Psychological Methods
2007, Vol. 12, No. 3, 336–358

1082-989X/07/$12.00 DOI: 10.1037/1082-989X.12.3.336

Nonlinear Principal Components Analysis: Introduction and Application

Mariëlle Linting and Jacqueline J. Meulman
Leiden University

Patrick J. F. Groenen
Erasmus University Rotterdam

Anita J. van der Kooij
Leiden University

The authors provide a didactic treatment of nonlinear (categorical) principal components analysis (PCA). This method is the nonlinear equivalent of standard PCA and reduces the observed variables to a number of uncorrelated principal components. The most important advantages of nonlinear over linear PCA are that it incorporates nominal and ordinal variables and that it can handle and discover nonlinear relationships between variables. Also, nonlinear PCA can deal with variables at their appropriate measurement level; for example, it can treat Likert-type scales ordinally instead of numerically. Every observed value of a variable can be referred to as a category. While performing PCA, nonlinear PCA converts every category to a numeric value, in accordance with the variable's analysis level, using optimal quantification. The authors discuss how optimal quantification is carried out, what analysis levels are, which decisions have to be made when applying nonlinear PCA, and how the results can be interpreted. The strengths and limitations of the method are discussed. An example applying nonlinear PCA to empirical data using the program CATPCA (J. J. Meulman, W. J. Heiser, & SPSS, 2004) is provided.

Keywords: nonlinear principal components analysis, PCA, CATPCA, categorical variables

In the social and behavioral sciences, researchers are often confronted with a large number of variables that they wish to reduce to a small number of composites with as little loss of information as possible. Traditionally, principal components analysis (PCA) is considered to be an appropriate way to perform such data reduction (Fabrigar, Wegener, MacCallum, & Strahan, 1999). This widely used method reduces a large number of variables to a much smaller number of uncorrelated linear combinations of these variables, called principal components, that represent the observed data as closely as possible. However, PCA suffers from two important limitations. First, it assumes that the relationships between variables are linear, and second, its interpretation is only sensible if all of the variables are assumed to be scaled at the numeric level (interval or ratio scale of measurement). In the social and behavioral sciences, these assumptions are frequently not justified, and therefore, PCA may not always be the most appropriate method of analysis. To circumvent these limitations, an alternative, referred to as nonlinear PCA, has been developed. A first version of this method was described by Guttman (1941), and other major contributions to the literature on this subject are from Kruskal (1965); Shepard (1966); Kruskal and Shepard (1974); Young, Takane, and de Leeuw (1978); and Winsberg and Ramsay (1983). For a historical overview, see Gifi (1990). This alternative method

---

Mariëlle Linting, Jacqueline J. Meulman, and Anita J. van der Kooij, Data Theory Group, Leiden University, Leiden, The Netherlands; Patrick J. F. Groenen, Econometric Institute, Erasmus University Rotterdam, Rotterdam, The Netherlands.

The data used in the application were collected by the National Institute of Child Health and Human Development (NICHD) Early Child Care Research Network, supported by the NICHD through a cooperative agreement that calls for scientific collaboration between the grantees and NICHD staff. We acknowledge the generous way in which the NICHD Study on Early Child Care has made this unique data set available for secondary analyses.

Special thanks are due to Louise Fitzgerald, Department of Psychology, University of Illinois at Urbana–Champaign. The combination of her technical and verbal skills was invaluable for the final version of this article.

Leiden University holds the copyright of the procedures in the SPSS Categories package, and the Department of Data Theory receives the royalties.

Correspondence concerning this article should be addressed to Mariëlle Linting, Data Theory Group, Leiden University, P.O. Box 9555, 2300 RB Leiden, The Netherlands. E-mail: linting@fsw.leidenuniv.nl

has the same objectives as traditional PCA but is suitable for variables of mixed measurement levels (nominal, ordinal, and numeric) that may not be linearly related to each other. In the type of nonlinear PCA that is described in the present article, all variables are viewed as categorical, and every distinct value of a variable is referred to as a category. Accordingly, the method is also referred to as categorical PCA.

The objective of this article is to provide a didactic introduction to the method of nonlinear PCA. In the first section, we discuss how nonlinear PCA achieves the goals of linear PCA for variables of mixed scaling levels by converting category numbers into numeric values. We then describe the different analysis levels that can be specified in nonlinear PCA, from which some practical guidelines for choosing analysis levels can be deduced, and also discuss the similarities between nonlinear and linear PCA. The second section starts with a discussion of software that may be used to perform nonlinear PCA and then provides an application of nonlinear PCA to an empirical data set (NICHD Early Child Care Research Network, 1996) that incorporates variables of different measurement levels and nonlinear relationships between variables. The nonlinear PCA solution is compared with the linear PCA solution on these same data. In the final section, we summarize the most important aspects of nonlinear PCA, focusing on its strengths and limitations as an exploratory data analysis method.

The Method of Nonlinear PCA

The objective of linear PCA is to reduce a number of *m* continuous numeric variables to a smaller number of *p* uncorrelated underlying variables, called principal components, that reproduce as much variance from the variables as possible. Because variance is a concept that applies only to continuous numeric variables, linear PCA is not suitable for the analysis of variables with ordered or unordered (discrete) categories. In nonlinear PCA, categories of such variables are assigned numeric values through a process called optimal quantification (also referred to as optimal scaling or optimal scoring). Such numeric values are referred to as category quantifications; the category quantifications for one variable together form that variable's transformation. Optimal quantification replaces the category labels with category quantifications in such a way that as much as possible of the variance in the quantified variables is accounted for. Just as continuous numeric variables, such quantified variables possess variance in the traditional sense. Then, nonlinear PCA achieves the very same objective as linear PCA for quantified categorical variables. If all variables in nonlinear PCA are numeric, the nonlinear and linear PCA solutions are exactly equal because, in that case, no optimal quantification is required, and the variables are merely standardized.

In nonlinear PCA, the optimal quantification task and the linear PCA model estimation are performed simultaneously, which is achieved by the minimization of a least squares loss function. In the actual nonlinear PCA analysis, model estimation and optimal quantification are alternated through use of an iterative algorithm that converges to a stationary point where the optimal quantifications of the categories do not change anymore. If all variables are treated numerically, this iterative process leads to the same solution as linear PCA. For more details on the mathematics of nonlinear PCA, we refer the reader to Gifi (1990); Meulman, van der Kooij, and Heiser (2004); and Appendix A of this article.

Category Quantification

In this section, we first define the concept of categorical variables. Then, we discuss the process of category quantification in more detail, considering the different types of analysis levels that may be specified in nonlinear PCA, and conclude by showing how this method can be used to discover and handle nonlinear relationships between variables.

Categorical variables. Nonlinear PCA aims at analyzing so-called categorical variables. Often, the term *categorical* is used to refer to nominal variables that consist of unordered categories. A familiar example is religion, with possible categories being Protestant, Catholic, Jewish, Muslim, Buddhist, none, and other. Obviously, when variables consist of unordered categories, it makes no sense to compute sums or averages. As principal components are weighted sums of the original variables, nominal variables cannot be analyzed by standard PCA.

Ordinal variables are also referred to as categorical. Such variables consist of ordered categories, such as the values on a rating scale, for example, a Likert-type scale. Despite superficial appearance, such scale values are not truly numeric because intervals between consecutive categories cannot be assumed to be equal. For instance, one cannot assume that the distance on a 7-point scale between *fully agree* (7) and *strongly agree* (6) is equal to the distance between *neutral* (4) and *somewhat agree* (5). On such a scale, it is even less likely that *fully agree* (7) is 3.5 times as much as *strongly disagree* (2), and it is not clear where categories such as *no opinion* and *don't know* should be placed. In summary, although ordinal variables display more structure than nominal variables, it still makes little sense to regard ordinal scales as possessing traditional numeric qualities.

Finally, even true numeric variables can be viewed as categorical variables with *c* categories, where *c* indicates the number of different observed values. Both ratio and interval variables are considered numeric in nonlinear PCA. Reaction time is a prime example of a ratio variable, familiar to most in the social and behavioral sciences: If experimental participants respond to a stimulus in either 2.0, 3.0, 3.8, 4.0, or 4.2 seconds, the resulting variable has five different categories. The distance between 2.0 and 3.0 is equal to the distance between 3.0 and 4.0, and those who react in 2.0 seconds react twice as fast as the individuals with a 4.0-second reaction time. Within nonlinear PCA, no distinction is made between the interval and ratio levels of measurement; both levels of measurement are treated as numeric (metric) variables.

Given that the data contain only variables measured on a numeric level, linear PCA is obviously an appropriate analysis method. Even among such true numeric variables, however, nonlinear relationships may exist. For example, one might wish to examine the relationship between age and income, both of which can be measured on numeric scales. The relationship between age and income may be nonlinear, as both young and elderly persons tend to have smaller incomes than those between 30 and 60 years old. If we were to graph income on the vertical axis versus age on the horizontal axis, we would see a function that is certainly not linear or even monotonic (where values of income increase with values of age) but rather an inverted-U shape (∩), which is distinctly nonlinear. Nonlinear PCA can assign values to the categories of such numeric variables that will maximize the association (Pearson correlation) between the quantified variables, as we discuss below. Thus, nonlinear PCA can deal with all types of variables—nominal, ordinal, and (possibly nonlinearly related) numeric—simultaneously.

The objective of optimal quantification. So far, we have seen that nonlinear PCA converts categories into numeric values because variance can be established only for numeric values. Similarly, quantification is required because Pearson correlations are used in the linear PCA solution. For instance, in linear PCA, the overall summary diagnostic is the proportion of variance accounted for (VAF) by the principal components, which equals the sum of the eigenvalues of the principal components divided by the total number of variables. Although it could be argued that Pearson correlations may be computed between ordinal variables (comparable to Spearman rank correlations), it does not make sense to compute correlations between nominal variables. Therefore, in nonlinear PCA, correlations are not computed between the observed variables but between the quantified variables. Consequently, as opposed to the correlation matrix in linear PCA, the correlation matrix in nonlinear PCA is not fixed; rather, it is dependent on the type of quantification, called an analysis level, that is chosen for each of the variables.

In contrast to the linear PCA solution, the nonlinear PCA solution is not derived from the correlation matrix but iteratively computed from the data itself, using the optimal scaling process to quantify the variables according to their analysis level. The objective of optimal scaling is to optimize the properties of the correlation matrix of the quantified variables. Specifically, the method maximizes the first p eigenvalues of the correlation matrix of the quantified variables, where p indicates the number of components that are chosen in the analysis. This criterion is equivalent to the previous statement that the aim of optimal quantification is to maximize the VAF in the quantified variables.

Nominal, ordinal, and numeric analysis levels. Analyzing data by nonlinear PCA involves dynamic decision making, as decisions originally made by the researcher may need to be revised during the analysis process; indeed, trying out various analysis levels and comparing their results is part of the data-analytic task. It is important to note that the insight of the researcher—and not the measurement levels of the variable—determines the analysis level of a variable. To enable the researcher to choose an appropriate analysis level for each of the variables in the analysis, a description of the properties of each level is given below.

In general, it should be kept in mind that different analysis levels imply different requirements. In the case of a nominal analysis level, the only requirement is that persons who scored the same category on the original variable (self-evidently) should also obtain the same quantified value. This requirement is the weakest one in nonlinear PCA. In the case of an ordinal analysis level, the quantification of the categories should additionally respect the ordering of the original categories: A category quantification should always be less than or equal to the quantification for the category that has a higher rank number in the original data. When a nominal or ordinal analysis level is specified, a plot of the category quantifications versus the original category numbers (or labels) will display a nonlinear function, as shown in the so-called transformation plots in Figure 1. This ability to discover and handle nonlinear relations is the reason for using the term *nonlinear* for this type of analysis. A numeric analysis level requires quantified categories not only to be in the right order but also to maintain the original relative spacing of the categories in the optimal quantifications, which is achieved by standardizing the variable. If all variables are at a numeric analysis level, no optimal quantification is needed, and variables are simply standardized, in which case potential nonlinear relationships among variables are not accounted for. If one wishes to account for nonlinear relations between numeric variables, a nonnumeric analysis level should be chosen. In the following paragraphs, examples of different analysis levels are discussed.

