JOURNAL OF PERSONALITY ASSESSMENT, *85*(2), 209–217

Dual Scaling for the Analysis of Categorical Data

Michael D. Maraun, Kathleen Slaney, and Jarkko Jalava
Department of Psychology
Simon Fraser University

Dual scaling is a set of related techniques for the analysis of a wide assortment of categorical data types including contingency tables and multiple-choice, rank order, and paired comparison data. When applied to a contingency table, dual scaling also goes by the name "correspondence analysis," and when applied to multiple-choice data in which there are more than 2 items, "optimal scaling" and "multiple correspondence analysis." Our aim of this article was to explain in nontechnical terms what dual scaling offers to an analysis of contingency table and multiple-choice data.

Dual scaling is a term that was coined by Nishisato (1980) to describe a set of related techniques for the analysis of a wide assortment of categorical data types including contingency tables and multiple-choice, rank order, and paired comparison data. When applied to a contingency table, dual scaling also goes by the name "correspondence analysis" (e.g., Lebart, Morineau, & Warwick, 1984), and when applied to multiple-choice data in which there are more than two items, "optimal scaling" (Gifi, 1990) and "multiple correspondence analysis" (e.g., Greenacre, 1984). Nishisato's Dual3 program (Microstats, 1986) and the optimal scaling routines of the Categories module of SPSS (Version 10.0) are among many programs available to researchers who require a dual scaling analysis of their categorical data. Because categorical data arise so very frequently within the behavioral and social sciences, dual scaling should rightly be a technique at the front of every researcher's toolbox. However, for a variety of reasons, it has not yet become well known. Our aim of this article was to explain in nontechnical terms what dual scaling offers to an analysis of contingency table and multiple-choice data. Those interested in the mathematical basis for the technique can consult Nishisato (1980) or Gifi (1990), whereas those interested in the dual scaling of more esoteric data types can consult Nishisato (1980, 1993).

PCA FOR CONTINUOUS VARIATES

Dual scaling can be understood from a number of distinct but related perspectives. One such perspective can be developed via the recognition that dual scaling is really just a principal component analysis (PCA) carried out on categorical data. Recall what is involved in a PCA of continuous data, for example, the data produced when 200 individuals are scored with respect each of the 14 subscales of Version 3 of the Wechsler Adult Intelligence Scale (WAIS–III; Wechsler, 1997). For analysis, the set of 14 WAIS–III subscale scores of a given individual can be thought of as his or her 14 variate intelligence profile. By using the 14 subscale scores as coordinates, each of the 200 individuals can then be represented as a point in 14-dimensional euclidean space. In analogous fashion, each of the 14 subscales can be represented as a point in 200-dimensional euclidean space. The euclidean space that contains the objects under study (individuals, animals, groups, etc.) can be called *object space* (OS), and the space that contains the variates under study can be called the *variate space* (VS). The collection of object points forms a *point cloud* (OC) embedded in OS, and the collection of variate points forms a point cloud (VC) embedded in VS. Consider the simple case in which, for example, 100 individuals are scored with respect two variates, and a standard scatterplot is produced. The 100 points displayed in the scatterplot comprise the OC, and this OC is embedded in a two-dimensional OS (the space generated by placing a *y*-axis at a right angle to an *x*-axis). In the WAIS–III example, on the other hand, each individual is measured with respect to 14 variates, the individuals thus forming a 200-point OC embedded in a 14-dimensional OS. The subscales form a 14-point VC embedded in a 200-dimensional VS.

When the scientist inquires as to the possibility that the 200 individuals under study fall empirically into distinct clusters with respect to their intelligence profiles or whether Individual 3 is more similar to Individual 5 than to Individual 34 in regard to his or her intelligence profile, the scientist is expressing an interest in certain features of the OC. The former question, for example, is a question about the topology (organization) of the points that comprise this point cloud to wit, whether it is comprised of a number of smaller point clouds separated by empty spaces or alternatively, whether the points of which it is comprised are distributed smoothly throughout OS. The latter question centers on a comparison of the distances, within OS, between the points corresponding to Individuals 3 and 5 on one hand and Individuals 3 and 34 on the other. Analogous questions can be asked about the subscales, notably, whether they "go together" or alternatively, form a number of distinct clusters. In this case, the researcher's interest is in topological features of VC.

There is, however, a fundamental problem that must be overcome if these and similar questions are to be answered. If OS is two-dimensional, then the topology of OC can be described directly, as one does when one examines a scatterplot and describes its features. However, if OS and VS are of high dimensionality, then OC and VC cannot be examined and hence cannot be described directly. Humans can visualize objects located in three dimensions, but cannot examine the organization of a cloud of points located in, for example, 14 dimensions. Univariate and bivariate quantitative indexes such as means, variances, and covariances provide information about the organization of VC and OC in their embedding spaces, but what the researcher needs is a method by which low-dimensional projections (pictures) of VC and OC can be produced so that she or he can see what these point clouds actually look like. Data analytic tools such as PCA and multidimensional scaling were invented to address precisely this need.

