DOCUMENT RESUME

ED 235 204 TM 830 617

AUTHOR Koch, William R.

TITLE The Analysis of Dichotomous Test Data Using Nonmetric Multidimensional Scaling.

PUB DATE Apr 83

NOTE 25p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).

PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.

DESCRIPTORS Achievement Tests; *Cluster Analysis; *Latent Trait Theory; *Multidimensional Scaling; Multiple Choice Tests; *Research Methodology; *Test Items

IDENTIFIERS *Facet Analysis

ABSTRACT

The technique of nonmetric multidimensional scaling (MDS) was applied to real item response data obtained from a multiple-choice achievement test of unknown dimensionality. The goal was to classify the 50 items into the various subtests from which they were drawn originally, the latter being unknown to the investigator. Issues addressed in the research included dimensionality, choice of item proximity measures, and appropriateness of the MDS model for analyzing dichotomous item response data. Three coefficients were chosen to form proximity matrices that reflected the associations of each item with each of the other items. These matrices then served as input to both the ALSCAL and MINISSA computer programs for MDS analysis. A three dimensional solution was found to be optimal, based on stress, the multiple correlation coefficient, and coefficient of alienation values. Both cluster analysis and regional (facet theory) analysis of the spatial configuration were used to interpret the results. (Author)

THE ANALYSIS OF DICHOTOMOUS TEST DATA USING NONMETRIC MULTIDIMENSIONAL SCALING

William R. Koch
The University of Texas at Austin

Introduction

The problem of determining the "underlying structure" of a set of interrelated variables, in some data reduction sense, is extremely difficult, quite subjective, and has no unique solution. This formidable task becomes complicated even further when the variables of interest happen to be scored dichotomously. However, just such a situation confronts the researcher when analyzing data collected from a sample of examinees responding to multiple-choice, achievement test items.

In the past, typical reasons for attempting to determine the dimensionality of a test have included the desires to obtain scores for persons on each of its separate subscales or to provide evidence for the construct validity of the test. More recently, another motive has been the desire to form unidimensional subsets of items so that each subset may be subjected separately to latent trait item calibration (parameter estimation). The latter procedure is necessary to meet the basic assumption of item response models that only a single trait is being measured by the set of items that comprise a test.

Traditionally, factor analysis has been the preferred method for attempting to determine the underlying structure of a test. However, artificially dichotomized item response data present some problems for the linear factor analytic model. For example, the model assumes that the variables have continuous distributions and that the variables have linear relationships both with each other and with the derived factors. Furthermore, the coefficients used to form the input item intercorrelation matrix are typically either phi coefficients or tetrachoric coefficients. The former lead to artifactual item difficulty factors (McDonald & Ahlawat, 1974), while the latter present other problems such as non-Gramian matrices (Christofferson, 1975).

Another complication is that examinees have some probability of correctly answering multiple-choice test items simply by chance, with the incidence of guessing increasing for very difficult items and/or for examinees with low achievement levels. As a result, guessing is usually considered to be a nonlinear function of achievement level, a situation that cannot easily be accommodated by the linear factor analytic model. Finally, the results of factor analysis are presented in terms of composite linear vectors (factors) at various angles to each other in space which are labeled subjectively for interpretation.

Paper presented at the Annual Meeting of the American Educational Research Association, Montreal, March 1983.

Compared to factor analysis, the assumptions about the input data made by nonmetric multidimensional scaling (MDS) are much less strict. MDS requires only that order information be present in the data and that the variables be related to each other at least monotonically. A large variety of coefficients is available to measure the proximities, or associations, of each variable with every other variable, the only requirement being that the values of the proximities may be ordered in magnitude. No distributional assumptions are made about the variables.

The goal of the MDS algorithm is to approximate a one-to-one monotonic relationship between the ordinal proximity information extracted from the input data and the rankings of the corresponding distances among the items located as points in the final spatial configuration. Specifically, the rule to be met is the following: if the similarity between stimulus i and stimulus j is greater than the similarity between stimulus k and stimulus l, then the distance between points i and j in the solution space should be less than or equal to the distance between points k and l. In the case of multiple-choice data, the stimuli are items and the items are positioned as points in the solution space.

Similar to factor analysis, the MDS solution is represented using a set of linear dimensions. However, the dimensions serve simply as arbitrary coordinate axes to locate the variables in space and usually have no intrinsic meaning in and of themselves. Rather, several methods may be used for interpretation of the results, such as discovering meaningful regions of the space, identifying clusters of the variables, and finding directions (linear or nonlinear) of stimulus variation in the space. These techniques will be discussed later in the paper.

Unfortunately, MDS methodology has its own set of potential problems when applied to dichotomous item response data. First, the data in their original form are arranged in a rectangular array of 0,1 responses from n examinees to k items. In order to be appropriate for classical MDS input, these data must be transformed into a (k x k) proximity matrix by means of some chosen coefficient of association. Also, in order to minimize the effect of item difficulty levels on the dimensionality of the solution, the coefficient used must be relatively unaffected by variations in item difficulty. Furthermore, to the extent that guessing on items introduces substantial random error into the data, the fit of the MDS model to the data is jeopardized. Despite these problems, one would still expect dichotomous item response data, properly transformed, to meet the minimal assumptions of the MDS model.

However, MDS has rarely been applied in the context of analyzing dichotomous item response data. Consequently, very little is known about the properties or value of MDS methodology in such situations. The present research was intended to contribute to the knowledge base and experience in this area.

Dimensionality

The novel task confronting the investigators in this symposium was to undertake a "blind" analysis of some real test data in an attempt to determine its underlying dimensionality, or structure. The moderator constructed a matrix of item response data from a sample of 2,794 examinees to 50 assorted items. The items were chosen deliberately from diverse subtests and aggregated to form a realistic, inherently multidimensional test. The motivations behind these procedures were to seek a fair, rigorous challenge for the analytic techniques being compared and to provide a known criterion against which solutions could be evaluated. The analyses were "blind" in the sense that only the moderator and the discussant knew the actual source and content of the items and, therefore, the "true" dimensionality of the test.

Dimensionality is a particularly slippery concept, and it becomes a technical term only in specific contexts. For example, in the variance decomposition framework of factor analysis, it is common to think of the derived factors as the dimensions of the multivariate space. This notion is reasonable in the sense that the retained linear composite variables (factors) are the number of orthogonal dimensions necessary to represent adequately or to "account for" the common variance present in the original set of variables.

In the context of MDS, dimensionality is frequently thought of as being the minimum number of orthogonal coordinate axes necessary to accommodate the order relations present among the input variables by means of their interpoint distances in Euclidean space. Others refer to dimensionality when interpreting and labeling directions through the solution space.

Finally, in the context of latent trait theory, Lord (1980) suggests an empirical definition of unidimensionality to provide evidence that the basic assumption of the item response models is met:

> A rough procedure is to compute the latent roots of the tetrachoric item intercorrelation matrix with estimated communalities placed in the diagonal. If (1) the first root is large compared to the second and (2) the second root is not much larger than any of the others, then the items are approximately unidimensional (p. 21).

Thus, we have another factor analytic definition of dimensionality.

In light of the brief discussion above, it seems reasonable to reconsider the rationale behind the research being conducted for this symposium. Specifically, three methods for analyzing multidimensional item response data are being compared with respect to their abilities to recover the underlying dimensionality of some systematically constructed real test data. Being real test data, their true dimensionality cannot be known with certainty even by the moderator or discussant. For example, the subtests from which the items were drawn may not have themselves been purely unidimensional, or the subtests may have been intercorrelated to some degree. Because the content areas measured by the items are unknown to the investigators, subject matter cannot serve as a subjective guide for interpreting the results (often a self-fulfilling prophecy).

