

RELIABILITY AND METRIC MULTIDIMENSIONAL SCALING

George A. Barnett

Michigan State University

Department of Communication

A special note of thanks to Joseph Woelfel, Martin Mistretta, James Gillham, William Miles and Paul Karps without whom this paper would have not been possible.

RELIABILITY AND METRIC MULTIDIMENSIONAL SCALING

This study will attempt to examine the phenomenon of aggregate or group reliability for metric multidimensional scaling (MDS). More specifically, it will explore how reliability is affected by the change in sample size for a number of metric multidimensional instruments.

THEORY:

According to Helm, Messick and Tucker (1959:14):

The fundamental concept in multidimensional scaling is psychological distance, which is usually estimated in terms of judgments of similarity among stimuli; i.e., two stimuli judged to be very similar are considered to be psychologically closer together than two stimuli judged to be very different. Given judgments of similarity among all the stimuli in a set, mathematical models exist which provide an interpretation of these psychological distance in terms of Euclidean geometry. The stimuli are treated as points in a Euclidean space, and analytical techniques are available to obtain the dimensionality of the space as well as stimulus scale values determined within a rotation and translation.

Multidimensional scaling models can be used in situations where the stimuli may vary simultaneously with respect to several underlying dimensions or attributes. According to Warren G. Torgerson (1958:248):

The notion of a single unidimensional, underlying continuum is replaced by the notion of an underlying multidimensional space. Instead of considering the stimuli to be represented by points along a single dimension, the stimuli are represented by points in a space of several dimensions. Instead of assigning a single number (scale value) to represent the position of the point along the dimension, as many numbers are assigned to each stimulus as there are independent dimensions in the relevant multidimensional space. Each number corresponds to the projections (scale value) of the points on one of the axes (dimensions) of the space.

Several attributes are unidimensional, among them; weight and length. Spatial position, however, varies along three dimensions, height, length and width. Spatial position on the surface of the earth is also measured with

three dimensions; altitude, longitude and latitude. Knowledge of the position of an object along any one or two dimensions will not locate it precisely in the space. All three dimensions must be known. Likewise, color is said to possess several underlying qualities. The color green can vary simultaneously according to hue, chroma and any other of several different qualities.

Multidimensional scaling has been used in the past to investigate the dimensions underlying perception of colors (Helm, 1959; Messick, 1956a; Torgerson, 1951), the likelihood of two nations going to war (Klingburg, 1941), attitudes towards current social problems, i.e. war, communism, capital punishment and the handling of criminals (Messick, 1956b; Abelson, 1954), and the perception of personalities (Jackson, Messick and Solley, 1957). Data for these studies were all collected by methods which were non-metric in nature. They used either a method of triad combinations, where the subjects were given three stimuli and asked to report which two were similar; or a method of pair comparisons, where the subjects were told to record the similarity of the pair on a discrete Likert-type scale whose range was very limited, nine points at the most.

As Gosta Ekman (1963:33) insightfully wrote:

Most methods of multidimensional scaling have been indirect methods in that they have been based on certain assumptions intervening between the experimental data and the final scale. Most of them have been developed from the unidimensional method of pair comparisons and have been based on similar assumptions. One weakness of these methods is connected with the use of discriminant dispersion as a unit of measurement; although they operate on the confusions between distances and not between stimuli (as in the unidimensional case), the range that can be covered by a moderate number of stimuli is somewhat limited in certain areas, e.g., in typical laboratory work in perception.

Metric scaling deals only with continuous distances over an unlimited expanse--ratio level data by definition, determined by the subjects who receive

the stimuli. It assumes a linear factor model, and that the space in which the concepts will be plotted is homogeneous. Metric scaling is isotropic (non-directional). "In practice, however, only interval scale values are required if a procedure for estimating an 'additive constant' is used." (Green and Carmone, 1971:10-11) There is reason to suspect that this "additive constant" approach is wrong and that a system should be developed which would allow each concept to vary in size by a variable-added system, rather than one which simply adds a constant value to all the unique stimuli.

