Multivariate Behavioral Research, *38*(2), 247-283

Three-Mode Models and Individual Differences in Semantic Differential Data

Takashi Murakami
Graduate School of Education and Human Development
Nagoya University, Japan

Pieter M. Kroonenberg
Department of Education
Leiden University, The Netherlands

This article investigates how individual differences in semantic differential data can be modeled and assessed using three-mode models. Individual differences are important because their existence may affect the generality of conclusions based on such data. An overview is given how individual differences arise and how they can be handled in the analysis. The results of the investigation will be illustrated with semantic differential data on the characterization of Chopin's *Preludes* by a group of Japanese university students.

Introduction

In 1957 Osgood (Osgood, Suci, & Tannenbaum, 1957) introduced the idea that concepts could be universally evaluated by semantic descriptions, like GOOD versus BAD, HARD versus SOFT and ACTIVE versus PASSIVE. He developed so-called *semantic differential scales* which were bipolar seven-point scales with, for instance, GOOD at one end and BAD at the other. Osgood's basic contention was that all concepts could be characterized by three basic (latent) variables, that is, *Evaluation* (E), *Potency* (P), and *Activity* (A), of which the bipolar semantic differential scales function as markers. In the fifties and sixties, Osgood and many researchers across the world conducted massive research with the semantic differential to establish the universality of Osgood's ideas, and they created parallel versions of the

The research for this article was financially supported by the following grants: a grant from the Nissan Fellowship Programme of the Netherlands Organization of Scientific Research (NWO) awarded to the second author, a grant awarded to both authors by the Japanese Society for the Promotion of Science. The generous support of the Graduate School of Education and Human Development, Nagoya University, Nagoya, Japan is also acknowledged.

Correspondence concerning this article should be directed to Pieter M. Kroonenberg, Department of Education, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, The Netherlands; tel.: +31-71-5273446; fax: +31-71-5273945; email: kroonenb@fsw.leidenuniv.nl

fundamental scales in many languages (see Snider & Osgood, 1969 for a comprehensive overview of semantic differential research).

In this article the term "semantic differential" refers very generally to ratings of several concepts on a set of bipolar adjective scales by a sample of subjects, and the concepts and scales are not necessarily the standard ones used in the original articles by Osgood and his coworkers.

Semantic differential data consist of ratings of subjects on a number of bipolar scales for a set of concepts. Such data can be arranged in a three-dimensional block. During the data collection, a subject k produces scores x_{ijk} by evaluating concept i on bipolar scale j, and the scores for subject k can be collected in a matrix $\mathbf{X}_k$ of I concepts by J scales. The data of K subjects thus form a three-way array $\mathbf{X}$ of size $I \times J \times K$. The subjects are seen as judges who express themselves about the relationships between the concepts and scales, and we do not necessarily assume that they are a random sample from a well-defined population. In other words, we will not explicitly concern ourselves with sampling issues.

Individual differences in such data are important because they may contradict the generality of the method (e.g., Osgood, 1964). Several articles have looked at the question of the unidimensionality of subjects (e.g., Snyder & Wiggins, 1970), but a detailed study into the main sources of individual differences and their effects on the "semantic space", the main product of the method, has not yet been conducted. One of the problems is that there are many sources of individual differences and they can appear in various guises in semantic differential data.

Even though the investigation into individual differences in semantic differential data was one of the motivations to develop Three-mode factor analysis (Tucker, 1964, 1966), not many results using this method have so far been reported (for an overview of such studies see the annotated bibliography by Kroonenberg, 1983a). Parallel Factor Analysis (Harshman, 1970; Harshman & Lundy, 1984a, 1984b; see also Carroll & Chang, 1970) is another method for the analysis of three-mode data with similar objectives, but its use to semantic-differential data has been even more limited (see, however, Harshman & De Sarbo, 1984). One of the reasons for the limited use of three-mode methods for analyzing semantic differential data might be that their more complicated structure makes them less straightforward to use. For instance, one must tackle questions such as how the data should be preprocessed, that is centered and/or normalized, how many components[1] should be retained for the three modes, how the components can be represented in the most advantageous manner, which transformations (rotations) should be applied to

[1] In this article we will systematically refer to components rather than factors, because we will not assume a specific error structure for the residuals.

components or to the core array if any, etcetera. Such choices are also influenced by the characteristics of individual differences, and the choices can only be made if one has (conceptual) models of the individual differences in terms of three-mode data. On the other hand, not every type of individual differences that a researcher conceptualizes in semantic differential data can necessarily be treated in such a framework.

The discussion of individual differences will consist of four parts. First, we will characterize sources of individual differences which are related to differences in component scores and component loadings. Secondly, we will discuss three-mode modeling of individual differences in semantic differential data. Thirdly, we will discuss some specific interpretational devices for three-mode analysis which might be less familiar. Finally, we will treat in detail an example of the evaluation of Chopin *Preludes* by Japanese students which will include a discussion of choice of number of components, core transformations, and interpretation in general.

Individual Differences in Scores and Loadings

If all individual differences can be regarded as random variations around a set of values common to all individuals, then their data can be modeled via scores for concepts and loadings for scales derived from the concept by scale table averaged over individuals. Such an analysis makes sense if the number of concepts is sufficiently large, and it has been the standard procedure in most semantic differential studies. However, the literature has shown many results which are inconsistent with this model (see Heise, 1969; and Pinson, 1983, for overviews of general outcomes of semantic differential research).

One may distinguish three major classes of individual differences in semantic differential data, differences in (a) the number of components, (b) component scores, and (c) component loadings. These classes seem to correspond to the "three ways" of individual differences introduced by Snyder and Wiggins (1970): (a) The basic underlying dimensions, E, P, A are not general; (b) different semantic scales function differentially as markers of E, P, A for different individuals; (c) individuals utilize the basic dimensions differently in making scalar judgments across classes of concepts. However, from their analyses it is far from clear how these types of individual differences can be evaluated from real data.

If there are only a few individual differences in the scale loadings, we may postulate a common loading matrix shared by all the subjects. On the other hand, if there are few individual differences in the concept scores we may assume a common score array shared by all the subjects. Models with common components for only one of the modes can be solved by ordinary PCA given a proper normalization.

It is our contention that it is possible to evaluate the several kinds of individual differences simultaneously with three-mode methods, such as three-mode principal component analysis (Kroonenberg, 1983b; Kroonenberg & De Leeuw, 1980; Tucker, 1966) and parallel factor analysis (Harshman, 1970; Harshman & Lundy, 1984a, 1984b). The body of this article is devoted to the presentation of the theoretical basis for this contention, and to providing empirical evidence for the usefulness of this approach.

Individual Differences And Three-Mode Models

Because in general people who share the same language can communicate with each other at a high level of understanding, it seems reasonable to assume that the overall level of individual differences to be found in semantic differential data is not overly large. As a matter of fact, in most studies using three-mode component analysis, the most important component of the subject mode considerably dominates the remaining components (e.g., Levin, 1965; Takeuchi, Kroonenberg, Taya, & Miyano, 1986; Wiggins & Fishbein, 1969). One of the few clear exceptions is the reanalysis by Kroonenberg (1983b, 1985) of the multiple personality case described by Osgood and Luria (1954), but this was a rather unusual case. The dominance of the common usage of the semantic differential scales explains why the ordinary method of analysis using an averaged table works relatively well in many cases.

The implication of the overall consensus is that we have to take into account the possibility that the existence of individual differences may be primarily confined to specific scales and concepts rather than that they are manifest throughout the entire data set. It seems not unreasonable to think that more subjective judgments like UNATTRACTIVE - ATTRACTIVE and UNLIKEABLE - LIKABLE are more likely to be a source of individual differences than more objective judgments like SOFT - HARD. Also the ratings of ambiguous concepts might be a source of individual differences which obviously cannot be modeled.

To investigate these kinds of sources, we will show how different three-mode models represent different kinds of assumptions about individual differences in concept scores and scale loadings. The existence of different models makes it possible to evaluate which individual differences are most prominent, and how they can be described in a parsimonious way.

Three-Mode Models

For models to be classified as three-mode models, they have to have explicit parameters for all three modes. Even though in the illustrative example only the most complete three-mode model, the Tucker3 model, will be used, we will first discuss a somewhat less restricted model to illustrate how individual differences can be modeled.

Tucker2 Model

Suppose that there exists a complete catalogue of components from which each subject makes a selection to judge concepts with a given set of scales. We assume that this catalogue is not overly large in line with our earlier assumptions about the common base of understanding between people. We also assume that the component models of almost all the subjects are constructed by their choice of columns from the lists of concept scores and scale loadings and by the linear combinations of these columns.

This situation can be expressed formally as follows. Let **A** be a matrix of concept scores consisting of columns in the list explained above, and **B** be a matrix of scale loadings constructed similarly, then the data matrix of each individual can be expressed as

$$\mathbf{X}_k = \mathbf{A}\mathbf{H}_k\mathbf{B}' + \mathbf{E}_k \tag{1}$$

where $\mathbf{H}_k$ is the matrix representing the choice and the combinations of columns of **A** and **B**. For example, if data matrix of subject k is approximately recovered by the sum of products of the 2nd column of **A** and the 3rd column of **B**, and 3rd column of **A** and 1st column of **B**, then h_{23k} and h_{31k} are unity and the remaining elements of $\mathbf{H}_k$ are zero.

Of course, this description is an oversimplification. However, this idea will be the basis of a three-mode component analytical model, which makes it possible to describe a given data set parsimoniously, especially when individual differences occur in a systematic way. The model in Equation 1 is formally the same as the Tucker2 model (see Kroonenberg & de Leeuw, 1980). The "2" in Tucker2 indicates that components are present for two of the three ways of the data array.

In the usual formulation of the Tucker2 model, **A** is a I by P columnwise orthonormal matrix, **B** is a J by Q columnwise orthonormal matrix (P not necessarily equal to Q), and a P by Q matrix, $\mathbf{H}_k$, is called a frontal slice of the *extended core array* **H**, which is a three-way array of order $P \times Q \times K$.