However, these objections in no way negate the merit of the present research. They simply help to clarify what is meant by the underlying dimensionality of the dichotomous item response data being analyzed. A reasonable operational definition of dimensionality recovery might be stated as the degree to which a technique is able to classify accurately each of the items as belonging to the subtest from which it was drawn originally.

Choice of Proximity Coefficients

Typically, the data that serve as input to MDS procedures are scored on at least the level of a so-called ordinal scale. Also, in the classical MDS situation, the data are arranged as a square, symmetric matrix containing the proximities of each variable with every other variable, similar to an intercorrelation matrix. Normally only the set of lower off-diagonal elements is used as input because the upper off-diagonal elements are redundant and the proximities of variables with themselves along the diagonal are not of interest. However, with an array of dichotomous t t data, none of the above conditions hold. Because the items are scored as either 0 or 1, the data exist at the level of a so-called nominal scale, there being just two categories of responses. From an extrinsic point of view, order information is present in the binary response data in the sense that achievement level has been mapped into two elements of an ordered set. The element 1 is "better than" the element 0 in that it denotes a higher level of achievement. But intrinsically, because there are only two elements in the set, we may just as well call it an unordered set. Certainly the assignment of the numbers (0,1) is arbitrary. Moreover, measures of association between any given pair of items are insensitive to the ordering (1>0). So we have a situation in which the raw data do not contain the order information we require for MDS.

Another difficulty is that the binary data arranged in the rectangular matrix are more properly characterized as dominance data rather than proximity data. That is, the data values represent the dominance relations of persons versus items, persons versus persons, and items versus items. Therefore, the data must be transformed into a square matrix of proximities that reflect the associations among the items.

Fortunately, a large number of well known measures of association have been developed for the case of two nominal variables crossed in a 2 x 2 table. Unfortunately, we seek proximity measures that are relatively unaffected by variations in item difficulty levels and item guessing levels if we are to avoid artifacts in the MDS solutions.

Only one study was found in the literature (Reckase, 1982) which addressed the choice of a proximity coefficient and its effect on MDS analyses of dichotomous test data. Also investigated in the study were the effects of such variables as item difficulty, discrimination, and guessing levels and the factor structure of the item response data. Thirteen coefficients were studied using simulated 50 item data-sets (1,000 examinees) with one, two, three, or nine factors present, both with and without random guessing effects in the data. A real data-set with two factors was also investigated. The coefficients included the phi, phi/phi max, tetrachoric, agreement, approval, eta, kappa, koppa, Yule's Q, Yule's Y, Goodman and Kruskal's gamma, Kendall's tau B, and the Lijphart index.

With the simulated data-sets, the membership of the items to each of the factors was known, so MDS was evaluated in terms of its ability to classify correctly the items onto their proper factors. In all of the MDS runs, the solutions were restricted to two-dimensional Euclidean spaces. Stress values (Kruskal, 1964) reflecting the goodness of fit of the MDS model to the data were also observed.

In general, the study showed that the MDS results were in large part a function of the particular coefficient used to measure item proximities. Also, coefficients were identified which were sensitive to the effects of item difficulty and guessing levels relative to those which were not so affected. In particular, Yule's Y coefficient (to be described later) performed quite well for the simulated data sets in classifying the items correctly. However, none of the coefficients were able to classify properly the real data items without prior knowledge of their content areas. In addition, it was found that items with low discrimination or high difficulty levels were so dissimilar from the other items that they tended to compress the locations of the remaining items together, thereby distorting clear patterns that might otherwise have been revealed in the space. The presence of guessing in the data tended to reduce the sizes of the proximity coefficients among the items, further clouding the structure.

A frequently recommended procedure to transform rectangular data into proximity data has been to calculate the set of squared Euclidean distances from each variable to every other variable (Kruskal & Wish, 1978; Schiffman, Reynolds, & Young, 1981). In the case of dichotomous item response data, this procedure amounts to summing the squared differences between two column vectors at a time, each of which contains only 0's or 1's. If the patterns of response to the two items were identical, this distance would be equal to zero. If the patterns were exactly opposite, the distance would be equal to the sample size. Also, it is easy to see that this distance would be quite sensitive (inflated) by large differences in the item difficulty levels of the two items.

Working within the framework of partial order analysis to address the issue of dimensionality of dichotomous data, Wise (1982) has proposed making modifications to the squared Euclidean distance. The sample size is used as a scaling constant and the item difficulty levels are taken into account by this new measure, called the Relative Proximity Index (RPI). This index will be described in detail later.

Proximity measures are commonly classified into two basic types, namely similarity data and dissimilarity data. Similarity data are defined such that high positive values indicate that two variables are very similar, while low values indicate dissimilarity. Dissimilarity data are defined such that high positive values indicate a high degree of dissimilarity, while low values indicate similarity. The popular MDS computer routines require that the user specify the input data as belonging to either one type or the other.

This requirement presents a dilemma for proximity measures that range in value from +1.0 to -1.0, such as phi, tetrachoric, Yule's Y, and many other well known coefficients of association. If we specify the data as being similarities, we force the procedure to view a zero relationship between two variables as having greater proximity or association than a negative relationship between the variables. Yet most would agree that a high negative correlation, for instance, would provide as much an indication of the strength of association between two variables as the same correlation positive in sign, and both would indicate an association stronger than zero.

Fortunately, the item intercorrelations for most test data, including achievement tests, tend to contain predominantly non-negative values, a phenomenon so pervasive as to have been canonized as the First Law of Intelligence (Guttman, 1965). However, in situations where sizable negative values are encountered, even after reversing the values for variables scored in opposite directions, the MDS model will have serious problems in fitting the data, and the stress of the solution will be high. A remedy for the situation would be to use a coefficient of proximity having a lower limit of 0.0 and a positive valued upper limit.

METHOD

Proximity Measures

Three coefficients were chosen to transform the rectangular item response data into three separate square, symmetric matrices of proximities for input to the MDS analyses. These coefficients were the phi, Yule's Y, and the Relative Proximity Index (RPI), which are described below using the following 2 x 2 table.

		Item j 0	Item j 1	
Item i	0	a	b	a+b
	1	c	d	c+d
		a+c	b+d	N=a+b+c+d

Note: a,b,c, and d are cell frequencies and N is the sample size.

Phi Coefficient

The phi coefficient is the well known special case of the Pearson product moment correlation coefficient between two dichotomous variables. Its formula is given by

$$\text{phi} = (ad - bc) \,/\, [(a + c)(b + d)(a + b)(c + d)]^{\frac{1}{2}} \quad . \qquad (1)$$

The phi coefficient was included as a proximity measure for two main reasons -- it is known to be sensitive to variations in the respective difficulty levels of the two items being compared and, being so widely used, it was to serve as a benchmark against which the other two proximity measures could be compared. It was treated as similarity data.