Consider the general case of data produced when each of N objects is scored on p continuous variates. A PCA of such data can address the following issues:

PCA 1. The first step in a description of the arrangement (topology) of the points that comprise VC and OC is to determine the dimensionalities of these point clouds. It is a fact that VC and OC have precisely the same dimensionality, for example, t, and that $t \leq \min(N,p)$ in which $\min(N,p)$ is equal to the smaller of N, the number of objects, and p, the number of variates. Hence, in the WAIS–III example, the dimensionality t of OC and VC can be no greater than $\min(200,14) = 14$. On the other hand, the WAIS–III subscales were built to measure multiple facets of a single property, intellectual functioning, and hence, it is quite reasonable to conjecture that t will turn out to be less than 14. If, for example, t turned out to be equal to 1, then (a) the points representing the WAIS–III subscales would lie on a line, that is, a single dimension, within their 200-dimensional VS, and (b) the objects would lie on a line within their 14-dimensional OS. In such a case, the topology of each of OC and VC would be particularly simple. A larger value of t, such as in the case in which all Verbal subscales occupied one dimension and all Performance subscales another, would indicate a more complicated arrangement of points. The first question that must be answered then is "what is the value of t?"

The answer is the following: A PCA involves the extraction of the eigenvalue/eigenvector pairs, $\{\lambda_1,\boldsymbol{v}_1\}$, $\{\lambda_2,\boldsymbol{v}_2\}$, …, of the $p \times p$ covariance matrix, S, of the variates in which λ_i is the ith eigenvalue, and $\boldsymbol{v}_i$ is the ith eigenvector and in which $\{\lambda_1,\boldsymbol{v}_1\}$ contains the largest eigenvalue, $\{\lambda_2,\boldsymbol{v}_2\}$ the next largest, and so forth.[1] It can be proven that the value of t is equal to the number of non-zero eigenvalues extracted and can range from 1 to $\min(N,p)$.

PCA 2. How are the points that comprise VC arranged within VS? That is, what is the topology of VC?

The answer is the following: Once again, one cannot directly examine the topology of VC, and a low-dimensional picture is required. This picture is provided by the eigenvectors of S, the covariance matrix. Each variate has a value on each eigenvector, and the best r-dimensional picture of VC is produced by plotting each variate according to its values on the first r eigenvectors (those corresponding to the r largest eigenvalues). However, to allow the researcher to examine this picture, r must be chosen to be less than 3 and preferably, equal to 2. Plotting each variate according to its values on the first two eigenvectors gives the best two-dimensional picture of VC, and to the extent that this picture is a "good picture" of VC, a description of the arrangement of the points within this picture can be taken as a description of the topology of VC. For example, variates that are strongly linearly related will be situated at close proximity within this picture, whereas those that are unrelated will be plotted further apart. Subsets of variates that are mutually strongly related will appear within the picture as a cluster of points.

Unfortunately, if t is greater than 2, use of a two-dimensional picture will involve a loss of information and possibly, mistaken claims being made about the topology of VC. The question then becomes, "in using the two-dimensional picture produced in a PCA to make claims about the topology of a $t >$ two-dimensional VC, how

[1]It is also a common practice to employ the correlation matrix of the variates rather than the covariance matrix, the effect being to equate the standard deviations of the variates in the sample. If the aim is to make inferences about a population of interest, the covariance matrix should be employed.

much information is, in fact, lost (how good is the PCA approximation)?" The total variation in VC (and OC) is equal to the sum of the t nonzero eigenvalues of S, $\sum_{i=1}^{t}\lambda_i$, and the proportion of variance in VC accounted for by the best r-dimensional picture is equal to $P_r = \frac{\sum_{i=1}^{r}\lambda_i}{\sum_{i=1}^{t}\lambda_i}$. Thus, the proportion accounted for by the best two-dimensional picture is equal to $P_2 = \frac{(\lambda_1+\lambda_2)}{\sum_{i=1}^{t}\lambda_i}$. The larger is P_2, the better will be the quality of the picture and the more accurate will be the description of VC made on the basis of the picture.

PCA 3. How are the points that comprise OC arranged within OS? That is, what is the topology of OC?

