DOCUMENT RESUME

ED 204 394 TM 810 389

AUTHOR Reckase, Mark D.
TITLE Guessing and Dimensionality: The Search for a Unidimensional Latent Space.
INSTITUTION Missouri Univ., Columbia.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE Apr 81
CONTRACT N00014-77-C-0097
NOTE 24p.; Paper presented at the Annual Meeting of the American Educational Research Association (65th, Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cluster Analysis; *Factor Analysis; *Guessing (Tests); *Latent Trait Theory; Multidimensional Scaling; Statistical Analysis; Test Items
IDENTIFIERS *Unidimensionality (Tests)

ABSTRACT

The purpose of this paper is to examine the capabilities of various procedures for sorting dichotomously-scored items into unidimensional subjects. The procedures include: factor analysis, nonmetric multidimensional scaling, cluster analysis, and latent trait analysis. Both simulated and real data sets of known structure were used to evaluate the procedures. The analysis of the one-factor data sets with varied levels of guessing showed the detrimental effects guessing could have on these techniques. Application of the procedures to the two- and nine-factor data sets showed that factor analysis and multidimensional scaling have promise. The analysis of the real test data showed that the factor analysis procedure was the only one which could do a reasonably good job of sorting items into unidimensional sets under realistic conditions. (Author/BW)

Guessing and Dimensionality: The Search for a Unidimensional Latent Space

Mark D. Reckase
University of Missouri-Columbia

A well known assumption of most of the latent trait or item response theory models currently being used (e.g., Rasch model, three-parameter logistic) is that the responses to the items on a test are dependent on a unidimensional latent trait (Lord & Novick, 1968; Whitely & Dawis, 1974). This assumption not only holds for this new class of test theory models, but also holds for many of the traditional true-score theory based procedures such as KR-20 reliability and item/test correlation indices of discrimination. For some applications of the above procedures, the assumption of unidimensionality is not a serious problem, since techniques are available for developing unidimensional sets of items. However, when dichotomously scored items are used, as is typically the case for multiple choice items, no widely accepted techniques are available for forming a unidimensional space. The problem is further complicated by the addition of guessing when multiple-choice items are used.

The technique typically used to sort items into unidimensional sets is factor analysis, but factor analysis assumes continuous measures and its use with dichotomously scored items results in numerous difficulties (Kim & Mueller, 1978). Christofferson (1975) and Muthen (1978) have developed factor analysis procedures that are specifically designed for dichotomously scored items, but they can only be used with relatively few items (25 or less) and they do not take guessing into account. It is important, therefore, that the currently available procedures for sorting items into unidimensional sets be evaluated, and that the effects of guessing on these procedures be determined. That is precisely the purpose of this paper.

Procedures Available for Item Sorting

In the methodology literature, there are several techniques available that have the potential to sort items into unidimensional subsets on the basis of the responses of a sample of individuals to the items. These techniques differ in their assumptions and their basic underlying model. Along with the factor analysis procedures already mentioned, they include nonmetric multidimensional scaling and cluster analysis. Also, latent trait theory calibration programs in conjunction with goodness of fit tests may be useable to find a set that fits a latent trait model. For the research reported here, each of these techniques will be applied to data-sets of known structure to determine which will best recover the structure. However, before describing the research design in detail, each technique will be described to make clear which

---

Paper presented at the meeting of the American Educational Research Association, Los Angeles 1981. This research was supported by contract number N00014-77-C-0097 from the Personnel and Training Research Programs of the Office of Naval Research.

of the many variations available have been used. In most cases, the most commonly used variation was selected.

Factor Analysis The factor analysis methods most commonly used are the method of principal components and the method of principal factors. These two methods differ mainly in the values placed in the main diagonal of the correlation matrix---1.0 for principal components method, which results in an analysis of all of the item variance, and communality estimates for principal factors method, which results in an analysis of only the common variance. Both of these procedures were used to analyze the intercorrelations between the test items for this study.

In addition to the particular factor analysis procedure used for the study, a correlation coefficient had to be selected for use in determining the relationship between the items. Oftentimes phi coefficients are used for this purpose, but the magnitude of the phi coefficients are affected by the difficulty of the test items, sometimes resulting in artifactual difficulty factors. The common alternative to the phi coefficient in factor analytic work is the tetrachoric correlation. The use of this correlation also leads to some difficulties due mainly to the fact that it does not yield a product moment correlation matrix. As a result the correlation matrices produced using tetrachoric correlations are sometimes not positive-semidefinite, resulting in the failure of the factor analysis procedure to yield meaningful results.

Because of these problems, both coefficients were used in conjunction with the factor analysis procedures in this study. In addition, in an attempt to compensate for the effects of guessing on the test items, the four-fold tables used to compute the tetrachoric correlations were corrected for guessing using a procedure developed by Carroll (1945). Corrected tetrachoric correlations were then computed based on these revised four-fold tables. Thus, the factor analysis techniques were applied to correlation matrices developed using each of the above three procedures.

Once the basic factor loading matrices were obtained using the above procedures, they were then rotated in an attempt to increase the interpretability of the results. Two commonly used rotation methods were selected: varimax and oblimin. The rotations were performed after the number of factors was determined using the scree technique.

Nonmetric Multidimensional Scaling The purposes of multidimensional scaling, as defined by Shepard (1972), are "...(a) of somehow getting hold of whatever pattern or structure may otherwise be hidden in a matrix of empirical data and (b) of representing that structure in a form that is much more accessible to the human eye...". Since determining the underlying structure in a set of item responses will reveal the unidimensional subsets, this type of procedure was also used in this study.

The particular version of nonmetric multidimensional scaling used for this study was the widely available MDSCAL technique developed by Kruskal (1964). This method maps the items into a low dimensional space in such a way that the distances between the items in that space are monotonically related to values of a similarity coefficient applied to the items. Since only the ordinal relation of the distances between the similarity coefficients is used, many more coefficients are appropriate for multidimensional scaling than are appropriate for factor analysis. Because of this fact, and because the correlations used with factor analysis have several problems, such as the effect of difficulty level on the magnitude of phi coefficients and sample size on the stability of the tetrachoric correlation, several other similarity coefficients were used in conjunction with the MDSCAL procedure. Particularly, coefficients that were robust to guessing effects were desired. The total set of coefficients used in conjunction with the MDSCAL procedure included: phi coefficient, Yule's Q, Yule's Y, eta coefficient, approval score, Kendall's tau b, Goodman/Kruskal's gamma, and Lijphart index.