Various kinds of individual differences can be modeled via special structures in the extended core array. When there are only individual differences of degree but not in kind, and the standard three-dimensional EPA-structure holds for all subjects, each $\mathbf{H}_k$ is approximately diagonal (Figure 1A). The three columns of the scale matrix **B** will have salient values in the rows corresponding to E, P, and A scales. If two components, for example, E and P, are fused into one component for several but not all subjects, their $\mathbf{H}_k$ are no longer diagonal, and two off-diagonal elements, for example, h_{12k} and h_{21k} will be large (Figure 1B).

In order to accommodate a peculiar usage of a scale, for example, BEAUTIFUL, the number of columns of the scale matrix **B** is increased with a separate component for BEAUTIFUL, and the corresponding element in the evaluation (first) component is diminished. Each frontal slice of the extended core array is no longer square but has order three by four. If a subject k uses the scale BEAUTIFUL in the usual way as a scale of E, the element, h_{14k} is salient as well as h_{11k} (Figure 1C). If another subjects, say k', uses BEAUTIFUL as a P-scale, $h_{24k'}$ rather than $h_{14k'}$ will be salient.

In the case where subjects rate concepts on two entirely unrelated dimensions, say students use the evaluation scales to rate their classmates on either academic achievement or physical ability, the number of columns of **A**, rather than **B**, will be increased. The concept scores for evaluative

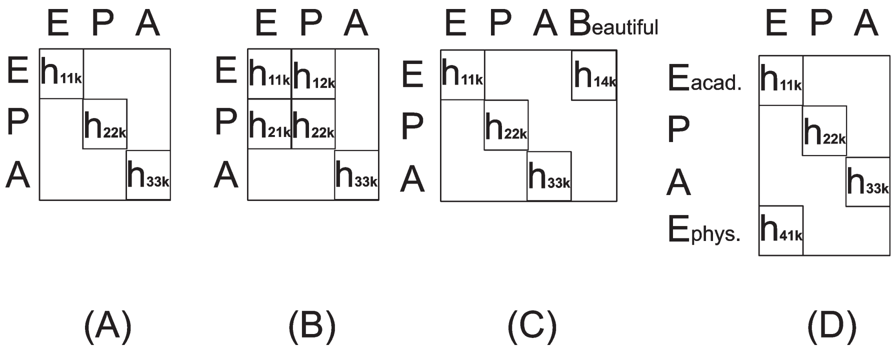

Figure 1
Effect of Individual Differences on the Core Slice of Subject k
(A) Pure EPA-structure; (B) Fusion of E and P; (C) Peculiar usage of BEAUTIFUL; (D) Evaluation rated on Academic ability or Physical ability.

ratings on academic achievement will appear in, say, the first column and the concept scores for evaluative ratings on physical ability in the fourth column (Figure 1D). In the frontal slices for students who rate according to academic achievement, h_{11k}, h_{22k}, and h_{33k} are salient, while those of students who rate according to physical ability, h_{41k} rather than h_{11k} is salient. When several subjects rated concepts on the basic E-dimension in a completely reversed order compared to other subjects, it is not necessary to increase the number of columns. Their value of h_{11k} is simply negative.

In the analysis of real data, there could be more complex combinations of several types of individual differences, and patterns emerging in the extended core array may be far more complicated. However, the basic patterns just described can be helpful in disentangling the more complex ones.

Tucker3 Model

Although the Tucker2 model may be the most comprehensive method for the purpose of description of three-mode data, a more parsimonious model in terms of parameters may be formulated. The Tucker3 model (Tucker, 1966), in which also components are computed for the subjects, can be conceived as a model in which the extended core array is modeled with several more basic core slices, as

(2) $$\mathbf{X}_k = \mathbf{A}\mathbf{H}_k\mathbf{B} + \mathbf{E}_k = \mathbf{A}(\boldsymbol{\Sigma}_r c_{kr}\mathbf{G}_r)\mathbf{B}' + \mathbf{E}_k$$

which may be written in more compact form by the use of Kronecker product as

(3) $$\mathbf{X} = \mathbf{AG}(\mathbf{C}'\otimes\mathbf{B}') + \mathbf{E}$$

where $\mathbf{X}$ is an I by JK matrix obtained by juxtaposition of the $\mathbf{X}_k$, $\mathbf{G}$ is a P by QR core matrix constructed in a similar fashion, $\mathbf{C}$ is an K by R columnwise orthonormal matrix consisting of subject scores, and $\mathbf{E}$ is a I by JK matrix of residuals. In sum notation this equation becomes

(4) $$x_{ijk} = \sum_{p=1}^{P} \sum_{q=1}^{Q} \sum_{r=1}^{R} g_{pqr}\left(a_{ip}b_{jq}c_{kr}\right) + e_{ijk}$$

which shows that the g_{pqr} function as weights for the combination of components. If a particular $g_{pqr} = 0$ then the p^{th}, q^{th} and r^{th} combination of components does not contribute towards the reconstruction or estimation of the data based on the model. By properly normalizing the data and the solution, the scale components can become loadings (variable-component correlations; principal coordinates) and the subject and concept components scores (normalized or standardized components); see below.

The Tucker3 model gives more parsimonious representation of three-way data than the Tucker2 model because it uses fewer parameters, but it seems less straightforward to evaluate some of the individual differences from this model. The cause may be the orthogonality of **C**, subject loading matrix. Because the centering is across concepts and not across subjects, and because the common base of understanding among people mentioned above, the first component has usually positive scores for all subjects. The orthogonality of **C** forces the other components to have negative signs for about half the scores. Although this may correspond well to the case where about the half of subjects rates concepts reversely to the other half on the specified dimension, it may not in the case where there are two or more independent dimensions of the concepts of which a subject agrees with only one of them.

PARAFAC/CANDECOMP *Model*

Both the Tucker2 and Tucker3 models have advantages in analyzing semantic differential data over their competitor the PARAFAC/CANDECOMP model

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}_k\mathbf{B}' + \mathbf{E}_k \tag{5}$$

where $\mathbf{D}_k$ is an S by S diagonal matrix of weights for subject k, **A** the I by S concept scores matrix, and **B**, the J by S scales loading matrix. As the PARAFAC/CANDECOMP model assumes individual differences in all the dimensions, it may have similar interpretations to the counterparts in the Tucker2 model, be it that there are no orthonormality constraints. However, the model will tend to produce degenerate results for many semantic differential data due to the limited individual differences in the subject mode. Even if this is not the case, highly correlated dimensions may occur which could be difficult to interpret. A further drawback is that the number of components in all modes are necessarily the same, while in the Tucker models, one may specify different numbers of components in different modes. Therefore, the latter models allow for discrimination between dimensions with considerable individual differences and those with little individual differences. In addition in the Tucker models, the core array can reveal the different patterns which arise from the individual differences due to the concepts or due to the scores or both, as explained previously. These characteristics seem to tip the advantage towards the Tucker models even if they require somewhat more complicated interpretations due to their core arrays. However, with respect to the latter, we will show that recently new procedures have been proposed to simplify such interpretations. A detailed theoretical and empirical model comparison falls outside the scope of the present article.

Semantic Differential Data and Preprocessing

In this section, we will discuss the preprocessing, that is centering and normalization of semantic differential data, and how this relates to individual differences and three-mode analysis.

In standard principal component analysis for two-mode data, the raw data are almost always analyzed as standard scores, that is the variables are first centered and then normalized by the average square root of the sums of squares, as the eigenvectors are those of the correlation matrix. This straightforward approach to centering and normalization which is done automatically by standard computer packages without interaction with the user, is not possible in three-mode data because there are many more ways to center and to normalize and not all methods have a desirable effect on the outcome of the analysis.

Harshman and Lundy (1984b) present extensive theoretical arguments why certain types of centering and normalizations are "appropriate" and some others are not. Apart from technical arguments in favor and against certain types of centering and normalization, there are also content-specific arguments related to the type of individual differences to be modeled, Luckily the substantive and theoretical arguments do generally not contradict each other. In this article we will concentrate on the content-specific arguments and refer to Harshman and Lundy (1984b; see also Kiers & Van Mechelen, 2001, for a summary) for the technical arguments, but our recommendations fall within the class of "appropriate" procedures for centering and normalization defined by Harshman and Lundy.

Centering

In two-way data, the standard practice is to center each variable so that the original scores are transformed into deviation scores per variable by subtracting the variable mean

(6) $$z_{ij} = x_{ij} - \bar{x}_{\bullet j}.$$

Centering of three-way data also consists of subtracting means from the original data, but due to the three-way nature, several different types of centerings are possible.

Centering Bipolar Scales. In semantic differential research, the choice of centering primarily determined by the nature of the scales which are bipolar and have opposite adjectives as end points. Which particular adjective is placed at which end of the scale is entirely arbitrary. For example for the seven-point BEAUTIFUL - UGLY scale, there is no particular reason to code the response "very beautiful" as 7 and "very ugly" as 1 or the other way around. Because of this, one cannot center across scales because in that case the direction of the scale matters. The mean value for a concept averaged across scales is dependent on the orientation of the scales, and as the orientation is arbitrary the concept means cannot be uniquely determined, so that centering using concept means is not meaningful. A more formal way to express this is that centering for semantic differential data can only be based on a two-factorial ANOVA decomposition per scale, that is

(7) $$x_{ik}^{j} = \bar{x}_{\bullet\bullet}^{j} + \left(\bar{x}_{\bullet k}^{j} - \bar{x}_{\bullet\bullet}^{j}\right) + \left(\bar{x}_{i\bullet}^{j} - \bar{x}_{\bullet\bullet}^{j}\right) + \left(x_{ik}^{j} - \bar{x}_{i\bullet}^{j} - \bar{x}_{\bullet k}^{j} + \bar{x}_{\bullet\bullet}^{j}\right)$$

where the scales index j is written as a superscript to indicate that all means are conditional on j. The first term in Equation 7 $\bar{x}_{\bullet\bullet}^{j}$ represents the mean of scale j taken over all measurements of the scale, the second term ($\bar{x}_{\bullet k}^{j} - \bar{x}_{\bullet\bullet}^{j}$) represents the deviations of subject k's scale mean compared to the overall scale mean, and the third term, ($\bar{x}_{i\bullet}^{j} - \bar{x}_{\bullet\bullet}^{j}$), represents the deviation of the mean rating of concept i across individuals compared to the overall mean of scale j. As the latter does not involve individual differences, nothing is lost from the individual differences point if this term is not included in an analysis.