To show the effect of using each analysis level, nonlinear PCA has been applied five times to an example data set. One of the variables (V1) was assigned a different analysis level in each of these analyses, whereas the other variables were treated numerically. Figures 1a, 1b, and 1c display the results of these analyses. (Figures 1d and 1e are discussed in the next subsection.) The horizontal axis (x) of the plots in Figure 1 displays the categories of V1, which range between

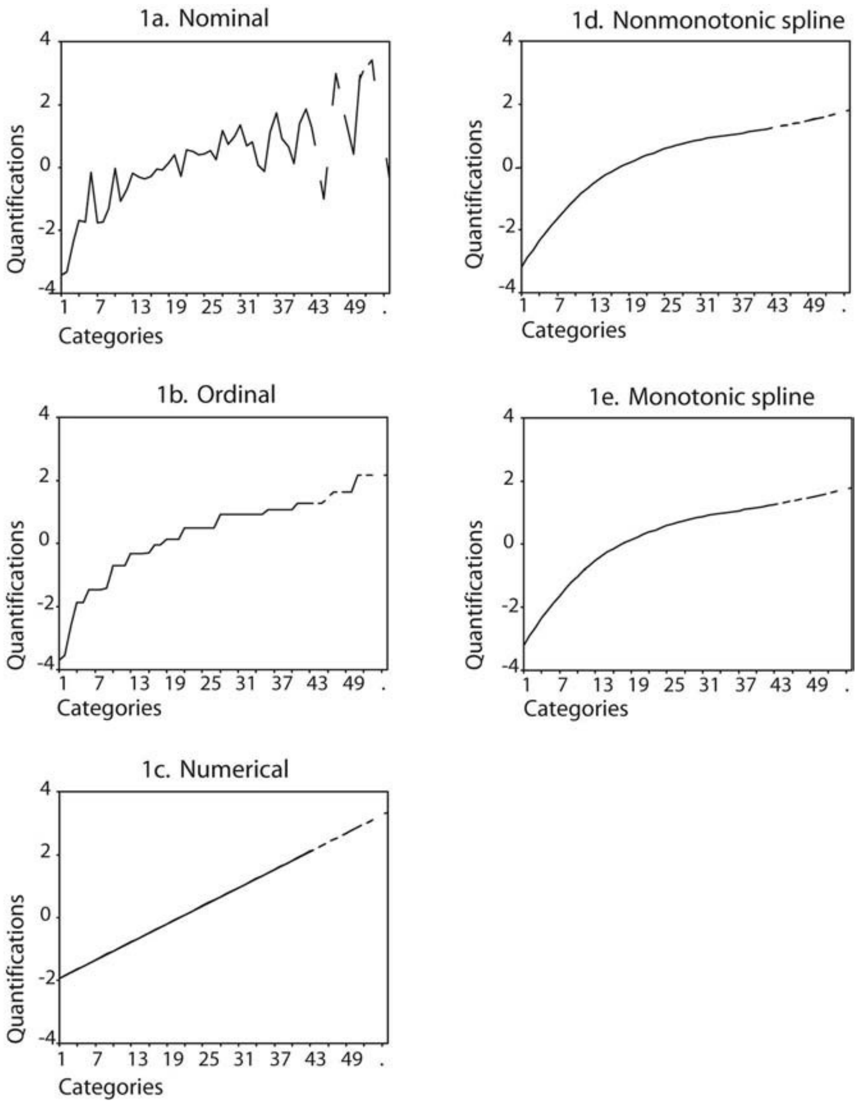

Figure 1. Transformation plots for different types of quantification. The same variable (V1) was assigned five different analysis levels, whereas the other variables were treated numerically. Observed category scores are on the *x*-axis, and the numeric values (standard scores) obtained after optimal quantification (category quantifications) are on the *y*-axis. The line connecting category quantifications indicates the variable's transformation. The gaps in the transformation indicate that some category values were not observed.

1 and 60; on the vertical axis (y), the category quantifications are shown. These quantifications are standard scores. The connecting line between the category quantifications indicates the variable's transformation. Because V1 has many categories, only every sixth category is displayed. A dot appearing as a label on the x-axis denotes that the corresponding value did not occur as a category in the data set. For example, instead of the value of 55, a dot appears on the x-axis because Category 55 did not occur in the data. Consequently, that value obtains no quantification, indicated by a gap in the transformation.

For a nominal analysis level (shown in Figure 1a), the optimal category quantifications may have any value as long as persons in the same category obtain the same score on the quantified variable. In this plot, we see that, although the overall trend of the nominal transformation is increasing, the quantifications are not in the exact same order as the original category labels. For example, between Categories 43 and 49, a considerable decrease in the quantifications occurs. In contrast to the order of the original category labels, the order of the nominal category quantifications is meaningful, reflecting the nature of the relationship of the variable to the principal components (and the other variables). If a nominal analysis level is specified and the quantifications are perfectly in the same order as the original categories, an ordinal analysis level would give exactly the same transformation.

It can be clearly seen in Figure 1b that the ordinal category quantifications are (nonstrictly) increasing with the original category labels (i.e., the transformation is monotonically nondecreasing). The original spacing between the categories is not necessarily maintained in the quantifications. In this example, some consecutive categories obtain the same quantification, also referred to as ties. For example, between Categories 25 and 37, we see a plateau of tied quantifications. Such ties may have two possible reasons. The first is that persons scoring in the tied categories do not structurally differ from each other considering their scores on the other variables and, therefore, the categories cannot be distinguished from each other. This can occur with ordinal, but also with nominal, quantifications. Ties can also occur because ordinal quantifications are obtained by placing an order restriction on nominal quantifications. If the nominal quantifications for a number of consecutive categories are in the wrong order, the ordinal restriction results in the same quantified value for these categories (the weighted mean of the nominal quantifications).

Finally, with a numeric (or linear) analysis level, the category quantifications are restricted to be linearly related to the original category labels, that is, the difference between the quantification of, for example, Categories 1 and 2 equals the difference between, for example, Categories 4 and 5. Then, the quantified values are simply standard scores of the original values and the transformation plot will show a straight line (see Figure 1c). This type of quantification is used when it is assumed that the relationship between a variable and the other variables is linear.

Nonlinear PCA has the most freedom in quantifying a variable when a nominal analysis level is specified and is the most restricted when a numeric analysis level is specified. Therefore, the method will obtain the highest VAF when all variables are analyzed nominally and the lowest VAF when all variables are analyzed numerically.

Smooth transformations. The nominal and ordinal analysis levels described above use step functions, which can be quite irregular. As an alternative, it is possible to use smooth functions—here, we use splines—to obtain a nonlinear transformation. A monotonic spline transformation is less restrictive than a linear transformation but more restrictive than an ordinal one, as it requires not only that the categories be in the same order but also that the transformation show a smooth curve. The simplest form of a spline is a function—usually, a second-degree polynomial (quadratic function) or third-degree polynomial (cubic function) of the original data—specified for the entire range of a variable. Because it is often impossible to describe the whole range of data with one such simple function, separate functions can be specified for various intervals within the range of a variable. Because these functions are polynomials, the smoothness of the function within each interval is guaranteed. The interval end points where two functions are joined together are called interior knots. The number of interior knots and the degree of the polynomials specify the shape of the spline and, therefore, the smoothness of the transformation. (Note that a first-degree spline with zero interior knots equals a linear transformation and that a first-degree spline with the number of interior knots equal to the number of categories minus two results in an ordinal transformation.)[1]

Nonmonotonic as well as monotonic splines can be used. Nonmonotonic splines yield smooth nonmonotonic transformations instead of the possibly very irregular transformations that result from applying a nominal analysis level. In Figure 1d, such a nonmonotonic spline transformation for V1 is displayed. A nonmonotonic spline analysis level is appropriate for variables with many categories that have either a nominal analysis level or an ordinal or numeric level combined with a nonmonotonic relationship with the other variables (and thus with a principal component).

In the example in Figure 1e, a monotonic spline transformation for V1 is shown using a second-degree spline with two interior knots, so that quadratic functions of the original values within three data intervals are obtained. The spline transformation resembles the ordinal transformation but follows a smooth curve instead of a step function. In general, it is advisable to use an ordinal analysis level when the number of categories is small and a monotonic spline analysis level when the number of categories is large compared with the number of persons. This monotonic spline transformation is identical to the nonmonotonic spline transformation in Figure 1d for this specific variable (V1) because the overall trend of the transformation is increasing and the number of knots (two) is small. For other variables and a larger number of knots, this equality will mostly not hold.

Analysis levels and nonlinear relationships between variables. To demonstrate the possible impact of choosing a particular analysis level, we assigned the same five analysis levels to another variable in the example data set (V2) that is nonmonotonically related to the other variables. The five panels in Figure 2 display transformation plots revealing that different analysis levels may lead to rather different transformations. The nominal transformation in Figure 2a and the nonmonotonic spline transformation in Figure 2d both show an increasing function followed by a decreasing function describing the relationship between the original category labels and the quantifications. The ordinal and

[1] For more details about the use of splines in nonlinear data analysis, we refer to Winsberg and Ramsay (1983) and Ramsay (1988).

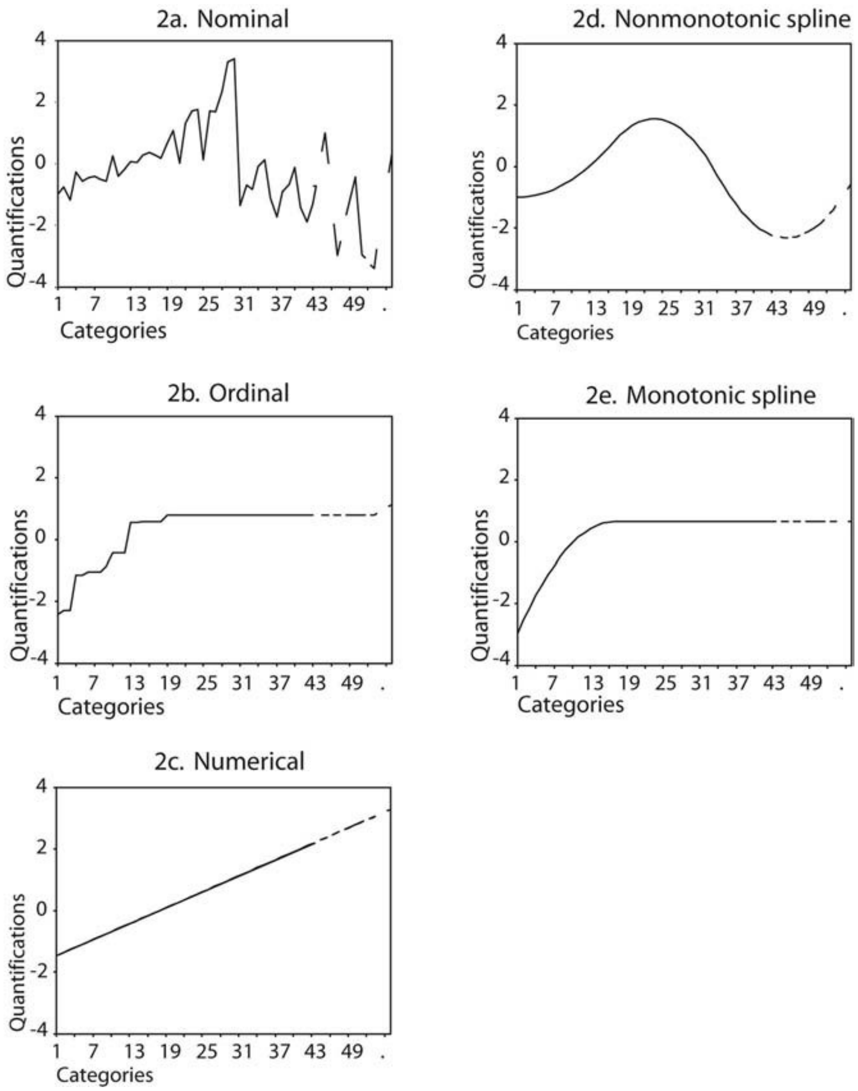

Figure 2. Transformation plots for different types of quantification. The same variable (V2) was assigned five different analysis levels, whereas the other variables were treated numerically. V2 is nonmonotonically related to the other variables. Observed category scores are on the *x*-axis, and the numeric values obtained after optimal quantification (category quantifications) are on the *y*-axis. The line connecting category quantifications indicates the variable's transformation. The gaps in the transformation indicate that some category values were not observed.

monotonic spline transformations in Figures 2b and 2e show an increase of the quantifications for Categories 1 to approximately 20, but all the quantifications for the higher category labels (except the last) are tied because the nominal quantifications did not increase with the category labels, as required by the ordinal analysis level. The numeric quantification in Figure 2c shows (by definition) a straight line.