Cluster Analysis Cluster analysis is another procedure that has the potential for sorting items into unidimensional subsets. This procedure also uses similarity coefficients as an indication of the distance between variables (items in this case). Items that have high similarity are considered to be close together and are, therefore, clustered together. Those with low similarity are considered far apart and are not included in the cluster.

Two different clustering algorithms were used in the study. The first, called simply CLUSTER, builds clusters one at a time. It first searches for the two most similar items. These form the beginning of a cluster. Next the item with the greatest similarity with the items in the cluster is found and added to the cluster. This procedure continues until no item has a similarity greater than a pre-set cutoff value. At that point the two most similar items not in the cluster are found to form the beginning of a new cluster. The clustering procedure continues as above until either all items are clustered, or none can be found above the pre-set criterion level for inclusion.

The second clustering procedure used, called HICLUSTER, is a hierarchical clustering procedure. In the procedure, the most similar pair of items are clustered and the pair is considered as a new item. Then the next most similar pair are chosen and clustered. This process continues until all items are paired. These initial clusters are combined to form larger clusters when all points in one cluster are paired with those in another. Clustering in this procedure continues until all of the items are combined into one large cluster.

Since neither of these two procedures require any special properties of the similarity coefficients, all of the coefficients listed in MDSCAL section were used in analyzing the test items. As with the MDSCAL procedure, it was hoped that using coefficients that were not as dependent on metric information would make the procedure less sensitive to guessing and difficulty effects.

Latent Trait Analysis Although the application of latent trait models is not usually thought of as an item sorting technique, some properties of the LOGIST program (Wood, Wingersky & Lord, 1976) for maximum likelihood estimation of the parameters of the three-parameter logistic model suggest that it might be used to find items that measure a unidimensional trait. Reckase (1979) has found that items with high discrimination parameter estimates from the three-parameter model tend to be from the same factor. This is not surprising since Lord & Novick (1968) have shown the latent trait discrimination parameters to be related to the loadings of the test items on the first principal factor of a test. Based on these findings it was hypothesized that repeated application of the LOGIST program to a test, with low discriminating items deleted after each stage of the analysis, would yield a unidimensional set of items. Therefore, this procedure was also used to try to form homogeneous item sets. No similarity coefficients were required for this procedure.

Data-sets Used in this Study

In order to evaluate the procedures listed above on the ability to form unidimensional item sets, types of data-sets were required. First, data-sets with known dimensionality and well controlled guessing levels were required for an initial evaluation of the procedures. These data-sets were produced using simulation techniques. Although these data-sets do not totally match real data, the fact that they have known structure is very helpful for evaluating the item sorting procedures.

The second type of data-set required is the actual responses of individuals to test items. It would be helpful if something were also known about the structure of this data-set, but accurate knowledge is seldom possible. Data from the Iowa Tests of Educational Development, ITED (Lindquist & Feldt, 1972) were used to produce this data-set. More detailed descriptions of these data-sets are given below.

Simulated Data-Sets All of the simulated data-sets used in this study were produced using a modification of a method developed by Wherry, Naylor, Wherry & Fallis (1965). This method randomly generates z-scores to match any desired factor pattern matrix using the linear factor analysis model. For the purpose of this study, the z-scores generated for each item where dichotomized at points corresponding to specified proportion correct difficulty indices to form 0, 1 scores. These dichotomous responses were further modified to reflect the effects of guessing. If a wrong response were generated, it was changed to a correct response with probability equal to the guessing level specified for the item.

Data-sets with three different levels of complexity were generated for this study. These included one-factor, two-factor, and nine-factor data. These different levels of complexity were required to gain understanding of the operation of the item sorting procedures using simple data-sets initially, followed by the analysis of more complex situations.

Guessing level and the distribution of item difficulty were also varied in producing these data-sets. Guessing levels of 0, .05, .15, .25, .35, .45, .55, .65, and .75 were used in the study. Normal and rectangular distributions of item difficulty were also used. The full set of data-sets produced for this study are given in Table 1, along with a label describing each. Data were generated for 1,000 simulated subjects for all data-sets.

Insert Table 1 about here

Real Data-Set. Only one real data-set was produced for this study. This data-set was constructed by randomly sampling 33 items from the 69 items on the Expression subtest of the ITED and combining them with a random sample of 17 items from the 36 on the Quantitative Thinking subtest from the battery. These two subtests were selected as being the most distinct. It was hoped, therefore, that the resulting 50-item test would have two factors. The item sorting procedures were applied to the data-set to see if the two factors could be identified.

Analyses

Two types of analyses were performed on the simulated and real data-sets described above. First, the four analysis procedures were applied to the numerous one factor data-sets to determine the effect of guessing on the procedures in this relatively pure case. Data-sets with guessing varying from 0.0 to .75 were used for this part of the study. The results of the application of the item sorting procedures were used as a basis for interpretation of the further analyses on the multidimensional data sets.

The second type of analysis performed was the application of each of the item sorting procedures to the multidimensional data-sets. In each case the procedures were evaluated on their ability to determine the underlying structure of the data when guessing was a factor. Both simulated and real data-sets were used in this part of the study.

Results

One Factor Data

The first analysis performed was the application of the principal components factor analysis technique to the one-factor simulated data-sets. The analyses were performed on tetrachoric correlations. A total of nine data-sets, each with 60 items, a rectangular distribution of traditional difficulty and guessing levels varying from 0.0 to .75, were used for the initial analyses. All of the items in the data-sets were generated using a factor loading matrix having .9 loadings for each item on the first factor and the remaining variance attributed to error.

The effect of guessing on the size of the first factor in the principal component analyses of these data-sets is shown in Figure 1. The plot shows the percent of total test variance accounted for by the first factor on the test as a function of the guessing level. The figure also shows the KR-20 reliability for each test and the percent of variance accounted for by the first factor when data-sets generated using normally distributed traditional item difficulties were used.

From the figure, it can be seen that the percent of variance accounted for by the first factor of the data sets drops off fairly quickly with increased guessing. The effect is somewhat greater for rectangularly distributed difficulty values. Note that guessing has a much smaller effect on the KR-20 statistic. The results based on phi coefficients had essentially the same pattern, but with a slightly lower percentage of variance accounted for in the first factor.

