Timmerman & Kiers, 2003; Van Deun, Smilde, van der Werf, Kiers, & Van Mechelen, 2009), results in a single set of bundles for all data blocks, implying that only similarities between data blocks can be traced (see Wilderjans et al., in press).

Data analysis

Aim

Given a set of $I_i \times J$ binary data matrices \mathbf{D}_i ($i = 1, \dots, N$), a number of clusters K , and a number of bundles P , the aim of a Clusterwise HICLAS analysis is to estimate the binary partition matrix \mathbf{P} and the binary object bundle matrices \mathbf{A}^i and variable bundle matrices \mathbf{B}^k such that the loss function

$$f = \sum_{i=1}^N \left\| \mathbf{D}_i - \sum_{k=1}^K p_{ik} (\mathbf{A}^i \otimes (\mathbf{B}^k)') \right\|_F^2 \quad (4)$$

is minimized, with $\|\cdot\|_F$ denoting the Frobenius norm (i.e., the square root of the sum of squared values).

Algorithm

In this section, we will introduce the Clusterwise HICLAS algorithm, which is an alternating least squares (ALS) algorithm that is based on the principles of the K-Means (MacQueen, 1967) and the HICLAS (Leenen & Van Mechelen, 2001) algorithms. Because the Clusterwise HICLAS loss function may be prone to many local optima,

two extra procedures are needed to lower the risk of the algorithm ending in a local optimum. First, a multistart procedure is proposed, along with a smart way to obtain “high-quality” initial parameter estimates. Second, a (time-consuming) procedure is discussed that improves the final estimates of the bundles, given the final clustering of the data blocks.

Clusterwise HICLAS ALS algorithm Identifying the globally optimal solution for the Clusterwise HICLAS loss function in Eq. 4 is a hard nut to crack, because all possible partitions of the data blocks, together with all possible bundle matrices, need to be sieved through (with the partitioning problem on its own already being a NP-hard problem; see, e.g., Brusco, 2006; van Os & Meulman, 2004). Therefore, we propose to use a fast relocation algorithm that can handle large numbers of data blocks, but that may end in a local minimum. In particular, we developed an ALS procedure (for more information on ALS, see de Leeuw, 1994; ten Berge, 1993) in which the cluster memberships of the data blocks (in partition matrix \mathbf{P}) are alternately updated until there is no improvement in the loss function value. Specifically, the Clusterwise HICLAS algorithm consists of the following five steps:

1. Initialize the partition matrix \mathbf{P} by assigning the N data blocks to one of the K clusters, such that there are no empty clusters (see below).
2. For each cluster k ($k = 1, \dots, K$), estimate the cluster-specific variable bundle matrix \mathbf{B}^k and the object bundle matrices \mathbf{A}^i for all data blocks belonging to the cluster in question. To this end, for each cluster, a HICLAS analysis is performed (for details about the HICLAS algorithm, see Leenen & Van Mechelen, 2001) on the data matrix that is obtained by column-wise concatenating (Kiers, 2000) the data blocks that belong to the cluster in question. At the end, compute the loss function value (Eq. 4).
3. Update the partition matrix \mathbf{P} row-wise and reestimate the bundles as in Step 2. To determine the optimal cluster for data block \mathbf{D}_i (i.e., updating row i of \mathbf{P}), compute for each cluster k an object bundle matrix $\tilde{\mathbf{A}}^{i(k)}$ ($I_i \times P$) by means of a Boolean regression (Leenen & Van Mechelen, 1998; Mickey et al., 1983). In this regression, the P columns of the variable bundle matrix \mathbf{B}^k of cluster k figure as the binary predictors, the I_i rows of data block \mathbf{D}_i as the criteria, and $\tilde{\mathbf{A}}^{i(k)}$ as the Boolean regression weights. Next, for each cluster k , the partition criterion $L_{ik} = \left\| \mathbf{D}_i - \tilde{\mathbf{A}}^{i(k)} \otimes (\mathbf{B}^k)' \right\|_F^2$, which denotes the extent to which data block \mathbf{D}_i does not fit in cluster k , is computed. Finally, data block \mathbf{D}_i is

reassigned to the cluster for which L_{ik} is minimal. After updating \mathbf{P} (and reestimating the bundles), check whether one of the clusters is empty. When this is the case, assign the data block that fits its current cluster the worst to the empty cluster (and again reestimate the bundles, as in Step 2).

4. Compute the loss function value (Eq. 4). When it has decreased, return to Step 3; otherwise, the algorithm has converged.
5. Perform a closure operation (Barbut & Monjardet, 1970; Birkhoff, 1940) on (each) \mathbf{A}^i and \mathbf{B}^k . This is necessary because the bundle matrices obtained at the end of Step 4 do not yet represent the quasi-order relation in \mathbf{M}_i correctly. This closure operation consists of changing each 0 value in \mathbf{A}^i and \mathbf{B}^k to 1 iff this modification does not alter \mathbf{M}_i (and, consequently, does not change the loss function value).

Multistart procedure In order to minimize the probability of ending up at a suboptimal solution, a multistart procedure is advised; such a procedure consists of running the Clusterwise HICLAS algorithm (see above) with different initializations of the partition matrix \mathbf{P} (Ceulemans, Van Mechelen, & Leenen, 2007; Steinley, 2003) and retaining the solution with the lowest value of the loss function (Eq. 4). Because suboptimal solutions may be omnipresent, it is of utmost importance to identify a set of “high-quality” initial partition matrices that are already close to the optimal one. To obtain Q such high-quality partition matrices, we propose using the following procedure:

- a. Determine a rational initial partition matrix \mathbf{P}^{rat} by, first, performing a HICLAS analysis with P bundles on each data block \mathbf{D}_i separately. Next, compute the kappa coefficient (Cohen, 1960), which can be conceived of as a measure of similarity, between each pair of obtained variable bundle matrices \mathbf{B}^i . Finally, obtain a rational clustering of the N data blocks into K clusters by performing a single-linkage (i.e., nearest-neighbor) hierarchical cluster analysis (Gordon, 1981) on the matrix of kappa coefficients and cutting the resulting tree at the desired number of clusters.
- b. Generate $Q \times 10$ different pseudorational partition matrices $\mathbf{P}^{\text{P-rat}}$ (with no empty clusters). Starting from \mathbf{P}^{rat} , each $\mathbf{P}^{\text{P-rat}}$ is obtained by reassigning each data block to another cluster with a probability equal to .20 (with all “other” clusters having the same probability of being assigned to).
- c. Compute for \mathbf{P}^{rat} and for each $\mathbf{P}^{\text{P-rat}}$ the Clusterwise HICLAS model (as in Step 2 above) and compute the loss function value (Eq. 4). Rank order all obtained

initial \mathbf{P} matrices, based on the loss function value, and select the best Q ones.