Two centering options are open to the researcher, that is the three-mode analysis is carried out on $z_{ijk} = x_{ik}^{j} - \bar{x}_{\bullet\bullet}^{j}$ or $z_{ijk} = x_{ik}^{j} - \bar{x}_{\bullet k}^{j}$. The first option has the disadvantage that this type of centering does not center the components of the scales, while the second type does (as explained in detail by Harshman & Lundy, 1984b). The second option is therefore preferred and the three-mode analysis will thus typically be applied to

(8) $$z_{ijk} = x_{ik}^{j} - \bar{x}_{\bullet k}^{j} = \left(\bar{x}_{i\bullet}^{j} - \bar{x}_{\bullet\bullet}^{j}\right) + \left(x_{ik}^{j} - \bar{x}_{i\bullet}^{j} - \bar{x}_{\bullet k}^{j} + \bar{x}_{\bullet\bullet}^{j}\right).$$

In their application, Harshman and De Sarbo (1984, p. 606) used double centering, including centering across scales but, as argued above, in general this will not be advisable.

The second term in Equation 8, $\left(x_{ik}^{j} - \bar{x}_{i\bullet}^{j} - \bar{x}_{\bullet k}^{j} + \bar{x}_{\bullet\bullet}^{j}\right)$, indicates the subject by concept interaction for each scale, and describes the relative differences between the individuals in the concept usage for each scale. If it is sufficiently small, the equation is almost equal to Equation 6, and the results will closely resemble the results of a two-mode analysis of the ratings averaged over all individuals. Therefore, it is the size of this term, which is crucial for the meaningful application of three-mode analyses to semantic differential data.

Normalization

There are three options for normalization in semantic differential data: (a) equalizing the sum of squares of the scales for each subject k, that is dividing by

$$s_{jk} = \sqrt{\sum_i z_{ijk}^2} \;;$$

(b) equalizing the sums of squares of scales across all measurements of the scale j, that is dividing by

$$s_j = \sqrt{\sum_{ik} z_{ijk}^2} \;;$$

and (c) equalizing the sums of squares of the subjects, that is dividing each subject's data by the square root of its total sum of squares,

$$s_k = \sqrt{\sum_{ij} z_{ijk}^2} \;.$$

After normalization, the data can have the following properties: (a) Differences in scale usage of subjects are equalized, in the sense that subjects who use the extremes of the scales and subjects who only score in the middle of the scales have the same sum of squares. Generally, the first component of the subjects then correlates highly with the model's goodness of fit to the subjects' data. However, fit comparisons can always be carried out in three-mode models, irrespective of the normalization, so that there is no particular reason to normalize to achieve this end; (b) Given that the proper centering has been carried out, the elements of the component matrix of the scales, **B,** can be interpreted as correlations; (c) In cases where all scores on a scale are equal, their variance is zero, and normalized scores are undefined.

In the first option, the columns of each data matrix $\mathbf{Z}_k$ are normalized, that is $z_{ijk}^{*} = z_{ijk}/s_{jk}$. When this option is carried out all three properties of normalization, mentioned above do or may occur. This normalization is especially sensitive to property (c) as subjects frequently use the same scale value for each concept. Moreover, the procedure is sensitive to outliers, which occur when all concepts are rated (more or less) the same and a single concept is rated very differently. In addition for each subject, standard deviation units will have different sizes with respect to the original scale. These properties makes the first option an unattractive choice.

The second option, normalization within scales, that is $z^*_{ijk} = z_{ijk}/s_j$, only has the, desirable, property (b). In addition, it has the property that differences in sizes of variances between scales do not affect the results of the analysis. However, if a small variance in a scale only reflects the unreliability of judgments, it may be harmful to enlarge it through this normalization, as the procedure increases the random component in the variation. The advantage over the first option is that differences in variability in scale usage by the subjects and their influence on the ratings remain in the analysis, and property (c) cannot occur (variables without variance will be eliminated from the outset)

In the third option, the sums of squares of the complete concept by scale matrix produced by each subject, that is $z^*_{ijk} = z_{ijk}/s_k$ are equalized. This is, for instance, standard practice in individual differences multidimensional scaling (Carroll & Chang, 1970) where subjects are also treated as judges. This option only has property (a) and thus eliminates to a certain extent response bias in extremity of scale usage. An argument against this option is that the variability of the scores of the subjects is related to their judgments of the relationships between concepts and scales and therefore should be analyzed together with the rest of the data. Moreover, this normalization does not have property (b).

Harshman & Lundy (1984b, pp. 247-248; see also Harshman & De Sarbo, 1984, p. 606) introduce the fourth option of simultaneous normalization within both subjects and scales, that is combining options 2 and 3. This approach needs an iterative algorithm, however the process is guaranteed to converge (see also ten Berge, 1989). This option allows one to benefit from the advantages of the separate normalizations, however, it does not take away any of the objections against any of the options. Given that there has been very little experience with this procedure (see, however, the application by Harshman & De Sarbo, 1984) and that one eliminates a large amount of interesting variability from the data, it is difficult to recommend the procedure at present.

Recommendations

On the basis of the previous discussion, it is recommended that semantic differential data should be centered across concepts for each scale-subject-combination (*jk*), that is the $\bar{x}_{\bullet jk}$ are removed from the raw data before three-mode analysis, and that the so centered data are either not normalized when differences in variability between scales is meaningful and interesting, or are normalized with the square root of the scale sum of squares s_j when these differences are not of interest.

Intepretational Devices

In this section, we will discuss briefly three devices which will assist in selecting models and in interpreting the outcomes of three-mode component analyses, in particular model-fit plots, joint biplots, and component and core transformations. Special attention is paid to these aspects because their use in three-mode analysis is either relatively unknown or has only recently been introduced.

Model Fit

Timmerman and Kiers (2000) suggested a model-selection procedure analogous to Cattell's scree plot for two-mode component analysis. In particular, they based the selection on choosing the model with the highest proportion fitted sum of squares, V_S, within the class of models with the same sum of numbers of components ($S = P + Q + R$). To compare classes with different S, they computed $dif_S = V_S - V_{S-1}$. Only those dif_S are considered which are sequentially highest. Timmerman and Kiers defined a *salience value*: $b_S = dif_S / dif_{S*}$, where dif_{S*} has the next highest value after dif_S. They proposed to select the model for which b_S has the highest value, and they call this the *DifFit-criterion*. Finally, the authors define a lower bound for dif_S to be taken into account. The dif_S should be greater than the average proportion explained variability taken over all feasible values of $S[s_{min} = \min(I, JK) + \min(J, IK) + \min(K, IJ) - 3]$ (In Timmerman & Kiers, 2000, inadvertently, the word "max" is printed instead of "min".) Timmerman and Kiers' procedure comes down to creating a three-mode version of Cattell's scree plot in which the residual sums of squares of each model is plotted versus the sum of numbers of components S (*Three-mode scree plot*). An alternative to their approach is constructing a deviance plot with the residual sum of squares of each model plotted versus degrees of freedom *df* (*Deviance plot*). In both plots, a convex hull can be drawn to connect favored models. The general idea is that models clearly within the hull are disfavored compared to the "hull models" who have similar or better fit and equal *df* (or *S*), equal fit with more *df* (or smaller *S*), or a combination of both. The procedure described by Timmerman and Kiers is essentially a way to define the convex hull in the Three-mode scree plot and to outline the procedure of choosing a model on the convex hull.

Joint Biplots

After the components have been computed, the core array provides the information about the relationships between these components. It is very instructive to investigate the component loadings of the concepts jointly with those of the scales, by projecting them together in one space, as it then becomes possible to evaluate the interaction between concepts and scales. The plot of the space common to scales and concepts is now generally referred to as a *joint biplot*.

Joint biplots can be constructed between one pair of component matrices, say **A** and **B**, for each component *r* of the third (or reference) mode, say **C**. They are constructed such that the columns $\mathbf{a}_p$ of **A,** and the columns $\mathbf{b}_q$, of **B** are close to each other in the joint biplot, where closeness is measured as the sum of all $P \times Q$ squared distances $d^2(\mathbf{a}_p, \mathbf{b}_q)$ over all *p* and *q*.

The plots are constructed as follows (Kroonenberg, 1983b, chap. 6.10). For each component *r* of the reference mode **C**, the components **A** and **B** are adjusted by dividing the core slice, $\mathbf{G}_r$, associated with that component between them. This is done using the singular value decomposition, and **A** and **B** are further adjusted by the relative number of elements in the modes to make the distances as comparable as possible.

$$(9) \quad \mathbf{D}_r = \mathbf{A}\mathbf{C}_r\mathbf{B}' = \mathbf{A}(\mathbf{U}_r\mathbf{\Lambda}_r\mathbf{V}_r')\mathbf{B}' =$$

$$\left[\left(\frac{I}{J}\right)^{1/4}\mathbf{A}\mathbf{U}_r\mathbf{\Lambda}_r^{1/2}\right]\left[\left(\frac{J}{I}\right)^{1/4}\mathbf{B}\mathbf{V}_r\mathbf{\Lambda}_r^{1/2}\right]' = \mathbf{A}_r^*\mathbf{B}_r^{*\prime}$$

with

$$(10) \qquad \mathbf{A}_r^* = \left(\frac{I}{J}\right)^{1/4}\mathbf{A}\mathbf{U}_r\mathbf{\Lambda}_r^{1/2} \text{ and } \mathbf{B}_r^* = \left(\frac{J}{I}\right)^{1/4}\mathbf{B}\mathbf{V}_r\mathbf{\Lambda}_r^{1/2}$$

When $\mathbf{G}_r$ is not square there are only $l = \min(P, Q)$ dimensions (or non-zero singular values) available. The procedure can be interpreted as rotating the component matrices by an orthonormal matrix, followed by a stretching (or shrinking) of the rotated components. There will be as many joint biplots as there are components for the *C*-mode, that is *R*. The interpretation of these plots is analogous to standard biplots with respect to the relations between the elements of the *A*-mode and those of the *B*-mode. However, an additional complication is that these relationships have to be interpreted in relation to the interpretation of the *C*-mode. In the present example, the subjects all have positive scores on the first subject component so that the pattern to be described by the concepts, the scales, and the first core slice $\mathbf{G}_1$ is shared by all subjects. The second subject component shows a contrast between two groups of subjects, so that the relationships described in the second joint biplot using $\mathbf{G}_2$ is characteristic for the group with positive scores on the second subject component, while the reverse is true for the group of subjects with negative scores on the second subject components (for further details, see the example).