Insert Figure 1 About Here

The effect of guessing on the factor pattern of the principal component results can be seen in Table 2. The table presents the factor loadings for the first three factors for the unrotated solution derived from the tetrachoric correlations. The items in this test were arranged from hard to easy starting with a .01 theoretical proportion correct and progressing to a .99 theoretical proportion correct at .02 intervals.

Insert Table 2 About Here

The loadings on the first factor shown in Table 2 demonstrate two effects present in all of the other factor analyses. First, the loadings are reduced from the theoretical value of .9 over the entire range. The loadings for the easy items show a reduction to the upper .70's. Secondly, the loadings for the hard items dramatically show the effects of guessing. As the difficulty of the items increase (lower numbered items), the factor loadings decline. The decline starts at around item 30, which has a theoretical percent correct of .59.

The other two factors shown in Table 2 seem to be two guessing factors. Factor II has moderate positive loadings for the moderately difficult items, while Factor III has moderately positive loadings for some of the very difficult items. It seems that these two factors were required to account for the curvilinear nature of the guessing effect. The phi coefficient results were similar, but with lower factor loadings. Also, phi coefficients resulted in difficulty factors defined by the very easy and very difficult items.

The factor analysis procedure was also run on the tetrachoric correlations obtained using Carroll's (1945) correction for guessing. The first three factors of the factor pattern matrix from the principal component analysis are presented in Table 3. As can be seen from this table, the first factors are now closer in magnitude to the .9 used to generate the data. The guessing effect for the difficult items has also deminished. Further, the third factor no longer seems to be related to guessing. Thus, the correction for guessing does seem to have some effect, but the presence of the second factor that still seems to be related to guessing indicates that all of the guessing component has not been removed.

Insert Table 3 About Here

The MDSCAL and cluster analysis procedures gave conceptually similar results to that obtained by the factor analysis when applied to the one-factor data with a guessing component, but of course the representation was different due to the different form of the analysis model. The effect of guessing on the MDSCAL results is shown by the dispersion of the hard items from the tight grouping of the easier items on the test. Figure 2 shows this effect for a 50 item simulated test with rectangular distribution of item difficulties and .25 guessing level. The MDSCAL analysis was performed on tetrachoric correlations. Figure 3 shows the same plot when no guessing effect was present in the data. For both plots, the items were numbered from hard to easy with the low numbered items being the hard items.

Insert Figures 2 & 3 About Here

The CLUSTER and HICLUSTER procedures resulted in similar guessing effects to those shown for the MDSCAL procedure. The 32 easiest items (with one or two exceptions) were grouped together into one major cluster, while the difficult items that were affected by guessing were included in several smaller clusters. These smaller clusters were absorbed into the large cluster when no guessing component was present. All of these results for the other similarity coefficients were much the same so they have not been presented here for the sake of brevity.

The application of the LOGIST program to the one factor data with a guessing component gave very good results. The program gave uniformly high estimates for the discrimination parameters, evenly spread difficulty parameter estimates, and accurate estimates of the guessing level. This fact was not surprising since the data met all of the assumptions of the three-parameter logistic model. Since guessing is built into this model, guessing effects are of no concern. The program estimates the guessing level rather than having the results distorted by a guessing effect.

Two Factor Data

The real test of a procedure designed to sort items into unidimensional sets is whether the procedure can recover the known structure of a set of data. The simplest data-set that can be used to evaluate the procedures is one that is composed of two factors of roughly equal size. Three data-sets of this type were used for the initial evaluation of the procedures used in this study (See Table 1). All three were generated using a factor loading matrix that had an item factor loading of .9 on one factor and a 0.0 loading on the other. Half of the items loaded on the first factor and half on the second. Rectangular and normal distributions of guessing were used. The guessing levels used were .0, .2, and .25.

Under the no guessing condition, all of the procedures except the CLUSTER procedure using eta coefficients did a good job of sorting the items into groups based on their membership on the two factors. In the case of the CLUSTER procedure using eta coefficients, many small clusters were formed instead of two large ones. A similar result was obtained when the procedures were applied to the two factor data-set with normal distribution of difficulties and guessing components, and with an average guessing level of .2. All of the procedures except the CLUSTER procedure using eta coefficients could easily sort the items into unidimensional sets.

When the two-factor data with rectangular distribution of item difficulties and guessing constant at .25 was used, quite variable results were obtained. The principal component and principal factor results on tetrachoric correlations were able to separate the two sets of items, but the results were not nearly as clear as for the previous two data-sets.

Figure 4 presents a plot of the items from the two factors in the space defined by two VARIMAX factors from the principal component analysis of the tetrachoric correlations. As can be seen from the plot, the items from the two factors are arranged along the two axes of the solution. The more difficult items with a large guessing component are closer to the origin. The closeness of the two sets of items in the area around the origin could indicate possible problems in item classification under less optimal circumstances.

Insert Figure 4 About Here

The results of the MDSCAL analysis of this data varied in quality depending on the similarity coefficient used in the analysis of the data. The configurations based on the eta coefficient, approval score, agreement coefficient, and Lijphart index resulted in poor discrimination of the two factors present in the data. The configurations based on the Yule's Y, Yule's Q, Goodman/Kruskal's gamma, phi coefficient, Kendall's tau b, tetrachoric correlation and the tetrachoric correlation corrected for guessing were easily divided into the sets of items for each factor. Figure 5 presents a plot of the MDSCAL results for the two dimensional solution based on the tetrachoric correlations corrected for guessing.

Insert Figure 5 About Here

The cluster analysis procedures did not do nearly as good a job of sorting the items into the two sets as did the MDSCAL and factor analysis procedures. The major problem in using the cluster analysis procedures was in determining the number of clusters to use as a result. When two clusters were specified for the HICLUSTER procedure the approval score correctly classified most of the items. From eight to 22 items were misclassified using the other coefficients with phi coefficients giving poorest results. It is interesting that one of the worst coefficients for use with the MDSCAL procedures, the approval score, was the best for use with the cluster analysis procedure. The results of the CLUSTER procedure were much worse, with seven to 37 items misclassified. Many of the misclassification errors were for the more difficult items.

The application of the LOGIST procedure to the two factor data-sets resulted in very good results regardless of the difficulty distribution and the level of guessing. In all cases, the items from one factor had high discrimination parameter estimates while those from the other had low parameter estimates. There was no difficulty in sorting the items into the two data-sets.