Evidently, it is possible to treat V2 ordinally or even numerically, but in this example, because of the nonmonotonic relationship between V2 and the other variables, such treatment has a detrimental effect on the fit of the variable. Table 1 gives the fit measures for Variables V1 and V2 obtained from the different analyses. The component loadings are correlations between the quantified variables and the principal components, and the sum of squared component loadings indicates the VAF by the principal components (e.g., when ordinally treated, V1 obtains a loading of .615 on the first component and a loading of −.062 on the second component; then, the VAF of V1 equals $.615^2 + [-.062]^2 = .382$). Variable V1 is monotonically related to the other variables, and therefore, the VAF merely increases from .312 for the numeric treatment to .382 for the ordinal transformation and to .437 for a nominal transformation.

Table 1
Component Loadings and Total Variance Accounted for (VAF) for Two Exemplary Variables Analyzed on Five Different Levels, With the Other Variables Treated Numerically

Analysis level	V1 (monotonic)			V2 (nonmonotonic)		
	Loading 1	Loading 2	VAF	Loading 1	Loading 2	VAF
Nominal	.655	−.087	.437	−.655	.087	.437
Nonmonotonic spline	.598	−.053	.360	−.499	.058	.252
Ordinal	.615	−.062	.382	−.282	.150	.102
Monotonic spline	.598	−.053	.360	−.263	.140	.089
Numeric	.557	−.041	.312	−.055	.106	.014

Note. One variable (V1) is monotonically related to the other variables, and the second variable (V2) is nonmonotonically related to the other variables.

Variable V2 is nonmonotonically related to the other variables, and for this variable, the VAF is .437 for the nominal transformation and .252 for the nonmonotonic spline transformation. When this variable is treated numerically, the VAF essentially reduces to zero (.014). This difference between a numeric and a nominal analysis level for V2 makes clear that when nonlinear relationships between variables exist, nonlinear transformations are crucial to the outcome and the interpretation of the analysis: When a variable (like V2) has a clearly nonlinear relationship to the other variables, applying an inappropriate (numeric) analysis level not only has a detrimental effect on the VAF but, more importantly, leaves the nonlinear relationship between the variable and the other variables to remain unknown and uninterpreted.

Representation of variables as vectors. The plots in Figures 1 and 2 show the transformations of variables, which represent the relationship between the original category labels and the category quantifications. Next, we focus on the representation of the quantified variables themselves. For all of the analysis levels described so far, one way to represent a quantified variable is by displaying its category points in the principal component space, where the axes are given by the principal components. In this type of plot, a variable is represented by a vector (an arrow). A variable vector is a straight line, going through the origin (0,0) and the point with, as coordinates, the component loadings for the variable. The category points are also positioned on the variable vector, and their coordinates are found by multiplying the category quantifications by the corresponding component loadings on the first (for the *x*-coordinate) and the second (for the *y*-coordinate) components. The order of the category points on the variable vector is in accordance with the quantifications: The origin represents the mean of the quantified variable, categories with quantifications above that mean lie on the side of the origin on which the component loadings point is positioned, and categories with quantifications below that mean lie in the opposite direction, on the other side of the origin. Figure 3, explained below, shows an example.

Returning to the example of the inverted-U-shaped relationship between age and income, we assume that age has been divided into three categories (young, intermediate, and old) and has been analyzed nominally, whereas income is treated ordinally. Then, the starting point of the vector representing income indicates the lowest income, and the end point signifies the highest income. The vector representing age is displayed in Figure 3. The fact that the categories young and old lie close together on the low side of the origin and intermediate lies far away in the direction of the component loading point (indicated by the black circle) reflects the U-shaped relation between income and age: Younger and older people have a relatively low income, and people with intermediate age have a relatively high income. For nominal variables, the order of the category quantifications on the vector may differ greatly from the order of the category labels.

The total length of the variable vector in Figure 3 does not indicate the importance of a variable. However, the

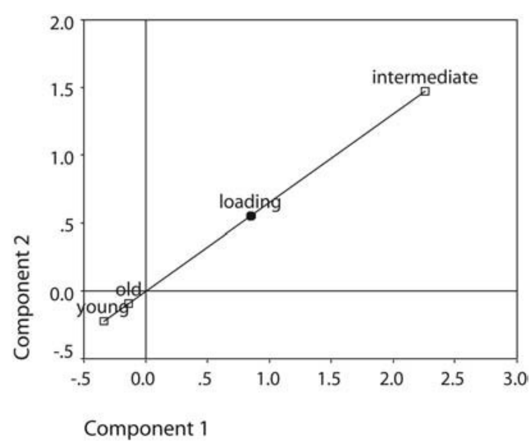

Figure 3. Category plot of the variable age from the fictional age–income example. The category points (which are standard scores) are positioned on a vector. This vector runs through the origin and the loading point (indicated by the black circle). The *x*-coordinate of the loading point is the component loading of the variable age on the first component, and the *y*-coordinate is the component loading of the variable age on the second component.

length of the variable vector from the origin up to the component loading point (loading vector) is an indication of the variable's total VAF. (In fact, the squared length of the loading vector equals the VAF.) In component loading plots, only the loadings' vectors are displayed (for an example, see Figure 4; we consider this figure in more detail in the section on component loadings below). In such a plot, variables with relatively long vectors fit well into the solution, and variables with relatively short vectors fit badly. When vectors are long, the cosines of the angles between the vectors indicate the correlations between the quantified variables.

Thus, the VAF can be interpreted as the amount of information retained when variables are represented in a low, say, two-dimensional, space. Nonlinear transformations reduce the dimensionality necessary to represent the variables satisfactorily. For example, when the original variables may be displayed satisfactorily only in a three-dimensional space, it may turn out that the transformed variables need only a two-dimensional space. In the latter case, nonlinear transformations enhance the interpretability of the graphical representation because it is much easier to interpret a two-dimensional space instead of a three-dimensional one, let alone a four-dimensional one. Of course, by allowing transformations, we replace the need of interpreting a high-dimensional space by an interpretation of the transformations. The latter is usually easier; if this is not the case, another approach to represent variables is to be preferred (see the next paragraph).

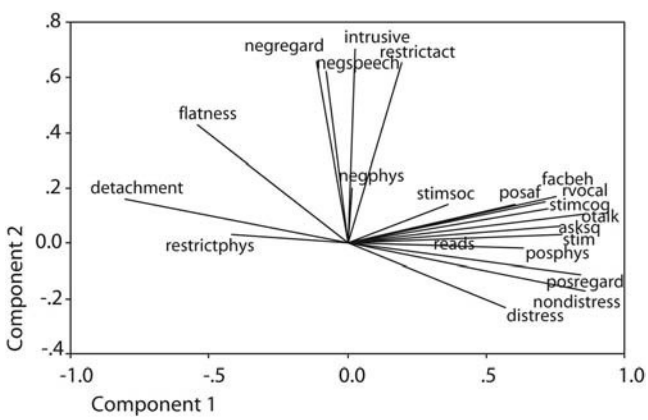

Figure 4. Component loadings of the 21 Observational Record of the Caregiving Environment behavior scales and ratings. Two nominal variables (type of care and caregiver education) are not depicted. The square of the length of the vectors equals the variance accounted for. Cosines of the angles between vectors approximate Pearson correlations between variables. asksq = asks question; distress = sensitivity to distress; facbeh = facilitates behavior; flatness = flatness of affect; intrusive = intrusiveness; negphys = negative physical actions; negregard = negative regard; negspeech = speaks negatively to child; nondistress = sensitivity to nondistress; otalk = other talk to child; posaf = positive affect; posphys = postive physical contact; posregard = positive regard; reads = reads aloud; restrictact = restricts activities; restrictphys = restricts in a physical container; rvocal = responds to vocalization; stim = stimulation; stimcog = stimulates cognitive development; stimsoc = stimulates social development.

Representation of variables as sets of points: Multiple nominal analysis level. In the quantification process discussed so far, each category obtains a single quantification, that is, one optimal quantification that is the same across all the components. The quantified variable can be represented as a straight line through the origin, as has been described above. However, the representation of the category quantifications of a nominal variable (or a variable treated as nominal) on a straight line may not always be the most appropriate one. Only when the transformation shows a specific form (e.g., an inverted-U shape, as in the income vs. age example) or particular ordering (as in Figure 1b) is this type of representation useful. In other words, we should be able to interpret the transformation.

When the transformation is irregular or when the original categories cannot be put in any meaningful order, there is an alternative way of quantifying nominal variables called multiple nominal quantification. The objective of multiple nominal quantification is not to represent one variable as a whole but rather to optimally reveal the nature of the relationship between the categories of that variable and the other variables at hand. This objective is achieved by assigning a quantification for each component separately. For example, imagine that we have a data set including a variable religion, with categories Protestant, Catholic, and Jewish, to which we assign a multiple nominal analysis level. Now suppose we find two principal components: The first indicates liberalism, and the second indicates membership of the religious denomination. Then, we may find that the order of the category quantifications for religion on the first component is Catholic (1), Protestant (2), Jewish (3) on which higher quantifications reflect greater liberalism. In contrast, the order of the category quantifications for religion on the second component may be Jewish (1), Catholic (2), Protestant (3) on which higher values reflect a larger number of members. For each component, the order of the quantifications reflects the nature of the relationship of religion to that component.

Multiple category quantifications are obtained by averaging, per component, the principal component scores for all individuals in the same category of a particular variable. Consequently, such quantifications will differ for each component (hence the term *multiple quantification*). Graphically, the multiple quantifications are the coordinates of category points in the principal component space. Because a categorical variable classifies the individuals in mutually exclusive groups or classes (the categories), these points can be regarded as representing a group of individuals. In contrast to variables with other measurement levels, multiple nominal variables do not obtain component loadings. The fit of a multiple nominal variable in a component is indicated by the variance of the category quantifications in that component. So, if all quantifications are close to the origin, the variable fits badly in the solution. It is important to realize that we define multiple quantifications only for variables with a nominal analysis level. Ordinal, numeric, and spline transformations are always obtained by a single quantification and can be represented as a vector. In the application discussed in the next section, two actual examples of multiple nominal grouping variables are shown.

Representation of individuals as points. Thus far, we have described the representation of the variables in the principal components space either by vectors or by a set of category points. In this paragraph, we address the representation of individuals in nonlinear PCA. Each individual obtains a component score on each of the principal components. These component scores are standard scores that can be used to display the individuals as person points in the same space as the variables, revealing relationships between individuals and variables. This representation is called a biplot in the statistical literature (Gabriel, 1971, 1981; Gifi, 1990; Gower & Hand, 1996). Multiple nominal variables can be represented as a set of category points in the principal components space, and these can be combined with the points for the individuals and the vectors for the other variables in a so-called triplot (Meulman, van der Kooij, & Heiser, 2004). When individuals and category points for multiple nominal variables are plotted together, a particular category point will be exactly in the center of the individuals who have scored in that category. For example, for the variable religion mentioned above, we can label the person points with three different labels: j for Jewish, p for Protestant, and c for Catholic persons. The category point labeled J for the category Jewish of the variable religion will be located exactly in the center of all the person points labeled j, the category point labeled P for the category Protestant will be exactly in the center of all the person points labeled p, and the category point labeled C for the category Catholic will be exactly in the center of all the person points labeled c.

Nonlinear and Linear PCA: Similarities and Differences

Nonlinear PCA has been developed as an alternative to linear PCA for handling categorical variables and nonlinear relationships. Comparing the two methods reveals both similarities and differences. To begin with the former, it can be seen that both methods provide eigenvalues, component loadings, and component scores. In both, the eigenvalues are overall summary measures that indicate the VAF by each component; that is, each principal component can be viewed as a composite variable summarizing the original variables, and the eigenvalue indicates how successful this summary is. The sum of the eigenvalues over all possible components equals the number of variables *m.* If all variables are highly correlated, one single principal component is sufficient to describe the data. If the variables form two or more sets and correlations are high within sets and low between sets, a second or third principal component is needed to summarize the variables. PCA solutions with more than one principal component are referred to as multidimensional solutions. In such multidimensional solutions, the principal components are ordered according to their eigenvalues. The first component is associated with the largest eigenvalue and accounts for most of the variance, the second accounts for as much as possible of the remaining variance, and so on. This is true for both linear and nonlinear PCA.

Component loadings are measures obtained for the variables and, in both linear and nonlinear PCA, are equal to a Pearson correlation between the principal component and either an observed variable (linear PCA) or a quantified variable (nonlinear PCA). Similarly, the sum of squares of the component loadings over components gives the VAF for an observed variable (linear PCA) or a quantified variable (nonlinear PCA). If nonlinear relationships between variables exist and nominal or ordinal analysis levels are specified, nonlinear PCA leads to a higher VAF than linear PCA because it allows for nonlinear transformations. For both methods, before any rotation, the sum of squared component loadings of all variables on a single component equals the eigenvalue associated with that component.