Improving the final bundles The HICLAS algorithm in Leenen and Van Mechelen (2001) appears to be prone to suboptimal solutions, especially when the data matrix is far from square (i.e., when there are more objects than variables, or vice versa). In Clusterwise HICLAS, however, often HICLAS analyses need to be performed on concatenated data blocks (see Step 2 of the algorithm), which are “far-from-square” matrices.

Therefore, to lower the risk of ending in a local minimum, it may be advisable to use a more time-consuming simulated annealing (SA) algorithm, called HICLAS^{SA}, to obtain the final bundle estimates (i.e., \mathbf{A}^i and \mathbf{B}^k), given the best encountered clustering (i.e., the clustering resulting from the multistarted ALS procedure). Note that SA has already successfully been applied to estimate the parameters of hierarchical classes models (see Ceulemans et al., 2007; Wilderjans et al., 2008, in press). A detailed description of the HICLAS^{SA} algorithm and the metaparameter setting that have been used is given in the Appendix.

Model selection

In general, the optimal number of clusters K and the optimal number of bundles P that underlie the coupled binary data set at hand are usually unknown. Therefore, one may perform different Clusterwise HICLAS analyses with increasing numbers of clusters and bundles. To select an appropriate model, one may rely on the interpretability of the solution and on a formal model selection heuristic that aims at selecting a model that has an optimal balance between, on the one hand, fit to the data (i.e., the loss function value), and on the other hand, the complexity of the model (i.e., the numbers of clusters and bundles). Specifically, we propose using a generalization of the well-known scree test (Cattell, 1966), which has already proved to be effective for determining the number of bundles in hierarchical classes analysis (see, e.g., Ceulemans, Van Mechelen, & Leenen, 2003; Leenen & Van Mechelen, 2001). This model selection strategy consists of plotting the loss function value (Eq. 4) of the different solutions against the number of bundles P for each value of K . Subsequently, the optimal value for K may be determined by examining the general (i.e., across numbers of bundles) increase in fit that is obtained by adding a cluster and choosing the number of clusters after which this general increase in fit levels off. Finally, considering solutions with K clusters only, one looks for an elbow in the scree plot for the selected K value (see Cattell, 1966) to determine the optimal number of bundles P .

Simulation study

Problem

In this section, we will present a simulation study to evaluate the performance of the Clusterwise HICLAS algorithm with respect to optimization and recovery (in the ideal situation in which the correct numbers of underlying bundles and clusters are known). Regarding optimization, we will study how sensitive the algorithm is to local minima. With respect to recovery, we will investigate the extent to which the algorithm succeeds in recovering the true structure underlying the data. For both aspects, we will also study whether and how the algorithm's performance depends on characteristics of the data. Specifically, we will focus on six data characteristics. The first three characteristics pertain to the clustering of the data blocks: (1) the number of underlying clusters, (2) the cluster size, and (3) the degree of congruence (i.e., similarity) between the bundles for each cluster. We expect the algorithm's performance to deteriorate when the number of underlying clusters increases (Brusco & Cradit, 2005; De Roover et al., 2011; Milligan, Soon, & Sokol, 1983), the clusters are of different sizes (Brusco & Cradit, 2001; Milligan et al., 1983; Steinley, 2003), and/or there is much congruence between the bundles for each cluster (De Roover et al., *in press*). Moreover, we expect the performance to be worst for certain combinations of these characteristics (i.e., many clusters of different sizes with much congruence between the cluster-specific bundles). The next factor, (4) the complexity of the underlying HICLAS model(s), is the number of underlying bundles. We conjecture that the algorithm's performance will decrease with an increasing number of bundles (De Boeck & Rosenberg, 1988; Wilderjans, Ceulemans, & Van Mechelen, 2008, 2009, *in press*). A further factor, (5) the sample size and amount of available information, is the number of observations per data block. We hypothesize that the performance of the Clusterwise HICLAS algorithm will improve when the algorithm has more information (i.e., more observations per data block) at its disposal (Brusco & Cradit, 2005; Hands & Everitt, 1987). Finally, with respect to (6) the amount of noise in the data, we expect the algorithmic performance to deteriorate when the amount of noise in the data becomes large (Brusco & Cradit, 2005; Wilderjans et al., *in press*).

Design and procedure

In the simulation study, the number of data blocks N was kept fixed at 30, and the number of variables J was fixed at 12. Furthermore, the six factors, which were introduced above, were systematically manipulated in a completely

randomized six-factorial design, with all factors considered random:

1. the number of clusters, K , at two levels: 2 and 4;
2. the cluster size, at three 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); and unequal with majority (70% of the data blocks in one cluster and the remaining data blocks distributed equally over the other clusters);
3. the degree of congruence between the cluster-specific variable bundle matrices \mathbf{B}^k , at two levels: low and high congruence;
4. the number of bundles, P , at two levels: 2 and 4;
5. the number of observations per data block, I_i , at two levels: $I_i = 50$ or 100;
6. the amount of noise in the data, ε , at three levels: .05, .15, and .25.

For each cell of the design, 10 coupled data sets $\tilde{\mathbf{D}}$ were generated as follows: A true partition matrix $\mathbf{P}^{(T)}$ was constructed by calculating the number of data blocks that belonged to each cluster (given the first and second factors and not allowing for empty clusters) and assigning the correct number of data blocks to each cluster randomly. Next, true object bundle matrices $\mathbf{A}^{i(T)}$ ($i = 1, \dots, N$) and a (common) base variable bundle matrix \mathbf{B}^{base} were simulated by independently drawing entries from a Bernoulli distribution with parameter value .50. Subsequently, in order to manipulate the degree of congruence between the different $\mathbf{B}^{k(T)}$ s, for each cluster k , a true variable bundle matrix $\mathbf{B}^{k(T)}$ was obtained by changing at random 5% or 25% of the cells of the common base bundle matrix \mathbf{B}^{base} ; this resulted in a set of $\mathbf{B}^{k(T)}$ matrices ($k = 1, \dots, K$) that were highly and lowly congruent, respectively.³ Next, true matrices \mathbf{T}_i ($i = 1, \dots, N$) were computed by combining $\mathbf{P}^{(T)}$, $\mathbf{A}^{i(T)}$, and $\mathbf{B}^{k(T)}$ by the Clusterwise HICLAS decomposition rule (Eq. 3). It should be noted that the matrices $\mathbf{A}^{i(T)}$ and $\mathbf{B}^{k(T)}$ (see above) were generated such that each possible bundle pattern that contained a single 1 (e.g., the patterns 1 0 and 0 1 in Tables 2 and 3) occurred at least once. This constraint was imposed to ensure that a unique decomposition of \mathbf{T}_i into $\mathbf{A}^{i(T)}$ and $\mathbf{B}^{k(T)}$ existed (Ceulemans & Van Mechelen, 2003; Van Mechelen et al., 1995). Finally, for