Based on the rather stringent evaluation of the procedures using the two factor data-set with rectangularly distributed item difficulties and .25 guessing level, it would seem that the factor analysis, nonmetric multidimensional scaling, and latent trait analysis have promise for sorting items into unidimensional sets. Cluster analysis does not look promising because of the difficulty in determining the appropriate number of clusters. The three selected classes of procedures will now be further evaluated using a nine factor data-set.

Nine Factor Data

In order to evaluate two procedures using simulated data-set that had more realistic properties than the two-factor data-sets, a nine-factor data-set was produced. This data-set was generated to have a general first factor with .5 loadings for each item and eight other factors with .5 loadings for five or six items and near zero loadings for all others. Item difficulties and guessing levels were normally distributed for this data-set. The mean guessing level was set at .2. The test data was generated using positive loadings on the factors for all items since it was felt that the items on the typical ability test were positively correlated.

The principal component analysis of the nine factor data using tetrachoric correlations resulted in a very clear sorting of the item sets present in the data. After the VARIMAX rotation was performed on the 12 factor principal component solution, each of the sets of five or six items was well defined by high loadings (.60 to .80) on its own factor. All other loadings on the factor were below .20. The analysis of phi coefficients gave a similar result, but with lower overall magnitude of the factor loadings. Since a normal distribution of item difficulties was used for this data-sets, eliminating the items of extreme difficulty, no difficulty factors seemed to be present in the results.

Most of the applications of the MDSCAL procedure also did a good job of sorting the items into unidimensional sets. As was found previously with the two dimensional data, analyses based on Yule's Q, Yule's Y, Goodman/Kruskal's gamma, Kendall's tau b, phi coefficient, tetrachoric correlation and corrected tetrachoric correlation performed well. Those based on the agreement score, the approval index, Lijphart's index and the eta coefficient did not adequately sort the items. Figure 6 shows the results for Yule's Q index as an example of these results.

Insert Figure 6 here

Application of the LOGIST procedure to the nine factor data also resulted in an accurate sorting of the items into the appropriate item sets, but the procedure was long and expensive in terms of computer time. The initial calibration of the items yielded discrimination parameter estimates for all items in the .4 to .8 range. By deleting the items with the lowest discrimination parameter estimates and recalibrating, a unidimensional subset of items was eventually obtained. Unfortunately, 10 iterations of item deletions and recalibrations were required to get one unidimensional set. To totally sort the items into sets required a prohibitive amount of computation. Therefore it seems that the LOGIST program does not offer a viable item sorting procedure when numerous dimensions are suspected to be in a test. LOGIST did accurately estimates the guessing level of the items, however, a fact that may prove useful in other situations.

ITED Data

The analysis of the two factor and nine factor data indicates that the factor analysis and nonmetric multidimensional scaling procedures using the appropriate statistics did a good job of sorting the dichotomously scored test items into undimensional sets. Unfortunately, simulated data-sets never have the same properties as real data, so the results given above are only tentative. The procedures must be evaluated using real data to give a true indication of their worth.

As described earlier, a real data-set composed of verbal expression and quantitative items from the ITED was produced from the full test battery for the purpose of performing a realistic evaluation of the procedures. The data-set produced was composed of 33 Expression items and 17 Quantitative items. These two subtests were combined to form a reasonable two-factor data-set using real data.

The first analysis performed on this data-set was factor analysis of the inter-item phi, tetrachoric and corrected tetrachoric correlations. The results of the principal component and principal factor analyses of the data were approximately the same, so only the VARIMAX rotation of the principal component results will be discussed. The results of this analysis were fairly good. Twenty-two of the 33 Expression items were correctly classified, while 14 of 17 Quantitative items were classified together. Thus 72% of the items were correctly placed into the content categories.

The MDSCAL procedure did not fair nearly as well in sorting the items into unidimensional item sets. In no case could a clear distinction be made between the content areas regardless of the coefficient being analyzed. The procedure did tend to cluster together items from the same content area, but these clusters could not be identified without knowledge of the dimensions themselves. Figure 7 shows the results of the MDSCAL procedure for the ITED data using Yule's Q statistic. As can be seen, the quantitative items cluster together, but it would be difficult to separate the content areas without some advanced knowledge of the content breakdown.

Insert Figure 7 about here

Discussion and Conclusion

The purpose of this paper has been to examine the capabilities of the various procedures for sorting items into unidimensional subsets. The procedures include: factor analysis, nonmetric multidimensional scaling, cluster analysis, and latent trait analysis. Both simulated and real data-sets of known structure were used to evaluate the procedures.

The analysis of the one-factor data-sets with varied levels of guessing showed the detrimental effects guessing could have on these techniques. Guessing reduced the magnitude of the factors in the data and sometimes induced other "guessing" factors not related to ability or item content. However, these detrimental effects were not noticeable when item difficulty and guessing were normally distributed. Items of more extreme difficulty were required before guessing had a serious effect.

Application of the procedures to the two- and nine-factor data-sets showed that factor analysis and multidimensional scaling had promise as techniques for sorting items into unidimensional sets. Cluster analysis performed well in some cases, but was dropped from consideration because no good way was known to determine the number of clusters. Latent trait analysis also performed well in some cases, but the computational requirements of the procedure made it an impractical procedure for general use.

The analysis of the real test data showed that the factor analysis procedure was the only one which could do a reasonably good job of sorting items into unidimensional sets under realistic conditions. Although the MDSCAL placed items from the same content area near each other in the space defined by the procedure, it did not adequately differentiate the items in such a way that they could be accurately classified. Thus, to the extent that the results of this study can be generalized beyond the simulated and real datasets used, factor analysis still seems to be the method most capable of sorting items into unidimensional sets.

References

Carroll, J. B. The effect of difficulty and chance success on correlations between items or between tests. Psychometrika, 1945, 10, 1-19.

Christofferson, A. Factor analysis of dichotomized variables. Psychometrika, 1975, 40(1), 5-32.

Kim, J. & Mueller, C. W. Factor analysis: statistical methods and practical issues. Beverly Hills, Calif.: Sage. 1978.

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

Lindquist, E. F. & Feldt, L. S. Iowa Tests of Educational Development. Iowa City, IA: Iowa Testing Programs, The University of Iowa, 1972.

Lord, F. M. & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Muthen, B. Contributions to factor analysis of dichotomous variables. Psychometrika, 1978, 43(4), 551-560.

Reckase, M. D. Unifactor latent trait models applied to multifactor tests: results and implications. Journal of Educational Statistics, 1979, 4(3), 207-230.

