

How to detect which variables are causing differences in component structure among different groups

Kim De Roover¹ · Marieke E. Timmerman² · Eva Ceulemans¹

Published online: 10 December 2015
© Psychonomic Society, Inc. 2015

Abstract When comparing the component structures of a multitude of variables across different groups, the conclusion often is that the component structures are very similar in general and differ in a few variables only. Detecting such “outlying variables” is substantively interesting. Conversely, it can help to determine what is common across the groups. This article proposes and evaluates two formal detection heuristics to determine which variables are outlying, in a systematic and objective way. The heuristics are based on clusterwise simultaneous component analysis, which was recently presented as a useful tool for capturing the similarities and differences in component structures across groups. The heuristics are evaluated in a simulation study and illustrated using cross-cultural data on values.

Keywords Multigroup data · Multilevel data · Simultaneous component analysis · Clustering · Invariance

Introduction

Assessing the covariance structures of a large set of variables across multiple groups is an important analysis step in behavioral research. To this end, dimension reduction methods are the methods of choice. In particular, if one has an a priori idea about how the covariances are caused by a few latent

variables, one usually resorts to the confirmatory factor analysis framework (Jöreskog, 1971; Kline, 2004; Sörbom, 1974). Often one has no such hypothesis, however, and then exploratory factor analysis (Dolan, Oort, Stoel & Wicherts, 2009; Hessen, Dolan & Wicherts, 2006) or component analysis (Jolliffe, 2002) may be used. In this article, we will focus on component analysis, which is more widely applicable than factor analysis, because it implies less stringent assumptions (e.g., no assumption of local independence of the variables, which often is unreasonable; see Borsboom, Mellenbergh & van Heerden, 2003).

When comparing component structures across groups, two types of differences may be revealed. On the one hand, one may find that subsets of groups have completely different component structures (see, e.g., the application in De Roover, Ceulemans, Timmerman, & Onghena 2013b). On the other hand, it often occurs that the component structures are very similar in general and differ in a few variables only (see the second application in De Roover, Ceulemans, Timmerman, Vansteelandt, Stouten & Onghena, 2012b, for an example). Such variables will be referred to as “outlying variables.” Detecting such outlying variables is important for two complementary reasons: First, it can reveal substantively interesting differences between the groups. Second, it helps to determine what is common across the groups. For instance, Krysinska et al. (2014) examined differences in the psychometric structures of the Post-Critical Belief Scale across samples that were measured many years ago as well as recent ones, to evaluate possible changes in the meanings of the 33 scale items over time. Comparing the component structures across the samples, two outlying items were found. On the one hand, these two outlying items indicated that an important shift in the interpretation of bible stories had taken place between the earlier and more recent samples. On the other hand, the part of the component structure that was stable across time

✉ Kim De Roover
Kim.DeRoover@ppw.kuleuven.be

¹ Research Group of Quantitative Psychology and Individual Differences, Tiensestraat 102, B-3000 Leuven, Belgium

² University of Groningen, Groningen, The Netherlands

was also of interest and was compared to the theoretically expected structure.

Identifying outlying variables can be cumbersome, however. It becomes increasingly difficult, the more groups are involved, because more structures have to be compared. Furthermore, the specific detection strategy followed may strongly impact the results, because component structures are highly sensitive to the specific sets of variables involved, and thus to which outlying items are sidelined step by step. To make these decisions in a more systematic and objective way, we propose and evaluate two formal detection heuristics. These heuristics are based on clusterwise simultaneous component analysis (clusterwise SCA; De Roover et al., 2012b). Clusterwise SCA was introduced to simplify the daunting task of finding between-group differences in component structures when the number of groups is large. Specifically, it assigns the groups to a few clusters and simultaneously conducts an SCA per cluster to summarize the within-cluster covariance structure. Consequently, the most important between-group differences in component structure are captured in the cluster-specific component loadings. Therefore, these loadings provide a good starting point to efficiently perform outlying-variable detection, even when the number of groups is large.

The remainder of the article is organized in five sections: First, the data structure and preprocessing are discussed. Then clusterwise SCA is discussed, followed by a description of the two detection heuristics, as well as a split-half procedure to improve the robustness of the detection results. The following section presents a simulation study to compare the performances of these heuristics, and the next illustrates the heuristics using cross-cultural data on values. To conclude, we describe some points for discussion and directions for future research.

Data structure and preprocessing

We assume that one disposes of I data blocks \mathbf{X}_i ($N_i \times J$) that each contain the scores of N_i subjects on the same J variables. For the sake of stable model estimates, each N_i is preferably larger than J . The I data blocks can be vertically concatenated into an $N \times J$ data matrix \mathbf{X} , where $N = \sum_{i=1}^I N_i$. To avoid between-block differences in the variable means being confounded with between-block differences in the within-block covariance structures, each variable is centered per data block. Since differences in the variances of the variables, both within and across blocks, affect the obtained component structure (Bro & Smilde, 2003; Harshman & Lundy, 1984;

Timmerman, Hoefsloot, Smilde & Ceulemans, 2015), the data may optionally be standardized. One may standardize across blocks (e.g., Timmerman & Kiers, 2003) or within blocks (e.g., De Roover, Ceulemans & Timmerman, 2012a), depending on whether one is interested in differences in covariance structures or correlation structures, respectively.

Method

In this section, we start by describing SCA and its clusterwise extension. Next, we introduce two heuristics for detecting outlying variables and a split-half procedure.

Simultaneous component analysis

In this article, we will use SCA-P (SCA with equal pattern matrices; Kiers & ten Berge, 1994), in which the I data blocks \mathbf{X}_i are modeled as follows:

$$\mathbf{X}_i = \mathbf{F}_i \mathbf{B}' + \mathbf{E}_i, \quad (1)$$

where \mathbf{F}_i ($N_i \times Q$) denotes the scores of the subjects in the i th group on the Q components, \mathbf{B} ($J \times Q$) denotes the loading matrix that is the same for all groups, and \mathbf{E}_i ($N_i \times J$) denotes the matrix of residuals. To partly identify the model, the variances of the component scores, computed across all groups, are fixed at one. The SCA-P model can be estimated via a principal component analysis of the $N \times J$ data matrix \mathbf{X} . Note that other variants of SCA exist, in which additional restrictions on the component scores of each group are imposed (Timmerman & Kiers, 2003). SCA-P solutions have rotational freedom, which can be used to facilitate interpretation. In this article, we will conduct a normalized VARIMAX rotation (Kaiser, 1958), but note that other criteria can be used equally well.

Although theoretical knowledge about the variables or interpretability of the solution will often drive how many components will be used, also formal model selection heuristics are available. A very popular heuristic is Cattell's scree test (1966) that selects the number of components after which the increase in model fit gained from additional components levels off: Q^{best} . This test may be conducted visually—that is, by looking for an elbow point in a scree plot (see, e.g., the Application section), or numerically—that is, by calculating scree ratios (see, e.g., Ceulemans & Kiers, 2006; Wilderjans, Ceulemans & Meers, 2013).

Clusterwise SCA

Clusterwise SCA-P (De Roover et al., 2013b) assigns each of the I groups to one of K clusters, while modeling the data within each cluster with SCA-P. Consequently, groups with

a similar component structure end up in the same cluster and differences in component structures can be examined by comparing the cluster-specific loading structures. Specifically, the model equation of clusterwise SCA-P is given by

$$\mathbf{X}_i = \sum_{k=1}^K p_{ik} \mathbf{F}_i \mathbf{B}^{(k)'} + \mathbf{E}_i. \quad (2)$$

Comparing Eqs. 1 and 2, we see that the loading matrix now has a superscript “(k)” that indicates its cluster-specific nature; p_{ik} indicates the estimated cluster membership of group i and equals one when group i is assigned to cluster k and zero when it is not. Note that clusterwise SCA-P has rotational freedom per cluster.