³ To evaluate how much the resulting cluster-specific variable bundle matrices differed in each level, the kappa coefficient (with 1 indicating perfect congruence/similarity; see Cohen, 1960) between the entries of each couple of cluster-specific variable bundle matrices was computed. When averaging this kappa coefficient over all couples of cluster-specific variable bundle matrices, mean kappas across all data sets of .84 ($SD = .03$) and .24 ($SD = .11$) were obtained for the high- and low-congruence conditions, respectively.

each true matrix \mathbf{T}_i , a data matrix \mathbf{D}_i was constructed by changing the value of exactly a proportion of ε of the cells of \mathbf{T}_i ($i = 1, \dots, N$).

As such, 10 (replications) \times 2 (number of clusters) \times 3 (cluster size) \times 2 (degree of congruence between the cluster-specific variable bundle matrices) \times 2 (number of bundles) \times 2 (number of observations per data block) \times 3 (amount of noise in the data) = 1,440 different coupled data sets $\tilde{\mathbf{D}}$ were obtained. Subsequently, a Clusterwise HICLAS analysis with the correct values for K and P was performed on each of these data sets $\tilde{\mathbf{D}}$ using a multistart procedure with 25 starts. These starts were obtained by selecting the best 25 initial partition matrices \mathbf{P} among (1) a rationally determined \mathbf{P} and (2) 125 pseudorational \mathbf{P} matrices (see the Algorithm section above). Note that the Clusterwise HICLAS algorithm was implemented in MATLAB code (version 7.12.0, R2011a) and is available upon request from the first author. Note further that the simulation study was run on a supercomputer consisting of INTEL XEON L5420 processors with a clock frequency of 2.5 GHz and with 8 GB RAM.

Results

Optimization performance: Goodness of fit and sensitivity to local minima In this section, we want to study the extent to which the Clusterwise HICLAS algorithm was able to find the global minimum of the loss function (Eq. 4). However, because error had been added to the data, this global optimum was unknown. Therefore, we used the true solution underlying the data (i.e., $\mathbf{P}^{(T)}$, $\mathbf{A}^{i(T)}$, and $\mathbf{B}^{k(T)}$) as a proxy of the global optimum, because this solution is always a valid solution with K clusters and P bundles for the data. As a consequence, we considered a Clusterwise HICLAS solution suboptimal when its loss value exceeded that of the proxy. This appeared to be the case for only 50 out of the 1,440 data sets (3.47%), which all contained a very low amount of noise.

In order to study how the optimization performance varied as a function of the manipulated data characteristics, we calculated the f_{diff} statistic, defined as the normalized (i.e., divided by the number of data entries) difference between the loss values of the proxy and the solution retained by the algorithm. Subsequently, we performed an analysis of variance with f_{diff} as the dependent variable and the six data characteristics as independent variables. Only taking main and interaction effects into account with an intraclass correlation $\hat{\rho}_I > .05$ (Haggard, 1958; Kirk, 1982), this analysis revealed large main effects of the number of bundles ($\hat{\rho}_I = .25$) and the amount of noise in the data ($\hat{\rho}_I = .43$): The (normalized) difference between the loss value of the retained solution and the proxy increased when the number of bundles and the amount of noise in the data increased. Both main effects were further qualified by a large interaction

between the factors ($Q^{\text{Te}_3^4} \subseteq Q^{\text{Te}_1^4}$): As one can see in the leftmost panel of Fig. 2, the effect of the number of bundles was more pronounced when the data contained a large amount of noise.

Recovery performance The recovery performance of the Clusterwise HICLAS algorithm was evaluated with respect to (1) the clustering of the data blocks and (2) the cluster-specific variable bundle matrices.

Recovery of the clustering of the data blocks To examine the extent to which the underlying clustering of the data blocks has been recovered, the adjusted Rand index (ARI; Hubert & Arabie, 1985) between the true partition of the data blocks (i.e., $\mathbf{P}^{(T)}$) and the estimated partition (i.e., \mathbf{P}) is computed. The ARI equals 1 if the two partitions are identical, and 0 when the overlap between the two partitions can be totally attributed to chance.

The mean ARI, across all data blocks, equaled .9417 ($SD = .1669$), implying that the Clusterwise HICLAS algorithm recovered the underlying clustering of the data blocks to a very large extent. For 1,219 out of the 1,440 data sets (84.65%), the clustering was recovered perfectly. To study how the recovery performance was influenced by the manipulated data characteristics, an analysis of variance was performed with ARI as the dependent variable and the six data characteristics as independent variables. Only taking into account effects with $\hat{\rho}_I > .05$, this analysis revealed that the recovery performance, as can be seen in the boxplots in the middle panel of Fig. 2, decreased when the cluster-specific variable bundle matrices were more congruent/similar ($\hat{\rho}_I = .10$).

Recovery of the cluster-specific variable bundle matrices To evaluate the recovery of the cluster-specific variable bundle matrices \mathbf{B}^k , we computed, for each cluster k , the kappa coefficient κ (Cohen, 1960) between the true and estimated variable bundle matrices; next, we obtained an overall κ_B^{all} statistic by averaging these kappa coefficients across all clusters:

$$\kappa_B^{\text{all}} = \frac{\sum_{k=1}^K \kappa(\mathbf{B}^{k(T)}, \mathbf{B}^k)}{K}, \quad (5)$$

with $\mathbf{B}^{k(T)}$ and \mathbf{B}^k being the true and estimated variable bundle matrices, respectively, for cluster k .⁴ The permuta-

⁴ To evaluate the recovery of the cluster-specific overlapping clustering of the variables that was implied by each \mathbf{B}^k , we defined the ω_B^{all} statistic as the average of the cluster-specific omega indices ω_B^k (Collins & Dent, 1988) between $\mathbf{B}^{k(T)}$ and \mathbf{B}^k (taking the permutational freedom of the clusters into account). It appears that κ_B^{all} and ω_B^{all} yielded very similar results.