Shepard, R. N. Introduction to volume I. In R. N. Shepard, A. K. Remney, & S. B. Nerlove (Eds.) Multidimensional scaling: Theory and applications in the behavioral sciences. New York, Seminar Press, 1972.

Whitely, S. E. & Dawis, R. V. The nature of objectivity with the Rasch model. Journal of Educational Measurement, 1974, 11(3), 163-178.

Wood, R. L., Wingersky, M. S. & Lord, F. M. LOGIST: A computer program for estimating examinee ability and item characteristic curve parameters (Research Memorandum RM-76-6). Princeton, NJ: Educational Testing Service, June 1976.

Table 1

List of Simulated Data-Sets

Dimensionality of Data Set	Label*
1-Factor	SD150N.CG00, SD150R.CG00, SD150R.CG05 SD150R.CG15, SD150R.NG15, SD150N.CG15 SD150R.CG25, SD150R.NG25, SD150N.CG25 SD150N.CG35, SD150R.CG35, SD150N.CG45 SD150R.CG45, SD150N.CG55, SD150R.CG55 SD150N.CG65, SD150R.CG65, SD150N.CG75 SD150R.CG75
2-Factor	SD250R.CG00, SD250N.NG20, SD250R.CG25
9-Factor	SD950N.NG20

*The label of the data set describes the data-set. The first two letters stand for simulation data. The next three or four digits tell the number of factors and the number of items. All data-sets contain 50 items. The letter following the 50 tells the distribution of traditional item difficulties: N or R meaning normal or retangular, respectively. Following the period is CG or NG standing for constant or normally distributed guessing. The final two digits give the guessing level. The values given are the guessing level for CG data-sets or the mean guessing level for NG data-sets.

Table 2

Factor Pattern Matrix for a Three-Factor Principal Component Solution to Data-Set SD150R.CG25

Item	Factor I	Factor II	Factor III
1	06	13	31
2	09	23	07
3	17	15	36
4	12	29	42
5	20	36	31
6	25	25	19
7	30	27	38
8	29	43	05
9	41	23	37
10	39	31	32
11	53	25	14
12	49	33	18
13	47	36	-14
14	54	32	14
15	55	40	-03
16	63	31	-01
17	55	36	-04
18	66	24	-15
19	62	32	-05
20	66	28	-07
21	63	34	-26
22	68	26	01
23	66	28	-11
24	74	19	-07
25	71	20	-19
26	72	18	-14
27	70	08	06
28	76	10	-11
29	73	14	-13
30	75	09	-14
31	78	-10	17
32	72	06	-21
33	77	-09	-07
34	79	-08	-09
35	75	-09	-15
36	81	-18	-01
37	77	-08	-07
38	78	-14	-10
39	78	-20	-13
40	78	-25	-05
41	77	-23	-01
42	77	-23	-05
43	79	-32	-01
44	79	-35	13
45	77	-39	01
46	77	-36	-04
47	76	-37	19
48	79	-49	17
49	75	-43	02
50	63	-60	55

Note: Values are presented without decimal points for an unrotated solution.

Table 3

Factor Pattern Matrix for a Three-Factor Principal Component Solution for Data-Set SD150R.CG25 Corrected for Guessing

Item	Factor I	Factor II	Factor III
1	64	75	117
2	78	92	-35
3	99	48	-02
4	66	63	-42
5	71	69	05
6	75	30	-16
7	91	14	21
8	64	51	-51
9	90	09	19
10	72	33	19
11	100	-06	09
12	95	-06	15
13	79	-01	-33
14	95	12	-02
15	94	10	-22
16	100	00	-22
17	80	17	09
18	98	-14	09
19	87	19	05
20	94	13	04
21	82	13	04
22	93	20	00
23	86	19	-26
24	97	03	12
25	93	08	-01
26	90	-01	18
27	89	06	15
28	96	00	11
29	91	12	-02
30	91	02	-37
31	92	-08	-19
32	84	-14	-31
33	88	-21	36
34	91	-17	07
35	82	-43	19
36	90	-48	-18
37	92	14	06
38	87	-22	-04
39	88	-17	-12
40	84	-54	-12
41	86	-13	-12
42	89	-23	-16
43	87	-46	-13
44	93	-14	29
45	86	-24	16
46	88	-29	-36
47	94	-03	16
48	95	-43	02
49	85	-71	-13
50	72	-52	66

Note: Factor loadings are presented without decimal points.

FIGURE 1

THE RELATIONSHIP BETWEEN THE PERCENT OF VARIANCE IN THE FIRST PRINCIPAL COMPONENT AND THE LEVEL OF GUESSING ON THE TEST ITEMS