To estimate a clusterwise SCA-P model with K clusters and Q components for a given data set, the sum of the squared residuals is minimized by means of an alternating least squares algorithm (see De Roover et al., 2013b). To reduce the probability of ending up in a local minimum, a multistart procedure is applied.¹ To determine the most appropriate number of clusters, clusterwise SCA-P analyses are performed with different numbers of clusters and Q^{best} components. Subsequently, a scree test may be performed, by visually inspecting a scree plot (see, e.g., the [Application](#) section) or by computing scree ratios, to determine the most appropriate number of clusters K^{best} . Yet, note that if the number of outlying variables is small, the differences in fit between solutions with different numbers of clusters may be very small, making the scree test less informative. In such cases, we recommend to explore solutions with different numbers of clusters in terms of outlying variables or one of them could be chosen on the basis of interpretability and/or (e.g., split-half) stability of the clustering and cluster-specific loading matrices. Of course, one should be aware that the more clusters, the more outlying variables will be detected. Indeed, a variable only needs to have a different loading structure in two of the clusters to be detected as outlying.

Other variants of clusterwise SCA exist, but are inappropriate for our present purposes. First, there are variants with equality restrictions across groups on the component variances and/or the correlations between the component scores (De Roover et al., 2012b; De Roover, Timmerman, Van Mechelen & Ceulemans, 2013c). Imposing these restrictions may lead to loading differences that are irrelevant for outlyingness. Furthermore, a variant exists that allows the number of components to differ across the clusters (De Roover, Ceulemans, Timmerman, Nezlek & Onghena, 2013a). We refrain from considering this variant, because we assume the component structure to be largely the same across clusters, and hence can safely impose an equal number of

components per cluster. This number can be chosen on the basis of the SCA-P analysis.

Outlying variable detection

To automate the detection of outlying variables, a so-called “outlyingness criterion” is needed. In this article we will focus on the proportional similarity of component loadings across clusters of groups, as quantified by the congruence coefficient (Tucker, 1951). This coefficient is computed per component (i.e., per column of loadings). It takes values between -1 and 1 , where the extreme values of -1 and 1 represent perfect proportional similarity between the two cluster-specific components (with and without reflection of the component in one of the clusters, respectively). According to Lorenzo-Seva and ten Berge (2006), a congruence value higher than .95 reflects virtual identity. Therefore, one might conclude that at least one outlying variable is present if the Tucker congruence value of at least one component is smaller than .95 for at least one cluster pair. Hence, in our first method, called “cutoff congruence,” we will discard variables until all congruence values exceed the .95 cutoff.

However, the correctness of the cutoff value can be debated. Indeed, Paunonen (1997) has shown that congruence values depend on the data characteristics (e.g., the number of variables, the variables-to-components ratio). Furthermore, it is plausible that the sensitivity of the congruence coefficient is affected by the nonoutlying-to-outlying variables ratio. Because it will probably be impossible to find a critical congruence value that works best in all conditions (i.e., when a certain value is ideal for one set of conditions, it may be too high—thus leading to false positives—in another set of conditions, and too low—inducing false negatives—in yet another set of conditions), the second heuristic uses the .95 value as a lower bound rather than a cutoff, and is therefore called the “lower-bound congruence” method.

In both methods, we have to resolve arbitrary differences between the cluster loading matrices in axis positions (rotational freedom), permutations, and reflections. To this end, we first estimate an SCA-P model (i.e., yielding a single loading matrix for all I groups under study) and rotate the SCA-P loadings toward a simple structure using normalized VARIMAX. Subsequently, we estimate the clusterwise SCA-P model and obliquely Procrustes rotate the cluster-specific loadings toward the normalized VARIMAX SCA-P ones. Note that we opt for oblique rotations of the cluster-specific loadings because we are not interested in differences in cross-loadings that are due to differences in the cluster-specific component correlations. The necessity of allowing for cluster-specific correlations between components precludes the use of state-of-the-art consensus rotations that simultaneously rotate all loading matrices to achieve both a simple structure of and maximal agreement between the

¹ The clusterwise SCA-P algorithm is implemented in a MATLAB R2014b function that can be obtained from the corresponding author.

loading matrices—for instance, consensus direct oblimin rotation (Lorenzo-Seva, Kiers & ten Berge, 2002), which outperformed other alternatives in a simulation study and pursues both a simple structure of and maximal agreement between the loading matrices, but does not allow for differences in component correlations across the clusters.

In the following paragraphs, we discuss the details of the two methods as well as a split-half procedure that can be used to obtain more robust results. As a guiding example, we will use the hypothetical loadings in Table 2 below. These loadings pertain to two component structures that are equal for Items 1 through 9 and differ for Items 10 through 13. The associated normalized VARIMAX-rotated SCA-P loadings and the obliquely Procrustes-rotated cluster-specific loadings are also given in Table 1.

Cutoff congruence method

The cutoff heuristic was recently used in De Roover, Timmerman, De Leersnyder, Mesquita and Ceulemans (2014a). It proceeds as follows:

1. For each cluster pair, component-specific congruence coefficients are computed and the minimum of these coefficients is retained as $\varphi_{k_1 k_2}^{\min}$, with k_1 and k_2 denoting the two clusters in the cluster pair. Following the rationale discussed above, we stop if the minimum $\varphi_{k_1 k_2}^{\min}$ value over cluster pairs exceeds .95, and thus indicates the virtual identity of all components; otherwise, we continue.
2. A set of variable-specific congruence-after-exclusion values is computed, by excluding each variable one by one. To this end, we compute per cluster pair the mean congruence value for the remaining variables across components, and retain the minimum value across cluster pairs.² Thus, we do not use the minimum value across components (see Step 1) (as De Roover et al., 2014a, and Krysinska et al., 2014, had done), because pilot simulation studies have shown that this value is, in some cases, quite prone to false positives. The variable for which this congruence after exclusion is the highest is considered the most outlying, and is therefore permanently removed. This step is repeated until the minimum congruence across all components and cluster pairs exceeds the .95 threshold.
3. The cluster-specific and overall SCA-P models are reestimated, using the retained variables only. The former

model is rotated to the latter, and all steps are repeated until no more outlying variables are found.

When applying this procedure to the hypothetical example, we start off with a congruence value of .73 in Step 1, suggesting the presence of at least one outlying variable. When tentatively removing the items one by one, the variable-specific congruence-after-exclusion values range between .81 and .90. The highest value is obtained for Item 11, which is therefore removed first in Step 2. Repeating this step leads to the removal of Items 10, 12, and 13. Finally, Step 3 does not yield additional outlying items.

Lower-bound congruence method

This method consists of the following steps:

1. For each cluster pair, both the minimum and mean congruence values across components are computed—that is, $\varphi_{k_1 k_2}^{\min}$ (see Step 1 of the cutoff congruence method) and $\varphi_{k_1 k_2}^{\text{mean}}$.
2. Variable-specific congruence-after-exclusion values are computed, and the most outlying variable is identified (see Step 2 of the cutoff congruence method). This variable is removed and its number is added to the outlyingness ranking matrix \mathbf{O} , together with the minimum $\varphi_{k_1 k_2}^{\min}$ and $\varphi_{k_1 k_2}^{\text{mean}}$ values from Step 1—thus, from before the variable's removal—and the cluster pair corresponding to the minimum $\varphi_{k_1 k_2}^{\text{mean}}$ value:

$$\mathbf{O} = \begin{bmatrix} \min(\varphi_{k_1 k_2}^{\text{mean}}) & k_1 & k_2 & \text{most outlying variable} & \min(\varphi_{k_1 k_2}^{\min}) \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

3. As in Step 3 of the cutoff congruence method, the cluster-specific and overall SCA-P models are reestimated, using the retained variables only, and the former is rotated to the latter. We keep alternating Steps 1 to 3, removing only one variable at a time, until only Q variables are left, implying that the (clusterwise) SCA-P models can no longer be reestimated.
4. To determine the number of outlying variables, the minimum $\varphi_{k_1 k_2}^{\text{mean}}$ values in the first column of \mathbf{O} are plotted against the number of removed variables (i.e., from 0 to $J - Q$). On this plot, the CHull procedure (Ceulemans & Kiers, 2006; Wilderjans et al., 2013) is performed to determine the number of removed variables J_{out} after which the increase in $\min(\varphi_{k_1 k_2}^{\text{mean}})$ levels off. However, to ensure that the retained variables have virtually identical structures in all clusters, we only consider scree ratios for selections of variables for which the $\min(\varphi_{k_1 k_2}^{\min})$ value is larger than the lower bound of .95. Finally, the first J_{out} variables in the fourth column of \mathbf{O} are considered to be the outlying variables.