The principal components in linear PCA are weighted sums (linear combinations) of the original variables, whereas, in nonlinear PCA, they are weighted sums of the quantified variables. In both methods, the components consist of standardized scores. In summary, nonlinear and linear PCA are very similar in objective, method, results, and interpretation. The crucial difference is that in linear PCA, the measured variables are directly analyzed, whereas, in nonlinear PCA, the measured variables are quantified during the analysis (except when all variables are treated numerically). Another difference concerns the nestedness of the solution, which is discussed separately in the next paragraph.[2]

Nestedness of the components. One way to view linear PCA is that it maximizes the VAF of the first component over linear transformations of the variables, then maximizes the VAF of the second component that is orthogonal to the first, and so on. This is sometimes called consecutive maximization. The success of the maximization of the VAF is summarized by the eigenvalues and their sum in the first *p* components. The eigenvalues amount to quantities that are equal to the eigenvalues of the correlation matrix. Another

[2] When multiple nominal variables are included, the relations between linear and nonlinear PCA are somewhat different. For more details, we refer the reader to Gifi (1990).

way to view linear PCA is that it maximizes the total VAF in p dimensions simultaneously by projecting the original variables from an m-dimensional space onto a p-dimensional component space (also see the section on graphical representation). In linear PCA, consecutive maximization of the VAF in p components is identical to simultaneous maximization, and we say that linear PCA solutions are nested for different values of p (e.g., corresponding components in p and $p + 1$ dimensions are equal).

In nonlinear PCA, consecutive and simultaneous maximization will give different results. In our version of nonlinear PCA, we maximize the VAF of the first p components simultaneously over nonlinear transformations of the variables. The eigenvalues are obtained from the correlation matrix among the quantified variables, and the sum of the first p eigenvalues is maximized. In this case, the solutions are usually not nested for different values of p. In practice, the differences between the components of a p-dimensional solution and the first p components of a $(p + 1)$-dimensional solution are often very small. They can be dramatic however, for example, if we try to represent a two- or three-dimensional structure in only one dimension. When one doubts whether p is the most appropriate dimensionality, it is advisable to look also at solutions with $p - 1$ and $p + 1$ components.

Choosing the appropriate number of components. In both types of PCA, the researcher must decide the adequate number of components to be retained in the solution. One of the most well-known criteria for this decision is the scree criterion (Fabrigar et al., 1999), which involves a scree plot with the components identified on the x-axis and their associated eigenvalues on the y-axis. Hopefully, such a plot shows a break, or an elbow, identifying the last component that accounts for a considerable amount of variance in the data. The location of this elbow indicates the appropriate number of components to be included in the solution.

Unfortunately, such elbows are not always easily discernible in the linear PCA scree plot. In nonlinear PCA, on the other hand, the fact that the sum of the first p eigenvalues is maximized automatically implies that the sum of the $m - p$ residual eigenvalues is minimized (because the sum of the eigenvalues over all possible components in nonlinear PCA remains equal to m, the number of variables in the analysis). Thus, the elbow in the nonlinear PCA scree plot (which is based on the eigenvalues of the correlation matrix of the quantified variables) may be clearer than that in linear PCA. Because nonlinear PCA solutions are not nested, scree plots differ for different dimensionalities, and the scree plots of the p-, the $(p - 1)$- and $(p + 1)$-dimensional solutions should be compared. When the elbow is consistently at Component p or $p + 1$, the p-dimensional solution may be chosen. There is some discussion in the literature as to whether or not the component where the elbow is located should be included in the solution (see Jolliffe, 2002). A reason for not including it is that it contributes only a little to the total VAF. If a different number of components than p is chosen, the nonlinear PCA should be rerun with the chosen number of components because the components are not nested.

Although the scree criterion may be convenient and is preferred to the "eigenvalues greater than 1" criterion (Fabrigar et al., 1999, p. 281), it is not an optimal criterion. More sophisticated methods, such as parallel analysis (Buja & Eyuboglu, 1992; Horn, 1965), were described by Zwick and Velicer (1986). Peres-Neto, Jackson, and Somers (2005) conducted an extensive simulation study in which they compared 20 stopping rules for determining the number of nontrivial components and developed a new approach. Such alternative methods are applicable to nonlinear PCA as well.

Rotation. PCA solutions may be rotated freely, without changing their fit (Cliff, 1966; Jolliffe, 2002). A familiar example is that of orthogonally rotating the solution so that each variable loads as highly as possible on only one of the two components, thus simplifying the structure (varimax). In a simple structure, similar patterns of component loadings may be more easily discerned. Variables with comparable patterns of component loadings can be regarded as a (sub)set. For example, a test designed to measure the concept intelligence may contain items measuring verbal abilities as well as others measuring quantitative abilities. Suppose that the verbal items correlate highly with each other, the quantitative items intercorrelate highly as well, and there is no (strong) relation between verbal and quantitative items; then, the component loadings from PCA will show two sets of variables that may be taken as verbal and quantitative components of intelligence, respectively. In a simplified structure, these two groups will concur with the components as closely as possible, allowing for a more straightforward interpretation. In nonlinear PCA, orthogonal rotation may be applied in exactly the same way. Note, however, that after rotation, the VAF ordering of the components may be lost.

Nonlinear PCA in Action

Having reviewed the elements and rationale for nonlinear PCA analysis, we are now ready to see how it performs on empirical data and compare the results with a linear PCA solution. Before turning to the actual application, however, we discuss the software used.

Software

Programs that perform nonlinear PCA can be found in the two major commercially available statistical packages: The SAS package includes the program PRINQUAL (SAS Institute, 1992), and the SPSS Categories module contains the program CATPCA (Meulman, Heiser, & SPSS, 2004).

Table 2
Descriptions of ORCE Ratings

Variable	Description	Range	*M*	*SD*
1. Distress	Is responsive to child's distress	1–4	3.15	0.77
2. Nondistress	Is responsive to child's nondistressed communication	1–4	2.89	0.70
3. Intrusiveness	Is controlling; shows adult-centered interactions	1–4	1.19	0.38
4. Detachment	Is emotionally uninvolved, disengaged	1–4	1.66	0.74
5. Stimulation	Stimulates cognitive development (learning)	1–4	1.94	0.68
6. Positive regard	Expresses positive regard toward child	1–4	3.08	0.74
7. Negative regard	Expresses negative regard toward child	1–3	1.03	0.12
8. Flatness	Expresses no emotion or animation	1–4	1.38	0.60

Note. ORCE = Observational Record of the Caregiving Environment.

In the following example, we apply the program CATPCA. The data-theoretical philosophy on which this program is based is defined on categorical variables with integer values.[3] Variables that do not have integer values must be made discrete before they can be handled by CATPCA. Discretization may take place outside of the program, but CATPCA also provides various discretizing options. If the original variables are continuous and we wish to retain as much of the numeric information as possible, we may use a linear transformation before rounding the result. This CATPCA discretizing option is referred to as multiplying. CATPCA contains a number of alternative discretizing options; for an elaborate description, we refer the reader to Meulman, van der Kooij, and Heiser (2004) and to the SPSS Categories manual (Meulman, Heiser, & SPSS, 2004). One of the main advantages of CATPCA is that it has standard provisions for the graphical representation of the nonlinear PCA output, including the biplots and triplots discussed above in the section on the representation of individuals. Another feature of the program is its flexible handling of missing data (see Appendix B). For further details on the geometry of CATPCA, we refer the reader to Meulman, van der Kooij, and Babinec (2002) and Meulman, van der Kooij, and Heiser (2004).

The ORCE Data

We analyzed a mixed categorical data set collected by the National Institute of Child Health and Human Development (NICHD) during their Early Child Care Study (NICHD Early Child Care Research Network, 1996). The subsample we used contains 574 six-month-olds who were observed in their primary nonmaternal caregiving environment (child-care center, care provided in caregiver's home, care provided in child's home, grandparent care, or father care). The Observational Record of the Caregiving Environment (ORCE; NICHD Early Child Care Research Network, 1996) was used to assess quality of day care. This instrument collects observations of the caregiver's interactions with a specific child and provides two types of variables: ratings of overall caregiver behavior during an observation cycle, ranging from *not at all characteristic* (1) to *highly characteristic* (4), and behavior scales that indicate the total number of times each of 13 specific behaviors (e.g., responding to the child's distress or asking the child a question) occurred during thirty 30-second observation periods. Typically, each child was observed four times, and the scores on the ratings and behavior scales were averaged over these four observation cycles. Descriptions of the ORCE variables used in this application appear in Tables 2 and 3. Note that all ratings range from 1 to 4, except negative regard (Variable 7), for which no rating of 4 occurred. The maximum frequency for the behavior scales differs per variable, the overall maximum being 76 for restriction in a physical container (Variable 19). For other talk (Variable 14), no frequency of zero was observed. Histograms for the ORCE variables are displayed in Figure 5. Because score ranges differ, the plots have been scaled differently on the *x*-axis. As Figure 5 depicts, the majority of the distributions are quite skewed, with more extreme scores (e.g., high scores for negative behaviors) displaying relatively small marginal frequencies.

Choice of Analysis Method and Options

One of the goals of the NICHD was to construct composite variables from the ORCE data that captured the information in the original variables as closely as possible. PCA fits this goal exactly. Because it makes no distributional assumptions (e.g., multivariate normality), the skewed distributions are not a problem. It could be argued that linear PCA is appropriate for the ORCE data, although the ratings were actually measured on an ordinal scale. However, it does not seem realistic to assume linear relationships among the variables a priori. In addition, the NICHD wished to identify relationships between the ORCE variables and relevant background characteristics, such as the caregiver's education and type of care. Neither of these

[3] This is a property not of the method of nonlinear PCA but of the CATPCA program.

Table 3
Descriptions of ORCE Behavior Scales

Variable	Description	Range	*M*	*SD*
9. Positive affect	Shared positive affect (laughing, smiling, cooing)	0–32	4.72	4.34
10. Positive physical	Positive physical contact (holding, touching)	0–55	19.89	10.45
11. Vocalization	Responds to child's nondistressed vocalization	0–26	4.70	4.59
12. Reads aloud	Reads aloud to child	0–11	0.37	1.16
13. Asks question	Directs a question to child	0–46	12.27	8.08
14. Other talk	Declarative statement to child	1–59	24.25	12.07
15. Stimulates cognitive	Stimulates child's cognitive development	0–34	2.99	3.85
16. Stimulates social	Stimulates child's social development	0–9	0.73	1.28
17. Facilitates behavior	Provides help or entertainment for child	0–55	18.80	9.68
18. Restricts activity	Restricts child's activities physically or verbally	0–21	1.26	1.96
19. Restricts physical	Restricts child in physical container (playpen)	0–76	20.87	14.35
20. Negative speech	Speaks to child in a negative tone	0–3	0.05	0.24
21. Negative physical	Uses negative physical actions (slap, yank, push)	0–2	0.01	0.11

Note. ORCE = Observational Record of the Caregiving Environment.

variables is numeric, and either might well have nonlinear relationships with the ORCE variables. Therefore, we decided to apply nonlinear rather than linear PCA. Before beginning the analysis, however, the appropriate analysis options have to be chosen. To aid researchers in such decisions, we discuss the decisions we made for the ORCE example below.

Analysis options. In the present data set, all behavior scales and ratings have a definite order, and we wished to retain this. As we did not wish to presume linearity, we treated all ORCE variables ordinally. As some of the behavior scales have numerous categories, it might have been useful to assign a monotonic spline analysis level to those variables; however, a spline analysis level is more restrictive and thus can lead to lower VAF. Because, for the ORCE data, the solution with monotonic spline quantifications differed only slightly from the solution with ordinal quantifications (which was most likely due to the large size of the sample), we decided that the latter was appropriate. In addition to the behavior scales and ratings, we included two background variables: type of care and caregiver education. Type of care was represented via six categories: father care, care by a grandparent, in-home sitter, family care, center care, and other. The caregiver's education was measured using six categories as well: less than high school, high school diploma, some college, college degree, some graduate, and graduate degree. Because the types of care were nominal (i.e., unordered) and we did not know whether the category quantifications of type of care and caregiver education would show an irregular pattern, we analyzed both background variables multiple nominally (as grouping variables). In summary, we specified an ordinal analysis level for the 21 ORCE variables and a multiple nominal level for the two background variables.

Second, all of the ratings and behavior scales had noninteger values, so we had to discretize them for analysis by CATPCA. As we wanted the option of handling the ORCE variables numerically for a possible comparison between ordinal and numeric treatment, we wished to retain their measurement properties to the degree possible. As the multiplying option (see the *Software* section, above) is designed for this purpose, we chose it for the behavior scales as well as the ratings.