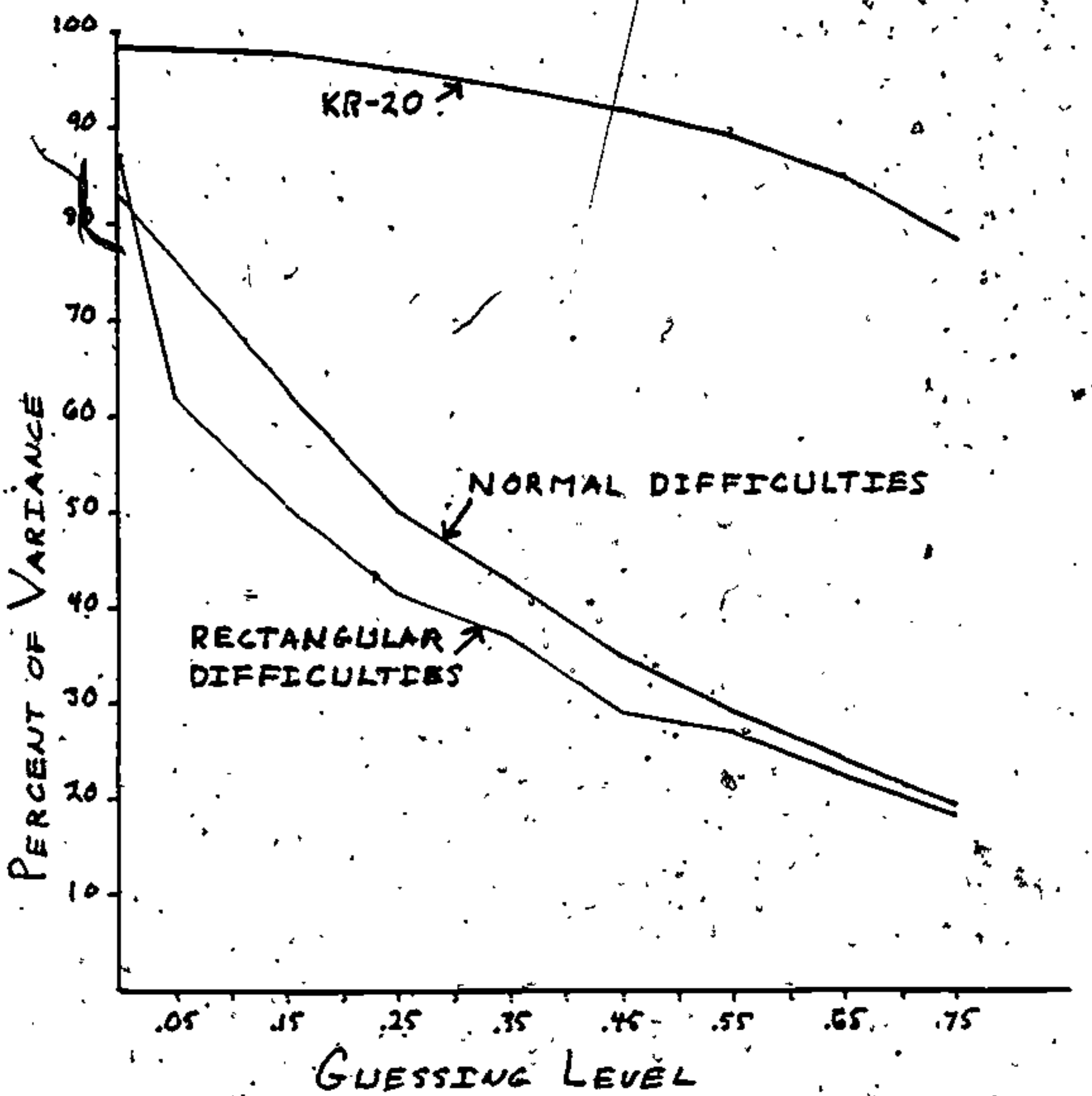

FIGURE 2
TWO-DIMENSIONAL REPRESENTATION OF ITEMS
FROM MDSCAL ANALYSIS
OF ONE-FACTOR, .25 GUESSING DATA

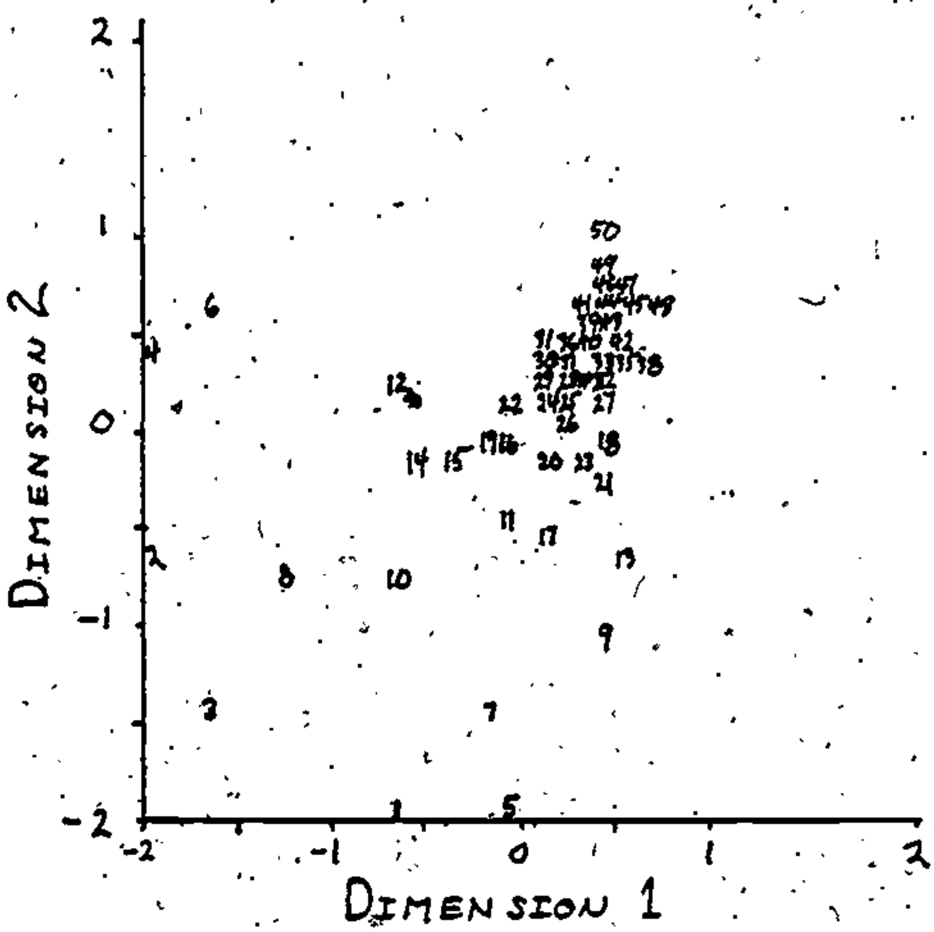

FIGURE 3

TWO DIMENSIONAL REPRESENTATION OF ITEMS FROM MDSCAL ANALYSIS OF ONE-FACTOR, .00 GUESSING DATA