² Before calculating the congruence values, the cluster-specific loadings of the considered subset of variables are rerotated toward the corresponding SCA-P loadings, which are also rerotated to a simple structure. This is done because outlying variables can adversely affect the rotations of the cluster-specific loadings toward the SCA-P loadings, as well as the simple-structure rotation of the SCA-P loadings itself.

Table 1 Hypothetical component loadings for two clusters, differing only with respect to the loadings of Items 10 to 13, the normalized VARIMAX-rotated SCA-P loadings for the associated hypothetical data

	Hypothetical Loadings						SCA-P			Clusterwise SCA-P					
	Cluster 1			Cluster 2						Cluster 1			Cluster 2		
Item 1	1	0	0	1	0	0				.99	-.07	-.05	1.00	.01	-.01
Item 2	1	0	0	1	0	0	.99	-.07	-.05	1.00	.01	-.01	1.00	-.13	.01
Item 3	1	0	0	1	0	0	.99	-.07	-.05	1.00	.01	-.01	1.00	-.13	.01
Item 4	0	1	0	0	1	0	.00	.99	.01	-.02	1.00	.13	.05	.99	-.05
Item 5	0	1	0	0	1	0	.00	.99	.01	-.02	1.00	.13	.05	.99	-.05
Item 6	0	1	0	0	1	0	.00	.99	.01	-.02	1.00	.13	.05	.99	-.05
Item 7	0	0	1	0	0	1	-.02	.04	.99	.08	-.01	1.01	-.02	.17	.99
Item 8	0	0	1	0	0	1	-.02	.04	.99	.08	-.01	1.01	-.02	.17	.99
Item 9	0	0	1	0	0	1	-.02	.04	.99	.08	-.01	1.01	-.02	.17	.99
<i>Item 10</i>	<i>0</i>	<i>1</i>	<i>0</i>	<i>0</i>	<i>0</i>	<i>1</i>	<i>-.04</i>	<i>.60</i>	<i>.54</i>	<i>-.02</i>	<i>1.00</i>	<i>.13</i>	<i>-.02</i>	<i>.17</i>	<i>.99</i>
<i>Item 11</i>	<i>0</i>	<i>0</i>	<i>1</i>	<i>1</i>	<i>0</i>	<i>0</i>	<i>.47</i>	<i>-.06</i>	<i>.52</i>	<i>.08</i>	<i>-.01</i>	<i>1.01</i>	<i>1.00</i>	<i>-.13</i>	<i>.01</i>
<i>Item 12</i>	<i>1</i>	<i>0</i>	<i>0</i>	<i>.77</i>	<i>.63</i>	<i>0</i>	<i>.92</i>	<i>.25</i>	<i>-.08</i>	<i>1.00</i>	<i>.01</i>	<i>-.01</i>	<i>.80</i>	<i>.56</i>	<i>-.03</i>
<i>Item 13</i>	<i>1</i>	<i>0</i>	<i>0</i>	<i>.77</i>	<i>0</i>	<i>.63</i>	<i>.90</i>	<i>-.01</i>	<i>.27</i>	<i>1.00</i>	<i>.01</i>	<i>-.01</i>	<i>.78</i>	<i>.01</i>	<i>.66</i>

set, and the thereto obliquely Procrustes-rotated loadings of the clusterwise SCA-P model with two clusters and two components for the hypothetical data

Applying this procedure to the hypothetical example results in the outlyingness ranking matrix in Table 2. We see that the congruences quickly increase when removing Items 11 and 10, but Items 12 and 13 also need to be removed to reach a $\min(\varphi_{k_1 k_2}^{\min})$ larger than .95—note that after removing Item 12, this value is actually .9458; thus, Item 13 also needs to be removed. After removing these four items, the congruences become 1.00 (because the data are errorless); therefore, this is the elbow selected by the CHull procedure (see Fig. 1).

Split-half procedure

To mitigate the effects of sampling fluctuations on the outlying variable detection, we propose using the following split-half procedure: First, split the data in two halves, by randomly selecting half of the rows of each data block and assigning them to the first half; the remainder of the data are collected in the second half. Next, the data blocks of both halves are

Table 2 Outlyingness ranking matrix that results from the lower-bound congruence method for the hypothetical example

$\min(\varphi_{k_1 k_2}^{\text{mean}})$	k_1	k_2	Most Outlying Variable	$\min(\varphi_{k_1 k_2}^{\min})$
.83	1	2	11	.73
.90	1	2	10	.85
.96	1	2	12	.94
.98	1	2	13	.95
1.00	1	2	4	1.00
1.00	1	2	5	1.00
1.00	1	2	6	1.00
1.00	1	2	8	1.00
1.00	1	2	7	1.00
1.00	1	2	1	1.00
1.00	1	2	2	1.00

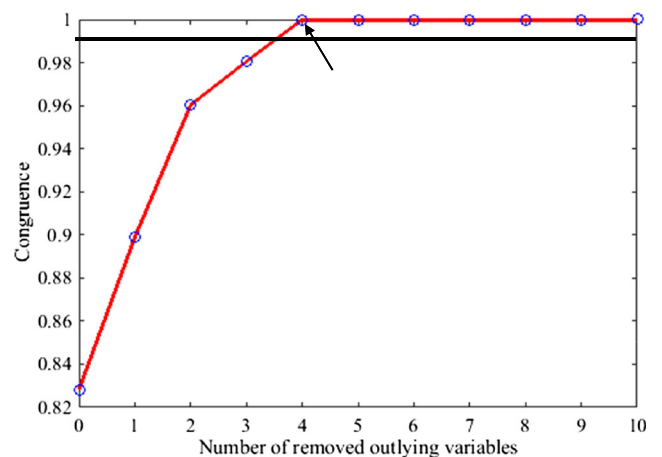


Fig. 1 CHull plot of the lower-bound congruence method for the values data. Specifically, the $\min(\varphi_{k_1 k_2}^{\text{mean}})$, labeled “Congruence,” is plotted against the number of variables already removed (the order wherein the variables are removed can be found in Table 2). The black horizontal line indicates where the $\min(\varphi_{k_1 k_2}^{\min})$ value (not depicted in the figure, but in Table 2) crosses the lower bound of .95. The arrow indicates the elbow after which the decrease in congruence levels off according to the CHull method

clustered according to the partition that resulted from the clusterwise SCA-P analysis on the entire data set, and SCA-P is performed per cluster as well as on the complete half. Subsequently, outlying-variable detection is performed using all of the half-specific loadings. Note that the clustering is not reestimated for each half of the data, for two reasons. First, the clustering is kept constant to avoid an entanglement of the stability of outlying-variable detection with the stability of the clustering. Second, for the procedure to make sense, the outlying-variable detection should be performed on clusters that are the same for both halves. The variables that are detected in both halves are considered to be outlying for the random split in question.

Of course, the random splits themselves are also very susceptible to sampling fluctuations. Therefore, we propose performing 20 different random splits and recording the 20 resulting sets of outlying variables. Afterward, the modus of the sets of outlying variables—that is, the set of outlying variables that is retained most often—is considered to be the final set of outlying variables. For the hypothetical example, the same set of outlying variables is obtained for each random split, because the data are error-free.

Simulation study

Problem

In this section, we present a simulation study in which the overall performances of the two heuristics are compared, as well as how the performance is influenced by five factors: (1) the number of nonoutlying variables, (2) the number of outlying variables, (3) the degree of outlyingness, (4) the number of clusters, and (5) the amount of error in the data. Factors 1 to 3 were chosen to assess whether the cutoff congruence method is sensitive to the critical congruence value used. With respect to Factors 4 and 5, we hypothesized that a higher number of clusters and larger amounts of error might complicate outlying variable detection. Finally, we explored the quality of the outlying-variable detection when too many clusters are used, because determining the appropriate number of clusters may be hard in empirical practice. When using too few clusters, performance will almost always be bad, due to the loss of information (i.e., the merging of clusters leads to mixing of the component structures; see De Roover et al., 2012b); therefore, we do not investigate this empirically. Note that, in contrast to previous clusterwise SCA simulations, we chose not to vary the numbers of data blocks, the numbers of rows per data block, and the cluster sizes, because we expect them to impact outlying-variable detection mostly indirectly, through the goodness of recovery of the clustering and the loading structures. For more detailed results on the goodness of recovery of clusterwise SCA-P models as a

function of these data characteristics, the reader is referred to De Roover et al. (2013b).