Yule's Y Coefficient

Yule's Y, also known as the coefficient of colligation, is formulated specifically to be independent of the marginals of the 2 x 2 table. Indeed, the Reckase (1982) study found Yule's Y to be relatively unaffected by item difficulty variations.

$$\text{Yule's } Y = [(ad)^{\frac{1}{2}} - (bc)^{\frac{1}{2}}]/[(ad)^{\frac{1}{2}} + (bc)^{\frac{1}{2}}] \qquad (2)$$

Yule's Y is jointly monotonic with Yule's Q, is a transformation of the tetrachoric coefficient, and is a special case of Goodman and Kruskal's gamma. Because of these properties, MDS results would be expected to be virtually identical for all four coefficients. Thus, Yule's Y was included because it represents a class of coefficients and because it performed very well with simulated data in the Reckase study. It was treated as similarity data also.

RPI

The RPI (Wise, 1982) is basically the squared Euclidean distance between the responses to two dichotomous items corrected for sample size and item difficulty.

$$\text{RPI} = (b+c)/N[((c+d)/N)(1-((b+d)/N))+((b+d)/N)(1-((c+d)/N))] \qquad (3)$$

The sample size is used simply as a scaling constant, while the other terms in the denominator equal the expected squared Euclidean distance for independent items with the same difficulty levels (marginals) as the values found in the 2 x 2 table. Small values of the RPI (its lower bound is 0.0) indicate close proximity between two items, so the RPI was treated as dissimilarity data. It was included because the squared Euclidean distance is often recommended for the transformation to proximity data, because it does not take on negative values, and because dissimilarity data are usually preferred to similarity data in MDS applications.

MDS Computer Programs

A wide variety of programs is available to perform MDS analyses. For an excellent review, see Schiffman, Reynolds, & Young (1982). Both ALSCAL (Takane, Young, & deLeeuw, 1977) and MINISSA (Roskam & Lingoes, 1970) were chosen for use in the present research. Through proper choice of program options, both perform classical nonmetric MDS, but there are several important differences. The ALSCAL program employs an iterative least squares monotonic transformation between the input data and their corresponding Euclidean distances in space, while attempting to minimize squared stress. The MINISSA program uses an iterative rank image monotonic transformation procedure, minimizes Guttman's coefficient of alienation, and rotates the final solution space along principal axes, although numerous other options are available.

Procedure

The ALSCAL program was run separately for the three forms of proximity data, specifying a range of solutions from one dimension to six dimensions for each. The objective was to obtain stress values as a function of increasing dimensionality for each of the three coefficients. Scree-like plots of stress versus number of dimensions were used to assist in the determination of the "true" dimensionality of the data. Comparisons of the plots were also made with published Monte Carlo results from MDS data-sets containing various levels of random error. In addition, rough comparisons of stress values were made with the expected values of stress both for random data and for data that fit the model (MacCallum, 1981).

The ALSCAL program also provides values for the squared correlation between the Euclidean distances based on the final solution and the corresponding disparities obtained during the monotonic transformation. Roughly speaking, the squared correlation indicates the proportion of variance in the original input data that is accounted for by the MDS model. The R-squared values were also plotted as a function of the number of dimensions in the solution.

Taking into consideration the magnitude of the stress values and the R-squared values, as well as the shape of the plots and Monte Carlo guidelines, the dimensionality of the optimal solution was determined. Once decided, this solution was pre-specified and submitted to MINISSA for replication purposes and to obtain a rotation of the configuration along principal axes. Guttman has found the latter procedure to facilitate interpretations in numerous empirical studies (1982).

Being derived from proximity information, the locations of the items in the final spatial configuration provide information about the similarities of all the items to each other. Therefore, cluster analysis of the item coordinates was performed to identify homogeneous groupings of items in the space. The goal was to classify each of the items into categories and/or regions of the space that could be surmised to represent subtest membership. The CLUSTER procedure in the SAS computer package was used to perform the cluster analyses, using hierarchical clustering based on Euclidean distances among the items.

Finally, in an attempt to ascertain the effects of traditional item discrimination and item difficulty levels (from item analysis) on the final spatial configuration, each was correlated with the set of coordinates locating each item on each of the dimensions.

RESULTS

Dimensionality

The results from running the ALSCAL program for each of the three types of proximity data are presented in Table 1 in terms of both stress and R-squared and are shown graphically in Figure 1. As expected, with increasing dimensionality of the solution, the stress values decreased and the R-squared values increased. The RPI coefficients consistently yielded the lowest stress and highest R-squared values, while Yule's Y always had the highest stress and lowest R-squared values for any particular number of dimensions. However, the general patterns were the same for all three coefficients.

Insert Table 1 about here.

Using the equations given by MacCallum (1981), the expected values for stress using random data were computed to be .460, .354, .294, and .255 for solutions with dimensionality equal to 2, 3, 4, and 5, respectively. Therefore, it appeared safe to conclude than the ALSCAL analyses in the present study did not capitalize on chance unduly. Based on MacCallum's tabled stress values for data that fit the MDS model, the present dichotomous item response data would best be described as having moderate levels of random error rather than low random error. For the RPI solution in three dimensions, stress was about .15 and R-squared was about .90, indicating a reasonably good fit of the MDS model to the data.

Insert Figure 1 about here.

The scree-like plots of stress and R-squared versus number of dimensions are illustrated in Figure 1. Two fairly obvious "elbows" appeared in the plots, one at the two dimensional solution and one at the three dimensional solution. For solutions of four dimensions and higher, the rate of change in both stress and R-squared was constant, implying that the "true" elbow in the plots was at the three dimensional solution. Also, the stress value for RPI in two dimensions was an unacceptably high value of about .20. Therefore, the three dimensional solution was deemed to be optimal for the data. Based on the RPI input matrix, the ALSCAL program yielded the coordinates that located the 50 items in three-dimensional space. These coordinates, along with the traditional item difficulty and item discrimination values from item analysis, are presented in Table 2.

Insert Table 2 about here.

The results of running MINISSA on the three coefficients, specifying solutions in three dimensions, paralleled the ALSCAL results. Kruskal's stress values were slightly lower than for the ALSCAL results, being .151, .160, and .137 for phi, Yule's Y, and RPI, respectively. The MINISSA program provides a measure of fit called the coefficient of alienation (K), sometimes preferred as a measure of fit because, unlike stress, it has both an upper and lower bound (stress has no upper bound). The K is defined as the square root of the quantity, one minus the coefficient of determination (R-squared). Guttman, while insisting that "no coefficient has anything to do with choosing dimensionality," still suggests a "rule of thumb" that K should be .15 or less (1982). Because K-squared plus R-squared equals 1.0, a value of .15 for K corresponds to a correlation coefficient of about .99. The K for the RPI data in three dimensions was .144, again suggesting that reasonable fit was attained in three dimensions.

Descriptive Correlations

The correlations between the traditional item difficulty and item discrimination levels from item analysis and the coordinates of the three dimensional MDS solutions are presented in Table 3. The main purpose of these analyses was to determine the degree to which the MDS configurations based on the phi, Yule's Y, and the RPI proximities were affected by variations in item difficulty and item discrimination levels. It was expected that the item difficulty levels would correlate highly with one of the dimensions of the MDS solution based on phi coefficients and that the dimensions of the solutions based on Yule's Y and the RPI would be less highly correlated with item difficulty.

Insert Table 3 about here.

Surprisingly, it was found that the first dimension of the RPI-based solution correlated quite highly (+.97) with item difficulty. A similar result was observed for the MDS solutions based on both phi and Yule's Y, but not to such an extent. Also, traditional item discrimination correlated moderately with several of the dimensions in each of the solutions. Thus, the attempt made in the present study to utilize proximity coefficients that were relatively insensitive to the effects of item difficulty (in particular) and item discrimination was not successful.