The answer is the following: Steps analogous to those previously described for viewing VC within VS can be taken to view OC within OS. The optimal picture of VC was produced by plotting the variates with respect to their values on eigenvectors. The best r-dimensional picture of OC, on the other hand, is produced by plotting individuals with respect to their first r "component scores." The jth component score of Individual i, c_{ij}, is produced by multiplying each of Individual i's scores on the p variates by the values of these variates of the jth eigenvector, $\mathbf{v}_j$ and summing the results: $c_{ij} = \sum_{k=1}^{p} v_{jk}x_{ik}$. Once again, r should be set to 2, a scatterplot produced of the pairs $[c_{i1}, c_{i2}]$, $i = 1 \ldots N$, and a description provided of the topological features of this plot.

DUAL SCALING FOR CATEGORICAL VARIATES

Now consider the case in which the researcher must answer questions analogous to PCA 1 through PCA 3, but in which his or her data are at the nominal or ordinal level of measurement. The researcher must now analyze data for which covariances and Pearson product–moment correlations cannot be meaningfully computed and hence on which a standard PCA cannot be carried out.[2] This is where dual scaling comes in because it is a variant of PCA carried out on categorical variates. Consequently, we consider the dual scaling analysis of a contingency table and some multiple-choice data.

Contingency Tables

Muntigl and Turnbull (1998) studied the structures of the arguments of N = 155 couples involved in long-term, positive, intimate relationships. Each argument was broken down into a number of turns, and the behavior occurring within each turn was assigned to one of a number of categories. Five categories of behavior described Turn 2: irrelevancy claim, challenge, contradiction, contradiction plus counterclaim, and counterclaim. The same five categories plus one additional category, orientation to Turn 1, described behavior that occurred during Turn 3. The first three turns of a hypothetical argument might then be as follows: T1→Individual A states proposition P; T2→Individual B challenges the correctness of P; T3→Individual A contradicts this challenge and makes a counterclaim. Interest was in the relationship between the behavior occurring in Turns 2 and 3.

Table 1 contains the contingency table, or cross-tabulation, of the 155 couples with respect T2 and T3 as reported in Muntigl and Turnbull (1998). The ijth entry of this table, n_{ij}, is equal to the number of couples in categories i of T2 and j of T3. The row marginals, n_{i*}, correspond to the number of couples in the categories of T2, and the column marginals, n_{*j}, to the number of couples in the categories of T3. This contingency table contains all of the sample information that is relevant to the making of inferences about the relationship between T2 and T3. Now, very often in the social sciences such an analysis would involve only a test of the null hypothesis, H_0: T2 and T3 are statistically independent using the Pearson chi-square statistic, $\chi^2 = N\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{(n_{ij} - \frac{n_{i*}n_{*j}}{N})}{n_{i*}n_{*j}}$. However, much more is required before the researcher can justifiably claim an understanding of the relationship between the variates. The following issues must be addressed (C = contingency):

C 1. When the elements of a given row of the contingency table are divided by their row marginal, they then become conditional probabilities. For example, dividing the elements of row 1 by n_{1*}, which is equal to 12, yields the set of conditional probabilities [0,0,0,.083,.167,.750], and these are the probabilities of occurrence of each T3 behavior conditional on "irrelevancy claim" having occurred in T2. The number .750, for instance, is the probability of an "orientation to Turn 1" occurring in T3 given that an irrelevancy claim has occurred in T2. The five rows of conditional

[2]Pearson product–moment correlations are invariant only up to affine transformations, whereas ordinal and nominal variates can be legitimately transformed by monotone transformations.

TABLE 1
Contingency Table of Turn 2 and Turn 3

	Turn 3						
Turn 2	*irr*	*ch*	*c*	*c&cc*	*cc*	*tlo*	n_{i*}
irr	0	0	0	1	2	9	12
ch	2	0	0	0	2	14	18
c	1	4	10	2	1	11	29
c&cc	1	1	1	1	4	1	9
cc	2	2	3	9	57	14	87
n_{*j}	6	7	14	13	66	49	

Note. N = 155. irr = irrelevancy claim; ch = challenge; c = contradiction; c&cc = contradiction plus counterclaim; cc = counterclaim, tlo = orientation to Turn 1; n_{i*} = n in row; n_{*j} = n in column.

probabilities thus created form a point cloud, for example, VC_r, in six-dimensional space. An analogous set of conditional probabilities can be created for each column of the contingency table by dividing the elements of each column by their respective column marginals. The six columns of conditional probabilities thus created also form a point cloud, for example, VC_c, in five-dimensional space.

The fact VC_r and VC_c are embedded in six- and five-dimensional spaces, respectively, means that the relationship between T2 and T3 is potentially multidimensional. If the hypothesis of statistical independence is, in fact, false, then the dimensionality, *t,* of this relationship must be between 1 and min(5,6) – 1 = 4 (i.e., 1 and the smaller number of categories minus 1). The loss of one dimension results from the fact that the elements within each of the five rows (and six columns) of conditional probabilities sum to unity, indicating the presence of a linear dependency. Given a rejection of the hypothesis of statistical independence, the next step is to make an inference as to the dimensionality, *t,* of the relationship between T2 and T3.