Design

The number of data blocks I was fixed at ten, and the number of observations N_i per data block at 75. Each simulated data set consisted of two or three equally sized clusters. The number of underlying components per cluster Q was set to three. Five factors were systematically varied in a complete factorial design:

1. the *number of nonoutlying variables* J_{no} , at two levels: 9, 12;
2. the *number of outlying variables* J_o , at three levels: 2, 4, 6^3 ;
3. the *degree of outlyingness*, at five levels: very high, high, medium, low, and very low;
4. the *number of clusters* K , at two levels: 2, 3;
5. the *error level* e , which is the expected proportion of error variance in the data blocks \mathbf{X}_i , at two levels: .20, .40;

For each cell of the factorial design, 100 data matrices \mathbf{X} were generated. We decided to use 100 replicates because this number corresponds to a maximal standard error for proportions—most results will be expressed as proportions—of .05. Each data matrix consisted of ten \mathbf{X}_i data blocks. For each data block, a component score matrix \mathbf{F}_i was randomly sampled from a multivariate normal distribution,⁴ of which the mean vector consists of zeros and of which the variance–covariance matrix was obtained by uniformly sampling the component correlations and variances between $-.5$ and $.5$ and between 0.25 and 1.75 , respectively. To construct the partition matrix \mathbf{P} , the groups were randomly assigned to the clusters, making sure the clusters had the same size. To generate the cluster-specific loading matrices $\mathbf{B}^{(k)}$, we determined randomly which of the J (equal to $J_{no} + J_o$) variables were outlying. To each of the three components, one third of the nonoutlying variables were assigned, by setting the corresponding loading to 1 and the others to 0. To simulate the different degrees of outlyingness (Factor 3), the outlying variables were randomly

³ Note that crossing Factors 1 and 2 manipulates the total number of variables (i.e., between 11 and 18) as well as the proportion of outlying variables (i.e., between .14 and .40).

⁴ Note that the component scores and residuals were sampled from a normal distribution. In component analysis, no specific distributional assumptions are made. On the basis of results obtained by Dudzinski, Norris, Chmura and Edwards (1975) and by Jolliffe (2002) with respect to principal component analysis, we would expect the results of the simulation to be comparable in the cases of other distributions, unless outliers are present (Jolliffe, 2002). For the latter case, a robust variant of clusterwise SCA-P would be needed, which could be developed along the lines of the work by Ceulemans, Hubert and Rousseeuw (2013) and Hubert, Rousseeuw and Vanden Branden (2005).

assigned to one component in Cluster 1, whereas in the other cluster(s) they received a loading b_{outl1} on the same component, but also a loading b_{outl2} on another component. The latter component differed between Clusters 2 and 3 in the case of three clusters (Factor 4). The sizes of these two loadings depended on the level of Factor 3: For a very high degree of outlyingness, b_{outl1} equals $\sqrt{.25}$ and b_{outl2} equals $\sqrt{.75}$, whereas for high, medium, low, and very low degrees of outlyingness, they equal $\sqrt{.50}$ and $\sqrt{.50}$, $\sqrt{.75}$ and $\sqrt{.25}$, $\sqrt{.85}$ and $\sqrt{.15}$, and $\sqrt{.95}$ and $\sqrt{.05}$, respectively. The error matrices \mathbf{E}_i were randomly sampled from the standard normal distribution. The cluster loading matrices $\mathbf{B}^{(k)}$ and the error matrices \mathbf{E}_i were rescaled by multiplying them by $\sqrt{(1-e)}$ and \sqrt{e} , respectively, such that the data contained the correct amount of error. Finally, each \mathbf{X}_i matrix was computed as $\mathbf{F}_i \mathbf{B}^{(k)'} + \mathbf{E}_i$.

The 12,000 simulated \mathbf{X} matrices were preprocessed such that each variable had a mean of zero per block and a unit variance over all blocks. Next, they were analyzed once with SCA-P and twice with clusterwise SCA-P, using K and $K + 1$ clusters; we always adopted the correct number of components Q . The clusterwise SCA-P algorithm was run 25 times, each time using a different random start, and the best solution out of the 25 runs was retained. Then, both heuristics as well as the split-half procedure were applied to the resulting clusterwise SCA-P loadings, using a critical congruence value of .95. On average, analyzing one data set with the correct number of clusters K took about 5 s (using MATLAB R2014b on an Intel Core i7-3770K processor of a personal computer, with a clock frequency of 3.4–3.9 GHz and a RAM speed of 1,600 MHz) without the split-half procedure, and 3 min when this procedure was also conducted.

Results

In this section, we first evaluate whether the clusterwise SCA-P analyses correctly recovered the underlying clustering and component structures in the case of K estimated clusters, because good outlying-variable detection is impossible otherwise. Next, the goodness of the outlying-variable detection is evaluated for both heuristics presented above. Then we report the results of the split-half procedure, focusing on the best-performing heuristic. Finally, the goodness of the outlying-variable detection when using one cluster too much is reported for the best heuristic.

Goodness of recovery of the clusterwise SCA-P clusters and loadings

To examine the recovery of the clustering, we computed the *adjusted Rand index* (ARI; Hubert & Arabie, 1985) between the true partition and the estimated one. The ARI equals 1 if

both are identical, and equals 0 when agreement is at chance level. The ARI was equal to 1 for 10,274 (86 %) out of the 12,000 data sets, with an overall mean of .91 ($SD = .23$). Thus, the clustering was recovered perfectly in most cases. Clustering mistakes were mainly made in the most difficult conditions. Specifically, 1,636 out of the 1,726 faulty clusterings occurred in the conditions with low or very low degree of outlyingness.

To evaluate how well the cluster-specific loading matrices were recovered, we calculated a goodness-of-cluster-loading-recovery statistic (*GOCL*) by computing congruence coefficients φ (Tucker, 1951) between the true and estimated component loadings and averaging these coefficients across components and clusters as follows:

$$GOCL = \frac{\sum_{k=1}^K \sum_{q=1}^Q \varphi(\mathbf{B}_q^{(k)T}, \mathbf{B}_q^{(k)M})}{KQ}, \quad (3)$$

with $\mathbf{B}_q^{(k)T}$ and $\mathbf{B}_q^{(k)M}$ indicating the q th component of the true and estimated cluster-loading matrices, respectively. Each estimated loading matrix $\mathbf{B}_q^{(k)M}$ was obliquely Procrustes-rotated toward its true counterpart $\mathbf{B}_q^{(k)T}$. To identify for each estimated loading matrix its associated true counterpart, the *GOCL* values were computed across all possible permutations, and the one that maximized the *GOCL* value was retained. The *GOCL* values ranged from 0 (no recovery at all) to 1 (perfect recovery). On average, the *GOCL* statistic amounted to .9951 ($SD = .0058$), indicating excellent recovery of the $\mathbf{B}^{(k)}$ matrices.

Goodness of outlying-variable detection

Table 3 shows, for both methods, the proportions of data sets with perfect outlying variable detection (i.e., data sets for which no false negatives or false positives occurred), the proportions of data sets without false negatives, and the numbers of false positives. Focusing on the overall performance, the lower-bound congruence method clearly outperformed the cutoff congruence method, with a proportion correct of .79 in comparison to only .37.