Finally, we treated the missing values passively (see Appendix B), deleting persons with missing values only for those variables on which they had missing values. The number of missing values in the ORCE data is relatively small: Two children had missing values on all eight ratings, 94 children had a missing value on responsiveness to distress, and 16 had missing values on caregiver education.

Number of components. Finally, we had to determine the adequate number of components to retain in the analysis. Because the ORCE instrument measures positive and negative interactions between caregivers and children, it seemed reasonable to assume that two components were called for. We generated a scree plot to check this assumption, using the eigenvalues of the correlation matrix of the quantified variables from a two-dimensional solution. From this plot, presented in Figure 6b, we concluded that the elbow is located at the third component. Remember that nonlinear PCA solutions are not nested, so a scree plot for a three-dimensional solution—in which the sum of the three largest eigenvalues is optimized—can be different from a scree plot for a two-dimensional solution, with the position of the elbow moving from the third to the second component. In the present analysis, different dimensionalities consistently place the elbow at the third component, as shown in Figure 6. Inspection of the three-dimensional solution revealed that the third component was difficult to interpret, suggesting that this solution is not theoretically sensible and

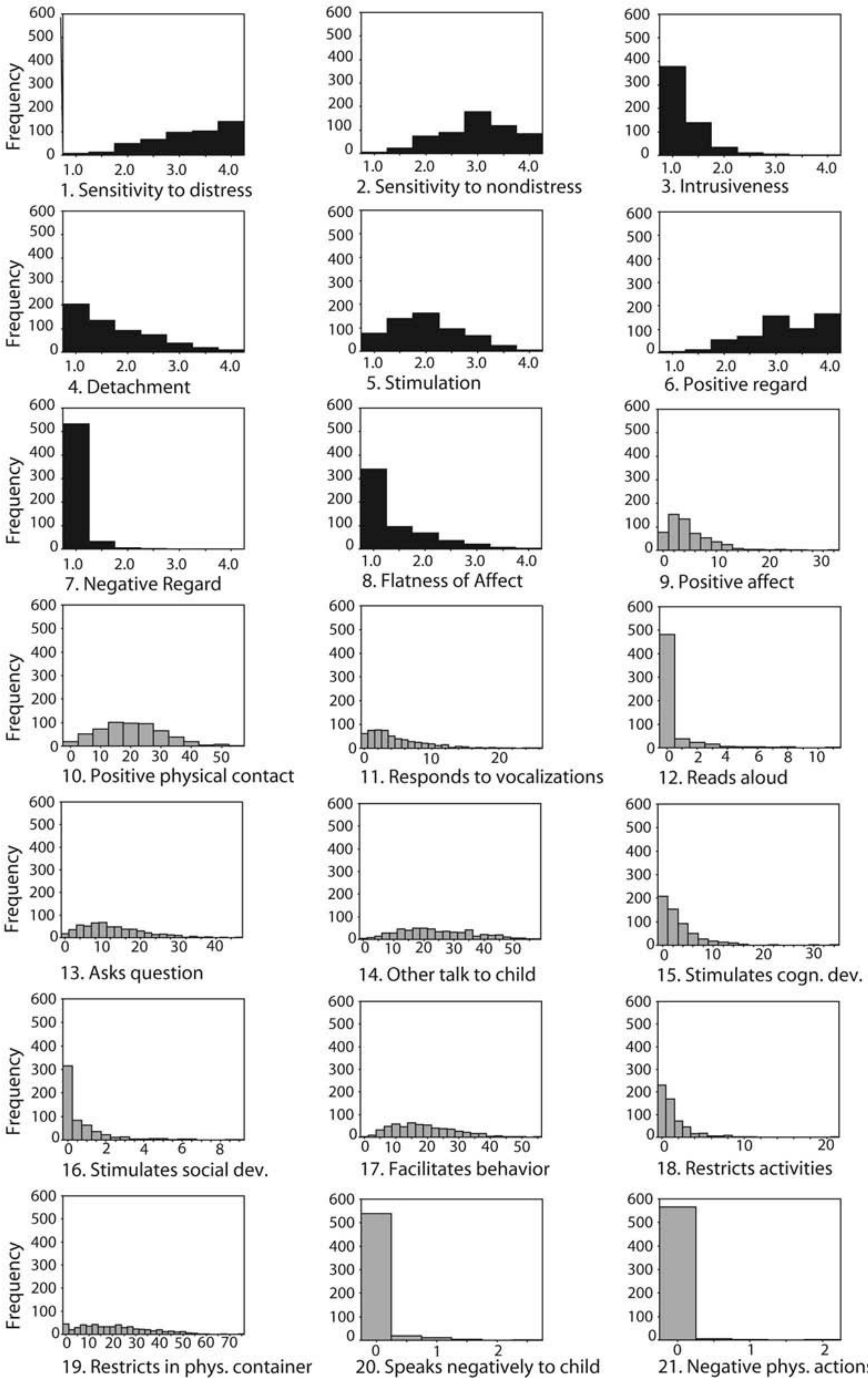

Figure 5. Histograms for the Observational Record of the Caregiving Environment variables. The black histograms represent ratings, and the grey histograms represent behavior scales. Note that the plots are scaled differently on the *x*-axis because the variables had different ranges. $N = 574$. Cogn. dev. = cognitive development; phys. = physical; social dev. = social development.

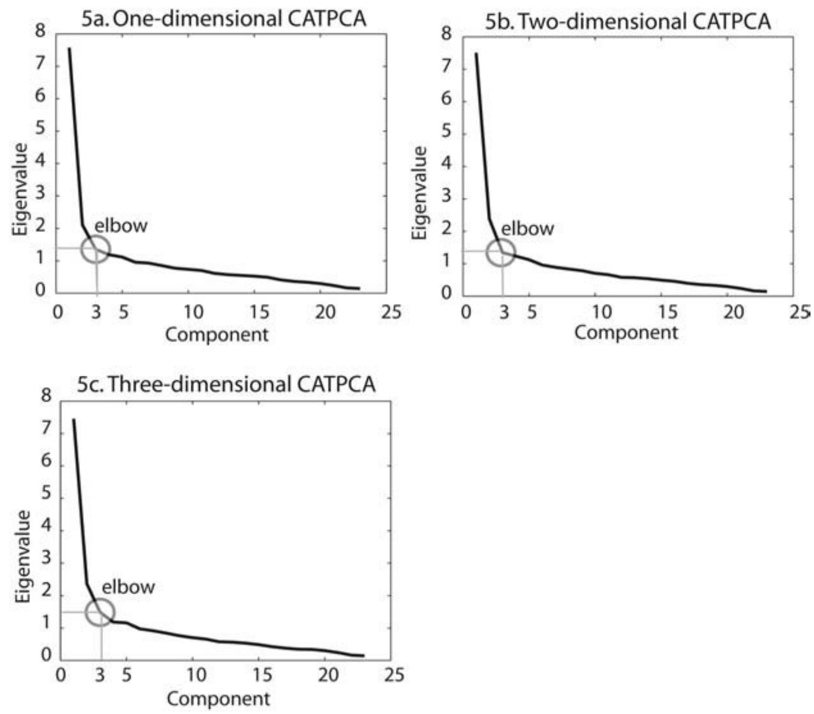

Figure 6. Scree plots from one-, two-, and three-dimensional nonlinear principal components analysis on the Observational Record of the Caregiving Environment data. On the y-axis are the eigenvalues of the correlation matrix of the quantified variables.

therefore has little value (Fabrigar et al., 1999). This lack of interpretability of the third component, combined with the information from the scree plot, suggests that the two-dimensional solution is most appropriate.

After deciding on the appropriate number of components, we checked whether we could simplify the structure of the solution by rotating the results. For the nonlinear PCA solution on the ORCE data, rotation was not called for, as most variables already loaded highly on only one component.

The Nonlinear PCA Solution for the ORCE Data

As we believe that the nonlinear PCA solution can best be represented graphically, the next part of this section mainly focuses on interpreting the plots from the CATPCA output.

Variance accounted for. The two-dimensional nonlinear PCA on the ORCE data yields an eigenvalue of 7.71 for the first component, indicating that approximately 34% (= 7.71/23, with 23 being the number of variables) of the variance in the transformed variables is accounted for by this first component. The eigenvalue of the second component is 2.41, indicating that its proportion of VAF is approximately 10%. Thus, the first and second components together account for a considerable proportion (44%) of the variance in the transformed variables.

Transformation plots. Figure 7 shows the optimal quantifications for the behavior scale responds to vocalization (with the quantifications on the y-axis vs. the original category labels on the x-axis). The labels on the x-axis are the

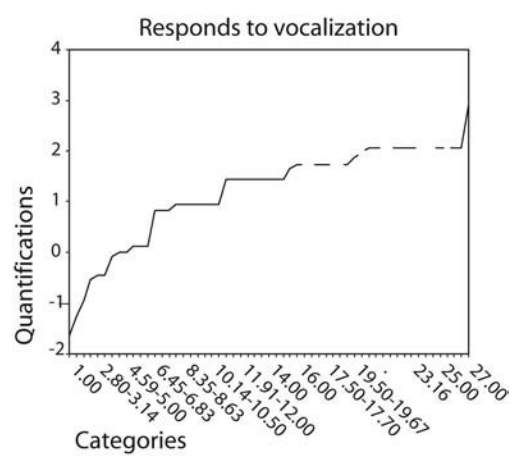

Figure 7. Ordinal transformation of the behavior scale responds to vocalization. The categories indicated on the x-axis are the result of discretizing using the multiplying option. To enhance legibility, only every fourth category is depicted. Numeric values obtained after optimal quantification (category quantifications) are on the y-axis.

ranges, in terms of the original category labels, of the categories that were constructed using the discretization option multiplying (see the *Software* section, above). Because this discretized variable had many categories, only one out of four category labels is displayed on the *x*-axis to keep the labels legible. Some possible categories do not occur, indicated by blank areas in the plot and by dots on the *x*-axis. In accordance with the variable's ordinal analysis level, the transformation shows a nondecreasing line. Some categories received the same quantification, as indicated by horizontal parts in the line. Possibly, for example, children in Categories 8.35–8.63 and 10.14–10.50 of responds to vocalization did not differ structurally in their patterns on the other variables and thus could not be distinguished from each other by nonlinear PCA. Another possibility is that the unrestricted nominal quantifications of consecutive categories were incorrectly ordered, raising the possibility of a nonlinear relationship between this variable and the others. In such a case, nominal treatment of a variable might be considered. However, because the ties do not extend over a large interval of categories, ordinal treatment seems quite adequate for these data.

The transformation plots of most of the other ORCE variables are comparable to that in Figure 7, with most variables showing one or more ties but never over a long sequence of categories. Variables with only a few categories (such as negative speech [Variable 20] and negative physical action [Variable 21]) have only a few quantifications and usually do not show ties.

Component loadings. Component loadings for the ordinally treated data are presented in Figure 4, displayed as vectors. (To enhance clarity, the grouping [multiple nominal] variables appear in a separate graph; see Figure 8.) As component loadings indicate Pearson correlations between the quantified variables and the principal components, they range between −1 and 1. The coordinates of the end point of each vector are given by the loadings of each variable on the first and second components. Because the cosines of the angles between the vectors equal the correlations between the quantified variables and because vectors are long (indicating good fit), variables that are close together in the plot are closely and positively related. Variables with vectors that make approximately a 180° angle with each other are closely and negatively related. Vectors making a 90° angle indicate that variables are not related.