Cluster Analysis

The cluster analysis procedure used to interpret the MDS results and to classify items was restricted to the spatial configuration resulting from the RPI input matrix, mainly due to the fact that the MDS model had the best fit to the RPI proximity data. Also, the three-dimensional configurations were roughly the same for all three types of proximity data. However, because the proportion of variance in common between item difficulty and the RPI coordinates on the first dimension was roughly .94, the decision was made to perform the cluster analysis only on dimensions two and three from the ALSCAL solution. In effect, this procedure removed the effect of item difficulty from the clustering of items into homogeneous subsets. The plot of the items in two-dimensional space is shown in Figure 2.

Insert Figure 2 about here.

Because hierarchical clustering was performed, it was possible to follow the progress of the formation of item clusters one step at a time. This progression of clustering provided a great deal more information than simply picking a stage of the hierarchical clustering results with some fixed number of clusters and attempting to classify the items by their cluster membership at that stage. In particular, the patterns of interrelations among the items could be more accurately discerned by observing items that clustered together at one stage and then either did or did not cluster with other items at subsequent stages. Figure 3 presents the results when 19 clusters remained, with previous clusterings indicated.

Insert Figure 3 about here.

Facet Theory / Regional Analysis

Because of the distinctive pattern of clusters, the decision was made to use the Guttman facet theory/regional analysis method of interpreting the space of the MDS solution (see Levy, 1981, for an excellent treatment). However, strict regional hypotheses expressed in the form of mapping sentences could not be formulated due to the absence of knowledge about the content of the items. Even so, extrapolating from the results of Guttman in the area of intelligence testing (1965), it was believed that two basic facets would partition and describe the regionality of the space, one facet playing a polar role and the other playing a modular role.

The polar facet was expected to divide the items into separate regions of the space in terms of subject matter or content area. Because no implied ordering is present, in the sense that no content area is better or ranked higher than another, this facet would simply divide the space using axes emanating from a pole or origin, much like slicing up a pie. Based on the cluster patterns, it appeared that four content areas (subtests of items) were present in the original item response data. Figure 4 illustrates the polar axes and shows the items classified into each content area.

Insert Figure 4 about here.

Achievement tests typically are constructed in consideration of the fact that test items measure performance at different cognitive levels. For example, items are sometimes written deliberately to measure performance at the knowledge, comprehension, and application levels of Bloom's taxonomy. In such cases there is a fairly clear hierarchy that orders the items. When rank order information is present, the facet corresponding to this aspect of the variables tends to play a modular role rather than a polar role, where the modular role can be represented as a series of concentric circles spreading out from the origin. The items that grouped together in the innermost circle would be expected to measure quite basic knowledge or general information, perhaps including fundamental rules. Moving outward, successive circles would contain items that were increasingly specific and that might require applications of basic information, up to problem solving and decision making.

Insert Figure 5 about here.

The modular regions derived from the present data are represented in Figure 5, with the polar facet still present in the space. Depending on the content of the items, these three circles might represent the knowledge, comprehension, and application levels of cognitive performance, or perhaps some other taxonomic combinations. At any rate, the general hierarchy of moving from general to specific would be expected in terms of the actual item content and the level of task performance required of the examinee. The results of the regional analysis are summarized in Table 4.

Insert Table 4 about here.

Finally, if Dimension I of the three-dimensional MDS solution were brought back into the analysis, we might say that a third facet of the item response data were present, namely item difficulty. This facet would be expected to play an axial role, being perpendicular to the other two facets on the circular base. Thus, we would think of the MDS space as being in the basic shape of a cylinder, as shown in Figure 6. In other words, items could be located at various heights above the two-dimensional surface previously described, depending on their difficulty levels. This shape has been described as a cylindrex (Levy, 1981).

Insert Figure 6 about here.

DISCUSSION

One major objective of the present research was to attempt to find a proximity coefficient for dichotomous variables that was relatively unaffected by variations in item difficulty. The rationale was that item difficulty would otherwise lead to the presence of an artifactual or nuisance dimension in the MDS results. Two coefficients that attempted to correct for item difficulty, Yule's Y and the RPI, were compared to phi, a coefficient that is known to be sensitive to item difficulty. The results showed that item difficulty effects were present in the MDS solutions regardless of the particular coefficient used. In fact, the RPI proximity measure, ostensibly including an adjustment to take item difficulty into account, produced coordinates on Dimension I that correlated almost perfectly with item difficulty.

These results suggest that the search for proximity measures that are insensitive to item difficulty may be misguided and futile, and that a shift in thinking is required. Perhaps it should be acknowledged that item difficulty is indeed a legitimate and reasonable dimension along which items should be expected to vary. Item difficulty may have to be regarded as one important aspect of dichotomous data that characterizes the similarities or dissimilarities of items to each other. In other words, from an MDS perspective, item difficulty may be a real dimension and not an artifact.

The use of MDS to perform the classification of items into subtests was largely an exercise in subjectivity. Lingoes (1981) has argued that MDS analysis in the absence of content is bound to be sterile, and some would claim that it is anti-scientific to perform a "blind" analysis that is devoid of a theoretical or conceptual base. However, such methods as factor analysis and MDS have frequently been criticized as being used in ways to prove whatever it is that the researcher wishes to prove. Therefore, the present research paradigm could be argued to comprise a relatively objective and fair test of three analytic procedures, albeit an extremely frustrating one to the investigators.

In the absence of knowledge of item content, some popular methods of interpreting MDS results had to be ruled out. These methods included regression or canonical correlation of external criteria on the MDS stimulus coordinates, as well as the use of the magnitudes and signs of the coordinates themselves to attempt to describe meaningful directions in the space.

The Guttman approach using facet theory and regional analysis to partition the space could be applied, but not in standard fashion. This was because knowledge of the content of the variables was lacking. Therefore, no specific hypotheses could be formulated nor mapping sentences generated. Despite these limitations, sufficient results have been published based on intelligence test data that the possibility existed to extrapolate to unknown achievement test data. The degree to which the generalization was successful can be determined only with knowledge of the item contents and subject matter areas from which the original items were drawn.

REFERENCES

Christofferson, A. Factor analysis of dichotomized variables. Psychometrika, 1975, 40, 5-32.

Guttman, L. The structure of interrelations among intelligence tests. Proceedings of the 1964 Invitational Conference on Testing Problems. Princeton, NJ: Educational Testing Service, 1965, 25-36.

Guttman, L. Personal communication. Seminar in Nonmetric Data Analysis, Psychology Department, University of Texas, Austin, Fall 1982.

Kruskal, J.B. Nonmetric multidimensional scaling: A numerical method. Psychometrika, 1964, 29, 115-129.

Kruskal, J.B., & Wish, M. Multidimensional scaling. Beverly Hills, CA: Sage, 1978.

Levy, S. Lawful roles of facets in social theories. In I. Borg (Ed.), Multidimensional Data Representations: When and Why. Ann Arbor, MI: Mathesis Press, 1981.

Lingoes, J.C. Testing regional hypotheses in multidimensional scaling. In I. Borg (Ed.), Multidimensional Data Representations: When and Why. Ann Arbor, MI: Mathesis Press, 1981.

Lord, F.M. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

MacCallum, R. Evaluating goodness of fit in nonmetric multidimensional scaling by ALSCAL. Applied Psychological Measurement, 1981, 5, 377-382.

McDonald, R.P., & Ahlawat, K.S. Difficulty factors in binary data. British Journal of Mathematical and Statistical Psychology, 1974, 27, 82-99.