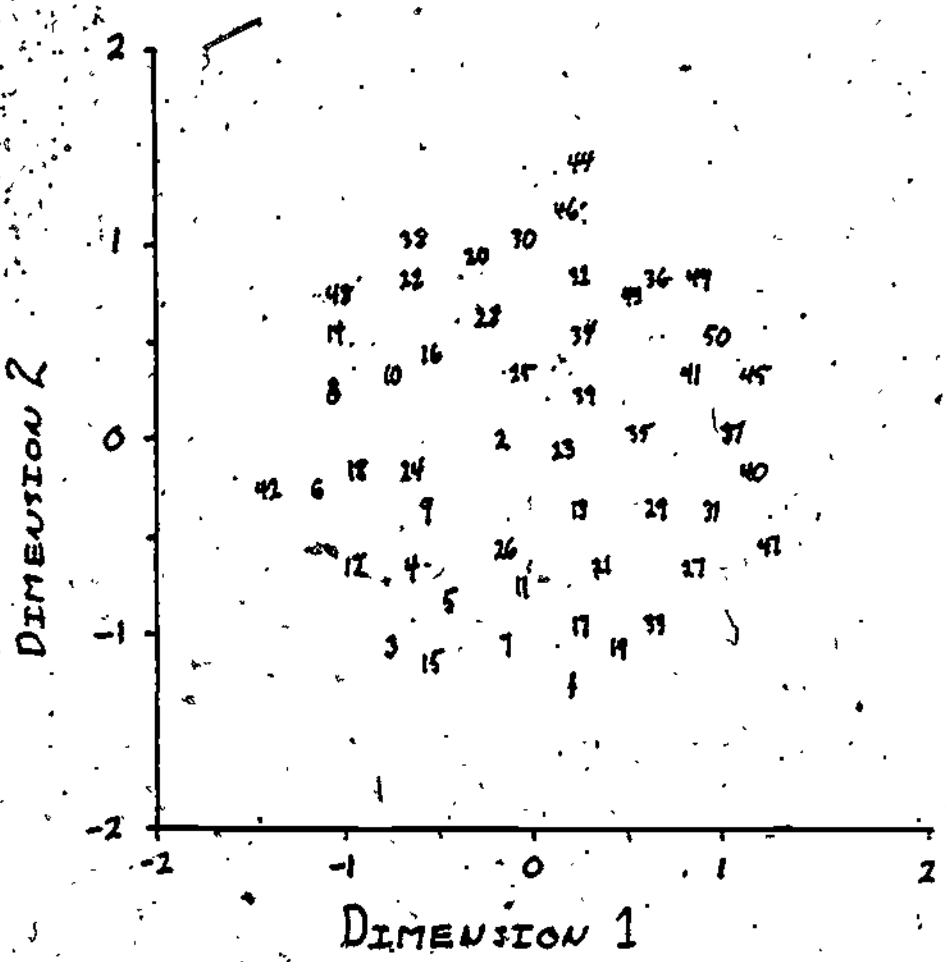

Figure 4

Principal Factor with Varimax Rotation Solution for Two Factor Data with .25 Guessing

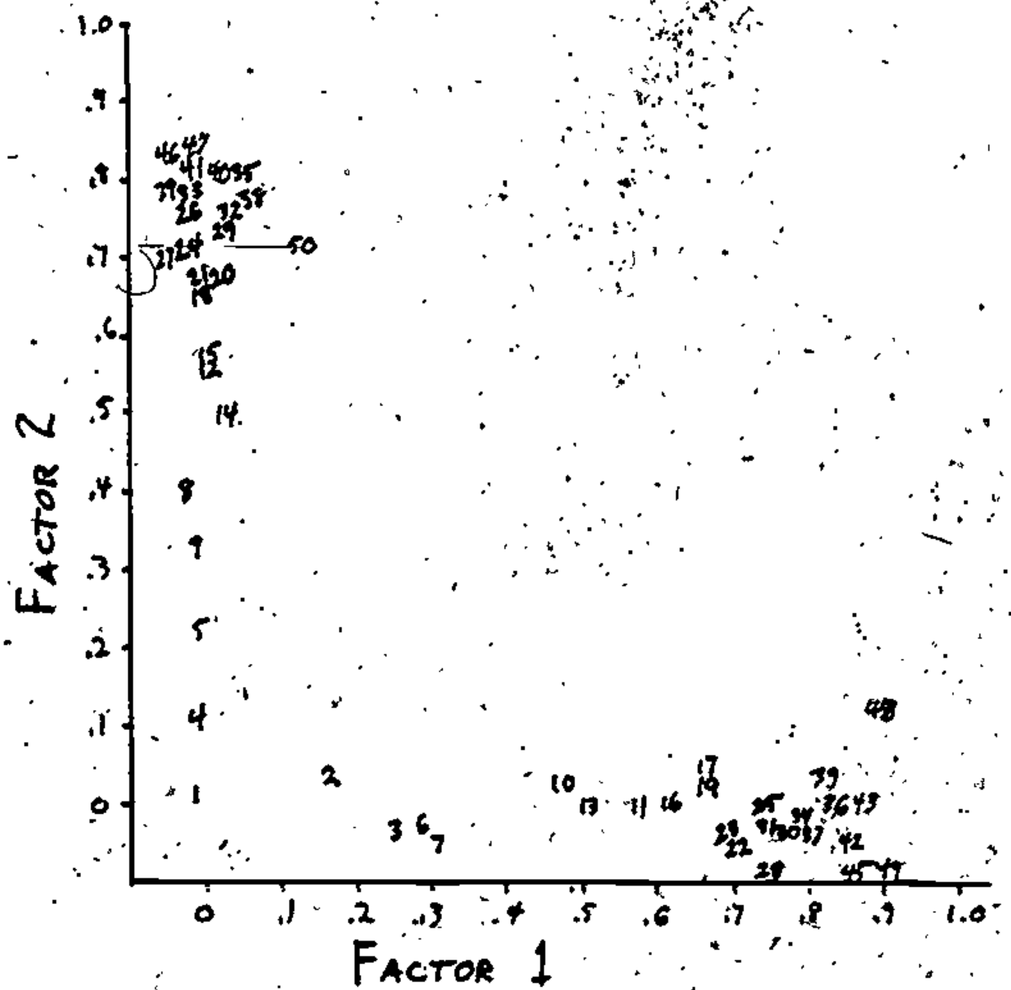

Figure 5
MDSAL Results
for Corrected and Uncorrected
Tetrachoric Correlations
on the Two-Factor Data-Set