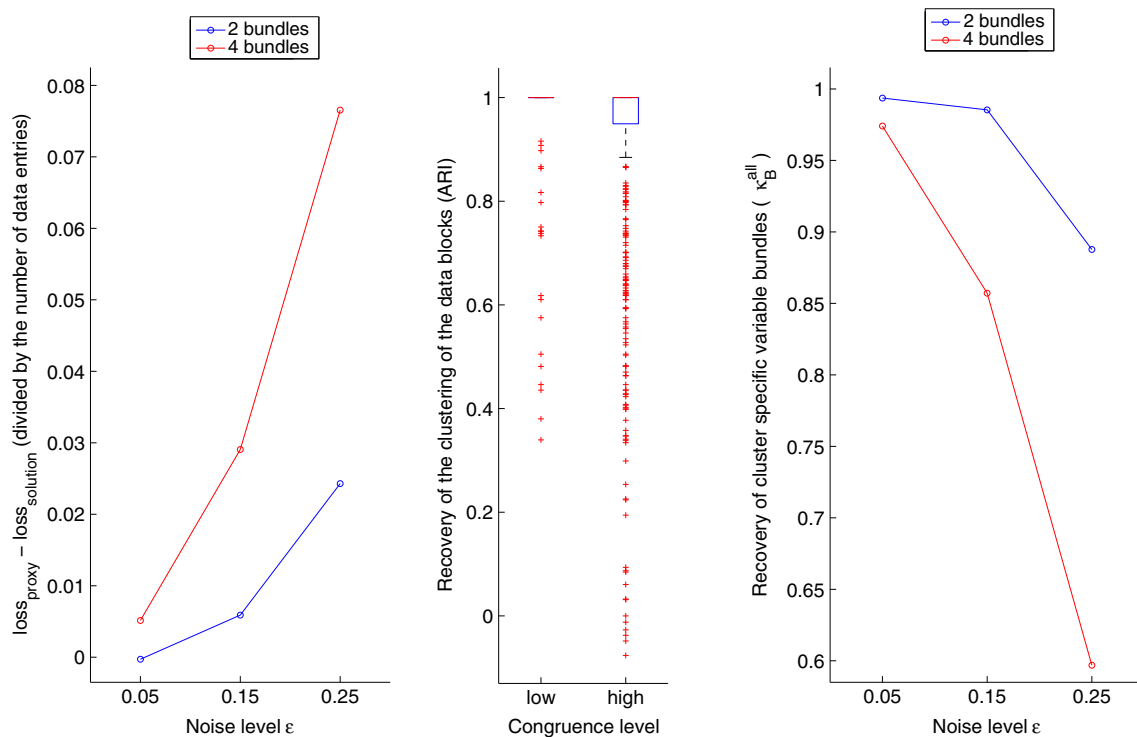


Fig. 2 (Left) Mean f_{diff} as a function of the number of bundles and the amount of noise in the data. (Middle) Boxplot of the adjusted Rand index (ARI) for different levels of the degree of congruence between

the cluster-specific variable bundle matrices. (Right) Mean κ_B^{all} as a function of the number of bundles and the amount of noise in the data

tional freedom of the bundles was dealt with by selecting the permutation of the bundles that optimized the kappa coefficient. To take the permutational freedom of the clusters into account, the cluster permutation was chosen that maximized κ_B^{all} . The κ_B^{all} statistic ranges between 0 (no recovery at all) and 1 (perfect recovery).

In the simulation study, the mean κ_B^{all} value, across all data sets, equaled .8825 ($SD = .1578$), indicating good recovery of the cluster-specific variable bundle matrices. An analysis of variance was performed with κ_B^{all} as the dependent variable and the six data characteristics as independent variables. When only effects with $\hat{\rho}_I > .05$ were taken into account, this analysis revealed that the recovery performance decreased when the number of bundles ($\hat{\rho}_I = .20$) and the amount of noise in the data ($\hat{\rho}_I = .29$) increased. Moreover, in this analysis, as can be seen in the rightmost panel of Fig. 2, it appears that the effect of the number of bundles was more pronounced when the data contained a large amount of noise ($\hat{\rho}_I = .26$).

Discussion of the results In the simulation study, we demonstrated that the Clusterwise HICLAS algorithm succeeded well in optimizing the loss function when using 25 multistarts (see above). Furthermore, the algorithm appeared to recover both the clustering of the data blocks and the cluster-specific variable bundles to a very large extent.

This implies that, under ideal situations (i.e., the correct number of clusters and bundles being used), a multistart procedure with 25 starts is sufficient, in that it results in good-quality solutions. Therefore, when analyzing empirical data, a two-stage procedure may be advised: In the first stage, when exploring different numbers of clusters and bundles, 25 multistarts may be sufficient. In the second stage, in order to improve the quality of the solutions (i.e., some particular combinations of P and K) retained for further investigation, these solutions may be reestimated with a larger number of multistarts (say, 50 or 100).

The simulation study further demonstrated that one should be cautious when the number of bundles is large, when the data contain a large amount of noise, and/or when there is much congruence among the cluster-specific variable bundles. However, these extreme situations, which almost never occur in practice, were included in the study in order to make it hard for the algorithm to find a good solution. When such an extreme situation is encountered in empirical applications, a way out would be to increase the number of multistarts.

Illustrative application

In this section, we will perform a Clusterwise HICLAS analysis to coupled data on emotion differentiation and

emotion regulation. “Emotion differentiation” refers to the degree to which individuals discriminate among the experiences of different emotions. In particular, individuals with greater emotional differentiation are able to clearly distinguish among (and thus experience) a variety of negative and positive discrete emotions, whereas individuals with lower levels of emotional differentiation tend to describe their emotions in an overgeneralized way, such as simply either good or bad (Kashdan, Ferrisizidis, Collins, & Muraven, 2010). Emotion differentiation is thought to form an indicator of psychological maladjustment and is, for instance, a central feature of alexithymia, a risk factor for depression (Bagpy, Taylor, Quilty, & Parker, 2007; Suvak et al., 2011). As a result, a key question is how the use of different emotion regulation strategies (i.e., the ways people try to change their emotions, such as, e.g., social sharing of emotions, distraction, or cognitive change) is related to emotion differentiation (Wranik, Barrett, & Salovey, 2007). Barrett et al. (2001) hypothesized that individuals who show larger emotion differentiation should also display stronger emotion regulation, especially for negative emotions.