The variables in Figure 4 form roughly three groups that seem to coincide with three basic orientations that are often used in the psychological literature and are based on the theory of Horney (1945): moving toward, moving away from, and moving against another individual. The first group of variables has high positive loadings on the first component and low loadings on the second and denotes positive behaviors, or moving toward the child. The second group, with high negative loadings on the first component and low loadings on the second component, contains variables that represent the caregiver's disengagement, or moving away from the child, such as showing flatness of affect and detachment. The orientation of the vectors for the

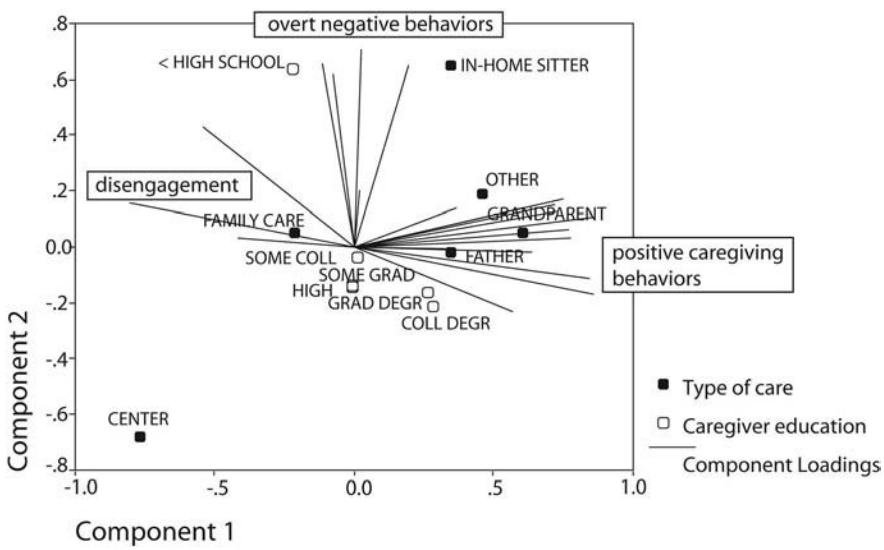

Figure 8. Component loadings from CATPCA on 23 variables. 21 Observational Record of the Caregiving Environment variables are indicated by vectors, and the multiple nominal variables type of care and caregiver education are represented by category points. Labels of the category points are in capital letters. COLL DEGR = college degree; GRAD DEGR = graduate degree; HIGH = high school diploma; SOME COLL = some college; SOME GRAD = some graduate school.

variables in the first and second groups is roughly the same, but the vectors point in opposite directions, indicating a strong negative relationship between these groups of variables. The vector of a variable points to the highest category of that variable; thus, for the first group, the vectors point to moving toward the child, and for the second group, the vectors point to the caregiver's moving away from the child. The third group of variables, representing overtly negative caregiver behaviors, or moving against the child, loads highly on the second component and hardly on the first.

Category points for multiple nominal variables. To clarify the relationship of the multiple nominal (grouping) variables type of care and caregiver education to the other variables in the data set, their category points are shown in Figure 8, along with the vectors for the ORCE variables from Figure 4. As can be seen, the ordinal variables' names have been replaced in the figure with summary labels derived from Figure 4. The group points are widely spread, indicating good fit. The locations of the group points reflect the relationship of each category to the ORCE variables: Orthogonal projection of a multiple nominal category point onto a vector for an ordinal variable reflects to what score on that ordinal variable the category is most related.

In Figure 8, orthogonal projection of the category point for center care on the variable vectors shows that children in center care experience relatively much disengagement from their caregivers, along with relatively little positive or overtly negative behavior. Caregivers in family day care show a bit more positive engagement with the children but still less than in other types of care. Fathers, grandparents, and other caregivers display a relatively high degree of positive behavior, along with few negative or disengaging behaviors. Finally, in-home sitters display about as many positive behaviors as fathers and grandparents. The fact that the point for the sitters falls close to the highest categories of overt negative behavior demonstrates that such behaviors are more likely with sitters than with other types of caregivers. However, we cannot conclude that such negative behaviors are characteristic of sitters because the high categories of the negative behaviors have very small marginal frequencies; that is, these behaviors occurred only rarely.

Examining group points for caregiver education in Figure 8, we see that caregivers with less than high school education (displayed at the upper left side of the plot) show a relatively high degree of overtly negative behavior and relatively little positive behavior toward their charges. The categories high school, some college, and some graduate lie close to the center of the plot, indicating that caregivers at these levels of education show moderate levels of both positive and negative behaviors. Caregivers with a college or graduate degree show the most positive and least overtly negative behavior. The categories of education are almost perfectly ordered from lower to higher education and (except for < high school) on a virtually straight line; thus, ordinal treatment would also have been appropriate for this variable. With respect to the relationship between type of care and caregiver education, the most notable conclusion is that in-home sitters are more likely to have less than a high school education than are other types of caregivers.

Person scores. Figure 9 displays the points for the children in a so-called object scores plot. (In CATPCA, person scores are referred to as object scores because individuals or other entities in a study are neutrally referred to as objects.) If a solution is dominated by outliers, all of the points in an object scores plot are positioned very close to the origin of the plot, whereas one or a few points lie very far away. In this example, we see that some points lie far away from the other points, but the child points are spread out in both components, indicating that our solution is not dominated by outliers. The points lying far outside the center of the plot correspond to children with extreme scoring patterns. For example, the points labeled 1 and 2 are exceptional on the second component. Other points with uncommon scoring patterns are indicated by the labels 3 to 7.

A child point is located as close as possible to the categories the child has scored in, given the restrictions placed by the analysis level of the variables, the fact that the component scores are standardized, and the fact that the components themselves are uncorrelated. To display children's scores on the quantified ORCE variables, the child points can be displayed in the same space as the variables

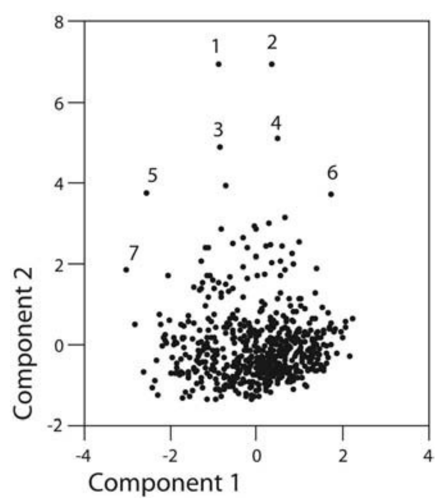

Figure 9. Object (component) scores from CATPCA on 23 variables: 21 Observational Record of the Caregiving Environment variables, type of care, and caregiver education. Object scores are standard scores for the persons on the components. In this case, these standard scores range from about −3 to 3 on the first component. On the second component, the points labeled 1 to 4 obtained very high standard scores, such that the scores on the second component range from about −1.8 to 7.

(represented by vectors). Such a joint plot, also called a biplot, is shown in Figure 10. We adjusted the scale of this figure to the range of the person scores. As person scores are standard scores, their range is usually larger than the range of the component loadings (between −1 and 1). As, in this data set, the range of the object scores is substantially larger than that of the loadings, the loadings would appear as very short vectors in the biplot. For reasons of clarity, we elongated the variable vectors by a factor of 4. Consequently, the scales of Figures 10 and 4 differ from each other.

To display a child's score on a particular variable, its point should be projected orthogonally onto the appropriate variable vector. In general, the children can be sorted into two groups on the basis of their scores on the second component: Those with an object score of 1 or higher on this component experienced, on average, a higher degree of intrusiveness, negative regard, flatness, negative speech, negative physical actions, and especially restriction of activity. For all of these variables, prior to quantification, *t* tests for the comparison of means between these two groups were significant at the .01 level. To show more detailed examples, the points for four of the children (those labeled 1, 2, 6, and 7) can be projected onto the elongated variable vectors. It can be seen that Children 1 and 2 experienced a relatively high degree of overt negative caregiving behaviors. Child 6 experienced a relatively high degree of overt negativity but also a good many positive caregiving behaviors (and a relatively low degree of disengagement). On the left side of the plot, Child 7 encountered a relatively high degree of disengagement (and relatively little positivity), in combination with some overt negative behaviors. As can be seen, such a biplot provides a clear and comprehensive view of the relationships between persons and variables. It is possible to enrich this view even further, as CATPCA can label persons by the categories of a particular (background) variable (rather than simply an anonymous case number) to help determine how well groups of persons can be distinguished on the basis of that particular variable.

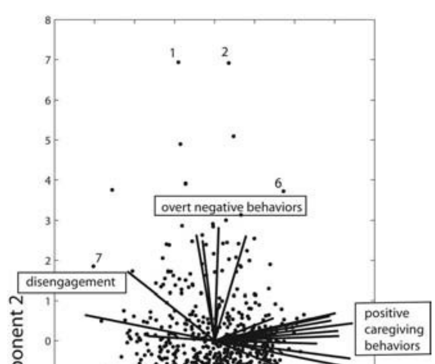

Figure 10. Biplot of the object scores (indicated by points) and component loadings (indicated by vectors) from CATPCA on the 23 variables. To show the relation between the object scores and component loadings more clearly, the vectors for the component loadings have been adjusted to the range of the objects (see Figure 9) by elongating them by a factor of 4. Note that the scale of the plot is equal to that of Figure 9 and changes from Figure 4.

Comparison of the Nonlinear and Linear PCA Solutions

As a comparison, we also conducted a linear PCA on the ORCE data. In the current example, the most important advantage of nonlinear PCA is that the variables type of care and caregiver education could be included in the analysis. In standard linear PCA, it is not possible to include nominal variables, and the information potentially yielded by these variables is lost. A second advantage of nonlinear PCA is that nominal or ordinal treatment may provide a greater proportion of VAF because the analysis allows more freedom in the quantifications of the data. The loadings and VAF of the two principal components appear in Table 4 for both types of solutions.[4]

As consecutive category points for caregiver education are not equidistant from each other (see Figure 8), numeric treatment (implying equal distances between the categories) of this variable by linear PCA would not be the best choice. We therefore excluded this variable, as well as the nominal variable type of care, from the linear PCA. To provide a fair comparison, the results from nonlinear PCA in Table 4 involve only the ordinally treated ORCE variables. Neither solution was rotated, to keep the comparison as straightforward as possible. Furthermore, the PCA solution with varimax rotation was very similar to the unrotated solution.

Table 4 shows that for 18 out of the 21 ORCE variables, more variance is accounted for in the nonlinear PCA solution than in the linear PCA solution, whereas only two variables (Variables 1 and 5) have less VAF in the nonlinear PCA solution and one variable (Variable 3) has the same. Only responds to distress (Variable 1) showed a decrease in VAF of more than 5% by analyzing it ordinally, whereas eight variables (Variables 7, 9, 10, 11, 13, 15, 19, and 20) showed an increase in VAF of 5% or more. The sum of VAF over variables is equal to the sum of the eigenvalues. In the linear PCA solution, the percentage of VAF by the components for all variables (the eigenvalue divided by the number of variables) was 32.7% for the first component and 9.9% for the second component. In the nonlinear PCA solution, the first component accounted for approximately 35.7% of the variance in the variables, and the second for approximately 11.0%, indicating that there is a small in-

[4] The VAF of a variable is the sum of squared component loadings across components.

Table 4
Component Loadings and VAF for Linear and Nonlinear PCA With Two Components

Variable	Linear PCA			Nonlinear PCA		
	Loading 1	Loading 2	VAF	Loading 1	Loading 2	VAF
1. Distress	.608	−.257	.435	.570	−.228	.377
2. Nondistress	.853	−.190	.764	.861	−.155	.766
3. Intrusiveness	.010	.671	.450	.009	.671	.450
4. Detachment	−.787	.155	.643	−.807	.140	.672
5. Stimulation	.785	.024	.616	.773	.048	.600
6. Positive regard	.830	−.136	.708	.841	−.091	.716
7. Negative regard	−.095	.643	.422	−.094	.705	.505
8. Flatness	−.558	.275	.387	−.555	.352	.432
9. Positive affect	.564	.065	.322	.603	.132	.381
10. Positive physical	.538	−.049	.292	.635	−.046	.406
11. Vocalization	.660	.097	.445	.704	.124	.511
12. Reads aloud	.273	.134	.092	.337	−.009	.113
13. Asks question	.727	.148	.551	.770	.090	.601
14. Other talk	.828	.186	.719	.858	.129	.753
15. Stimulates cognitive	.630	.181	.429	.723	.146	.544
16. Stimulates social	.376	.110	.153	.373	.198	.179
17. Facilitates behavior	.729	.182	.565	.749	.154	.584
18. Restricts activity	.169	.657	.461	.184	.679	.496
19. Restricts physical	−.283	−.050	.083	−.420	.040	.178
20. Negative speech	−.074	.604	.370	−.070	.691	.483
21. Negative psysical	.015	.179	.032	.019	.213	.046
Total (sum of eigenvalues)		8.941			9.793	

Note. PCA = principal components analysis; VAF = variance accounted for.

crease of VAF on both components. In the present analysis, we see only a small increase in the total VAF because variables turned out to be only slightly nonlinearly related. In cases when strongly nonlinear relationships between variables occur, the increase in VAF will be much larger than in the present case. For the current data, the increase in VAF is secondary to the major advantage of being able to deal with nominal variables. (Note that the inclusion of nominal variables could also increase the total VAF of the solution.)