Component and Core Transformations

Once a basic three-mode analysis has been carried out, transformations or rotations may be applied to the components or the core array to enhance interpretability. If the components of a mode have been rotated then the core array has to be counter-rotated in an appropriate way (and vice versa), to allow for a coherent interpretation of the results. Arguments for rotating components follow directly from those for two-mode analysis and will not be repeated here. In three-mode analysis often orthonormal components are rotated, which is equivalent to Harris and Kaiser's (1964) independent cluster rotation. The rationale for rotating the core array stems from the desire to create a very simple core such that instead of $P \times Q \times R$ combinations of components only a much smaller number have to be interpreted. Procedures to simplify core arrays have been described by Kiers (1997, 1998) and Henrion and Andersson (1999). Murakami, ten Berge, and Kiers (1998) describe conditions under which an extremely simple core array can be created via orthonormal rotations on the core array (called an orthomax core rotation). The latter rotational procedure was used in the example. It should be noted that very simple core arrays may lead to uninterpretable results due to the creation of highly correlated components. Component rotations used in connection with joint biplots are only really useful when the rotations are applied to the reference mode **C**, because prior rotations of **A** and **B** will be lost in the construction of the joint biplot.

Software

All analyses presented in this article were carried out with 3WayPack, a suite of three-mode programs developed by the second author. Information about this package can be obtained from the website: *http://three-mode.leidenuniv.nl/*. On the same website in the Data Set section the data analyzed in this article have been made available for re-analysis. A set of Matlab-routines for carrying out three-mode analysis was developed by Rasmus Bro and Claus Andersson, and is available as the *N*-way toolbox from their website: *http://www.models.kvl.dk/source/nwaytoolbox*. A copy of the MatLab program is necessary to run the programs of the toolbox. The site also contains an interactive tutorial on using the package for three-mode principal component analysis.

Example: Chopin Data

In the following, applied part of this article, we will demonstrate that in real data there really exist several types of individual differences. We will show how a three-mode analysis of semantic differential data may be conducted using, amongst others, several measures of fit to assess the adequacy of solutions and core transformations procedure to facilitate interpretation.

Music appreciation and the evaluation of the characteristics of musical pieces have in the past been frequently researched with semantic differential scales (see e.g., Nielzen & Cesarec, 1981; Swanwick, 1973). Following this line of research, the study used as an example of the analysis of individual differences in semantic differentials deals with judgments by Japanese students of Chopin's *Preludes*.

Method

Subjects

Thirty-eight Japanese university students (21 males and 17 females) participated in the experiment. For students to be eligible for the study they had to be familiar with classical music so that they could be expected to provide appropriate judgments about the music.

Concepts

The concepts were the 24 short piano solo pieces (or preludes) making up Op. 28 composed by Frederic Chopin, played by Samson François on Angel, AA-8047. They were copied on a cassette tape, and edited for the experiment. To avoid the experiment to be too long, of some longer preludes (No. 2, 4, 6, 8, 13, 15, 17, 21) only the first one minute and twenty seconds were presented.

Scales[2]

Eighteen pairs of adjectives were selected from previous research on semantic differential studies of music. Also included were expressions often used to describe Chopin's music in the literature. It was expected that the students would judge rather objectively on all but two scales on the basis of key (mainly, major or minor) and tempo of each prelude. In contrast, it was expected that they would express subjective evaluations on the scales, UNATTRACTIVE - ATTRACTIVE and UNINTERESTING - INTERESTING. The set of scales employed here is not necessarily a typical one used in semantic differential studies, where in general the scales can be divided into three categories, Evaluation, Activity, and Potency (Osgood, Suci, & Tannenbaum, 1957). Before the final analysis, the codings of some scales were reversed to induce as many positive correlations among them as possible.

Procedure

The experiment was conducted individually for each subject in a sound-proof room. Before the beginning of the session, subjects were given the opening part of the first movement of Mozart's *13th Piano Sonata* (K. 331) in order to practice rating piano music. All preludes were presented to the subjects in the same order as they were arranged by the composer. A monaural speaker was used, and the subjects were asked to rate each prelude on 20 scales immediately after the music stopped. Since the total duration of the music was about 26 minutes and most subjects finished their ratings for each prelude in one minute, the total time of the experiment was generally about one hour.

---

[2] Bipolar scales are always referred to in such a way that the low-end marker will be mentioned first and the high-end marker last.

Analysis Methods

Individual Differences

In the theoretical section, several sources of individual differences were discussed. For the Chopin data, we eliminated response styles and modeled the remaining individual differences with the Tucker3 model. This implies that we assumed that the subjects had a common scale space, a common prelude space, but that they potentially differed in the way they judged the relationships between the common scale and prelude spaces. In addition, we assumed that differences between subjects were sufficiently systematic that they could be modeled as a linear combination of a limited number of subject components, or types of subjects. By allowing subjects to have zero weights on the components, we accommodated the situation that they had less components than other subjects.

Preprocessing

The multivariate two-factorial ANOVA design (Equation 7) was used to base the centering on and the second option was applied for the normalization, that is the three-mode analysis is based on the centered and normalized data z_{ijk} with:

(11) $$z_{ijk} = \left(x_{ijk} - \bar{x}_{\bullet jk}\right) / s_j$$

where

$$s_j = \sqrt{\sum_{ik}\left(x_{ijk} - \bar{x}_{\bullet jk}\right)^2}.$$

Number of Components

The numbers of components for each of the three modes were determined empirically. Via model comparison based on the relative fit of the model to the data and via an inspection of the interpretability of the results, an adequate model was selected. This selection was based on several measures of fit such as overall fit, fit per component of each mode and whether all scales were more or less adequately part of the model. For comparisons of the overall fit, both Timmerman and Kiers' (2000) DifFit measure was inspected via a three-mode scree plot and the deviance-*df* ratio was considered via a deviance plot. For the preferred solution, the residuals were inspected for both the levels of the modes and for the individual data points.

Scaling of Parameter Matrices

To improve the interpretability of the solution we exploited both the transformational and scaling freedom of the solution of the three-mode model. In particular, the solution was scaled such that both the concept and subject components had unit lengths (i.e. $\mathbf{A'A} = \mathbf{I}$ and $\mathbf{C'C} = \mathbf{I}$). The core slices corresponding to the scale components were normalized, that is

$$\sum_{r=1}^{R} \mathbf{G}_r'\mathbf{G}_r = \mathbf{I}_Q .$$

As a result, the sums of squares of elements of the scale components (**B**) corresponded to the size of explained variances of the data by the model. In addition, this transformation makes the matrix **B** into the loading matrix of variables, that is each element of **B** is the correlation between the observed scores and the scores predicted by the model using all combinations of concepts and subjects.

Rotation of Parameter Matrices

Initially, we rotated the matrix of loadings for scale-mode with Kaiser's (1958) varimax method, and counterrotated the core array accordingly. The remaining two modes were rotated via Kiers' (1997) orthomax core rotation procedure. Even though this produced readily interpretable results, an even simpler result could be obtained by using the orthomax core rotation procedure in a different way.

By applying the orthomax core rotation for all three modes simultaneously, a very simple core array could be obtained (see Table 7). This astonishingly simple core was possible because the numbers of components are such that $PR - 1 = Q$, and the core was normalized as above (see Murakami et al., 1998). Since the rotated loadings and scores obtained by using the simultaneous core rotation procedure are very similar to the first approach, we will report only the results with the extremely simple core.

Joint Biplots

Apart from presenting the solution numerically, joint biplots for each of the components of the subject mode were constructed so that for each 'subject type' the joint relationships between scales and preludes could directly be inspected rather than only via the components.

Results

First, we will present the two-factorial MANOVA table to get a general idea of the extent of the individual differences in the data, followed by the selection of an adequate Tucker3 model plus the results of the analysis of residuals, then we will present the rotated components for each of modes, as well as the core array. Next, the joint biplots are presented for each subject component.

Size of Individual Differences

The first task in any analysis dealing with individual differences is to establish whether such individual differences exist at all. Towards this end, the multivariate two-factorial analysis of variance with a single observation per cell was carried out for the Chopin data (Table 1).

The conclusions from this multivariate two-factorial analysis of variance were that the subjects' differences were comparatively small (10% of the total sum of squares), but the deviations of the preludes means from their overall mean were rather different (41% of the total sum of squares). Furthermore, the students effectively used the scales to distinguish between the preludes as the Subjects × Preludes interaction sum of squares accounted for 49% of the total sum of squares. The associated interaction mean square

Table 1
Multivariate Two-Factorial Analysis of Variance of Chopin Data

Source	SS	%	*df*	MS	F
Main effects					
Subjects	2649	10	740	3.58	4.4
Preludes	11329	41	460	24.62	30.4
Two-way interaction					
Subjects × Preludes	13768	49	17020	.81	
Total Sum of Squares	27747	100			

is very much smaller, but it should be realized that such large interactions generally contain systematic variability associated with only few degrees of freedom while most of the degrees of freedom are associated with (random) errors.