We examined this hypothesis on the basis of experience sampling data on the experience of emotions and the use of emotion regulation strategies collected from 31 psychology students at Katholieke Universiteit Leuven (mainly between 18 and 21 years old; 15 males and 16 females). Experience sampling methods allow for collecting data during participants’ ongoing daily activities, enabling research to capture “life as it is lived” (Bolger, Davis, & Rafaeli, 2003). Using programmed palmtop computers, participants were beeped at 10 random times a day over the course of one week.⁵ At each beep (or measurement occasion), the subjects were asked to rate the extents to which they were experiencing seven emotions and using nine emotion regulation strategies to deal with them (see Table 4 for an overview of the emotions and the regulation strategies); for this purpose, a scale ranging from 0 (*not at all*) to 5 (*to a very large extent*) was used (a more detailed description of the data set can be found in Vande Gaer et al., *in press*).

The data were dichotomized by performing a mean split on each emotion/regulation strategy (with a score on or above the mean being recoded to 1, and a score below the mean to 0). Next, different Clusterwise HICLAS analyses (with 25 multistarts, being the best ones out of 125 initial partition matrices, and 111 SA chains—1 rational, 10 random, and 100 pseudorational—in Step 2 of the algorithm) were performed on the dichotomized data, with

the number of clusters ranging from one to six and the number of bundles from one to five. Taking into account the interpretability of the solution and applying the generalized scree test (as we explained in the [Model Selection](#) section) to the obtained loss function values (see Fig. 3), we selected the solution with two underlying clusters and two underlying bundles as the final solution.

In Table 4, the emotion/regulation strategy bundle matrix for the first (11 subjects) and second (20 subjects) clusters is displayed. In this table, it can be seen that both groups experienced negative (i.e., first bundle) and positive (i.e., second bundle) emotions, and that negative emotions called for a whole spectrum of regulation strategies (e.g., ranging from calmly reflecting on the feelings to actively trying to avoid thinking about and expressing these negative emotions). However, regarding negative emotions, both groups differed in their emotion differentiation: The first group only experienced stress and anxiety, whereas individuals from the second group displayed a more differentiated set of emotional reactions, additionally including depression and anger. When relating these differences in emotion differentiation between groups to differences in regulation strategies, it turned out that individuals from this second cluster, who were characterized by larger emotion differentiation, also used a wider range of regulation strategies than did individuals from the first cluster, who displayed less emotion differentiation. For positive emotions, the groups did not differ in emotion differentiation or in emotion regulation (i.e., they did not show any specific regulation strategy in these instances). These results support the results reported in Barrett et al. (2001). These authors explained the link between emotion differentiation and emotion regulation by the availability of discrete emotion knowledge that may become activated during the representation process. This knowledge, which is more elaborated in the case of larger emotion differentiation, may provide lots of information as how to deal with the specific situation. This is especially true, they argued, in the case of very intense negative emotions, since the call for emotion regulation in these situations is the greatest.

For comparative purposes, we also performed a standard HICLAS analysis (with two bundles) on the concatenated data. From this analysis, as one can see in Table 4, it appears that a standard HICLAS resulted in a too simplistic picture of the underlying mechanisms, in that the variable bundles for both groups got mixed up, with the largest group dominating the solution. In particular, as in the second group, all subjects displayed a very differentiated set of emotional reactions (i.e., all positive and negative emotions), and in the case of negative emotions, they employed almost all emotion regulation strategies (as in the second group) except talking about feelings with others (as in the first group).

⁵ Although compliance was good (i.e., 71% of the beeps were returned), there were missing data. For each participant, we removed the measurement moments in which they did not rate all emotions and all regulation strategies.

Table 4 Emotion/regulation strategy bundle matrix with two bundles for both clusters of subjects and for the standard HICLAS analysis

Emotion or Regulation Strategy	Cluster 1 (11 Subjects)		Cluster 2 (20 Subjects)		Standard HICLAS	
Angry	0	0	1	0	1	0
Depressed	0	0	1	0	1	0
Anxious	1	0	1	0	1	0
Stressed	1	0	1	0	1	0
Relaxed	0	1	0	1	0	1
Happy	0	1	0	1	0	1
Feel good about myself	0	1	0	1	0	1
Could not stop thinking about feelings	1	0	1	0	1	0
Calmly reflected on feelings	1	0	1	0	1	0
Talked about feelings with others	0	0	1	0	0	0
Avoided thinking about feelings	0	0	1	0	1	0
Changed thinking cause of feelings	0	0	1	0	1	0
Avoided expressing emotions	0	0	1	0	1	0
Focused on feelings	1	0	1	0	1	0
Focused on problems	1	0	1	0	1	0
Activities to distract feelings	1	0	1	0	1	0

Concluding remarks

In this article, we proposed Clusterwise HICLAS, a generic modeling strategy for tracing differences and similarities between binary object \times variable data blocks that have the variables in common, which is a type of

data often encountered in psychology (see the examples mentioned in the introduction). Using Clusterwise HICLAS, the data blocks are clustered and for each cluster a set of bundles is derived, which represents the structural mechanisms underlying the data. As a consequence, differences and similarities between the data blocks can be studied easily by comparing the cluster-specific bundles. In an extensive simulation study, we demonstrated that the Clusterwise HICLAS algorithm performs well with respect to, on the one hand, minimizing the loss function, and on the other hand, disclosing the true clustering of the data blocks and the true cluster-specific bundles. Analyzing emotion data, we further showed that Clusterwise HICLAS is able to reveal groups of individuals who differ regarding emotion differentiation and emotion regulation strategies.

Although we introduced Clusterwise HICLAS as a model for coupled binary data blocks that have the variables in common, Clusterwise HICLAS can also be applied without further adaptations to coupled binary data blocks that share the objects. Such a Clusterwise HICLAS approach would also have interesting applications in the behavioral sciences. Take, as an example, a researcher who administers different questionnaires to the same set of persons (i.e., each data block pertains to a different questionnaire). In this case, a Clusterwise HICLAS analysis would result in a clustering of the questionnaires (e.g., intelligence and personality questionnaires) and cluster-specific bundles, which would reveal the personality and the intelligence of the persons under study. As such, one could study how individual differences in intelligence relate to individual differences in personality.