Reckase, M.D. The use of nonmetric multidimensional scaling with dichotomous test data. Paper presented at the Annual Meeting of the American Educational Research Association, New York, March 1982.

Roskam, E.E., & Lingoes, J.C. MINISSA-I: A Fortran IV (G) program for the smallest space analysis of square symmetric matrices. Behavioral Science, 1970, 15, 204-205.

Schiffman, S.S., Reynolds, M. L., & Young, F.W. Introduction to multidimensional scaling. New York: Academic Press, 1981.

Takane, Y., Young, F.W., & deLeeuw, J. Nonmetric individiual differences multidimensional scaling: An alternating least squares method with optimal scaling features. Psychometrika, 1977, 42, 7-67.

Wise, S.L. Using partial orders to determine unidimensional item sets appropriate for item response theory. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York, March 1982.

Table 1

Stress and R-squared values from ALSCAL Analyses

	Stress			R-squared		
Dimensions	Phi	Yule's Y	RPI	Phi	Yule's Y	RPI
1	.374	.400	.334	.669	.626	.724
2	.222	.239	.195	.809	.780	.853
3	.163	.174	.146	.861	.836	.891
4	.136	.146	.125	.882	.859	.905
5	.114	.123	.103	.903	.883	.922
6	.096	.104	.089	.920	.904	.933

Table 2
Traditional Item Difficulty and Discrimination Values
and ALSCAL Coordinates based on RPI Coefficients

Item	Diff.	Discrim.	Dim. I	Dim. II	Dim. III
1	.929	.269	1.828	-1.443	1.112
2	.863	.250	1.528	1.450	1.309
3	.859	.292	1.633	-0.810	1.059
4	.856	.424	1.252	0.373	-0.281
5	.832	.410	1.151	0.495	-0.102
6	.817	.235	0.424	-2.278	0.818
7	.795	.450	1.065	-0.003	0.469
8	.793	.487	0.860	0.244	-0.266
9	.792	.479	0.999	0.361	0.263
10	.788	.309	1.912	0.334	0.234
11	.772	.431	1.340	0.372	-0.101
12	.762	.455	0.844	0.437	0.687
13	.756	.421	1.018	-0.264	0.162
14	.755	.362	1.349	0.684	-0.463
15	.748	.410	1.085	0.231	-0.498
16	.742	.362	0.491	-1.353	0.432
17	.733	.558	0.644	0.202	-0.118
18	.723	.356	0.260	-0.969	-0.945
19	.721	.402	0.881	0.869	-0.814
20	.714	.285	1.163	-0.274	-1.542
21	.702	.472	0.740	0.166	-0.414
22	.700	.507	0.193	0.677	0.084
23	.700	.398	0.768	-0.565	0.362
24	.678	.361	0.755	0.263	-1.075
25	.672	.519	0.245	-0.147	0.209
26	.646	.483	0.274	0.364	-0.689
27	.645	.408	0.640	0.963	0.573
28	.634	.429	0.228	0.976	0.593
29	.627	.514	0.044	0.474	0.340
30	.572	.297	-0.579	-1.527	-0.973
31	.563	.504	-0.270	0.666	-0.192
32	.561	.219	-0.568	-2.551	0.152
33	.540	.447	-0.315	0.118	0.281
34	.528	.412	-0.670	0.916	0.627
35	.515	.461	-0.642	0.441	-0.321
36	.463	.494	-0.550	0.124	0.225
37	.456	.267	-0.577	-2.259	-0.266
38	.445	.395	-0.914	0.173	-0.668
39	.442	.405	-0.420	1.025	-0.456
40	.431	.429	-0.489	0.756	0.679
41	.373	.301	-1.279	1.460	-1.062
42	.373	.356	-1.481	0.548	0.870
43	.344	.353	-1.369	0.792	-0.767
44	.328	.282	-2.218	0.069	0.692
45	.294	.327	-1.631	1.332	0.349
46	.271	.207	-2.543	-0.422	1.281
47	.256	.333	-1.898	0.070	-0.341
48	.241	.262	-2.151	-0.852	-1.113
49	.226	.209	-2.389	-0.558	1.766
50	.146	.086	-2.660	-2.148	-2.162

Table 3

Correlations Among Traditional Item Parameters and the ALSCAL Coordinates of the MDS Solution Spaces

Variable	Item Difficulty	Item Discrimination
Dimension I (phi)	-.77	-.59
Dimension II (phi)	.49	-.27
Dimension III (phi)	.09	.01
Dimension I (Yule's Y)	.70	.61
Dimension II (Yule's Y)	-.48	.26
Dimension III (Yule's Y)	-.05	.06
Dimension I (RPI)	.97	.40
Dimension II (RPI)	.03	.53
Dimension III (RPI)	.16	.02

Table 4

Items Classified into Regions of the MDS Space

Level	Subtest I	Subtest II	Subtest III	Subtest IV
General	4,5,8,11, 15,17,21, 35,47	7,9,10,29, 33,36,44	13,23,25	--------
Applied	14,19,22, 24,26,31, 38,39,43	12,27,28, 34,40,42	1,3,16, 46	18,48
Specific	20,41	2,45	6,32,37, 49	30,50