Table 2 shows the partitioning of the sum of squares of each scale according to a two-factorial ANOVA on that scale, where the two-way interaction term is at the same time the residual sum of squares due to the

Table 2
Partitioning of the Sum of Squares of Each Scale into its Constituent Parts as defined by Equation 7

Nr.	Scales	Proportional $SS(Preludes)_j$	Proportional $SS(Subject)_j$	Proportional $SS(Residuals)_j$	Mean Sum of $Squares_j$
2	FAST	.69	.06	.26	2.34
5	HEAVY	.61	.05	.34	2.44
4	SEVERE	.52	.04	.44	1.57
14	RESTLESS	.51	.08	.41	1.84
1	DARK	.49	.05	.46	1.57
18	GLOOMY	.46	.07	.47	1.41
9	VEHEMENT	.45	.08	.47	1.68
17	NOISY	.44	.09	.48	1.26
13	SAD	.42	.07	.51	1.49
19	DRAMATIC	.42	.08	.50	2.17
7	STRONG	.41	.08	.51	1.31
12	LOUD	.40	.11	.49	1.42
	Overall	**.40**	**.10**	**.50**	**1.52**
6	CLOUDY	.36	.12	.52	1.48
11	HARD	.33	.09	.58	1.33
8	COLD	.27	.14	.59	1.04
10	LARGE	.22	.11	.68	1.16
15	COARSE	.21	.25	.55	1.23
16	THICK	.20	.15	.65	1.09
20	ATTRACTIVE	.10	.16	**.75**	1.38
3	INTERESTING	.07	.19	**.74**	1.21

Note. $Prop.SS(Preludes)_j + Prop.SS(Subjects)_j + Prop.SS(Residuals)_j = 1.00$

single observation per cell (see Equation 7). The major factor contributing to the variability of the scales is that of the preludes with differences due to subjects only contributing around 10%. Furthermore, it becomes evident that in particular the evaluative scales UNINTERESTING - INTERESTING and UNATTRACTIVE - ATTRACTIVE did not fit as well as the other scales.

Model Selection

Table 3 provides an overview of all models with less than or equal to three components in any mode. Applying the Timmerman and Kiers' (2000) DifFit criterion, indicated a model with 2 components for preludes, 2 for scales and 1 component for subjects. However, such a model only allowed for individual differences in degree, that is the size of the coefficient for a

Table 3
Summary of Analyses: Overall Fitted and Residual Sum of Squares

Model Size $P \times Q \times R$	Sum of Components	SS(Res) = Deviance	*df*	Prop. SS(Fit)	Difference in Prop.Fit	DifFit Ratio
1 × 1 × 1	3*	14.21	17380	.289	**.289**	**2.44**
1 × 2 × 2	5	13.96	17325	.302		
2 × 1 × 2	5	13.82	17321	.309		
1 × 3 × 3	7	13.82	17272	.309		
3 × 1 × 3	7	13.47	17264	.327		
2 × 2 × 1	5*	11.84	17339	.408	**.119**	**10.11**
2 × 2 × 2	6*	11.61	17300	.420	.012	1.01
3 × 3 × 1	7	11.55	17300	.422		
2 × 3 × 2	**7**	**11.47**	**17281**	**.426**		
2 × 2 × 3	7	11.47	17263	.426		
3 × 2 × 2	7*	11.44	17277	.428	.009	-
2 × 3 × 3	8	11.28	17242	.436		
3 × 3 × 2	8	11.28	17256	.436		
3 × 2 × 3	8*	11.23	17238	.439	.011	-
3 × 3 × 3	9*	11.00	17214	.450	.012	-

Note. An asterisk in the second column indicates that this model is the best model in the class of models with the same *S* (= Sum of number of components).

subject on the first subject component. As we were especially interested in more subtle individual differences and the first component largely reflected the size of the variances of the subjects (r = 0.89), it was decided to go beyond the simple model and search for an adequately fitting model with more components which also included information on further individual differences.

The deviance plot (Figure 2) indicated that both the 3 × 2 × 2 and 2 × 3 × 2 models were possible candidates on the basis of their deviance-*df* ratio and the three-mode scree plot (Figure 3) showed that both models were close to the convex hull of best models. Inspection of the fit of the scales showed that Interesting and Attractive, the only scales which measured a personal opinion with respect to the music, did not fit in the 3 × 2 × 2-solution (fit = 0.039 and 0.029, compared to 0.426 overall), while these scales had some fit in the 2 × 3 × 2-solution (fit = 0.131 and 0.140 compared to 0.426 overall). On this heuristic basis, the 2 (subjects) × 3 (scales) × 2 (preludes) solution was chosen.

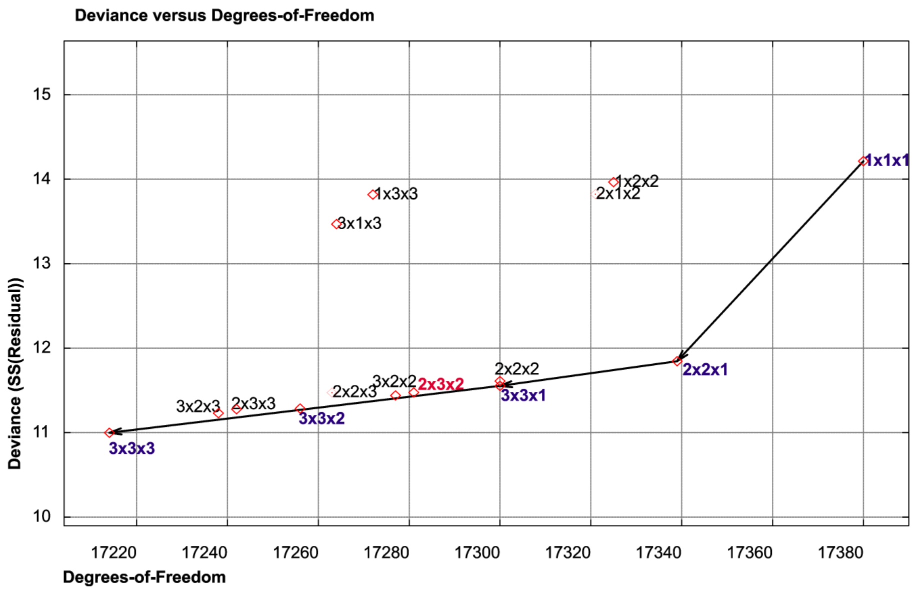

Figure 2
Deviance Plot Showing the Residual Sums of Squares versus the Degrees of Freedom Model 2 × 3 × 2 is the model selected.

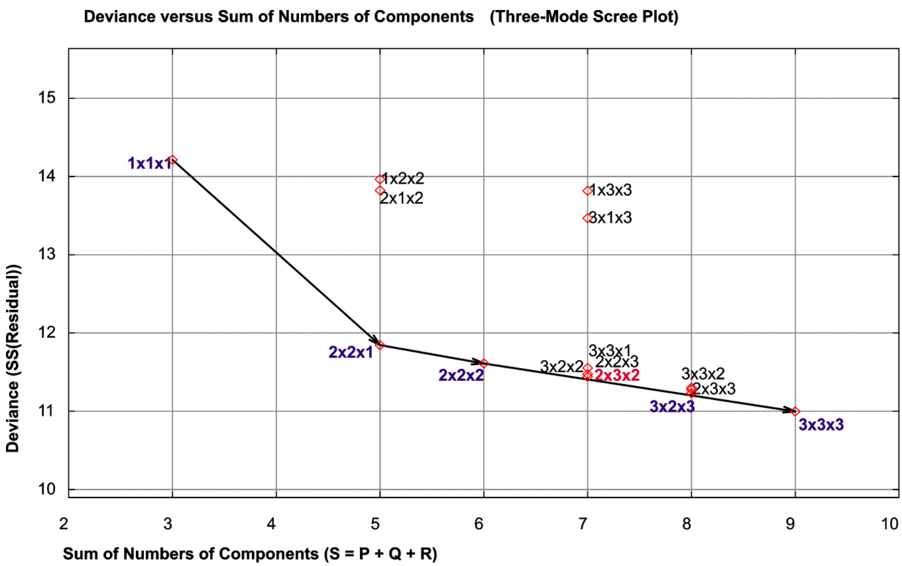

Figure 3
Three-Mode Scree Plot Showing the Residual Sums of Squares versus the Total Number of Components
Model 2 × 3 × 2 is the model selected.

To check the stability of the 2 × 3 × 2-solution a bootstrap analysis with 100 replications was carried out. The mean fit of the solutions was 0.438 with a standard error of the mean of 0.013, while the original fit value was .426 or a standardized value of -0.903 in the sampling distribution of the mean, showing that there is good stability for the fit. Inspecting a plot of the scale space with all 100 bootstrap solutions included, showed that the third scale component was well defined and the two scales expressing the students' personal opinions about the preludes were well separated from the other scales and from each other as was evident from their confidence ellipses.

Interpretation of the 2 × 3 × 2-Solution

The three-mode orthomax transformation of the core array was applied as described above, and the component matrices were inversely transformed. The preludes are presented as unit-length components (Table 6), and the scales (Table 5) are presented as an oblique pattern matrix. Owing to many zeroes in the core array and the rather simple structures of the scales and preludes components, the interpretation was relatively clear, and interactions of three modes could be fairly easily interpreted. We will start with the interpretations of the components for each mode.

Subject Mode

The component scores of the 38 subjects are shown in Table 4, along with the variances of their responses and the proportions of their total sums of squares which were not fitted by the model, $SS(Res_k)/SS(Total_k)$. All individuals had positive scores on the first component, and the correlation coefficient of the scores with the total sum of squares was 0.89, indicating that the individual differences on the first component are associated with the extent of their scale usage. Subjects with large values (highest = 0.24) on the first component differentiate clearly between preludes on the scales, while subjects with small scores (lowest = 0.07) differentiate much less and use the values around their averages on the scales. However, the differences between the subjects on the first component are not very large. About half the scores on the second component are negative. As will be shown below, this component could be interpreted as denoting individual differences in preference for certain types of music.