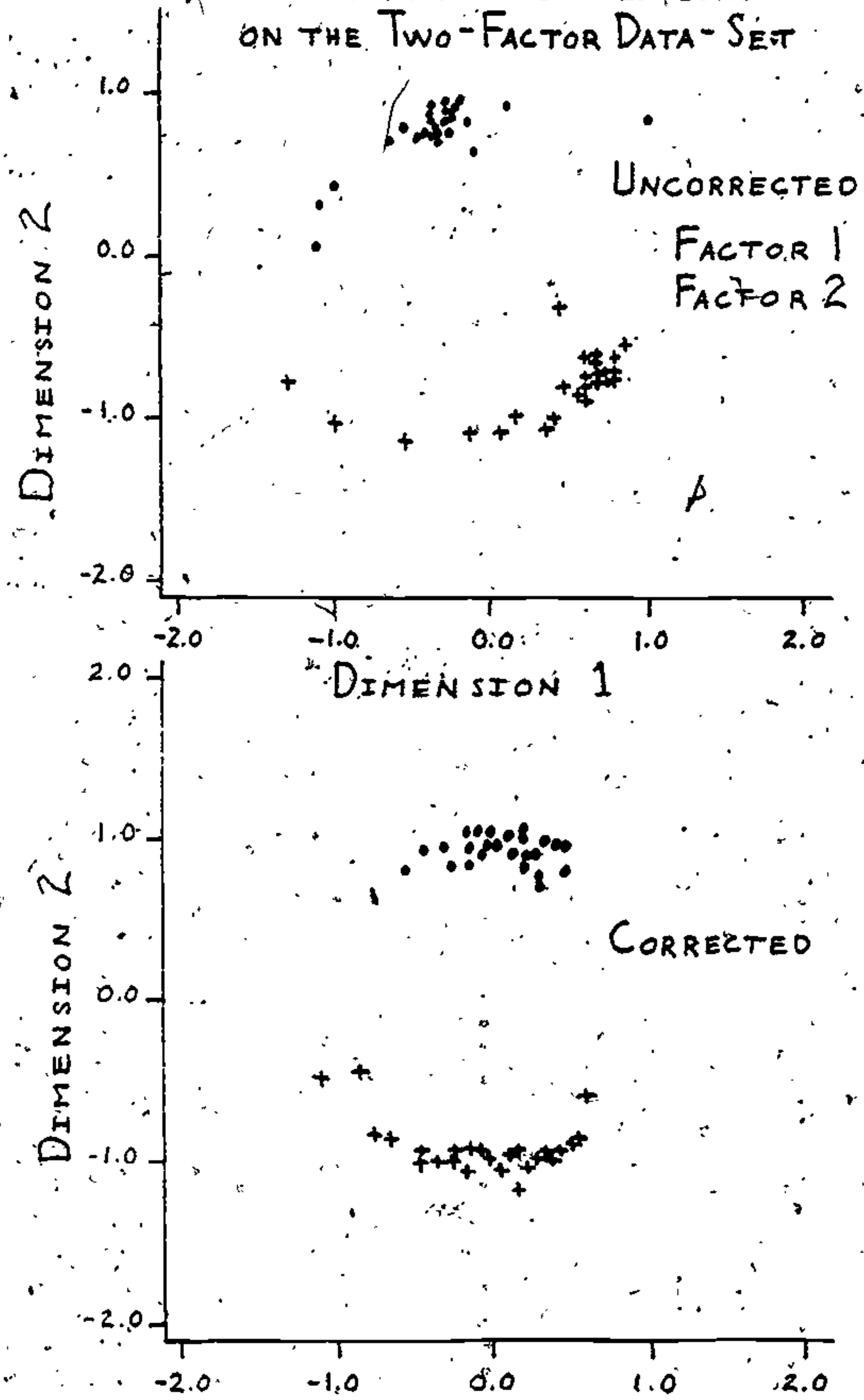

FIGURE 6

TWO DIMENSIONAL REPRESENTATION OF ITEMS FROM MDSCAL ANALYSIS OF NINE FACTOR, .20 GUESSING DATA USING YULE'S Q STATISTIC

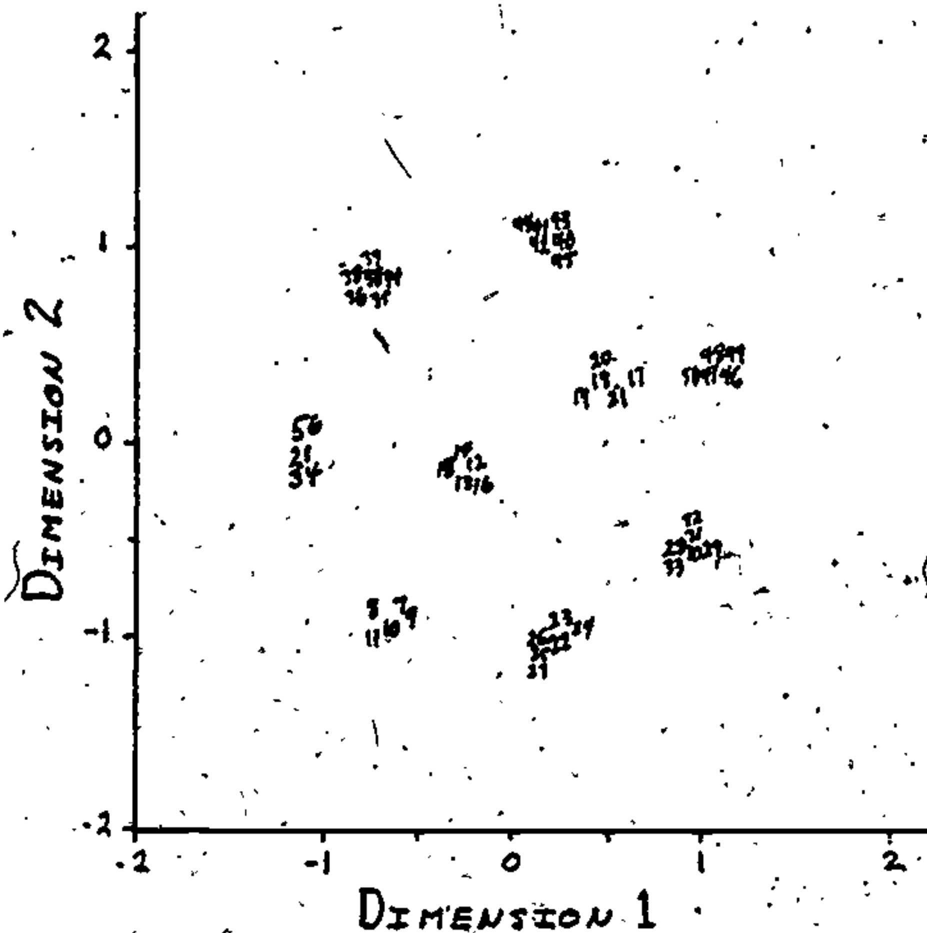

FIGURE 7

TWO DIMENSIONAL REPRESENTATION OF ITEMS BASED ON MDSCAL ANALYSIS OF ITED DATA USING YULE'S Q STATISTIC