To get an indication of why the cutoff congruence method fell short, we examine the influence of the five manipulated factors. Not unexpectedly, we see that the performance was mostly influenced by the degree of outlyingness of the outlying variables. In *Outlying variable detection* section, we already hypothesized that the critical congruence value may not be ideal for all conditions. Indeed, it is clearly not sensitive enough to detect subtle loading differences. Therefore, we also evaluated the performance of the cutoff congruence method when a critical congruence value of .96 was applied. Table 4 shows that performance increased somewhat for the

Table 3 Proportions of correct data sets (i.e., data sets without false negatives or false positives), proportions of data sets without false negatives, and numbers of false positives for each method and for each level of the manipulated factors of the simulation study

	Correct Data Sets		No False Negatives		Number of False Positives	
	Cutoff Congruence	Lower-Bound Congruence	Cutoff Congruence	Lower-Bound Congruence	Cutoff Congruence	Lower-Bound Congruence
Nine nonoutlying	.40	.77	.40	.86	733	3,473
Twelve nonoutlying	.34	.81	.34	.90	454	3,760
Two outlying	.38	.80	.39	.90	421	2,257
Four outlying	.36	.80	.37	.88	280	2,189
Six outlying	.35	.76	.36	.85	486	2,787
Very high degr. outlyingness	.99	1.00	1.00	1.00	20	25
High degr. outlyingness	.78	.98	.78	.99	93	172
Medium degr. outlyingness	.05	.90	.06	.96	230	793
Low degr. outlyingness	.00	.76	.01	.90	287	1,649
Very low degr. outlyingness	.00	.29	.01	.54	557	4,594
Two clusters	.32	.76	.32	.90	182	5,339
Three clusters	.41	.81	.42	.86	1,005	1,894
20 % error	.36	.89	.36	.95	178	2,080
40 % error	.38	.69	.39	.81	1,009	5,153
Overall	.37	.79	.37	.88	1,187	7,233

high and medium degree-of-outlyingness conditions when using this slightly higher critical congruence value, but it remained substandard for the medium, low, and very low degrees of outlyingness. Applying an even higher value to improve the results for the lowest degrees of outlyingness would

Table 4 Proportions of correct data sets, proportions of data sets without false negatives, and numbers of false positives by the cutoff congruence method when a higher critical congruence value of .96 is applied

	Correct Data Sets	No False Negatives	False Positives
Nine nonoutlying	.45	.46	985
Twelve nonoutlying	.41	.42	634
Two outlying	.45	.46	566
Four outlying	.44	.44	409
Six outlying	.41	.42	644
Very high degr. outlyingness	1.00	1.00	20
High degr. outlyingness	.96	.96	93
Medium degr. outlyingness	.18	.19	299
Low degr. outlyingness	.02	.03	438
Very low degr. outlyingness	.00	.01	769
Two clusters	.39	.39	238
Three clusters	.47	.49	1,381
20 % error	.41	.41	247
40 % error	.45	.47	1,372
Overall	.43	.44	1,619

be too strict—leading to an excessive number of false positives—for some data sets, and especially for real data. It thus seems impossible to find a critical congruence value that would be ideal for all cases.

The selected critical congruence value is hardly an issue for the lower-bound congruence method, since it only uses the value as a lower bound in the CHull procedure. This method results in markedly higher performance. Specifically, comparing the results for the different degrees of outlyingness shows that the lower-bound congruence method broke down only for the very low degree of outlyingness, whereas the cutoff congruence method completely failed from the medium degree of outlyingness onward. The detection mistakes made were mainly false positives; specifically, for the 48,000 outlying variables that were present in the entire simulation, 7,233 false positives occurred, and 3,454 false negatives. False positives or negatives may occur either because of a faulty outlyingness ranking resulting from Steps 2 and 3 of the procedure, or because of a faulty number-of-outlying-variables selection in Step 4. The former type of mistake was encountered for 1,471 (12 %) of all simulated data sets (resulting in 4,101 out of the 7,233 false positives, as well as 2,922 out of the 3,454 false negatives); 1,382 of these 1,471 ranking mistakes occurred in the case of a very low degree of outlyingness and/or 40 % error. The latter type of mistake was found for 1,082 data sets (9 %), in which mostly (i.e., in 818 cases) too many outlying variables were detected, explaining the remaining 3,132 false positives. Note that this overselection is a documented characteristic of the CHull procedure (Wilderjans et al., 2013),

which can be mitigated by using the split-half procedure (see [Goodness of outlying-variable detection by means of the split-half procedure](#) section).

For the 10,526 data sets with a correct outlyingness ranking, we inspected the $\min(\varphi_{k_1 k_2}^{\min})$ values before and after the removal of the final outlying variable. The value before removal of the final outlying variable ranged from .91 to .99, with an overall mean of .9766 ($SD = .02$). Note that this value was larger than .95 in 9,632 of the 10,526 cases. The value after removal of the final outlying variable ranged from .9449 to .9996, with an overall mean of .9956 ($SD = .003$). This value was smaller than .95 for only one out of the 10,526 data sets. These results confirm that the guideline proposed by Lorenzo-Seva and ten Berge (2006) is unsuitable as a cutoff but works very well as a lower bound.

Goodness of outlying-variable detection by means of the split-half procedure

The to-be-preferred method according to the results above—that is, the lower-bound congruence method—did result in quite a number of false positives (i.e., 7,233; see Table 3). Since these may be caused by sampling fluctuations, it is certainly interesting to look into the performance of the lower-bound congruence method when the split-half procedure is also used. The results of the split-half lower-bound congruence method are given in Table 5. Comparing Tables 3 and 5, the most striking differences are (1) that the proportions of correct data sets are equal or higher for the medium to very high levels of outlyingness, but lower for the low and very low degrees of outlyingness (see the first column of Table 5)—which is due to a decrease in the false positives for all levels and a drop in the proportions of data sets without false negatives for the lowest degrees of outlyingness (see the second column of Table 5)—and (2) that the decrease of the number of false positives is spectacular (i.e., only 558 instead of 7,233 false positives; see the third column of Table 5). Thus, if one wants to be more conservative in the outlying-variable detection (i.e., avoiding false positives at the cost of more false negatives) or wants to obtain more robust results with respect to sampling fluctuations, the split-half procedure is definitely recommended.

To inspect the stability of the detection results over the 20 random splits, we checked for each data set in how many splits the resulting set of outlying variables was the correct one (without false negatives or false positives). The frequency of the correct set of outlying variables depended mostly on the degree of outlyingness: On average, the correct set was found for 19, 19, 16, 12, and 2 out of the 20 random splits for the very high, high, medium, low, and very low degrees of outlyingness, respectively.

Goodness of outlying-variable detection in the case of $K + 1$ clusters

Again we focused on the lower-bound congruence method, because this is clearly the best according to the results in [Goodness of outlying-variable detection](#) section. When this method was applied using one additional cluster, it still performed perfectly for 42 % of all simulated data sets, whereas for 60 % at least the outlyingness ranking was correct. Not surprisingly, these percentages were much higher when the error level was lower—59 % correct detection and 75 % correct outlyingness rankings when only 20 % error was present in the data—or when the degree of outlyingness was high—65 % correct detection and 87 % correct outlyingness rankings—or very high—78 % correct detection and 94 % correct outlyingness rankings. Overall, 13,875 false positives and 10,106 false negatives occurred. The error in the data seems to be an important causal factor behind the false positives, since 9,278 of them occurred in the conditions with 40 % error variance. With respect to the false negatives, the degree of outlyingness is again the most important factor, with 8,129 out of the 10,106 false negatives occurring for the low and very low degrees of outlyingness.