The third column of Table 4 shows the extent to which the model succeeded in fitting the data of the individual students. The largest proportional residual fitted sum of squares was 0.77 indicating that only 23% of the data of subject 27 were fitted by the model, while the model succeeded in fitting 58% for the best fitting subject. However, overall the model fitted most subjects in a comparable manner.

Scale Mode

The loadings for the scales is shown in Table 5. The scales, CALM - RESTLESS, GENTLE - SEVERE, QUIET - NOISY, LYRIC - DRAMATIC, TRANQUIL - VEHEMENT, WEAK - STRONG, SLOW - FAST, and STILL - LOUD, had salient loadings on the first component, while LIGHT - HEAVY, CHEERFUL - GLOOMY, BRIGHT - DARK, SOFT - HARD, HAPPY - SAD, and CLEAR - CLOUDY, had salient loadings on the second component. The scale, SLOW - FAST, had high loadings on two components. The first component could be identified as the bipolar dimension of 'SOOTHING VS. AGITATING' while the second one as 'CHEERFUL versus MOURNFUL'. These interpretations were strengthened by interpreting them in combination with the dimensions of concept-mode (see below). The

Table 4
Subjects: Total Sums of Squares, Proportional Residual Sums of Squares and Component Scores, which were Counter-Rotated with Inverse of Rotation Matrix of Core Array (Sorted on Second Component)

Subject	SS(Total)	SS(Res)/ SS(Total)	Component 1 'Consensus'	Component 2 'Contrast'
Range min-max	0.16-1.06	0.42-0.77	0.07-0.24	-0.44-0.37
19	.839	.528	.209	-.441
27	.502	.766	.112	-.257
30	.595	.499	.190	-.235
10	.394	.444	.162	-.213
18	.235	.618	.101	-.190
12	.723	.542	.202	-.163
35	.558	.460	.193	-.152
11	.430	.447	.171	-.147
20	.344	.533	.140	-.131
3	.680	.480	.209	-.128
16	.357	.568	.138	-.062
24	.448	.631	.143	-.049
17	.602	.483	.196	-.047
6	.415	.582	.147	-.034
22	.337	.669	.118	-.033
2	.540	.607	.162	-.022
5	.510	.438	.188	-.020
23	.159	.765	.068	-.013
32	.954	.674	.196	-.008
28	.494	.497	.175	.012
4	.364	.643	.126	.030
29	.162	.679	.079	.037
34	.754	.633	.183	.041
37	.366	.717	.111	.064
31	.701	.424	.221	.073
26	1.059	.562	.235	.103
7	.654	.570	.182	.104
9	.352	.596	.128	.104
33	.358	.595	.129	.110
38	.872	.521	.222	.136
21	.375	.645	.121	.141
13	.390	.704	.111	.158
8	.476	.609	.144	.163
14	.365	.779	.088	.181
25	.489	.687	.127	.191
1	.577	.546	.169	.214
15	.590	.612	.149	.293
36	.979	.558	.208	.371

Table 5
Scale Components Counter-Rotated with Inverse of Rotation Matrix of Core Array

Scale	1 'SOOTHING - AGITATING'	2 'CHEERFUL - MOURNFUL'	3 'DISLIKE - LIKE'	SS(Res)/ SS(Data)
CALM - RESTLESS	**0.765**	0.124	0.101	.390
GENTLE - SEVERE	**0.680**	0.369	0.072	.403
QUIET - NOISY	**0.660**	-0.288	0.053	.479
LYRIC - DRAMATIC	**0.650**	0.155	-0.026	.553
TRANQUIL - VEHEMENT	**0.652**	0.294	0.083	.482
WEAK - STRONG	**0.625**	0.184	0.104	.564
SLOW - FAST	**0.606**	**-0.558**	0.058	.319
STILL - LOUD	**0.572**	0.390	0.083	.513
LIGHT - HEAVY	0.141	**0.790**	-0.063	.353
CHEERFUL - GLOOMY	0.232	**0.671**	-0.073	.491
BRIGHT - DARK	0.234	**0.669**	-0.052	.497
SOFT - HARD	-0.034	**0.617**	-0.059	.615
HAPPY - SAD	0.237	**0.604**	-0.052	.576
CLEAR - CLOUDY	0.346	**0.545**	-0.150	.561
WARM - COLD	**0.411**	0.383	0.020	.684
SMALL - LARGE	0.318	0.342	0.040	.780
DELICATE - COARSE	0.333	0.354	-0.087	.756
THIN - THICK	0.372	0.336	-0.146	.728
UNATTRACTIVE - ATTRACTIVE	-0.063	-0.132	**0.344**	.860
UNINTERESTING - INTERESTING	0.073	-0.189	**0.300**	.869

Note. Values greater than 0.40 have been set in bold, except in the last component, where the largest elements have been set in bold. Because the rotation was applied to the core array, the simple structure in not optimal.

two scales, UNINTERESTING-INTERESTING and UNATTRACTIVE - ATTRACTIVE had the largest loadings on the third component. This dimension was interpreted as the expression of 'DISLIKE versus LIKE' by subjects, and it had also non-negligible loadings for the scales THIN-THICK and CLEAR - CLOUDY. As will be detailed below, this dimension was related to certain specific preludes.

The last column of Table 5 shows the extent to which the model succeeded in fitting the data of the various scales. The two scales expressing appreciation of the music had the largest proportional residuals and did not fit well. Our assumption is that such appreciation is much less well defined than the reaction of the students to the physical characteristics of the music as expressed in the other scales and thus gives rise to a large amount of random error. As indicated above, from the bootstrap analyses, it became clear that notwithstanding the low level of explanation, the position of the scales with respect to other scales was clearly defined in the analysis. Furthermore, our impression is that semantic differential scales such as SLOW - FAST, which are very directly related to the music had better explained variability than those which are somewhat further away from the common way to describe music, such THIN - THICK.

Prelude Mode

The component scores for concepts (preludes) are shown in Table 6. Examples of preludes with large positive values in the first component were No. 16 (*Presto con fuoco*, bb minor), No. 18 (*Allegro molto*, f minor), and No. 24 (*Allegro appasionato*, d minor), and they have fast tempos and minor keys. On the other hand, typical preludes with large negative scores were No. 7 (*Andantino*, A major), and No. 15 (*Sostenuto*, Db major), with slow tempos and major keys. Preludes with high positive values on the second component are No. 3 (*Vivace*, G major), No. 5 (*Allegro molto*, D major), and No. 23 (*Moderato*, F major), who generally have fast tempos and major keys while the ones with high negative values were No. 2 (*Lento*, a minor), No. 6 (*Lento assai*, b minor), and No. 20 (*Largo*, c minor), with slow tempos and minor keys. Hence the first component corresponded the dimension of 'slow-major versus fast-minor' while the second component described the 'slow-minor versus fast-major' contrast.

From the last column of Table 6 it can be seen that the variabilities of preludes 1, 8, 10, 17, and 21 were not very well represented by the model, and at the same time that their variabilities are somewhat smaller than the other preludes. Small variability with low fit generally indicates that a concept, or prelude in this case, received more or less average scores on most scales and that there is not much system in the variability around these averages.

Core Array

As mentioned in the previous sectin, after transformation the core array was extremely simple as $PR - 1 = Q$ (Table 7). There were only four non-zero elements and they indicate which components are linked together and with what weight.

Table 6
Preludes 2 × 3 × 2-Solution Counter-Rotated with Inverse of Rotation Matrix of Core Array

No.	Key	Tempo	Component 1 'fast + minor-slow + major	Component 2 'fast + major-slow + minor'	Variance	SS(Res)/SS(Total)
16.	b^b minor	*Presto con fuoco*	**0.358**	0.174	1.042	0.543
18.	f minor	*Allegro molto*	**0.299**	-0.119	0.953	0.412
24.	d minor	*Allegro appasionato*	**0.261**	-0.101	0.935	0.549
22.	g minor	*Molto agitato*	**0.245**	-0.135	0.836	0.471
14.	e^b minor	*Allegro*	**0.237**	-0.044	0.735	0.613
12.	g$^\#$ minor	*Presto*	**0.232**	-0.011	0.714	0.662
11.	B major	*Vivace*	**-0.208**	0.174	0.829	0.480
13.	F$^\#$ major	*Lento*	**-0.227**	-0.053	0.682	0.716
7.	A major	*Andantino*	**-0.346**	0.107	1.213	0.437
15.	D^b major	*Sostenuto*	**-0.360**	0.086	1.126	0.388
3.	G major	*Vivace*	0.103	**0.368**	0.953	0.482
23.	F major	*Moderato*	-0.128	**0.318**	1.008	0.376
5.	D major	*Allegro molto*	0.006	**0.286**	0.777	0.548
19.	E^b major	*Vivace*	-0.080	**0.245**	0.749	0.538
4.	e minor	*Largo*	-0.170	**-0.240**	0.813	0.711
9.	E major	*Largo*	-0.003	**-0.276**	0.868	0.632
6.	b minor	*Lento assai*	-0.157	**-0.318**	0.828	0.553
20.	c minor	*Largo*	-0.004	**-0.319**	0.913	0.524
2.	a minor	*Lento*	-0.155	**-0.356**	0.972	0.528
10.	c$^\#$ minor	*Allegro molto*	0.040	0.127	0.621	**0.906**
1.	C major	*Agitato*	0.153	0.109	0.693	**0.860**
21.	B^b major	*Cantabile*	-0.163	0.040	0.559	**0.746**
8.	f$^\#$ minor	*Molto agitato*	0.167	0.018	0.658	**0.802**
17.	A^b major	*Allegretto*	-0.098	-0.044	0.522	**0.933**

Note. The component coefficients larger than 0.30 have been set in bold. Relative residual proportions larger than 0.80, indicating a bad fit, are also bold.