Discussion

In this article, we have provided an elaborate explanation of the method of nonlinear PCA. We have discussed how, by using optimal quantification, the method achieves the same objectives as linear PCA for nominal, ordinal, and numeric variables. We have then described an extensive example of a nonlinear PCA analysis, explaining the different analysis options we chose and interpreting the results. Hopefully, this article will provide a didactic and practical guide for researchers interested in applying nonlinear PCA to their own data. For further applications of nonlinear PCA, we refer the reader to Vlek and Stallen (1981); Eurelings-Bontekoe, Duijsens, and Verschuur (1996); Beishuizen, Van Putten, and Van Mulken (1997); De Haas, Algera, and Van Tuijl (2000); Huyse et al. (2000); Hopman-Rock, Tak, and Staats (2001); Zeijl, Te Poel, Du Bois-Reymond, Ravesloot, and Meulman (2000); Arsenault, Tremblay, Boulerice, and Saucier (2002); De Schipper, Tavecchio, Van IJzendoorn, and Linting (2003); and Theunissen et al. (2004).

Comparing the nonlinear PCA results from the application section with the linear PCA results, we found that, for the data we used, the most important advantage of nonlinear PCA over linear PCA is its ability to deal with nominal variables. Another important advantage that is not as evident for the ORCE example is the ability of nonlinear PCA to discover and handle nonlinear relationships between variables (as shown in Figures 1 and 2). Nonlinear PCA enables the researcher to deal with variables in accordance with their measurement level, as well as to explore whether using analysis levels other than the measurement level provides more insight into the relationship of a variable with the other variables in the data set. In addition, nonlinear PCA can help demonstrate whether numeric (linear) treatment of variables is justified.

A secondary advantage of nonlinear PCA is that it may be able to account for more of the variance in the data than linear PCA if nominal or ordinal analysis levels are used. For nonnumeric analysis levels, the method is less restrictive and can therefore reach a higher proportion of VAF. Comparing the nonlinear with the linear PCA solutions, we indeed found that for some of the variables in our example data set, the VAF increased considerably. Usually, in scale construction, variables that are closely related are simply summed (each variable with weight of 1) to obtain a composite variable. If that strategy is followed, no advantage is taken of the increased VAF for variables in nonlinear PCA. Alternatively, the advantage of increased VAF can be used by the researcher by using the component loadings as variable weights in the computation of composite variables. This approach may be especially advantageous when a data set is heterogeneous, that is, contains variables that measure different types of characteristics and do not all have the same range. However, when applying this strategy, it is important that the nonlinear PCA results are stable (thus, there is little sampling error in estimating the component loadings). Stability can be determined by performing a bootstrap study on the nonlinear PCA results (see below). If sampling error is of considerable influence, a coarser weighting strategy than using the component loadings directly as weights could be applied. For example, assign a weight of 1 to variables with high positive loadings, a weight of -1 to variables with high negative loadings, and a weight of 0 to variables with low loadings on a particular component. Somewhat more complex weighting alternatives (weights of -2, -1, 0, 1, and 2) may also be considered (Grice, 2001).

Although increasing VAF seems profitable, in cases in which variables have only slightly nonlinear relationships with each other (which is often the case when only Likert-type scales are measured), a nonlinear PCA will not add a great deal to the linear solution. For any data set, we recommend first checking whether nonlinear treatment is warranted by trying out nominal or ordinal analysis levels for the variables at hand and looking at the VAF compared with numeric treatment. Also, alternative measures of fit may be considered: For example, Gower and Blasius (2005) proposed a fit measure based on the multivariate predictability of the data from the nonlinear PCA solution. In addition, transformation plots should be considered. If all plots show a regularly ascending line, numeric treatment may be just as satisfactory as ordinal or nominal treatment in terms of both VAF and interpretability.

One should also consider whether the gain in VAF outweighs any possible increase in complexity of the solution. If the solution becomes much more difficult to interpret, for example, because of very irregular quantifications, one might want to use linear PCA regardless of any gain in VAF. In addition, the risk that the results are too much dependent on the data and cannot be generalized to the population increases when less restrictive analysis levels are used (with a nominal analysis level being the least restrictive). However, if a variable is strongly nonlinearly related to the other variables, one will fail to recognize and interpret that relationship when using linear PCA. Consequently, VAF should always be considered in the light of interpretability of the solution, and it might be useful to try different analysis levels for the variables at hand and compare solutions before deciding which gives the clearest representation of the data.

Because applying nonlinear PCA involves trying out different options to find the best possible results, users may feel that the method is not very straightforward. In addition, all of this exploration may lead to searching for structure that could be dependent on the sample at hand. According to Diaconis (1985), in general, using exploratory analyses may easily lead to magical thinking, that is, seeing structure that is not generalizable to a larger population. Researchers using methods incorporating optimal quantification have also been accused of capitalizing on chance. As an exploratory method, neither linear nor nonlinear PCA provides standard options for assessing whether the solution depends too heavily on sample characteristics, and thus, it is important to develop methods for assessing measures of stability and statistical significance for such exploratory techniques. Procedures like the nonparametric bootstrap (Efron, 1982; Efron & Tibshirani, 1993) and permutation tests (see, e.g., Good, 2000) may be useful for overcoming this limitation. A companion article (Linting, Meulman, Groenen, & van der Kooij, 2007) discussing the use of the bootstrap to assess stability in the nonlinear PCA context is provided. Such studies will further increase the usefulness of this method.

In conclusion, nonlinear PCA can be a useful alternative to the more familiar linear method. As Buja (1990) pointed out, this methodology is “one of the most powerful and universal tools for the analysis of multivariate data due to its ability to recover frequent types of nonlinear structure and its applicability to categorical data” (p. 1034). For researchers dealing with (large) data sets including different types of categorical variables that are possibly nonlinearly related, nonlinear PCA can be a valuable addition to their methodological toolbox.

References

Arsenault, L., Tremblay, R. E., Boulerice, B., & Saucier, J. (2002). Obstetrical complications and violent delinquency: Testing two developmental pathways. *Child Development, 73,* 496–508.

Beishuizen, M., Van Putten, C. M., & Van Mulken, F. (1997). Mental arithmetic and strategy use with indirect number problems up to hundred. *Learning and Instruction, 7,* 87–106.

Buja, A. (1990). Remarks on functional canonical variates, alternating least squares methods and Ace. *Annals of Statistics, 18,* 1032–1069.

Buja, A., & Eyuboglu, N. (1992). Remarks on parallel analysis. *Multivariate Behavioral Research, 27,* 509–540.

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31,* 33–42.

De Haas, M., Algera, J. A., & Van Tuijl, H. F. J. M. (2000). Macro and micro goal setting: In search of coherence. *Applied Psychology, 49,* 579–595.

De Schipper, J. C., Tavecchio, L. W. C., Van IJzendoorn, M. H., & Linting, M. (2003). The relation of flexible child care to quality of center day care and children's socio-emotional functioning: A survey and observational study. *Infant Behavior & Development, 26,* 300–325.

Diaconis, P. (1985). Theories of data analysis: From magical thinking through classical statistics. In D. C. Hoaglin, F. Mosteller, & J. W. Tukey (Eds.), *Exploring data tables, trends, and shapes* (pp. 1–36). New York: Wiley.

Efron, B. (1982). *The jackknife, the bootstrap, and other resampling plans.* Philadelphia: Society for Industrial and Applied Mathematics.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap.* New York: Chapman & Hall.

Eurelings-Bontekoe, E. H. M., Duijsens, I. J., & Verschuur, M. J. (1996). Prevalence of DSM-III-R and ICD-10 personality disorders among military conscripts suffering from homesickness. *Personality and Individual Differences, 21,* 431–440.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Gabriel, K. R. (1971). The biplot-graphic display of matrices with application to principal component analysis. *Biometrika, 58,* 453–467.

Gabriel, K. R. (1981). Biplot display of multivariate matrices for inspection of data and diagnosis. In V. Barnett (Ed.), *Interpreting multivariate data* (pp. 147–174). Chichester, England: Wiley.

Gifi, A. (1990). *Nonlinear multivariate analysis.* Chichester, England: Wiley.

Good, P. I. (2000). *Permutation tests: A practical guide to resampling methods for testing hypotheses.* New York: Springer-Verlag.

Gower, J. C., & Blasius, J. (2005). Multivariate prediction with nonlinear principal components analysis: Theory. *Quality & Quantity, 39,* 359–372.

Gower, J. C., & Hand, D. J. (1996). *Biplots.* London: Chapman & Hall.

Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods, 6,* 430–450.

Guttman, L. (1941). The quantification of a class of attributes: A theory and a method of scale construction. In P. Horst (Ed.), *The prediction of personal adjustment* (pp. 319–348). New York: Social Science Research Council.

Hopman-Rock, M., Tak, E. C. P. M., & Staats, P. G. M. (2001). Development and validation of the Observation List for early signs of Dementia (OLD). *International Journal of Geriatric Psychiatry, 16,* 406–414.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185.

Horney, K. (1945). *Our inner conflicts: A constructive theory of neurosis.* New York: Norton.

Huyse, F. J., Herzog, T., Lobo, A., Malt, U. F., Opmeer, B. C., & Stein, B. (2000). European consultation-liaison psychiatric services: The ECLW collaborative study. *Acta Psychiatrica Scandinavica, 101,* 360–366.

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

Kruskal, J. B. (1965). Analysis of factorial experiments by estimating monotone transformations of the data. *Journal of the Royal Statistical Society, Series B, 27,* 251–263.

Kruskal, J. B., & Shepard, R. N. (1974). A nonmetric variety of linear factor analysis. *Psychometrika, 39,* 123–157.

Linting, M., Meulman, J. J., Groenen, P. J. F., & van der Kooij, A. J. (2007). Stability of nonlinear principal components analysis: An empirical study using the balanced bootstrap. *Psychological Methods, 12,* 359–379.

Meulman, J. J., Heiser, W. J., & SPSS. (2004). *SPSS Categories 13.0.* Chicago: SPSS.

Meulman, J. J., van der Kooij, A. J., & Babinec, A. (2002). New features of categorical principal components analysis for complicated data sets, including data mining. In W. Gaul & G. Ritter (Eds.), *Classification, automation, and new media* (pp. 207–217). Berlin, Germany: Springer.

Meulman, J. J., van der Kooij, A. J., & Heiser, W. J. (2004). Principal components analysis with nonlinear optimal scaling transformations for ordinal and nominal data. In D. Kaplan (Ed.), *The Sage handbook of quantitative methodology for the social sciences* (pp. 49–70). London: Sage.

NICHD Early Child Care Research Network. (1996). Characteristics of infant child care: Factors contributing to positive caregiving. *Early Childhood Research Quarterly, 11,* 269–306.

Peres-Neto, P. R., Jackson, D. A., & Somers, K. M. (2005). How many principal components? Stopping rules for determining the number of non-trivial axes revisited. *Computational Statistics and Data Analysis, 49,* 974–997.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Science, 3,* 425–441.

SAS Institute. (1992). *SAS/STAT software: Changes and enhancements.* Cary, NC: Author.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177.

Shepard, R. N. (1966). Metric structures in ordinal data. *Journal of Mathematical Psychology, 3,* 287–315.

SPSS. (2007). *CATPCA algorithm.* Retrieved January 22, 2007, from the SPSS Statistical Support—Algorithms Web page at http://support.spss.com/Tech/Products/SPSS/Documentation/Statistics/index.html

Theunissen, N. C. M., Meulman, J. J., Den Ouden, A. L., Koopman, H. M., Verrips, G. H., Verloove-Vanhorick, S. P., & Wit, J. M. (2004). Changes can be studied when the measurement instrument is different at different time points.

Health Services and Outcomes Research Methodology, 4, 109–126.

Vlek, C., & Stallen, P. J. (1981). Judging risks and benefits in the small and in the large. *Organizational Behavior and Human Performance, 28,* 235–271.

Winsberg, S., & Ramsay, J. O. (1983). Monotone spline transformations for dimension reduction. *Psychometrika, 48,* 575–595.

Young, F., Takane, Y., & de Leeuw, J. (1978). The principal components of mixed measurement level multivariate data: An alternating least squares method with optimal scaling. *Psychometrika, 43,* 279–281.

Zeijl, E., Te Poel, Y., Du Bois-Reymond, M., Ravesloot, J., & Meulman, J. J. (2000). The role of parents and peers in the leisure activities of young adolescents. *Journal of Leisure Research, 32,* 281–302.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99,* 432–442.