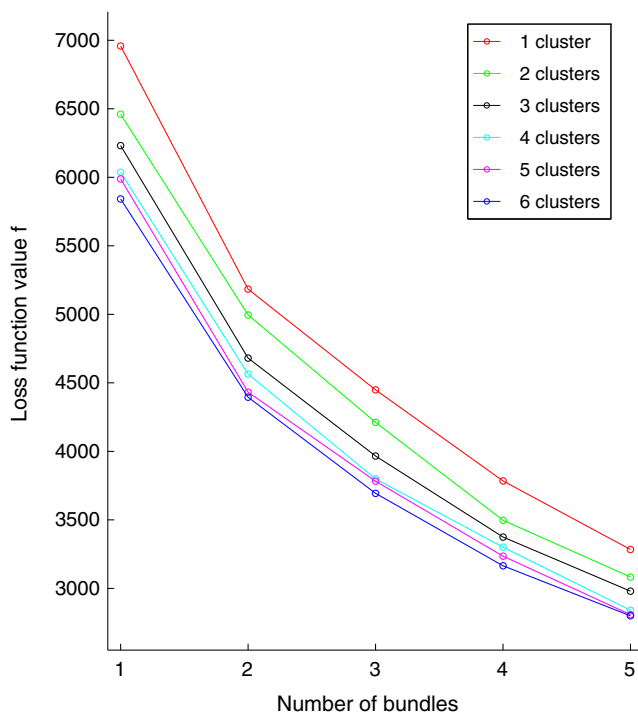


Fig. 3 Loss function value for Clusterwise HICLAS solutions, with the number of bundles varying from one to five and the number of clusters ranging from one to six

In the future, it might be useful to further extend the Clusterwise HICLAS approach with respect to two factors. First, in the Clusterwise HICLAS model, the number of bundles is assumed to be the same across clusters. Often, however, this is not a tenable assumption. For example, there are no substantive reasons why the number of bundles needed to describe the individual differences in personality should equal the number of bundles that underlie individual differences in intelligence. Therefore, it might be interesting to allow the number of bundles to vary across clusters. However, such an extension would not be trivial, because the estimation of the clustering of the data blocks could be a hard nut to crack. In particular, the algorithm might be tempted to assign blocks to those clusters that have more bundles (i.e., parameters), because in these clusters there are more parameters to describe the data, which might result in a better fit of the data. Moreover, a challenging model selection problem would arise, because the number of clusters and the number of bundles for each cluster would need to be determined.

Second, in Clusterwise HICLAS, the cluster-specific bundles are obtained by performing a HICLAS analysis on (a concatenation of) all data blocks belonging to the cluster in question, implying that each data entry influences the estimation of the bundles to the same extent. However, data entries from different blocks may differ in their information value. One reason for this might be that the data were measured in a more reliable way in some blocks than in others, resulting in the former containing less noise than the latter (i.e., noise heterogeneity between data blocks). Another reason might be that some blocks were better instances of a particular cluster than were others (i.e., some blocks might fit a cluster better than others). It has been demonstrated that not accounting (properly) for these differences in information value between data blocks can hamper the disclosure of the true structure underlying the data (see, e.g., Wilderjans et al., 2008, 2009, *in press*). Therefore, an interesting extension of the Clusterwise HICLAS modeling strategy would be to estimate the information value of each data block and to down-weight in the analysis entries from less-informative blocks in favor of entries belonging to more-informative blocks, which is the key idea behind SIMCLAS (Wilderjans et al., *in press*), in which the bundles are assumed to be equal across data blocks. This extension of Clusterwise HICLAS, however, also would not be a trivial one, because estimating the information value of a data block may be a very difficult task.

Author note The research reported in this article was partially supported by the Research Fund of Katholieke Universiteit Leuven (PDM, grant to T.F.W.); by the fund for Scientific Research (FWO)–Flanders (Belgium), Project G.0477.09 awarded to E.C., Marieke Timmerman, and Patrick Onghena; and by the Research Council of K.U. Leuven (Grant GOA/2010/02).

Appendix: HICLAS^{SA} algorithm

In the HICLAS^{SA} algorithm, only the values of the variable bundle matrix are considered as parameters that have to be estimated, while the values in the object bundle matrices are always kept at their optimal value, conditional on the current estimates of the values of the variable bundle matrix (i.e., each time the variable bundle matrix is updated, associated optimal object bundle matrices are calculated). In particular, calculating the optimal object bundle matrices conditional on the variable bundle matrix can be done by performing a Boolean regression with the (columns of the) variable bundle matrix as the predictor, the (rows in the) data matrix as the criterion, and the (rows of the) object bundle matrix being the Boolean regression weights (for the application of Boolean regression in the context of hierarchical classes analysis, see De Boeck & Rosenberg, 1988; Van Mechelen & De Boeck, 1990; Van Mechelen, Lombardi, & Ceulemans, 2007).

To estimate the variable bundle matrix that optimizes the loss function (i.e., the number of discrepancies), a simulated annealing procedure (SA; for an introduction, see Aarts, Korst, & van Laarhoven, 1997; Kirkpatrick, Gelatt, & Vecchi, 1983) was adopted. The goal of SA, which is a local search technique, is to walk through the solution space (i.e., all possible variable bundle matrices) in a more or less intelligent way (Aarts & Korst, 1990; Aarts et al., 1997; Kirkpatrick et al., 1983). The best-fitting solution that is encountered during this “walk” is retained as the final solution. Specifically, a chain of solutions is constructed that consists of several subchains, by the following procedure: First, on the basis of the current solution, a (new) candidate solution is generated. Next, if the candidate solution yields a better loss function value than the current one, the candidate solution is accepted and becomes the (new) current solution. However, sometimes a worse candidate solution is accepted in order to escape local minima; this depends on (1) the relative quality of the candidate solution (i.e., the difference in loss function value between the current and candidate solutions) and (2) a quantity, called the current “temperature”, that steers the probability of accepting worse solutions. At the end of each subchain, the temperature is decreased, and subchains are generated until a prespecified subchain stop criterion is met.