C 2. Given a decision about the value of *t,* what is the form of the relationship between T2 and T3 within this *t*-dimensional space? This question can also be stated as “which categories of T2 ‘go with’ which categories of T3?”

Both of issues C 1 and C 2 can be addressed in a dual scaling analysis. Dual scaling involves the extraction of the eigenvalues and eigenvectors of a matrix M_1, calculated from the contingency table, whose elements quantify the departure of two categorical variates from statistical independence (see Nishisato, 1980). Matrix M_1 plays a role analogous to that of the covariance matrix for continuous variates in that it quantifies the associations among either the rows (T2 categories) or columns (T3 categories) of the contingency table. In analogy to standard PCA, the answer to question C 1 is that *t* is equal to the number of nonzero eigenvalues of matrix M_1. In fact, the sum of the *t* nonzero eigenvalues of M_1 is equal to $\frac{\chi^2}{N}$, a measure of the departure from independence of two categorical variates. Thus, one can think of the departure from statistical independence of T2 and T3 as distributed in a *t*-dimensional space. Question C 2 is also addressed in a fashion analogous to that of standard PCA, in that the best $r < t$ dimensional picture of the relationship between T2 and T3 is produced by plotting the categories of each variate according to the first *r* coordinates output in a dual scaling analysis, these coordinates based on the first *r* eigenvectors extracted from matrix M_1. Once again, it is most useful to produce a two-dimensional plot (i.e., choose *r* to be 2), and if the plot is of good quality, then a description of the organization of the categories as displayed in the plot can be taken as a description of the relationship between the two variates.

Application of Nishisato’s (1980) Dual3 (Microstats, 1986) program to the contingency table of Table 1 yields the (edited) output presented in Table 2. As can be seen, the hypothesis that T2 and T3 are statistically independent in the population should be rejected ($\alpha = .05$; $\chi^2(20, N = 155) = 96.742$, $p < .05$. Hence, the inference can be made that T2 and T3 are associated and that the population dimensionality of this relationship is between 1 and 4. Output from Dual3 is presented one dimension/solution at a time beginning with output associated with the largest eigenvalue of M_1. Each eigenvalue of M_1 is labeled in the output as “squared correlation ratio,” and the sum of these eigenvalues is equal to $\frac{\chi^2}{N}$, in this case, $\frac{96.742}{155} = .624$. Thus, .624 is the total variance to be accounted for. Within the output for Solution 2, one finds cumulative = 94.91%; this is equal to 100 times the sum of the first two eigenvalues over the sum of all eigenvalues, $100 \times P_2$, and calculated as $100 \times \frac{(.374 + .218)}{.624} = 94.91$. This indicates that the best two-dimensional picture accounts for roughly 95% of the association between T2 and T3 and hence that *t* is, essentially, equal to 2.

In the output, the categories of T2 are referred to as “rows,” and those of T3, “columns.” Dual3 produces two types of coordinates: “weighted” and “normed.” The normed coordinates of Category 1 of T2 (i.e., an irrelevancy claim in Turn 2) are, for example, [.740, –1.638], whereas the weighted coordinates for Category 1 of T3 (i.e., an irrelevancy claim in Turn 3) are [.308, –.257]. The normed coordinates of each solution/dimension are normalized to have a sum of squares that is equal to *N,* whereas the weighted categories of solution/dimension *j* are normed categories that have been multiplied by the square root of the *j*th eigenvalue of M_1. Because solutions/dimensions that account for a large proportion of the total variance are those

TABLE 2
Output From Dual3 Analysis

	Row-Column Association[a]			
	Turn 2 (Rows)		*Turn 3 (Columns)*	
	Normed	*Weighted*	*Normed*	*Weighted*
Solution 1[b]				
irr	0.7403	0.4528	0.5038	0.3082
ch	1.0300	0.6300	1.0428	0.6378
c	1.5586	0.9533	1.5176	0.9282
c&cc	–0.1262	–0.0772	–0.4608	–0.2818
cc	–0.8217	–0.5026	–1.0464	–0.6400
tlo			0.8875	–0.6400
Solution 2[c]				
irr	–1.6376	–0.7651	–0.5508	–0.2573
ch	–1.9276	–0.9005	1.9027	0.8889
c	1.3037	0.6091	2.1638	1.0109
c&cc	0.7910	0.3696	0.4504	0.2104
cc	0.1083	0.0506	0.1138	0.0532
tlo			–1.0954	–0.5118