Appendix A

Description of Nonlinear PCA in Mathematical Terms

In this appendix, the way that nonlinear principal components analysis (PCA) is performed in CATPCA is described mathematically. Suppose we have the $n \times m$ data matrix $\mathbf{H}$, consisting of the observed scores of n persons on m variables. Each variable may be denoted as the jth column of $\mathbf{H}$, $\mathbf{h}_j$, a vector of size $n \times 1$, with $j = 1, \ldots, m$. If the variables $\mathbf{h}_j$ are not of numeric measurement level or are expected to be nonlinearly related to each other, nonlinear transformation of the variables is called for. During the transformation process, each category obtains an optimally scaled value, called a category quantification. Nonlinear PCA can be performed by minimizing a least squares loss function in which the observed data matrix $\mathbf{H}$ is replaced by the $n \times m$ matrix $\mathbf{Q}$, containing the transformed variables $\mathbf{q}_j = \phi_j(\mathbf{h}_j)$. In the matrix $\mathbf{Q}$, the observed scores for the persons are replaced by the category quantifications of the categories a person scored in. The CATPCA model is equal to the linear PCA model, capturing the possible nonlinearity of relationships between variables in the transformations of the variables. We start by explaining how the objective of linear PCA is achieved in CATPCA by minimizing a loss function and then show how this loss function is extended to accommodate weights to deal with missing values, person weights, and multiple nominal transformations. In this appendix, we assume all variable weights to be 1.

The scores of the persons on the principal components obtained by PCA are called component scores (or object scores in CATPCA). PCA attempts to retain the information in the variables as much as possible in the component scores. The component scores, multiplied by a set of optimal weights, called component loadings, should approximate the original data as closely as possible. Usually, in PCA, component loadings and component scores are obtained from a singular value decomposition of the standardized data matrix or an eigenvalue decomposition of the correlation matrix. However, the same results can be obtained through an iterative process in which a least squares loss function is minimized. The loss to be minimized is the loss of information due to representing the variables by a small number of components: In other words, the difference between the variables and the component scores weighted by the component loadings. If $\mathbf{X}$ is considered to be the $n \times p$ matrix of component scores (or object scores), with p the number of components, and if $\mathbf{A}$ is the $m \times p$ matrix of component loadings, with its jth row indicated by $\mathbf{a}_j$, the loss function that can be used in PCA for the minimization of the difference between the original data and the principal components can be expressed as $L(\mathbf{Q}, \mathbf{X}, \mathbf{A}) = n^{-1}\Sigma_j\Sigma_n(q_{ij} - \Sigma_s x_{is}a_{js})^2$. In matrix notation, this function can be written as

$$L(\mathbf{Q}, \mathbf{X}, \mathbf{A}) = n^{-1}\sum_{j=1}^{m} \mathrm{tr}(\mathbf{q}_j - \mathbf{X}\mathbf{a}_j)'(\mathbf{q}_j - \mathbf{X}\mathbf{a}_j), \quad \text{(A1)}$$

where tr denotes the trace function that sums the diagonal elements of a matrix, so that, for example, tr $\mathbf{B}'\mathbf{B} = \Sigma_i\Sigma_j b_{ij}^2$.

It can be proven that Loss Function A1 is equivalent to

$$L_2(\mathbf{Q}, \mathbf{A}, \mathbf{X}) = n^{-1}\sum_{j=1}^{m} \mathrm{tr}(\mathbf{q}_j\mathbf{a}_j' - \mathbf{X})'(\mathbf{q}_j\mathbf{a}_j' - \mathbf{X}) \quad \text{(A2)}$$

(see Gifi, 1990, pp. 167–168, for the deduction of this function, including missing values). Loss Function A2 is used in CATPCA instead of Loss Function A1 because in Loss Function A2, vector representations of variables as well as representations of categories as a set of group points can be incorporated, as is shown shortly.

Loss Function A2 is subjected to a number of restrictions. First, the transformed variables are standardized, so that $\mathbf{q}_j'\mathbf{q}_j = n$. Such a restriction is needed to solve the indeterminacy between $\mathbf{q}_j$ and $\mathbf{a}_j$ in the inner product $\mathbf{q}_j\mathbf{a}_j'$. This normalization implies that $\mathbf{q}_j$ contains z scores and ensures that the component loadings in $\mathbf{a}_j$ are correlations between variables and components. To avoid the trivial solution $\mathbf{A} = 0$ and $\mathbf{X} = 0$, the object scores are restricted by requiring

$$\mathbf{X}'\mathbf{X} = n\mathbf{I}, \tag{A3}$$

with I the identity matrix. We also require that the object scores are centered; thus,

$$\mathbf{1}'\mathbf{X} = \mathbf{0}, \tag{A4}$$

with $\mathbf{1}$ indicating a vector of ones. Restrictions A3 and A4 imply that the columns of $\mathbf{X}$ (the components) are orthonormal z scores: Their mean is zero, their standard deviation is one, and they are uncorrelated. For a numeric analysis level, $\mathbf{q}_j = \phi_j(\mathbf{h}_j)$ implies a linear transformation, that is, the observed variable $\mathbf{h}_j$ is merely transformed to z scores. For nonlinear analysis levels (nominal, ordinal, spline), $\mathbf{q}_j = \phi_j(\mathbf{h}_j)$ denotes a transformation according to the analysis level chosen for variable j.

Loss Function A2 is minimized in an alternating least squares way by cyclically updating one of the three sets of parameters $\mathbf{X}$, $\mathbf{Q}$, and $\mathbf{A}$, while keeping the other two fixed. This iterative process is continued until the improvement in subsequent loss values is below some user-specified small value, called the convergence criterion. In CATPCA, starting values of $\mathbf{X}$ are random.

Loss Function A2 is specified for the simple situation without missing values or the possibility of different person weights. However, weights for missing values and person weights can be easily incorporated into the loss function. To accommodate for the passive treatment of missing values (see Appendix B), a diagonal $n \times n$ matrix $\mathbf{M}_j$ is introduced, with the ith main diagonal entry ii, corresponding to person i, equal to 1 for a nonmissing value and equal to 0 for a missing value for variable j. Thus, for persons with missing values in variable j, the corresponding diagonal elements in $\mathbf{M}_j$ are zero, so that the error matrix premultiplied by $\mathbf{M}_j$, $\mathbf{M}_j(\mathbf{q}_j\mathbf{a}_j' - \mathbf{X})$, contains zeros for the rows corresponding to persons with a missing value on variable j. Therefore, for variable j, the persons with missing values do not contribute to the CATPCA solution, but these same persons do contribute to the solution for the variables for which they have a valid score (this is called passive treatment of missings; see Appendix B). We allow for person weights by weighting the error by a diagonal $n \times n$ matrix $\mathbf{W}$ with nonnegative elements w_{ii}. Usually these person weights, w_{ii}, are all equal to 1, with each person contributing equally to the solution. For some purposes, however, it may be convenient to be able to have different weights for different persons (e.g., replication weights).

Incorporating the missing data weights $\mathbf{M}_j$ and the person weights $\mathbf{W}$, the loss function that is minimized in CATPCA can be expressed as $L_3(\mathbf{Q}, \mathbf{A}, \mathbf{X}) = n^{-1}\sum_{j=1}^{m}\sum_{i=1}^{n} w_{ii}m_{iij} \sum_{s=1}^{p}(q_{ij}a_{js} - x_{is})^2$, or, equivalently, in matrix notation as

$$L_3(\mathbf{Q}, \mathbf{A}, \mathbf{X}) = n_w^{-1}\sum_{j=1}^{m} \operatorname{tr}(\mathbf{q}_j\mathbf{a}_j' - \mathbf{X})'\mathbf{M}_j\mathbf{W}(\mathbf{q}_j\mathbf{a}_j' - \mathbf{X}). \tag{A5}$$

Then, the centering restriction becomes $\mathbf{1}'\mathbf{M}_*\mathbf{W}\mathbf{X} = \mathbf{0}$, with $\mathbf{M}_* = \sum_{j=1}^{m}\mathbf{M}_j$, and the standardization restriction becomes $\mathbf{X}'\mathbf{M}_*\mathbf{W}\mathbf{X} = mn_w\mathbf{I}$.

Loss Function A5 can be used for nominal, ordinal, numeric, and spline transformations, where the category points are restricted to being on a straight line (vector). If categories of a variable are to be represented as group points (using the multiple nominal analysis level)—with the group point in the center of the points of the persons who scored in a particular category—categories will not be on a straight line, but each category will obtain multiple quantifications, one for each of the principal components. In contrast, if the vector representation is used instead of the category point representation, each category obtains one single category quantification, and the variable obtains a different component loading for each component. To incorporate multiple quantifications into the loss function, we reexpress $L_3(\mathbf{Q}, \mathbf{A}, \mathbf{X})$ into a convenient form for introducing multiple nominal variables. Consider, for each variable, an indicator matrix $\mathbf{G}_j$. The number of rows of $\mathbf{G}_j$ equals the number of persons, n, and the number of columns of $\mathbf{G}_j$ equals the number of different categories of variable j. For each person, a column of $\mathbf{G}_j$ contains a 1 if that person scored in a particular category and a 0 if that person did not score in that category. So, every row of $\mathbf{G}_j$ contains exactly one 1, except when missing data are treated passively. In the case of passive missing values, each row of the indicator matrix corresponding to a person with a missing value contains only zeros. In the loss function, the quantified variables $\mathbf{q}_j$ can now be written as $\mathbf{G}_j\mathbf{v}_j$, with $\mathbf{v}_j$ denoting the quantifications for the categories of variable j. Then, the loss function becomes

$$L_3(\mathbf{v}_1, \ldots, \mathbf{v}_m, \mathbf{A}, \mathbf{X}) = n^{-1}\sum_{j=1}^{m} \operatorname{tr}(\mathbf{G}_j\mathbf{v}_j\mathbf{a}_j' - \mathbf{X})'\mathbf{M}_j\mathbf{W}(\mathbf{G}_j\mathbf{v}_j\mathbf{a}_j' - \mathbf{X}). \tag{A6}$$

The matrix $\mathbf{v}_j\mathbf{a}_j'$ contains p-dimensional coordinates that represent the categories on a straight line through the origin, in the direction given by the component loadings $\mathbf{a}_j$. As $\mathbf{q}_j = \mathbf{G}_j\mathbf{v}_j$ for all variables that are not multiple nominal, Loss Function A6 is the same as Loss Function A5.

Appendixes continue

The advantage of Formulation A6 is that multiple nominal transformations can be directly incorporated. If a multiple nominal analysis level is specified, with categoriesrepresented as group points, $\mathbf{v}_j\mathbf{a}_j'$ is replaced by $\mathbf{V}_j$, containing the group points, the centroids of the object points for the persons in p dimensions. Thus, the loss function can be written as

$$L_4(\mathbf{V}_1, \ldots, \mathbf{V}_m, \mathbf{X}) = n^{-1}\sum_{j=1}^{m} \mathrm{tr}(\mathbf{G}_j\mathbf{V}_j - \mathbf{X})'\mathbf{M}_j\mathbf{W}(\mathbf{G}_j\mathbf{V}_j - \mathbf{X}), \tag{A7}$$

where $\mathbf{V}_j$ contains centroid coordinates for variables given a multiple nominal analysis level and $\mathbf{V}_j = \mathbf{v}_j\mathbf{a}_j'$ contains coordinates for the category points located on a vector for the other analysis levels. For more information on these issues and a detailed description of the CATPCA algorithm, we refer the reader to the SPSS Web site (SPSS, 2007).

Appendix B

Missing Data in CATPCA

A reasonable amount of literature provides sophisticated ways of handling missing data in general (see, e.g., Schafer & Graham, 2002). CATPCA provides, in addition to several simple, well-known ways of dealing with this problem (e.g., listwise deletion and simple imputation), two methods worth describing. The first, referred to as passive treatment of missing data, guarantees that a person with a missing value on one variable does not contribute to the solution for that variable but does contribute to the solution for all the other variables. Note that this type of treatment differs from pairwise deletion, in that the latter deletes pairs of values in pairwise computations, whereas passive treatment preserves all information. This treatment of missing data does not use assumptions such as MAR (missing at random) or MCAR (missing completely at random). Passive treatment of missings is possible in nonlinear PCA because its solution is not derived from the correlation matrix (which cannot be computed with missing values) but from the data themselves.

Additionally, CATPCA offers the possibility of treating missing values as an extra category. This option implies that the *missing* category will obtain a quantification that is independent of the analysis level of the variable. For example, the *missing* category of a variable with an ordinal analysis level will obtain an optimal position somewhere among the ordered categories. The greatest advantage of this option is that it enables the researcher to deal with variables that include numeric or ordered categories plus categories like no response, don't know, or not applicable. The option may also be useful if persons omit some questions for a specific reason that distinguishes them from persons who do answer the question. When the *missing* category obtains a quantification that clearly distinguishes it from the other categories, the persons with missing data structurally differ from the others (and this will be reflected in the person scores). If the *missing* category obtains a quantification close to the (weighted) mean of the quantifications, the persons having missing values cannot be considered as a homogeneous group, and treating missing data as an extra category will give approximately the same results as treating missing data as passive.

Received May 5, 2003
Revision received September 25, 2006
Accepted January 3, 2007 ■