A single SA chain of the HICLAS^{SA} algorithm was implemented as follows:

1. Initial values for the variable bundle matrix were generated by independently sampling values from a Bernoulli distribution with parameter value .50;
2. To obtain the initial temperature, a subchain of solutions was generated, always accepting the new solution; the initial temperature then equaled the

average increase in the loss function across those steps in which worse solutions were accepted, divided by $\ln 0.8$;

3. A candidate solution was constructed by changing one randomly selected parameter of the current solution (i.e., one value of the variable bundle matrix), with each cell having the same probability of being changed;
4. A worse candidate solution was accepted if $p < \exp[-(f_{\text{curr}} - f_{\text{cand}})/T_{\text{curr}}]$, with T_{curr} being the current temperature, f_{curr} and f_{cand} indicating the loss function values for the current and the candidate solution, respectively, and p being sampled from a uniform (0, 1) distribution;
5. A subchain stopped if $(J \times 2^P) \times 5$ solutions had been generated or if 10% of this number had been accepted;
6. After performing each subchain, the temperature was decreased by a factor of .9 (i.e.,);
7. A SA chain stopped when the current temperature became smaller than .000001 or when, for the last five subchains, there was no change in the loss function value of the last accepted solution in each subchain.

Because the HICLAS^{SA} algorithm might end up in a local rather than the global optimum, we recommend using a multistart procedure, which boils down to running different SA chains that start from different initial variable bundle matrices (and different initial temperatures).

References

- Aarts, E. H. L., & Korst, J. H. M. (1990). *Simulated annealing and Boltzmann machines. A stochastic approach to combinatorial optimization and neural computing*. Chichester: Wiley.
- Aarts, E. H. L., Korst, J. H. M., & van Laarhoven, P. J. M. (1997). Simulated annealing. In E. H. L. Aarts & J. K. Lenstra (Eds.), *Local search in combinatorial optimization* (pp. 91–120). Chichester: Wiley.
- Ashmore, R. D., Deaux, K., & McLaughlin-Volpe, T. (2004). An organizing framework for collective identity: Articulation and significance of multidimensionality. *Psychological Bulletin, 130*, 80–114.
- Bagpy, R. M., Taylor, G. J., Quilty, L. C., & Parker, J. D. A. (2007). Reexamining the factor structure of the 20-item Toronto Alexithymia Scale: Commentary on Gignac, Palmer and Stough. *Journal of Personality Assessment, 89*, 258–264.
- Barbut, M., & Monjardet, B. (1970). *Ordre et classification: Algèbre et combinatoire*. Paris: Hachette.
- Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion, 15*, 713–724.
- Birkhoff, G. (1940). *Lattice theory*. Providence: American Mathematical Society.
- Bolger, N., Davis, A., & Rafaeli, E. (2003). Diary methods: Capturing life as it is lived. *Annual Review of Psychology, 54*, 579–616.
- Brusco, M. J. (2006). A repetitive branch-and-bound algorithm for minimum within-cluster sums of squares partitioning. *Psychometrika, 71*, 347–363.
- 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 meaning and strategic use of factor analysis. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 174–243). Chicago: Rand McNally.
- Ceulemans, E., & Storms, G. (2010). Detecting intra- and inter-categorical structure in semantic concepts using HICLAS. *Acta Psychologica, 133*, 296–304.
- Ceulemans, E., & Van Mechelen, I. (2003). Uniqueness of *n*-way *n*-mode hierarchical classes models. *Journal of Mathematical Psychology, 47*, 259–264.
- Ceulemans, E., & Van Mechelen, I. (2008). CLASSI: A classification model for the study of sequential processes and individual differences therein. *Psychometrika, 73*, 107–124.
- Ceulemans, E., Van Mechelen, I., & Leenen, I. (2003). Tucker3 hierarchical classes analysis. *Psychometrika, 68*, 413–433.
- 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*, 377–391.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.
- Collins, L. M., & Dent, C. W. (1988). Omega: A general formulation of the Rand index of cluster recovery suitable for non-disjoint solutions. *Multivariate Behavioral Research, 23*, 231–242.
- De Boeck, P. (2008). Random item IRT models. *Psychometrika, 73*, 533–559.
- De Boeck, P., & Rosenberg, S. (1988). Hierarchical classes: Model and data analysis. *Psychometrika, 53*, 361–381.
- de la Torre, J. (2011). The generalized dina model framework. *Psychometrika, 76*, 179–199.
- de Leeuw, J. (1994). Block-relaxation algorithms in statistics. In H. Bock, W. Lenski, & M. M. Richter (Eds.), *Information systems and data analysis* (pp. 308–325). Berlin: Springer.
- De Roover, K., Ceulemans, E., & Timmerman, M. E. (in press). How to perform multiblock component analysis in practice. *Behavior Research Methods*.
- De Roover, K., Ceulemans, E., Timmerman, M. E., & Onghena, P. (2011). *A clusterwise simultaneous component method for capturing within-cluster differences in component variances and correlations*. Manuscript submitted for publication.
- De Roover, K., Ceulemans, E., Timmerman, M. E., Vansteelandt, K., Stouten, J., & Onghena, P. (in press). Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. *Psychological Methods*. doi:10.1037/a0025385
- DeSarbo, W. S., Oliver, R. L., & Rangaswamy, A. (1989). A simulated annealing methodology for clusterwise linear regression. *Psychometrika, 54*, 707–736.
- Gordon, A. D. (1981). *Classification*. New York: Chapman & Hall.
- Haggard, E. A. (1958). *Intraclass correlation and the analysis of variance*. New York: Dryden.
- 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.
- Hart, D., & Fegley, S. (1995). Prosocial behavior and caring in adolescence: Relations to self-understanding and social judgment. *Child Development, 66*, 1346–1359.

- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2, 193–218.
- Ip, E. H., Wang, Y. J., De Boeck, P., & Meulders, M. (2004). Locally dependent latent trait models for polytomous responses with application to inventory of hostility. *Psychometrika*, 69, 191–216.
- Kashdan, T. B., Ferssizidis, P., Collins, R. L., & Muraven, M. (2010). Emotion differentiation as resilience against excessive alcohol use: An ecological momentary assessment in underage social drinkers. *Psychological Science*, 21, 1341–1347.
- Kiers, H. A. L. (2000). Towards a standardized notation and terminology in multiway analysis. *Journal of Chemometrics*, 14, 105–122.
- Kiers, H. A. L., & ten Berge, J. M. F. (1989). Alternating least squares algorithms for simultaneous components analysis with equal component weight matrices for all populations. *Psychometrika*, 54, 467–473.
- Kiers, H. A. L., & ten Berge, J. M. F. (1994). Hierarchical relations between methods for simultaneous component analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical and Statistical Psychology*, 47, 109–126.
- Kirk, R. E. (1982). *Experimental design: Procedures for the behavioral sciences* (2nd ed.). Belmont: CA Brooks/Cole.
- Kirkpatrick, S., Gelatt, C. D. J., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220, 671–680.
- Kroonenberg, P. M. (2008). *Applied multiway data analysis*. Hoboken: Wiley.
- Leenen, I., & Van Mechelen, I. (1998). A branch-and-bound algorithm for boolean regression. In I. Balderjahn, R. Mathar, & M. Schader (Eds.), *Classification, data analysis, and data highways* (pp. 164–171). Berlin: Springer.
- Leenen, I., & Van Mechelen, I. (2001). An evaluation of two algorithms for hierarchical classes analysis. *Journal of Classification*, 18, 57–80.
- Leenen, I., Van Mechelen, I., Gelman, A., & De Knop, S. (2008). Bayesian hierarchical classes analysis. *Psychometrika*, 73, 39–64.
- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In L. M. Le Cam & J. Neyman (Eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, pp. 281–297). Berkeley: University of California Press.
- Maris, E., De Boeck, P., & Van Mechelen, I. (1996). Probability matrix decomposition models. *Psychometrika*, 61, 7–29.
- Mickey, M. R., Mundle, P., & Engelman, L. (1983). Boolean factor analysis. In W. J. Dixon (Ed.), *BMDP statistical software manual* (pp. 538–545, 692). Berkeley: University of California Press.
- 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.
- Millsap, R. E., & Meredith, W. (1988). Component analysis in cross-sectional and longitudinal data. *Psychometrika*, 53, 123–134.
- Reich, W. A., Kressel, K., Scanlon, K. M., & Weiner, G. A. (2007). Predicting the decision to pursue mediation in civil disputes: A hierarchical classes analysis. *Journal of Psychology*, 141, 627–635.
- Reich, W. A., Lounsbury, D. W., Zaid-Muhammad, S., & Rapkin, B. D. (2010). Forms of social support and their relationships to mental health in HIV-positive persons. *Psychology, Health & Medicine*, 15, 135–145.
- Rosenberg, S., Van Mechelen, I., & De Boeck, P. (1996). A hierarchical classes model: Theory and method with applications in psychology and psychopathology. In P. Arabie, L. J. Hubert, & G. De Soete (Eds.), *Clustering and classification* (pp. 123–155). River Edge: World Scientific.
- Spath, H. (1982). Algorithm 48: A fast algorithm for clusterwise linear regression. *Computing*, 29, 175–181.
- Steinley, D. (2003). Local optima in *k*-means clustering: What you don't know may hurt you. *Psychological Methods*, 8, 294–304.
- Stirratt, M. J., Meyer, I. H., Ouellette, S. C., & Gara, M. A. (2008). Measuring identity multiplicity and intersectionality: Hierarchical classes analysis (HICLAS) of sexual, racial, and gender identities. *Self and Identity*, 7, 89–111.
- Suvak, M. K., Litz, B. T., Sloan, D. M., Zanarini, M. C., Barrett, L. F., & Hofmann, S. G. (2011). Emotional granularity and borderline personality disorder. *Journal of Abnormal Psychology*, 120, 414–426.
- ten Berge, J. M. F. (1993). *Least squares optimization in multivariate analysis*. Leiden: DSWO Press.
- ten Berge, M., & de Raad, B. (2001). The construction of a joint taxonomy of traits and situations. *European Journal of Personality*, 15, 253–276.
- Timmerman, M. E., & Kiers, H. A. L. (2003). Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, 68, 105–121.
- 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 and Statistical Psychology*, 62, 299–318.
- Van Deun, K., Smilde, A. K., van der Werf, M. J., Kiers, H. A. L., & Van Mechelen, I. (2009). A structured overview of simultaneous component based data integration. *BMC Bioinformatics*, 10, 246.
- Van Mechelen, I., & De Boeck, P. (1989). Implicit taxonomy in psychiatric diagnosis: A case study. *Journal of Social and Clinical Psychology*, 8, 276–287.
- Van Mechelen, I., & De Boeck, P. (1990). Projection of a binary criterion into a model of hierarchical classes. *Psychometrika*, 55, 677–694.
- Van Mechelen, I., De Boeck, P., & Rosenberg, S. (1995). The conjunctive model of hierarchical classes. *Psychometrika*, 60, 505–521.
- Van Mechelen, I., Lombardi, L., & Ceulemans, E. (2007). Hierarchical classes modeling of rating data. *Psychometrika*, 72, 475–488.
- Van Mechelen, I., & Van Damme, G. (1994). A latent criteria model for choice data. *Acta Psychologica*, 87, 85–94.
- van Os, B. J., & Meulman, J. J. (2004). Improving dynamic programming strategies for partitioning. *Journal of Classification*, 21, 207–230.
- Vande Gaer, E., Ceulemans, E., Van Mechelen, I., & Kuppens, P. (in press). The CLASSI-N model for the study of sequential processes. *Psychometrika*.
- Wang, C., & Chang, H. (2011). Item selection in multidimensional computerized adaptive testing-gaining information from different angles. *Psychometrika*, 76, 363–384.
- Wilderjans, T. F., Ceulemans, E., & Van Mechelen, I. (2008). The CHIC model: A global model for coupled binary data. *Psychometrika*, 73, 729–751.
- Wilderjans, T. F., Ceulemans, E., & Van Mechelen, I. (2009). Simultaneous analysis of coupled data blocks differing in size: A comparison of two weighting schemes. *Computational Statistics and Data Analysis*, 53, 1086–1098.
- Wilderjans, T. F., Ceulemans, E., & Van Mechelen, I. (in press). The SIMCLAS model: Simultaneous analysis of coupled binary data matrices with noise heterogeneity between and within data blocks. *Psychometrika*.
- Wilson, S. J., Bladin, P. F., Saling, M. M., & Pattison, P. E. (2005). Characterizing psychosocial outcome trajectories following seizure surgery. *Epilepsy & Behavior*, 6, 570–580.
- Wranik, T., Barrett, L. F., & Salovey, P. (2007). Intelligent emotion regulation: Is knowledge power? In J. Gross (Ed.), *Handbook of emotion regulation* (pp. 393–428). New York: Guilford Press.