Note. irr = irrelevancy claim; ch = challenge; c = contradiction; c&cc = contradiction plus counterclaim; cc = counterclaim; tlo = orientation to Turn 1.
[a]$\chi^2(20, N = 155) = 96.73873$*; total variance accounted for = .6241208.
[b]Squared correlation ratio = .37409; delta (total variance accounted for): partial = 59.94%, and cumulative = 59.94%. [c]Squared correlation ratio = .21827; delta (total variance accounted for): partial = 34.97%, cumulative = 94.91%.
*Significant at $p = .05$ level.

with large eigenvalues, weighted coordinates reflect the relative importance of each dimension in accounting for the total variance to be explained. The choice as to which types of coordinates should be used to plot the row and column categories has been the subject of some controversy (see, e.g., Greenacre, 1989), the issue being what a plot of each type of coordinates allows the researcher to conclude about the associations amongst the categories. Nishisato (1988) recommended that one set of categories (e.g., those of T2) should be plotted using normed coordinates and the other set (those of T3) weighted coordinates. Given this choice, first, the distances among the T3 points in the two-dimensional picture are approximations to the associations among the T3 categories. In particular, the greater is the association between two T3 categories, the closer these categories will be in the two-dimensional picture. Second, each T3 category is positioned in the two-dimensional picture in such a way that it is closest to the T2 categories for which the associated conditional probabilities are large (i.e., those T2 behaviors that have a high likelihood of occurrence given the T3 behavior in question).

The coordinates of the 11 categories (T2 categories normed, T3 categories weighted) are plotted in Figure 1, and hence, Figure 1 is a best picture of the (essentially) two-dimensional relationship between T2 and T3. The picture reveals the following: (a) The T3 behaviors form three distinct clusters, counterclaim and contradiction plus counterclaim, contradiction and challenge, and irrelevancy claim and orientation to Turn 1; (b) counterclaim in Turn 2 is positioned close to counterclaim in Turn 3 and contradiction plus counterclaim in Turn 3, indicating that a counterclaim or a contradiction and counterclaim in Turn 3 is often preceded by a counterclaim in Turn 2. In fact, counterclaim in Turn 2 has a high probability of occurrence conditional on each of counterclaims in Turn 3 ($\frac{57}{66} = .864$) and contradiction plus counterclaim in Turn 3 ($\frac{9}{13} = .692$); (c) contradiction in Turn 2 is positioned close to contradiction in Turn 3, indicating that a contradiction in Turn 3 is often preceded by a contradiction in Turn 2. In fact, contradiction in Turn 2 has a high probability of occurrence conditional on a contradiction in Turn 3 ($\frac{10}{14} = .714$); (d) orientation to Turn 1 and irrelevancy claim in Turn 3 are situated in the middle of the picture and are not close to any Turn 2 behaviors in particular, indicating that no Turn 2 behaviors are more likely than any others to precede an irrelevancy claim or an orientation to Turn 1 in Turn 3; and (e) challenge in Turn 3 is positioned between contradiction in Turn 2 and counterclaim in Turn 2 due to the fact that the former have high probabilities of occurrence conditional on the latter.

Multiple-Choice Data

Items 1, 4, 7, and 21 of the Beck Depression Inventory (BDI; Beck, 1996) ask respondents about their current degree of sadness, loss of pleasure, self-dislike, and loss of interest in sex, respectively. Each of these items has a four-category re-

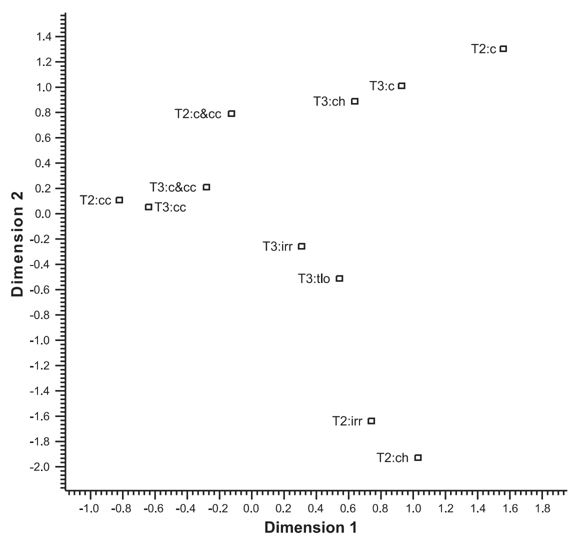

FIGURE 1 Best two-dimensional picture of Turns 2 and 3 categories. T2 = Turn 2; T3 = Turn 3; irr = irrelevancy claim; c = contradiction; c&cc = contradiction plus couterclaim; cc = counterclaim; tlo = orientation to Turn 1.