Table 7
Three-mode Orthomax Transformed Normalized Core Array

	Scale 1 SOOTHING - AGITATING	Scale 2 CHEERFUL - MOURNFUL	Scale 3 DISLIKE - LIKE
First subject component: 'Consensus'			
Prelude 1: fast + minor - slow + major	$g_{111} = 1.00$	0	0
Prelude 2: fast + major - slow + minor	0	$g_{221} = 0.99$	0
Second subject component 'Contrast'			
Prelude 1: fast + minor - slow + major	0	$g_{122} = 0.12$	0
Prelude 2: fast + major - slow + minor	0	0	$g_{223} = 1.00$

Note. g_{pqr} = the weight for the p^{th} prelude component, the q^{th} scale component and the r^{th} subject component.

Using this simple core, the Model Equation 4 for the centered data becomes

$$(12)\quad \hat{z}_{ijk} = 1.00{*}a_{i1}c_{k1}b_{j1} + 0.99{*}a_{i2}c_{k1}b_{j2} + 0.12{*}a_{i1}c_{k2}b_{j2} + 1.00{*}a_{i2}c_{k2}b_{j3}.$$

This shows that the first subject component on which all subjects scored positively and which had a high correlation with the individual differences in size of the variance of their scores, is exclusively linked the first scale component with the first prelude component and the second scale component with the second prelude component. In other words, the 'slow-major versus fast-minor' contrast in the preludes was directly related to the 'SOOTHING - AGITATING' contrast, and the 'slow-minor versus fast-major' contrast was directly related to the 'CHEERFUL VERSUS MOURNFUL' contrast. As all subjects had positive weights, this interpretation of musical tempo and keys was shared by all subjects, they only differed in the degree to which they expressed these contrasts. In particular, the relations between preludes and scales for this subject component, are primarily determined by the Prelude by Scales interaction term in Equation 7 and the individual differences on this components are primarily quantitative, rather than qualitative.

Some subjects had positive scores on the second subject component and some had negative scores indicating genuine individual differences (see Table 4). The core array showed that these differences were dominated by personal preferences for certain types of keys and tempos. In particular, the highest value in the second slice of the core array (g_{223} = 1.00) was associated with the third scale component, which strongly featured the scales UNATTRACTIVE - ATTRACTIVE and UNINTERESTING - INTERESTING and with the 'slow-minor versus fast-major' contrast; some students preferring fast-major pieces while others preferred slow-minor pieces. The smaller value (g_{122} = .12) indicated that students who tended to find fast-major preludes UNINTERESTING, also tended to judge fast-minor preludes somewhat more CHEERFUL and slow-major preludes somewhat more MOURNFUL than the general opinion.

Joint Biplots

One of the problems of the above descriptions via combinations of components, is that they are heavily dependent on the reification of the components and that they become rather abstract. By using joint biplots, a less abstract overview could be achieved because there was no need to explicitly define or name the components for the scale and the preludes. The relationships between the scales and preludes could be captured in two joint biplots: One representing the judgments about the Prelude-Scale relationship shared by all students (Figure 4), and another which depicted the individual differences (Figure 6).

Figure 4 displays in much detail which preludes were given which characteristics, for instance Nr. 16 (bb-Presto con fuoco) and Nr. 3 (G-Vivace) were considered particularly FAST and NOISY, while Nr. 20 (c-Largo) and Nr. 2 (a-Lento) were judged particularly hard, gloomy, heavy and dark. Because the scales ATTRACTIVE and INTERESTING had very small arrows they should not be interpreted in this plot. Preludes opposite to the arrows drawn were located at the other pole of the scale, for instance Nr. 7 (A-Andantino) and Nr. 15 (Db-Sostenuto) were judged to be particularly CALM as they were at the opposite end of the RESTLESS arrow.

To get some further insight in the nature of the relationships, we have labeled the preludes with the sequence number in which Chopin arranged them and coded them in three principal groups for which the relevant preludes are connected (Figure 5). Very clear in the plot is that the students via their scoring of the semantic differentials, recreated Chopin's alternating arrangements of the preludes in major and minor keys. They also reproduced Chopin's Circle of Fifths arrangement of the preludes, going from keys with few sharps, to those with few flats, to the most complex ones with either many sharps or flats. Because of the contents of the scales, the corresponding preludes in major and minor keys are generally diametrically located in the plot, indicating that the emotions described by the semantic differential scales alternate between subsequent preludes. The most notable incorrect placement is Prelude 9 (E-Largo), which was judged to be GLOOMY and HEAVY while it "should" have been judged CALM and LYRIC. Also its opposite prelude (Number 10) is located "incorrectly" because notwithstanding its a minor key, it is located with the major keys. The

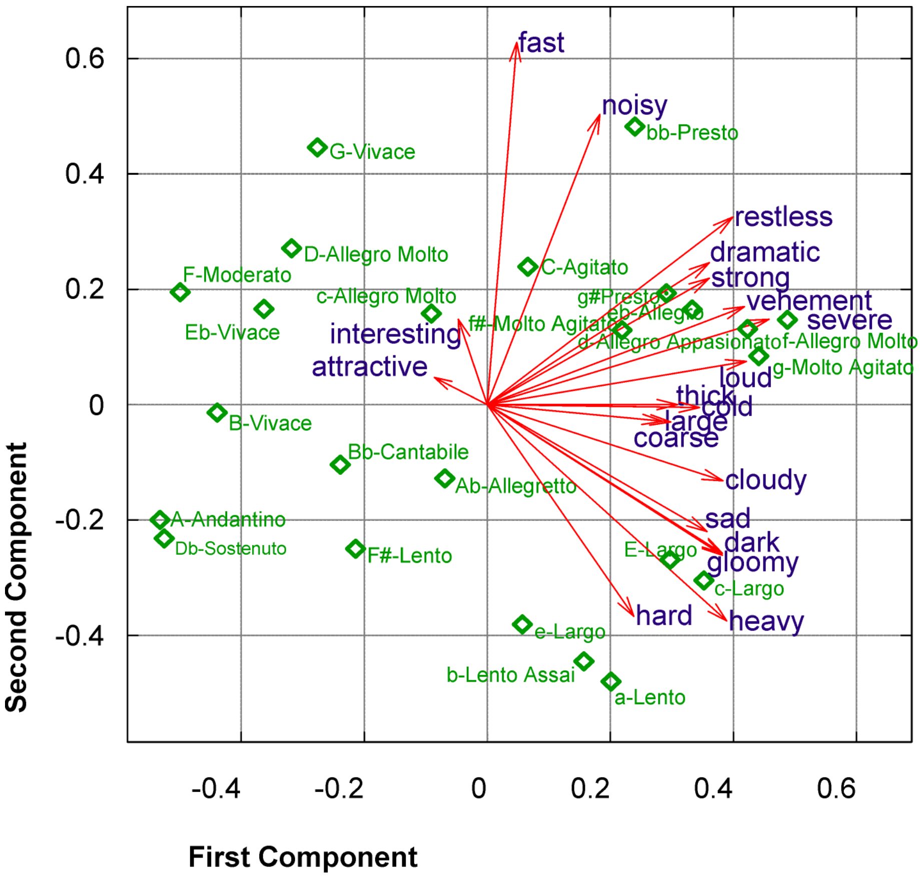

Figure 4
Joint Biplot for the Consensus (= first) Subject Component
For subjects with large values on the first component the above configuration is large, for subjects with small values, it is small.

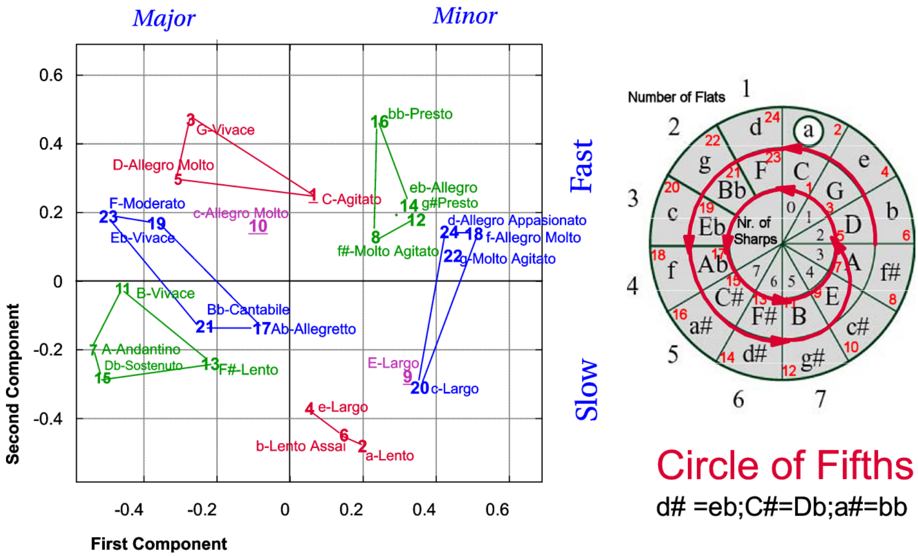

Figure 5
Figure 4 Labeled with Number, Key and Tempo
Dimensions labeled and groups of preludes labeled with different colors. On the circle of fifths, the curved path corresponds with the order of the Preludes in the figure.

explanation for the deviant position of Prelude 9 is that it shifts from major to minor in the fifth measure and does not really return to the major key. The anomaly for 10 is probably a lag effect of the different character of prelude 9, and also its key changes during the piece. Finally, Prelude 20 (set in a key and tempo typically used for funeral marches) "should" have been judged more LOUD and VEHEMENT, rather than HEAVY and GLOOMY. Apart from these deviations, the clear structure built into preludes can be clearly discerned from the judgments of the students using semantic differentials and even the deviations can be explained from their changing keys during the pieces.