When applying the split-half procedure with one cluster too many, the proportion of entirely correct detections dropped further, to .37, with proportions of .67, .63, .38, .15, and .00 for the respective degrees of outlyingness. More specifically, the number of false positives decreased

Table 5 Proportions of correct data sets (i.e., data sets without false negatives or false positives), proportions of data sets without false negatives, and numbers of false positives for the lower-bound congruence method when the split-half procedure is used, and for each level of the manipulated factors of the simulation study

	Correct Data Sets	No False Negatives	Number of False Positives
Nine nonoutlying	.75	.75	404
Twelve nonoutlying	.78	.78	154
Two outlying	.78	.78	147
Four outlying	.78	.78	112
Six outlying	.74	.74	299
Very high degr. outlyingness	1.00	1.00	3
High degr. outlyingness	.99	.99	21
Medium degr. outlyingness	.94	.94	101
Low degr. outlyingness	.72	.72	140
Very low degr. outlyingness	.18	.18	293
Two clusters	.77	.78	318
Three clusters	.76	.76	240
20 % error	.85	.85	201
40 % error	.68	.68	357
Overall	.76	.77	558

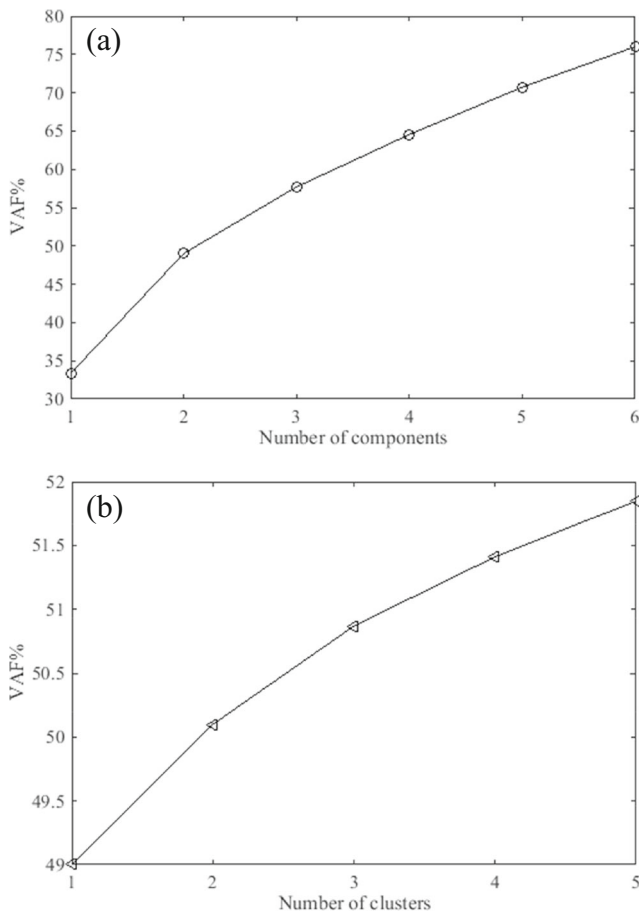


Fig. 2 Percentages of variance accounted for (VAF%) (a) by SCA-P solutions with the number of components varying from one to six, and (b) by clusterwise SCA-ECP solutions, with the number of clusters varying from one to five, for the values data

from 13,875 to 3,978, while the number of false negatives increased from 10,106 to 26,653.

Conclusion

On the basis of these simulation results, we advise researchers to use the lower-bound congruence method, rather than the cutoff congruence method, since the lower-bound method displayed a clearly superior performance. Because the lower-bound congruence method led to a fairly large number of false positives, we also advise using the split-half procedure whenever it is desirable to keep this number as low as possible.

Choosing the appropriate number of clusters may be hard, since increases in fit with additional clusters may be very small when only few outlying variables are present. The results in [Goodness of outlying-variable detection in the case of \$K + 1\$ clusters](#) section, indicate that this choice is indeed crucial for the performance of the outlying-variable detection. This conclusion should be put in perspective however, since (1) the false negatives largely pertain to loading differences that are so subtle that we would not be interested in them in the case of empirical data (because they would probably be error-driven), and (2) the outlyingness ranking remains correct, and thus informative, for the majority of the cases.

Application

In this section, we apply outlying-variable detection to cross-cultural data on values from the International College Survey (ICS) 2001 (Diener et al., 2001; Kuppens, Ceulemans, Timmerman, Diener & Kim-Prieto, 2006). The ICS study included 10,018 participants from 48 different countries. Each of them rated, among other things, how much they valued 11 aspects, listed in [Table 7](#) below, using a 9-point Likert scale (1 = *do not value it at all*, 9 = *value it extremely*). Of these participants, 330 with missing data were excluded. Between-country differences in means were removed by centering the aspects per country, and between-aspect differences in variability were eliminated by standardizing each aspect across countries. Consequently, only between-country differences in covariance structures were retained.

Regarding model selection, we first assessed the most appropriate number of components by performing SCA-P analyses with one to six components and comparing the resulting solutions in terms of complexity–fit balance. On the basis of the scree plot in [Fig. 2a](#) and the clear elbow therein, we retained two components. To determine the optimal number of clusters, we performed clusterwise SCA-P analyses with two components per cluster and one to five clusters. Since [Fig. 2b](#) does not display a clear elbow, we inspected the interpretability of the solutions with two and three clusters and retained the one with two clusters.

Next, we scrutinized the selected clusterwise SCA-P model. The partition of the clusterwise SCA-P model with two clusters and two components per cluster is given in [Table 6](#). As is discussed by De Roover et al. (2014a)—who present a

Table 6 Clustering of the clusterwise SCA-P model with two clusters and two components per cluster for the values data

Cluster 1	Cluster 2
Bangladesh, Cameroon, Chile, Croatia, Egypt, Georgia, Ghana, India, Indonesia, Iran, Kuwait, Malaysia, Nigeria, Philippines, Poland, South Africa, Thailand, Turkey, Uganda, Zimbabwe	Australia, Austria, Belgium, Brazil, Bulgaria, Canada, China, Colombia, Cyprus, Germany, Greece, Hong Kong, Hungary, Italy, Japan, Mexico, Nepal, Netherlands, Portugal, Russia, Singapore, Slovakia, Slovenia, South Korea, Spain, Switzerland, United States, Venezuela

Table 7 Cluster-specific component loadings of the clusterwise SCA-P model with two clusters and two components per cluster for the values data, obliquely Procrustes-rotated toward the SCA-P loadings, which are also included in the table

	Cluster 1		Cluster 2		SCA-P	
	Showing Success	Fun, Happiness, and Achievement	Showing Success and Benevolence	Fun, Happiness, and Achievement	Showing Success and Benevolence	Fun, Happiness, and Achievement
Material wealth	.74	.03	.71	.13	.72	.07
Physical attractiveness	.80	.05	.76	.12	.77	.09
Physical comforts	.75	.09	.74	.10	.75	.08
Excitement/arousal	.80	.04	.60	.09	.66	.09
Competition	.71	.01	.73	.11	.72	.05
Intelligence/ knowledge	-.10	.85	-.03	.68	-.08	.75
Happiness	.28	.73	.30	.42	.30	.53
<i>Success</i>	<i>.31</i>	<i>.59</i>	<i>.35</i>	<i>.74</i>	<i>.30</i>	<i>.66</i>
<i>Fun</i>	<i>.14</i>	<i>.48</i>	<i>-.21</i>	<i>.77</i>	<i>-.15</i>	<i>.72</i>
<i>Heaven/afterlife</i>	<i>.28</i>	<i>.21</i>	<i>.74</i>	<i>-.16</i>	<i>.60</i>	<i>-.05</i>
<i>Self-sacrifice</i>	<i>.35</i>	<i>.24</i>	<i>.63</i>	<i>.00</i>	<i>.54</i>	<i>.08</i>

The loadings with an absolute value larger than .40 are printed in boldface, and the outlying variables according to the lower-bound congruence method are printed in italics.

largely similar clustering—Cluster 1 contains preindustrial countries that are more traditional and more focused on the basic values necessary for survival, whereas the other countries are gathered in Cluster 2.

To find out which differences led to this clustering, we turn to the cluster-specific loading matrices in Table 7. Those cluster-specific loadings were obliquely Procrustes-rotated toward the normalized VARIMAX-rotated SCA-P loadings (also presented in Table 7). According to the strong SCA-P loadings, the first component captures the covariance among “material wealth,” “physical attractiveness,” “physical comforts,” “excitement/arousal,” “competition,” “heaven/afterlife,” and “self-sacrifice”; we therefore label it “showing success and benevolence.” The second SCA-P component captures “happiness,” “intelligence/knowledge,” “success,” and “fun”; it is thus labeled “fun, happiness, and achievement.” The cluster-specific loading structures largely resemble the SCA-P structure, and thus could be interpreted similarly. Some interesting between-cluster differences were found, however. For example, “heaven/afterlife” and “self-sacrifice” have a positive cross-loading on the second component for Cluster 1, whereas in Cluster 2 they have a negative and a very low cross-loading component, respectively. Thus, when inhabitants from the countries in Cluster 1 value “fun, happiness, and achievement,” they also value “heaven/afterlife” and “self-sacrifice” to some extent, whereas for Cluster 2 this is not the case. Also in Cluster 1, the loadings of “heaven/afterlife” and “self-sacrifice” on the first component are lower—therefore, the first component is merely labeled “showing success” in this cluster.