TABLE 3
BDI Raw Data

	BDI Items			
Sample	*1*	*4*	*7*	*21*
Depressed inpatients				
1	4	4	4	4
2	3	4	2	1
3	4	4	4	4
4	4	3	4	4
5	4	4	2	3
University undergrads				
1	3	2	1	1
2	1	1	2	1
3	2	1	2	3
4	1	1	1	1
5	1	1	2	2

Note. BDI = Beck Depression Inventory.

sponse scale, with Category 1 reflecting low depression and Category 4 reflecting a high depression. Although these response scales are at least at the ordinal level, one can treat them as nominal and let the data determine how the 16 item categories go together. Table 3 contains a small data set in which the first five individuals are depressed inpatients and the last five are university undergraduates.[3] The aim is to answer questions PCA 1 through PCA 3, but now because the items are viewed as being at the nominal or ordinal level of measurement, covariances and correlations and hence standard PCA can no longer be employed.

In a dual scaling of such data, the data matrix is first recoded to produce an $N \times c$ indicator matrix in which N is equal to the total number of participants, and c is equal to the total number of categories of the items. The four BDI items have a total of 16 categories, and the 10×16 indicator matrix for the data of Table 3 is presented in Table 4. In this indicator matrix, columns 1 through 4 represent the four categories of Item 1, the next four columns are the categories of Item 4, and so forth. Individuals receive unities for those categories in which they appear and zeros elsewhere. The indicator matrix itself is then transformed into matrix M_1, which, once again, quantifies the degree of association between the categories. Finally, the eigenvalue/eigenvector pairs of matrix M_1 are extracted and questions PCA 1 through PCA 3 answered in the usual way.

Now, for the BDI data, we have a 10-individual OC embedded in a 16-dimensional OS and a 16-category VC embedded in a 10-dimensional VS. The dimensionality, t, of each of VC and OC can be no greater than the smaller of $c - p$ (the total number of categories minus the number of items) and min(N, c). Hence, because $c - p = 16 - 4 = 12$, and min(N, c) = min(10,16) = 10, t can be no greater than 10. Because BDI items are viewed as being indicators of a single depression construct, the categories of any subset of BDI items should, in theory, be unidimensional (occupy but one dimension), and the categories of each item should have the same ordering along this single dimension. Yet of course, the dimensionality of a set of items is not a property of item content per se but rather of the responses of the individuals under study to this item content. In any case, aside from the upper bound of 10, the value of t is an empirical issue.

A dual scaling of the data in Table 3 was carried out using Homals, an optimal scaling routine found in the categories module of SPSS Version 10.0 (see Meulman & Heiser, 1999). Table 5 contains an edited version of the output. Matrix M_1 turns out to have eight nonzero eigenvalues, indicating that the dimensionality, t, of VC and OC is equal to 8. However, at this point a problem arises, for in the case of multiple-choice data, the sum of the eigenvalues of M_1 is always equal to $\frac{c}{p} - 1$ (the average number of categories per item minus 1). That is, the total information to be explained is determined not by features of the relationships among the item categories but merely by the total number of items and categories. In this example, the total information is equal to $\frac{16}{4} - 1 = 3$. On the other hand, the eigenvalues of M_1 must lie between 0 and 1, and therefore, unlike in a standard PCA, a two-dimensional dual scaling solution for Table 3 can account for at most $\frac{2}{3} \times 100 = 66\%$ of the variation in each of VC and OC. Hence, according to P_2, the traditional measure of variance explained, a two-dimensional solution can look poor regardless of the relationships that exist among the item categories. In response to this problem, Nishisato (1993) suggested a number of alternatives to P_2. One such alternative is derived by noting that the average of the eigenvalues of M_1 is equal to $\frac{1}{p}$, that solutions for which $\lambda_i < \frac{1}{p}$ have certain undesirable properties, and thus by redefining the total information to be explained as $P_{2c} = \sum_{i=1}^{t} \lambda_i - \frac{1}{p} = (\frac{c}{p} - 1) - \frac{1}{p} = \frac{(c-1)}{p} - 1$.

In our example, the two largest eigenvalues of M_1 are equal to .876 and .634, $P_2 = \frac{1.51}{3} = .503$, and $P_{2c} = \frac{1.51}{2.75} = .549$. The coordinates used to locate each of the 16 categories in their optimal two-dimensional picture are found in Table 5 under the heading of "Category Quantifications" and for the individuals under the heading "Object Scores." The categories are plotted in Figure 2, and the individuals are labeled by their group (D = depressed in-

[3]Typically, dual scaling would be used to analyze much larger data sets. The data set analyzed here is small for ease of viewing.