The structure of 'cognition' of the music in Figures 4 and 5 was essentially common to all music listeners who were more or less familiar with the western classical music. The more succinct patterns described through the core array are now presented in more detail and differences between the judgments of the preludes could now be more easily assessed.

The students' preferences with respect to music were partially described in Figure 6. The strength of the preference was primarily indicated by their score on the second subject component, and its nature could be seen by the joint biplot associated with the second subject component. The core array already showed that primarily the second component of the preludes was involved in determining the preferences of the students. Some students liked the mournful music with a slow tempo in a minor key [e.g. Nr. 2 (a-Lento) and c-Largo)], while others liked cheerful music with a fast tempo in a major key, [e.g. Nr. 3 (G-Vivace) and Nr. 23 (F-Moderato)]. Figure 6 depicts the preferences of the first type of students; to view the preferences of the second group the arrows of the scales should be mirrored around the origin.

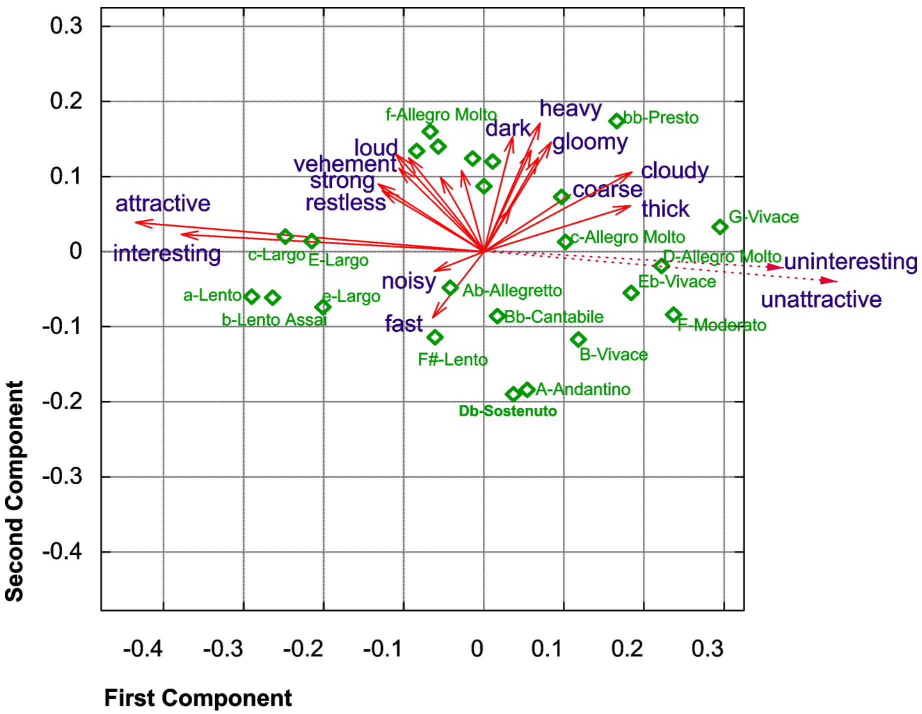

Figure 6

Joint Biplot for the Individual Differences (= second) Subject Component

For clarity, some preludes and scales were not labeled. The opposite poles of attractive and interesting are indicated. This picture corresponds to subjects score positively on the second subject component. Subjects with negative scores have the scales mirrored around the origin. Note that the scale of this figure is larger than that of Figure 4.

It should be noted that the explained variances of the two scales related to the preference dimension are relatively small, indicating that the judgments of the students are very diverse or not very well focused, so that a large part of the variability of these scales could not be modeled by the three-mode model chosen. There are undoubtedly many more characteristics of music that determine the individual preferences, but they were not contained in this set of semantic differential scales.

Conclusion

In this article we have given an overview and an illustration of the way individual differences in semantic differential research can be conceived and modeled. Some of these differences, such as differences in mean scale usage by subjects, were eliminated from the main analysis, but they can be further analyzed as well, should there be a specific interest in them. Other differences were explicitly modeled and a series of models were outlined each of which incorporated one of more types of individual differences.

As an alternative to the approach outlined here, one could perform separate analyses for the common part (first term of Equation 8) and for the individual differences part (second term of Equation 8), that is a (two-mode) singular value decomposition on the first and a three-mode analysis of some kind on the second term. The advantage of separate analyses is the simplicity of the analysis of the first term and possibly an easily interpreted three-mode solution with very few components for the second term. A disadvantage is that two independent analyses are performed without any explicit relationship between the two, which can create complications in interpretation. To avoid such a situation, we proceeded with the simultaneous analysis.

One of the advantages of the approach proposed here is that it becomes possible to assess whether individual differences are present, whether they are worth including in the analysis, and ways are shown to analyze their nature in some detail. The price to pay for increasing insight into these aspects of the data, is an increase in complexity of analysis, but that seems a natural consequence of the questions asked. Complex questions require complex methods to analyze them and generally do not supply very simple answers. On the other hand, due to extremely simple core array and the joint biplots, a fairly transparent description of the major patterns of similarities and differences of the Chopin preludes could be given.

References

Carroll, J. D. & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart-Young" decomposition. *Psychometrika, 35*, 283-319.

Harris, C. W. & Kaiser, H. F. (1964). Oblique factor analytic solutions by orthogonal transformations. *Psychometrika, 29*, 347-362.

Harshman, R. A. (1970). Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics, 16*, 1-84.

Harshman, R. A. & De Sarbo, W. S. (1984). An application of PARAFAC to a small sample problem, demonstrating preprocessing, orthogonality constraints, and split-half diagnostic techniques. In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 602-642). New York: Praeger.

Harshman, R. A. & Lundy, M. E. (1984a). The PARAFAC model for three-way factor analysis and multidimensional scaling. In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 122-215). New York: Praeger.

Harshman, R. A. & Lundy, M. E. (1984b). Data preprocessing and the extended PARAFAC model. In H. G. Law, C. W. Snyder, Jr., J. A. Hattie, & R. P. McDonald (Eds.), *Research methods for multimode data analysis* (pp. 216-284). New York: Praeger.

Heise, D. R. (1969). Some methodological issues in semantic differential research. *Psychological Bulletin, 72*, 406-422.

Henrion, R. & Andersson, C. A. (1999). A new criterion for simple-structure transformations of core arrays in *N*-way principal components analysis. *Chemometrics and Intelligent Laboratory Systems, 47,* 189-204.

Kaiser, H. F. (1958). The varimax rotation of analytic rotation in factor analysis. *Psychometrika, 23*, 187-200.

Kiers, H. A. L. (1997). Three-mode orthomax rotation. *Psychometrika, 62*, 579-598.

Kiers, H. A. L. (1998). Joint orthomax rotation of the core and component matrices resulting from three-mode principal components analysis. *Journal of Classification, 15*, 245-263.

Kiers, H. A. L. & Van Mechelen, I. (2001). Three-way component analysis: Principles and illustrative application. *Psychological Methods, 6*, 84-110.

Kroonenberg, P. M. (1983a). Annotated bibliography of three-mode factor analysis. *British Journal of Mathematical and Statistical Psychology, 36*, 81-113.

Kroonenberg, P. M. (1983b). *Three-mode principal component analysis: Theory and applications* (Errata, 1989; available from the author). Leiden: DSWO Press.

Kroonenberg, P. M. (1985). Three-mode principal components analysis of semantic differential data: The case of a triple personality. *Applied Psychological Measurement, 9*, 83-94.

Kroonenberg, P. M. & De Leeuw, J. (1980). Principal component analysis of three-mode data by means of alternating least squares algorithms. *Psychometrika, 45*, 69-97.

Levin, J. (1965).Three-mode factor analysis. *Psychological Bulletin, 64*, 442-452.

Murakami, T., ten Berge, J. M. F., & Kiers, H. A. L. (1998). A case of extreme simplicity of the core matrix in three-mode principal components analysis. *Psychometrika, 63*, 255-261.

Nielzen, S. & Cesarec, Z. (1981). On the perception of emotional meaning in music. *Psychology in Music, 9*, 17-31.

Osgood, C. E. (1964). The semantic differential technique in the comparative study of cultures. *American Anthropologist, 66*, 171-200.

Osgood, C. E. & Luria, Z. (1954). A blind analysis of a case of multiple personality. *Journal of Abnormal and Social Psychology, 49*, 579-591.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. (1957). *The measurement of meaning* Urbana, IL: University of Illinois Press.

Pinson, C. (1983). Pour une étude critique du différentiel sémantique (A critical assessment of the semantic differential). *Revue Française du Marketing, 95 (3)*, 3-25.

Snider, J. G. & Osgood, C. E. (Eds.) (1969). *Semantic differential technique: A sourcebook*. Chicago, IL: Aldine.

Snyder, F. W. & Wiggins, N. (1970). Affective meaning systems: A multivariate approach. *Multivariate Behavioral Research, 5*, 453-468.

Swanwick, K. (1973). Musical cognition and aesthetic response. *Bulletin of the British Psychological Society, 26*, 285-289.

Takeuchi, H., Kroonenberg, P. M., Taya, H., & Miyano, H. (1986). An analysis of Japanese language on thermal sensation. *Mathematical Linguistics, 15*, 201-209.

ten Berge, J. M. F. (1989). Convergence of PARAFAC preprocessing procedures and the Deming-Stephan method of iterative proportional fitting. In R. Coppi & S. Bolasco (Eds.), *Multiway data analysis* (pp. 53-63). Amsterdam: Elsevier

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

Tucker, L. R. (1964). The extension of factor analysis to three-dimensional matrices. In H. Gulliksen & N. Frederiksen (Eds.), *Contributions to mathematical psychology* (pp. 110-127). New York: Holt, Rinehart & Winston.

Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31*, 279-311.

Wiggins, N. & Fishbein, M. (1969). Dimensions of semantic space: A problem of individual differences. In J. G. Snider & C. E. Osgood (Eds.), *The semantic differential technique: A book of readings* (pp. 183-193). Chicago: Aldine.

Accepted January, 2003.