Finally, we performed the outlying-variable detection. On the basis of the simulation results, we applied the lower-bound

congruence method. The resulting outlyingness ranking matrix is given in Table 8, and the CHull plot in Fig. 3. Due to the saturation at the higher end of the convex hull plot, the automated CHull procedure suggests the presence of eight (out of 11) outlying variables. Upon visual inspection of Fig. 3 (and relying on the second-highest scree ratio given by CHull), we suspect that four outlying variables (i.e., “heaven/afterlife,” “self-sacrifice,” “success,” and “fun”; see Table 8) are present in the data, and that the other four are false positives.

To obtain more robust results and correct for the oversensitivity of CHull—thus, hopefully eliminating the supposed false positives—we performed the split-half procedure described in Split-half procedure section for the lower-bound congruence method. The 20 random splits resulted in seven different sets of outlying variables (see Table 9), with, as

Table 8 Outlyingness ranking matrix, as calculated by the lower-bound congruence method for the values data

$\min(\varphi_{k_1 k_2}^{\text{mean}})$	k_1	k_2	Most Outlying Variable	$\min(\varphi_{k_1 k_2}^{\text{min}})$
<i>.91</i>	<i>1</i>	<i>2</i>	<i>Heaven/afterlife</i>	<i>.88</i>
<i>.94</i>	<i>1</i>	<i>2</i>	<i>Fun</i>	<i>.92</i>
<i>.97</i>	<i>1</i>	<i>2</i>	<i>Success</i>	<i>.95</i>
<i>.97</i>	<i>1</i>	<i>2</i>	<i>Self-sacrifice</i>	<i>.96</i>
.99	1	2	Excitement/arousal	.99
1.00	1	2	Physical comforts	.99
1.00	1	2	Happiness	1.00
1.00	1	2	Competition	1.00
1.00	1	2	Material wealth	1.00
1.00	1	2	Intelligence/knowledge	1.00

The selected items are shown in italics.

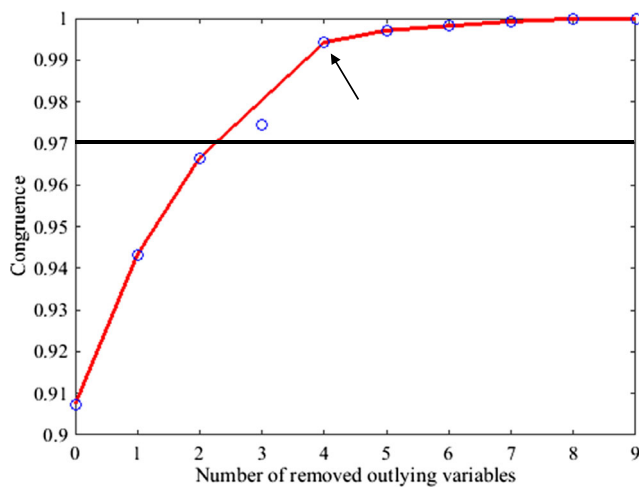


Fig. 3 CHull plot of the lower-bound congruence method for the values data. Specifically, the $\min(\varphi_{k_1 k_2}^{mean})$, labeled “Congruence,” is plotted against the number of variables already removed (the order wherein the variables are removed can be found in Table 8). The black horizontal line indicates where the $\min(\varphi_{k_1 k_2}^{min})$ value (not depicted in the figure, but in Table 8) crosses the lower bound of .95. The arrow indicates the elbow after which the decrease in congruence levels off

suspected, the above-mentioned set of four variables being found most frequently (i.e., nine times). Moreover, “heaven/afterlife,” “success,” and “fun” also occur in each of the other sets of outlying variables, and are thus always detected as outlying. “Self-sacrifice” is detected in no less than four out of the six other sets (thus, in a total of 18 of the 20 random splits).

Discussion

Researchers are often interested in differences in covariance structures across different groups. Clusterwise SCA-P explores such differences in an efficient manner. In many cases, the differences will pertain to a few variables only, which we have called “outlying variables.” Detecting such outlying variables is important for two reasons: First, it can reveal

structural differences that can help sharpen substantive theories on, for instance, cross-cultural differences. Second, in psychometrics, one often aims to find a set of variables that has a common structure across groups, since this is a prerequisite for comparing the scores of the subjects on the latent variables that summarize this common structure. This article has presented and evaluated two heuristics to detect such outlying variables, which can be applied with or without a split-half procedure. On the basis of a simulation study, we recommend to use the lower-bound congruence method, with the split-half procedure whenever the risk of false positives should be minimized.

One might argue that the outlying variable detection should be based on the individual group-specific loading matrices, rather than on the cluster-specific loading matrices resulting from a clusterwise SCA-P analysis, in order to conserve all of the information in the data. This alternative heuristic could be implemented straightforwardly. It does imply two problems, however. First, the huge number of pairwise comparisons will lead to more false positives. Second, including all of the idiosyncratic (and possibly error-driven) variations in the group-specific loading structures will also result in more false positives or in finding differences that are of less interest (e.g., differences that only occur in one of the many pairwise comparisons). Using clusterwise SCA-P has the advantage of focusing on the most important structural differences only.

The bootstrap method proposed by Chan and colleagues (1999) is a relevant method to consider with respect to the outlying-variables problem. Specifically, they proposed a resampling method to test whether a set of factor loadings is significantly different between a target and a replication group. The method can be applied per factor (i.e., column-wise in a loading matrix) to test whether or not it is different, but it can also be applied per variable (i.e., row-wise in a loading matrix) to detect which variables have different loadings in the two groups, and thus can be considered outlying. However, applying the Chan bootstrap method is not straightforward in our case, because the method is not directly suitable for comparing the loadings of more than two groups (or

Table 9 Results of the split-half procedure using 20 splits for the lower-bound congruence method for the values data

Frequency	Outlying Variables
9	Heaven/afterlife, self-sacrifice, success, & fun
5	Physical comforts, heaven/afterlife, self-sacrifice, success, & fun
2	Excitement/arousal, physical comforts, heaven/afterlife, self-sacrifice, success, & fun
1	Heaven/afterlife, success, & fun
1	Happiness, heaven/afterlife, success, & fun
1	Happiness, physical comforts, heaven/afterlife, self-sacrifice, success, & fun
1	Happiness, excitement/arousal, physical comforts, heaven/afterlife, self-sacrifice, success, & fun

The selected set of outlying variables is shown in boldface.

clusters of groups) simultaneously, and recurring to pairwise comparisons would lead to the problems listed in the previous paragraph. Moreover, the Chan procedure does not sequentially remove items and test again. As we argue in the present article, often some sort of iterative procedure is needed to identify the nonoutlying variables, because the initial loadings (i.e., of the full data set) can be severely distorted by the outlying ones. Finally, the Chan bootstrap approach is not yet adapted to comply with the assumptions of component analysis models (i.e., with respect to the rank of the residuals).

The present article has focused mainly on exploratory analyses, in which one has no a priori idea about the common covariance structure, but the presented heuristics may be helpful within the confirmatory context, and in the measurement invariance testing framework as well. Specifically, when configural and/or weak measurement invariance (Meredith, 1993) cannot be confirmed, one can apply the heuristics presented in this article to check for the presence of outlying variables (De Roover et al., 2014a). To this end, the a-priori-assumed latent variable structure can be used as a target structure when applying the detection methods, instead of the SCA-P loadings.

The CFA framework also offers some methods to trace which variables are causing measurement invariance tests to fail, such as the sequential model modification procedure (MacCallum, 1986; MacCallum, Roznowski & Necowitz, 1992) and item-level invariance testing (Cheung & Rensvold, 1999). These methods have some disadvantages, however, in that they require researchers to run a multitude of time-consuming analyses, and they imply assumptions that are often questionable (see De Roover et al., 2014a). Furthermore, applying them to many groups is not straightforward, since many of the typically used fit measures are unsuitable or need adjustment (Rutkowski & Svetina, 2014).