Figure 2

ALSCAL Coordinates for DIM. II vs. DIM. III Based on RPI Coefficients

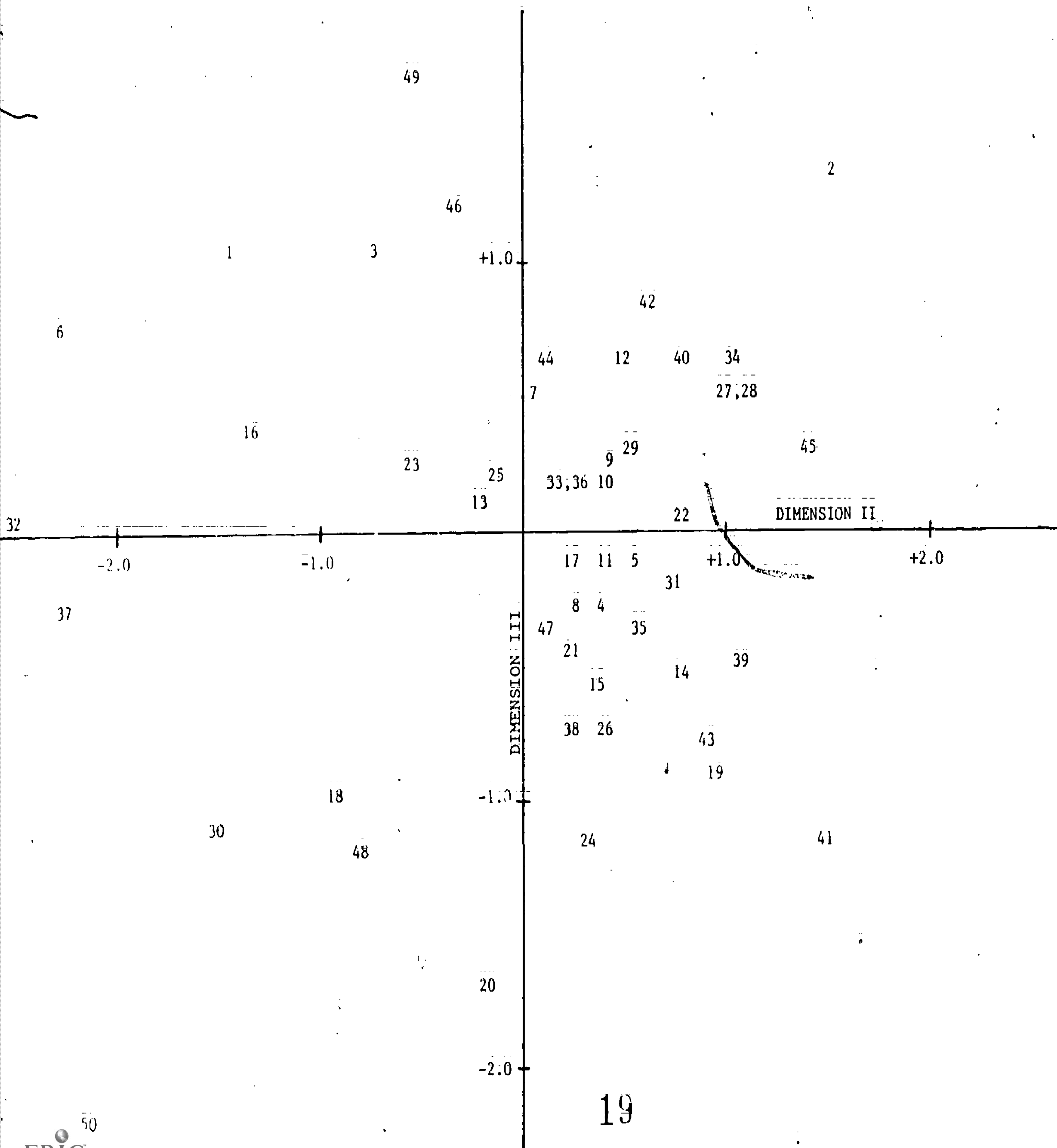


Figure 1

ALSCAL Stress and R-Squared vs. Number of Dimensions

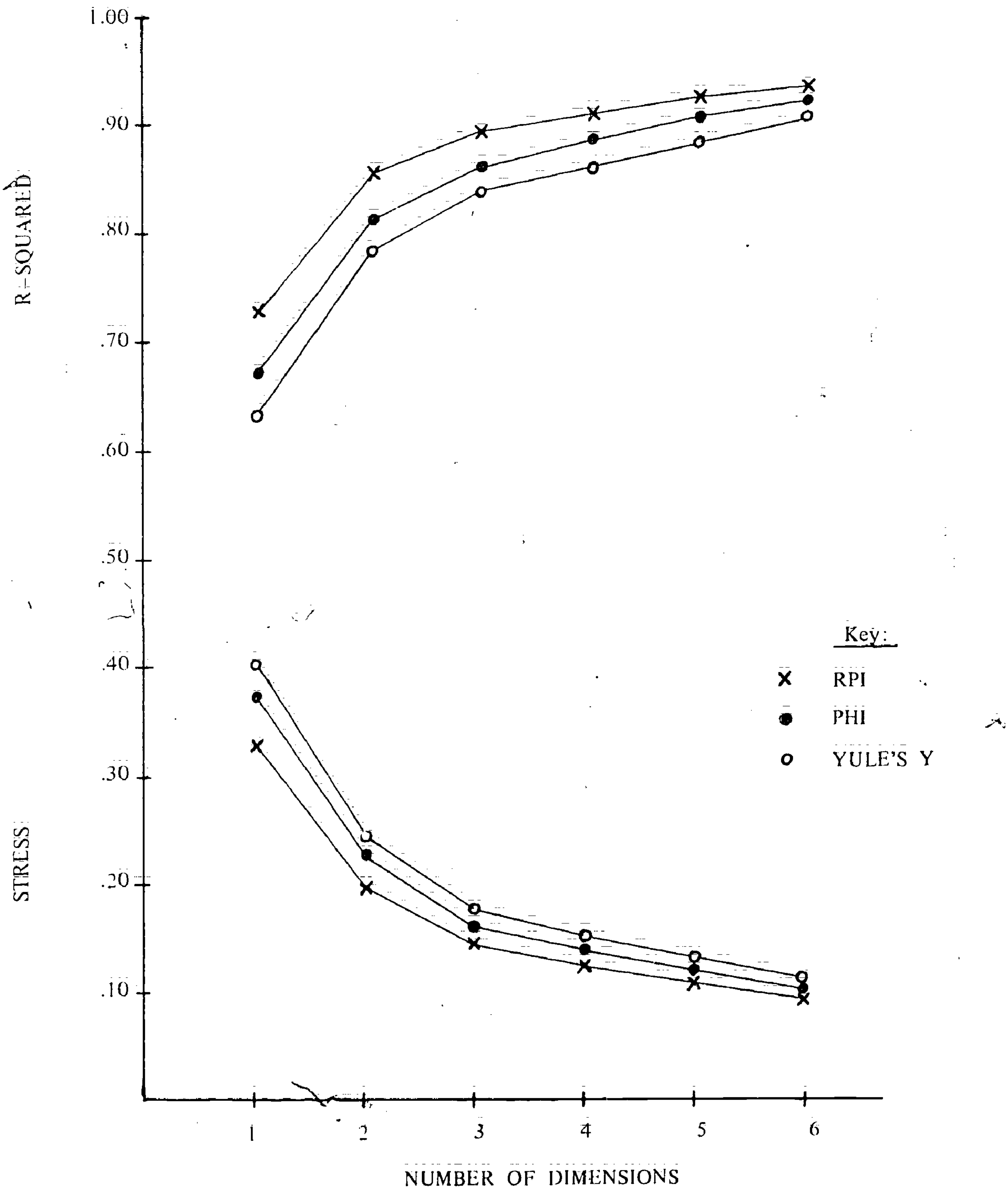


Figure 3

Hierarchical Cluster Analysis of the MDS Spatial Configuration

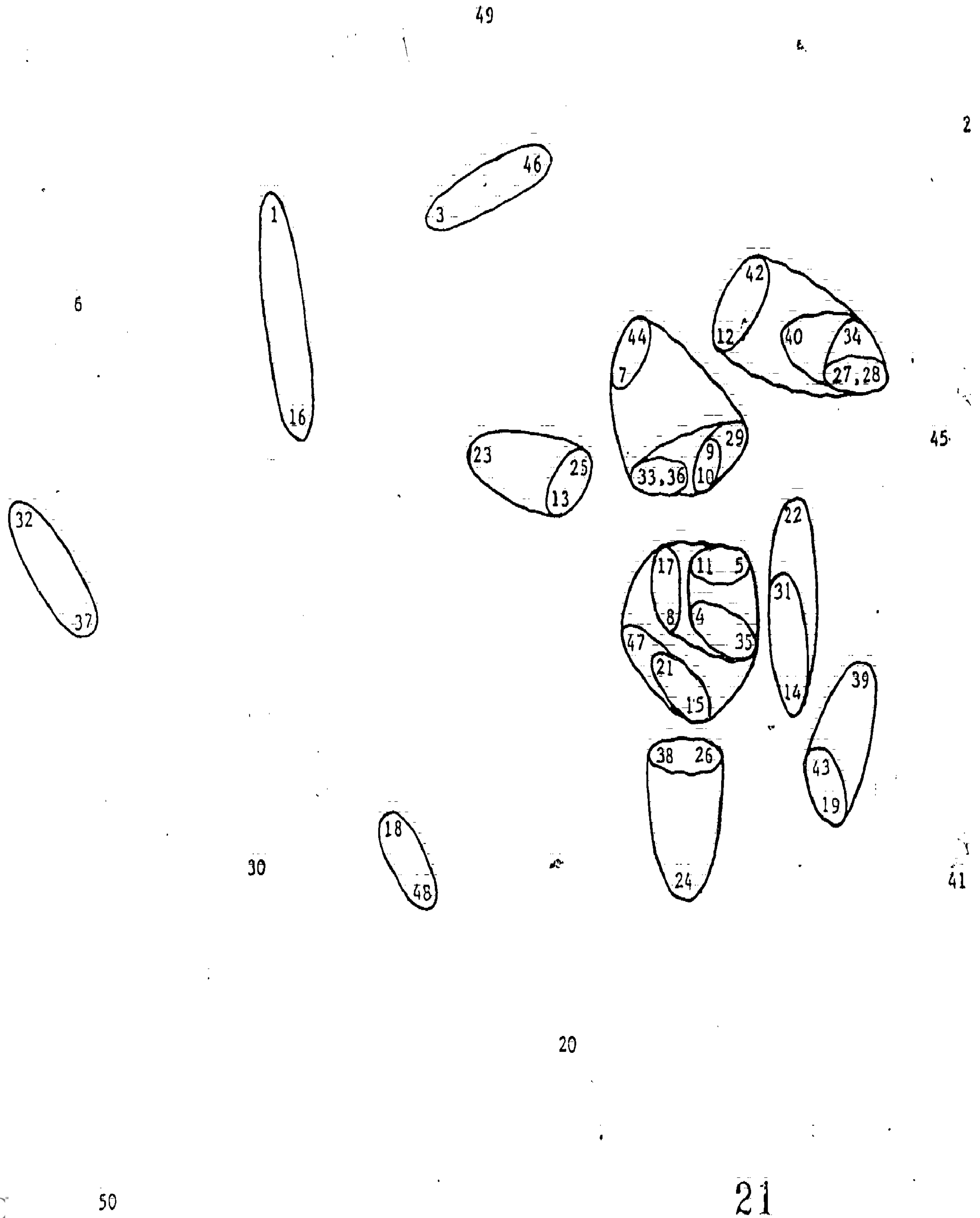


Figure 4

The Polar Facet, Dividing the Space into Four Regions

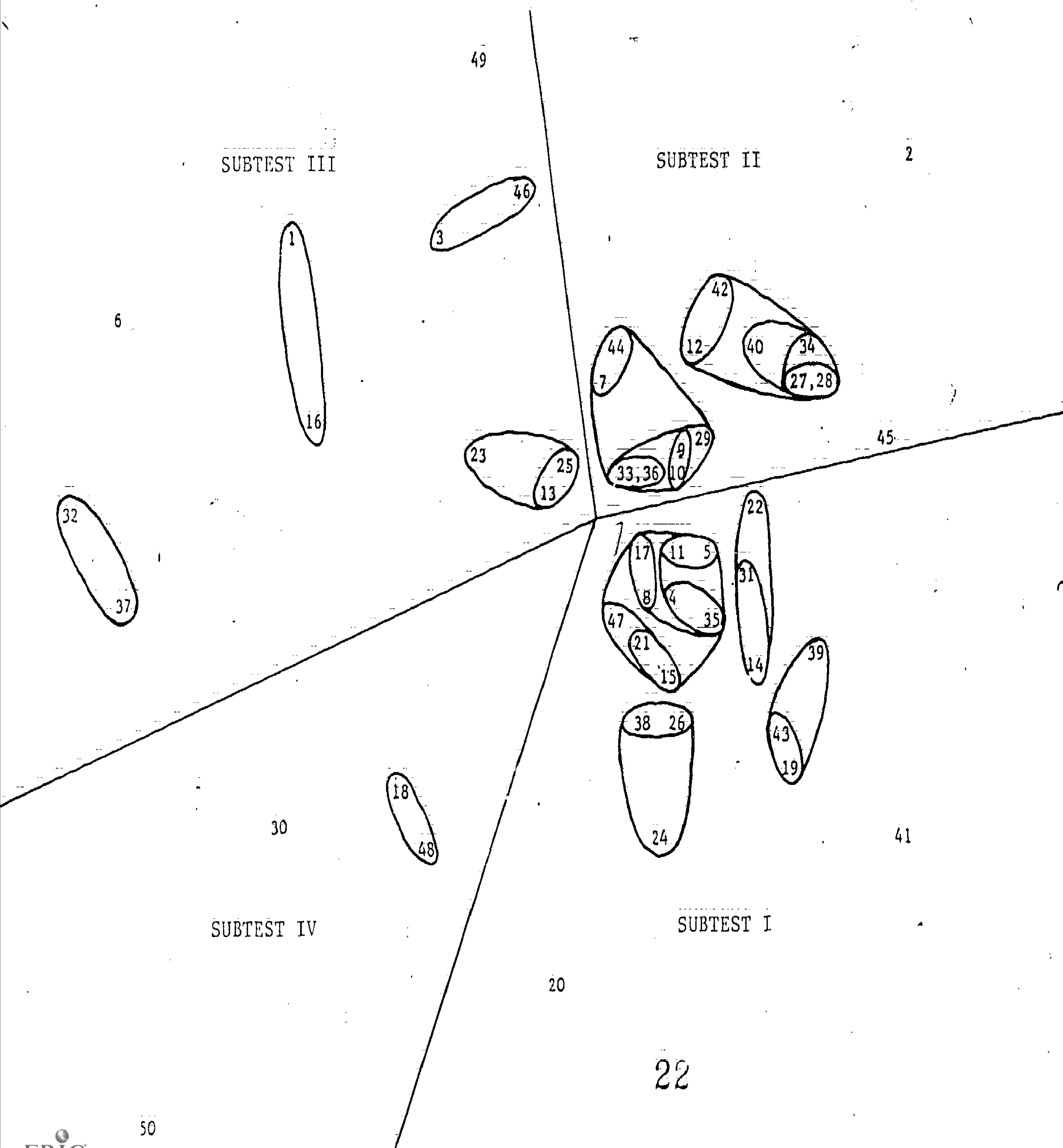


Figure 5

The Addition of the Modular Facet, Ordering Items from General to Specific

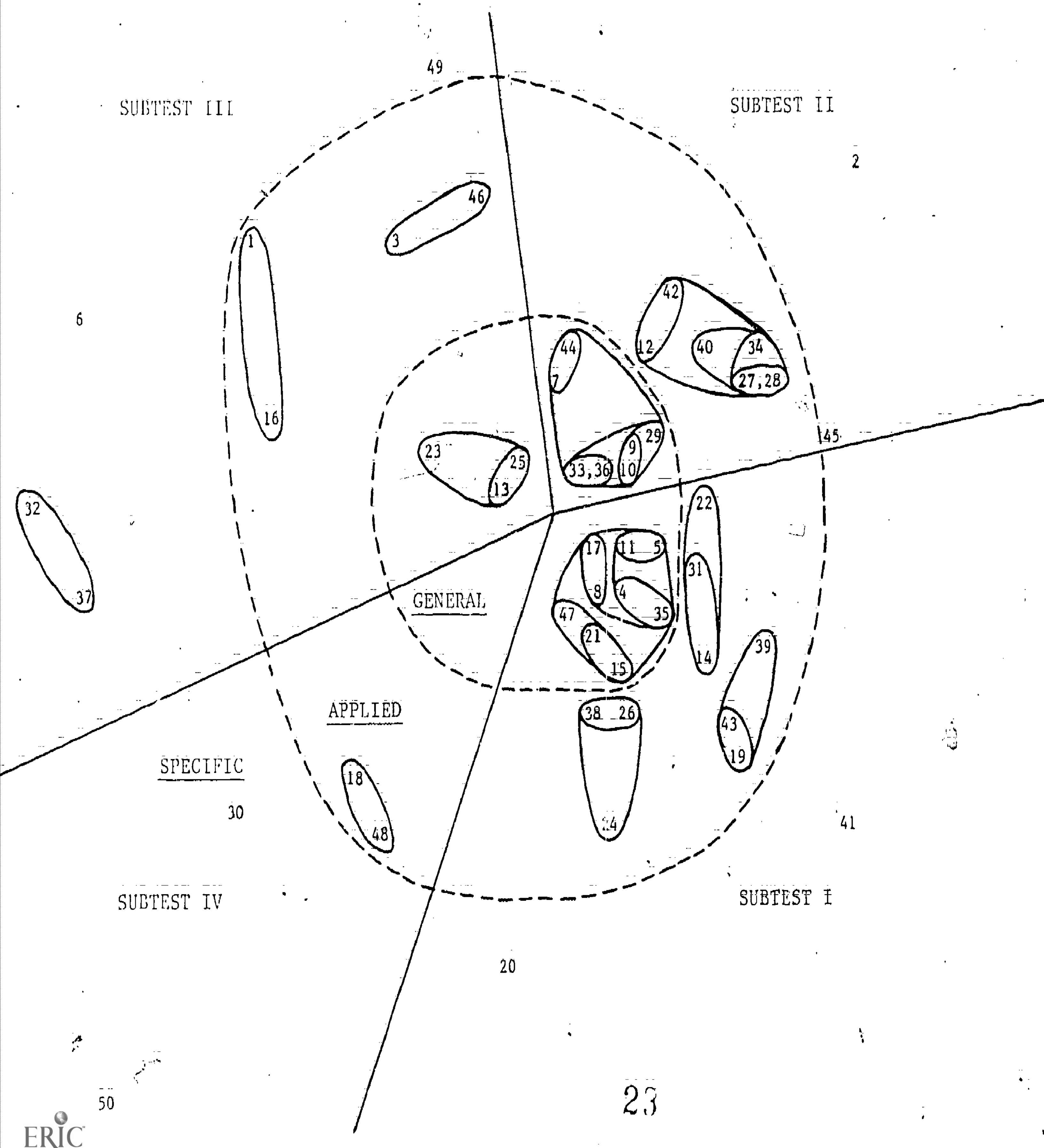