TABLE 4
Indicator Matrix

	BDI Items and Response Categories															
	1				4				7				21			
Sample	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*	*1*	*2*	*3*	*4*
Depressed inpatients																
	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
	0	0	1	0	0	0	0	1	0	1	0	0	1	0	0	0
	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0	1
	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	1
	0	0	0	1	0	0	0	1	0	1	0	0	0	0	1	0
University undergraduates																
	0	0	1	0	0	1	0	0	1	0	0	0	1	0	0	0
	1	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0
	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0
	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
	1	0	0	0	1	0	0	0	0	1	0	0	0	1	0	0

Note. BDI = Beck Depression Inventory.

TABLE 5
Dual Scaling Analysis Output

			Category Quantifications	
			Dimension	
BDI Items	*Response Category*	*Marginal Frequency*	*1*[a]	*2*[b]
1	1	3	0.905	0.387
	2	1	0.540	1.597
	3	2	0.671	–1.421
	4	4	–1.149	0.021
	Missing	0		
4	1	4	0.813	0.689
	2	1	0.983	–2.225
	3	1	–1.586	–0.240
	4	4	–0.663	–0.073
	Missing	0		
7	1	2	0.992	–1.293
	2	5	0.449	0.635
	4	3	–1.410	–0.197
	Missing	0		
21	1	4	0.797	–0.700
	2	1	0.866	1.120
	3	2	0.088	1.136
	4	3	–1.410	–0.197
	Missing	0		
Object scores				
Depressed inpatients	1		–1.322	–0.175
	2		0.358	0.616
	3		–1.322	–0.175
	4		–1.586	–0.240
	5		–0.364	0.676
University undergraduates	1		0.983	–2.225
	2		0.846	0.400
	3		0.540	1.597
	4		1.001	–0.360
	5		0.866	1.120

Note. BDI = Beck Depression Inventory.
[a]Eigenvalue = .876. [b]Eigenvalue = .634.

patient, U = university undergraduate) and response pattern (their values on the four items) in Figure 3.

Examining Figure 2, the following three clusters are evident: (a) Category 3 of Item 4 and Category 4 of Items 1, 4, 7, and 21; (b) Category 1 of Items 7 and 21 and Category 3 of Item 1; and (c) Category 1 of Items 1 and 4; Category 2 of Items 1, 7, and 21, and Category 3 of Item 21. Category 2 of Item 4 is off by itself and evidently is not strongly associated with any of the other categories. Category 3 of Item 7 does not appear in the representation because it was not endorsed by any of the respondents. Thus, although there is some evidence that as per expectation, categories of like kind go together, the data by no means squares perfectly with this expectation. In Figure 3, the optimal two-dimensional picture of OC, each individual is positioned closest to the four categories in Figure 2 that comprise his or her response pattern (i.e., the four categories that describe him or her). Note that the five depressed inpatients are located to the left of the picture and are somewhat separated from the university undergraduates who are found to the right. Hence, as would be expected, the depressed inpatients have a higher likelihood of being in Category 4 of each of the items. One can also see that the first undergraduate (U1) has an unusual response pattern and as a result is an outlier within the plot. This individual's higher scores on Items 1 and 4 are not accompanied by higher scores on Items 7 and 21. Moreover, this individual is the only one to provide the response of 2 to Item 4 (evidently, this is why Category 2 of Item 4 is an outlier in VC). Notice also that in the plot of VC, Category 4 of Items 1 and 4 are "pulled toward" Category 2 of Item 7 and Category 3 of Item 21 because the fifth depressed inpatient (D5) has the unusual response pattern (4423), a mixture of the response patterns found in Clusters 1 and 3.

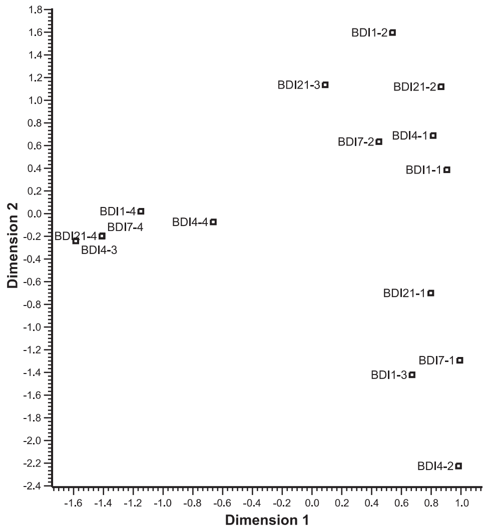

FIGURE 2 Best two-dimensional picture of Beck Depression Inventory (BDI) item categories from SPSS Version 10.0 Homals analysis. Each point corresponds to a particular pairing of BDI item (i.e., BDI 1, 4, 7, or 21) and response category (i.e., 1, 2, 3, or 4).