Finally, an advantage of the outlying-variable detection heuristics is that they are not limited to the clusterwise SCA-P case, but can also be used to compare any set of component or factor loading matrices for the same variables. As examples, one may think of the loading matrices that result from fitting mixtures of factor analyzers (McLachlan & Peel, 2000; Yung, 1997), a subspace k -means analysis (Timmerman, Ceulemans, De Roover & Van Leeuwen, 2013), or a switching principal component analysis (De Roover, Timmerman, Van Diest, Onghena & Ceulemans, 2014b).

Author note K.D.R. is a postdoctoral fellow of the Fund for Scientific Research Flanders (Belgium). The research leading to the results reported in this article was sponsored in part by the Belgian Federal Science Policy within the framework of the Interuniversity Attraction Poles program (IAP/P7/06), as well as by the Research Council of KU Leuven (Grant No. GOA/15/003).

References

- Borsboom, D., Mellenbergh, G. J., & van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, *110*, 203–219. doi:10.1037/0033-295X.110.2.203
- Bro, R., & Smilde, A. K. (2003). Centering and scaling in component analysis. *Journal of Chemometrics*, *17*, 16–33.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*, 245–276. doi:10.1207/s15327906mbr0102_10
- Ceulemans, E., Hubert, M., & Rousseeuw, P. (2013). Robust multilevel simultaneous component analysis. *Chemometrics and Intelligent Laboratory Systems*, *129*, 33–39.
- 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*, 133–150. doi:10.1348/000711005X64817
- Chan, W., Ho, R. M., Leung, K., Chan, D. K.-S., & Yung, Y.-F. (1999). An alternative method for evaluating congruence coefficients with Procrustes rotation: A bootstrap procedure. *Psychological Methods*, *4*, 378–402.
- Cheung, G. W., & Rensvold, R. B. (1999). Testing factorial invariance across groups: A reconceptualization and proposed new method. *Journal of Management*, *25*, 1–27.
- De Roover, K., Ceulemans, E., Timmerman, M. E., Nezlek, J. B., & Onghena, P. (2013a). Modeling differences in the dimensionality of multiblock data by means of clusterwise simultaneous component analysis. *Psychometrika*, *78*, 648–668. doi:10.1007/s11336-013-9318-4
- De Roover, K., Ceulemans, E., Timmerman, M. E., Vansteelandt, K., Stouten, J., & Onghena, P. (2012a). Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. *Psychological Methods*, *17*, 100–119. doi:10.1037/a0025385
- De Roover, K., Ceulemans, E., Timmerman, M. E., & Onghena, P. (2013b). A clusterwise simultaneous component method for capturing within-cluster differences in component variances and correlations. *British Journal of Mathematical and Statistical Psychology*, *86*, 81–102.
- De Roover, K., Ceulemans, E., & Timmerman, M. E. (2012b). How to perform multiblock component analysis in practice. *Behavior Research Methods*, *44*, 41–56. doi:10.3758/s13428-011-0129-1
- De Roover, K., Timmerman, M. E., De Leersnyder, J., Mesquita, B., & Ceulemans, E. (2014a). What's hampering measurement invariance: Detecting non-invariant items using clusterwise simultaneous component analysis. *Frontiers in Psychology*, *5*(604), 1–11. doi:10.3389/fpsyg.2014.00604
- De Roover, K., Timmerman, M. E., Van Diest, I., Onghena, P., & Ceulemans, E. (2014b). Switching principal component analysis for modeling means and covariance changes over time. *Psychological Methods*, *19*, 113–132.
- De Roover, K., Timmerman, M. E., Van Mechelen, I., & Ceulemans, E. (2013c). On the added value of multisets methods for three-way data analysis. *Chemometrics and Intelligent Laboratory Systems*, *129*, 98–107.
- Diener, E., Kim-Prieto, C., Scollon, C., & Colleagues. (2001). [International College Survey 2001]. Unpublished raw data.
- Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling*, *16*, 295–314. doi:10.1080/10705510902751416
- Dudzinski, M. L., Norris, J. M., Chmura, J. T., & Edwards, C. B. H. (1975). Repeatability of principal components in samples: Normal

- and non-normal data sets compared. *Multivariate Behavioral Research*, 10, 109–117.
- Harshman, R. A., & Lundy, M. E. (1984). Data preprocessing and the extended PARAFAC model. In H. C. Law, C. W. Snyder Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 122–215). New York: Praeger.
- Hessen, D. J., Dolan, C. V., & Wicherts, J. M. (2006). Multi-group exploratory factor analysis and the power to detect uniform bias. *Applied Psychological Research*, 30, 233–246.
- Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2, 193–218.
- Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal components analysis. *Technometrics*, 47, 64–79.
- Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). New York: Springer.
- Jöreskog, K. G. (1971). Simultaneous factor analysis in several populations. *Psychometrika*, 36, 409–426.
- Kaiser, H. F. (1958). The Varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23, 187–200. doi:10.1007/BF02289233
- Kiers, H. A. L., & ten Berge, J. M. F. (1994). Hierarchical relations between methods for simultaneous components analysis and a technique for rotation to a simple simultaneous structure. *British Journal of Mathematical and Statistical Psychology*, 47, 109–126.
- Kline, R. B. (2004). *Principles and practice of structural equation modeling* (2nd ed.). New York: Guilford Press.
- Krysinska, K., De Roover, K., Bouwens, J., Ceulemans, E., Corveleyn, J., Dezutter, J., ... Pollefeyt, D. (2014). Measuring religious attitudes in secularised Western European context: A psychometric analysis of the Post-Critical Belief Scale. *International Journal for the Psychology of Religion*, 24, 263–281.
- Kuppens, P., Ceulemans, E., Timmerman, M. E., Diener, E., & Kim-Prieto, C. (2006). Universal intracultural and intercultural dimensions of the recalled frequency of emotional experience. *Journal of Cross-Cultural Psychology*, 37, 491–515.
- Lorenzo-Seva, U., Kiers, H. A. L., & ten Berge, J. M. F. (2002). Techniques for oblique factor rotation of two or more loading matrices to a mixture of simple structure and optimal agreement. *British Journal of Mathematical and Statistical Psychology*, 55, 337–360.
- Lorenzo-Seva, U., & ten Berge, J. M. F. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, 2, 57–64. doi:10.1027/1614-2241.2.2.57
- MacCallum, R. (1986). Specification searches in covariance structure modeling. *Psychological Bulletin*, 100, 107–120. doi:10.1037/0033-2909.100.1.107
- MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, 111, 490–504. doi:10.1037/0033-2909.111.3.490
- McLachlan, G. J., & Peel, D. (2000). *Finite mixture models*. New York: Wiley.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543. doi:10.1007/BF02294825
- Paunonen, S. V. (1997). On chance and factor congruence following orthogonal Procrustes rotation. *Educational and Psychological Measurement*, 57, 33–59.
- Rutkowski, L., & Svetina, D. (2014). Assessing the hypothesis of measurement invariance in the context of large-scale international surveys. *Educational and Psychological Measurement*, 74, 31–57.
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27, 229–239.
- Timmerman, M. E., Ceulemans, E., De Roover, K., & Van Leeuwen, K. (2013). Subspace *K*-means clustering. *Behavior Research Methods*, 45, 1011–1023. doi:10.3758/s13428-013-0329-y
- Timmerman, M. E., Hoefsloot, H. C. J., Smilde, A. K., & Ceulemans, E. (2015). Scaling in ANOVA-simultaneous component analysis. *Metabolomics*, 11, 1265–1276. doi:10.1007/s11306-015-0785-8
- Timmerman, M. E., & Kiers, H. A. L. (2003). Four simultaneous component models of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika*, 68, 105–122. doi:10.1007/BF02296656
- Tucker, L. R. (1951). *A method for synthesis of factor analysis studies (Personnel Research Section Rep. No. 984)*. Washington: Department of the Army.
- Wilderjans, T. F., Ceulemans, E., & Meers, K. (2013). CHull: A generic convex-hull-based model selection method. *Behavior Research Methods*, 45, 1–15. doi:10.3758/s13428-012-0238-5
- Yung, Y. F. (1997). Finite mixtures in confirmatory factor-analysis models. *Psychometrika*, 62, 297–330.