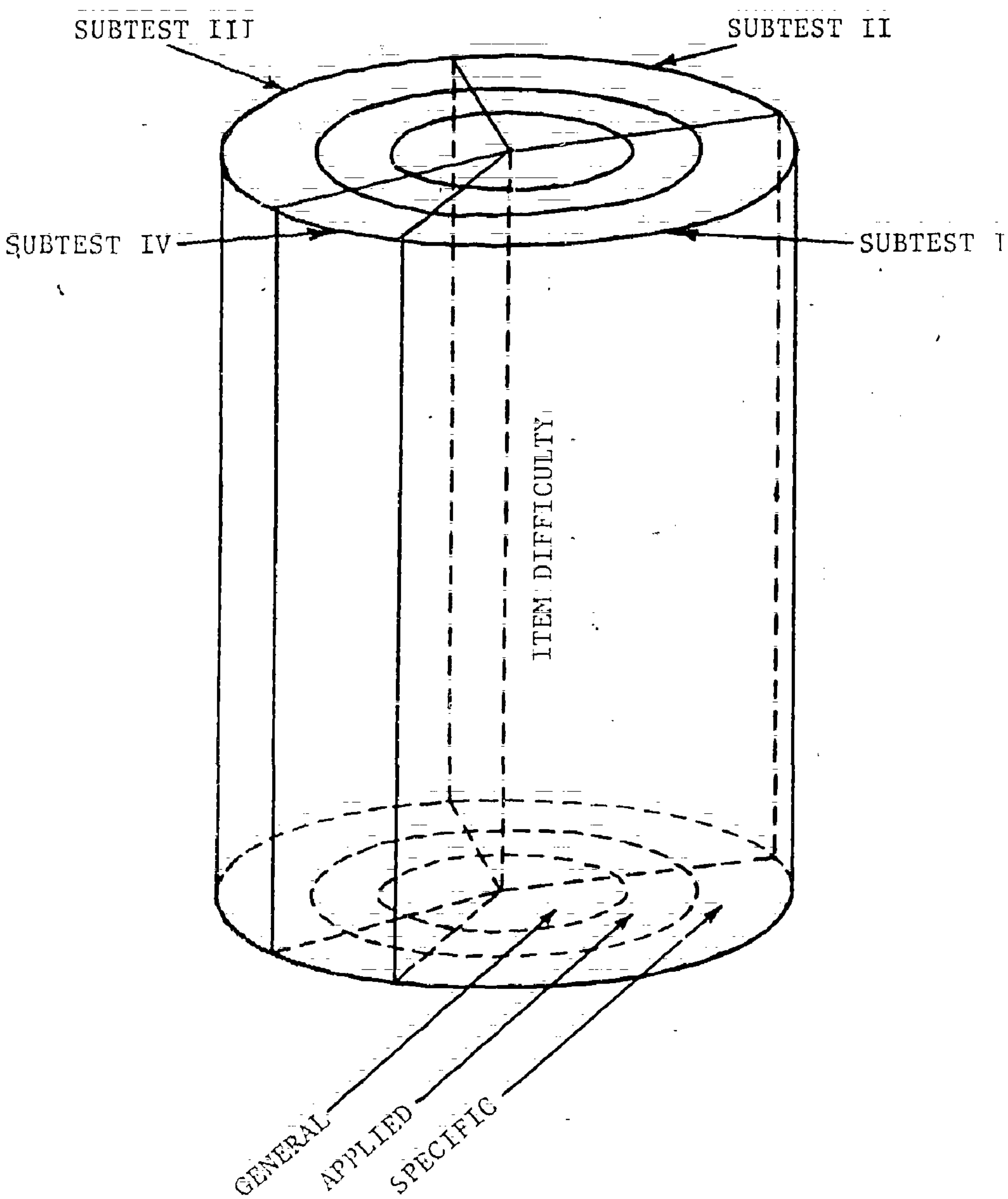


Figure 6

Cylindrex of Achievement Testing



THE ANALYSIS OF DICHOTOMOUS TEST DATA USING NONMETRIC MULTIDIMENSIONAL SCALING

William R. Koch
The University of Texas at Austin

ABSTRACT

The technique of nonmetric multidimensional scaling (MDS) was applied to real item response data obtained from a multiple-choice achievement test of unknown dimensionality. The goal, was to classify the 50 items into the various subtests from which they were drawn originally, the latter being unknown to the investigator. Issues addressed in the research included dimensionality, choice of item proximity measures, and appropriatness of the MDS model for analyzing dichotomous item response data. Three coefficients were chosen to form proximity matrices that reflected the associations of each item with each of the other items. These matrices then served as input to both the ALSCAL and MINISSA computer programs for MDS analysis. A three dimensional solution was found to be optimal, based on stress, R^2, and coefficient of alienation values. Both cluster analysis and regional (facet theory) analysis of the spatial configuration were used to interpret the results.