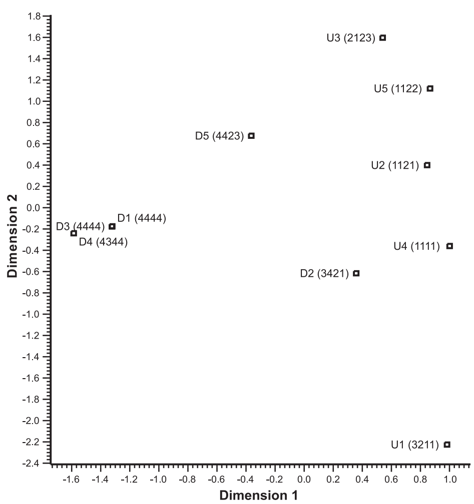

FIGURE 3 Best two-dimensional picture of depressed and nondepressed patients. The response patterns for each individual are represented in parentheses; D = depressed inpatient; U = university undergraduate student.

FINAL REMARKS

Dual scaling is not a factor analysis technique. It does not rest on talk of unobservable or latent entities. It is instead a component technique whose aim is the representation of the relationships that exist in a set of categorical variates. This is made possible through the recognition that when one inquires as to the nature of the relationships that exist in a set of variates, one is, in fact, inquiring as to the topology of the point cloud formed by the variates, this point cloud embedded in a high-dimensional space. Dual scaling projects information about such point clouds into a low-dimensional representation. The lack of popularity of the technique may well be a result of a lack of awareness as to what is involved in describing multivariate relationships.

Many treatments of component and factor analysis explain to researchers that they should examine results one dimension at a time and interpret each dimension by examining its correlations with each variate (these correlations often called "loadings"). However, if the aim is to come to an understanding of the relationships that exist in a set of variates, this approach is insufficient because to describe the relationships that exist in a set of variates is to describe the organization or topology of VC in VS, that is, its clusterings and empty spaces. However, the researcher can no more describe the topology of VC by considering the coordinates of the variates one dimension at a time than can the geographer meaningfully describe the topological features of a city by describing the positions of buildings, parks, and hills one direction at a time. In both cases, what is required is a high-quality, two-dimensional picture.

Note that we have said virtually nothing about statistical inference in this article. There have now been invented a range of inferential tools that can be employed in a dual scaling analysis including the construction of bootstrap confidence intervals for the parameters (the eigenvalues and category coordinates) of dual scaling solutions (see, e.g., Greenacre, 1984). However, the chief aim of a dual scaling analysis is not statistical inference but rather the description of the high-dimensional categorical data structures that often arise in psychological research. The researcher who believes he or she has found an interesting relationship through the employment of dual scaling should, as per sound scientific practice in general, attempt to replicate the finding at a later date.

ACKNOWLEDGMENT

This research was supported in part by an NSERC Grant awarded to M. D. Maraun.

REFERENCES

Beck, A (1996). *BDI–II.* San Antonio, TX: Psychological Corporation.

Gifi, A. (1990). *Nonlinear multivariate analysis.* New York: Wiley.

Greenacre, M. (1984). *Theory and applications of correspondence analysis.* London: Academic.

Greenacre, M. (1989). The Carroll–Green–Schaffer scaling in correspondence analysis: A theoretical and empirical approach. *Journal of Marketing Research, 26,* 358–365.

Lebart, L., Morineau, A., & Warwick, K. (1984). *Multivariate descriptive statistical analysis.* New York: Wiley.

Microstats. (1986). Dual3 Version 3.20 [Computer software]. Islington, Ontario, Canada: Author.

Meulman, J., & Heiser, W. (1999). SPSS categories 10.0 [Computer software]. Chicago: SPSS.

Muntigl, P., & Turnbull, W. (1998). Conversational structure and facework in arguing. *Journal of Pragmatics, 29,* 225–256.

Nishisato, S. (1980). *Analysis of categorical data: Dual scaling and its applications.* Toronto, Ontario, Canada: University of Toronto Press.

Nishisato, S. (1988). Assessing the quality of joint graphical display in correspondence analysis and dual scaling. In E. Diday (Ed.), *Data Analysis and Informatics V* (pp. 409–416). Amsterdam: North-Holland.

Nishisato, S. (1993). On quantifying different types of categorical data. *Psychometrika, 58,* 617–629.

Wechsler, D. (1997). *Manual for the Wechsler Adult Intelligence Scale—Third Ed.* San Antonio: TX: Psychological Corporation.

Michael D. Maraun
Department of Psychology
Simon Fraser University
Burnaby, British Columbia, Canada V5A 1S6
Email: maraun@sfu.ca

Received November 10, 2003
Revised January 21, 2005