In metric MDS, a pair of concepts is presented to subjects along with an arbitrary criterion of distance (red and white are 10 units apart) and they are asked to rate the similarity of the concepts. This format has the following advantage. The data is at least interval level making possible certain mathematical manipulations which legitimately cannot be performed with ordinal level data obtained with pair comparison on a directional finite point scale.

Metric multidimensional scaling came into prominence in psychometric theory during the 1950s with the writings of Torgerson, Messick, Abelson and others. However, it soon fell into disrepute because the method was considered extremely unreliable for prediction of the individual case. The use of pair comparisons with discrete point scales is favored today by psychologists because individual scores are more consistent. This method makes it impossible to use mathematical models which provide an interpretation of psychological distances in terms of Euclidean space. This paper will not concern itself with the problems of reliably measuring individual differences; instead it will stress the problem of the reliability of collective group perceptions of behaviors using an instrument that is a metric multidimensional scale.

The significance of this piece of research for the social sciences is to determine how parsimonious metric MDS could be to discover the relationship of various social objects as defined by their spatial locations and determined by collective presentations of a finite group of people. Thus, various social trends can be measured and recorded. This will have potential use in the development of a theory of social change, an area greatly ignored by contemporary social science.

James Gillham, in The Aggregation of Shared Information in a Sociology Department (1972), used a metric MDS instrument to describe the social structure of a finite cultural entity. There was a highly significant relationship between the amounts of information generated about an individual and his movement or change in measured position in the space overtime. The immediate implications to the fields of Sociology, Political Science and Anthropology are clear. Possibly, this instrument could have the greatest impact in the area of Communication theory and research. One would be able to show existing structures and the changes in the groups' shared definition over time as the result of new information entering into the social system.

METHODS:

Two separate pieces of research were conducted and will be reported here in. Both used MDS instruments as described below, but entirely different social spaces were generated.

Data for the first study were collected from a sample which consisted of three undergraduate classes in Sociology at the University of Illinois during the fall semester of 1971. Although the enrollment in these three classes was approximately 800 students, data were collected on about half. At time one,

there were 390 completed questionnaires, and at time two, there were 410; these being the students that came to class on the particular dates. There were only 127 students who participated in both waves.

The following instructions were given to the students:

"DEAR PARTICIPANT:

Recent research has shown that people see different actions as more or less "far apart" or distant from each other. For example, "sitting" is closer to "lying down" than it is to "running." Unlike physical distance which is measured in feet or miles or meters, etc., social distance is measured in "galileos." You are supposed to estimate how many "galileos" apart the following actions are.

Assume that red and white are 10 galileos apart.

How far apart are:"

The subjects were then given seventeen behaviors in a non-random pattern of pairs, creating 136 couples ($N \cdot N - 1/2 = 17 \cdot 16 / 2 = 136$ pairs) to make distance estimates upon. The behaviors were:

- | | |
|-----------------------|------------------------------|
| 1. walking | 10. buying gifts |
| 2. running | 11. using drugs |
| 3. sitting | 12. being anti-materialistic |
| 4. sleeping | 13. anti-government feelings |
| 5. reading | 14. anti-war activity |
| 6. eatin | 15. sexual freedom |
| 7. working | 16. wearing long hair |
| 8. smoking cigarettes | 17. committing suicide |
| 9. driving a car | |

Thus a sample pair would read, "How far apart are: walking and running?
walking and sitting?

. . .
walking and committing
suicide?

Along with the distance estimates, the students were asked the standard demographic items.

Data were collected at two points in time, the initial wave during the first week in November and the second in the first week in December.

Nunnally says (1971:123):

...test-retest reliability should be performed only when, 'there is little reason to believe that memory has a significant effect in making scores on the two administrations similar.' This would be the case when (1) there are many test items, say, over 100; (2) the items are difficult to remember; and (3) the retesting is done after a considerable period of time, say, after at least two months.

This instrument had well over 100 items, but only seventeen stimuli.

Although only five weeks passed between the administrations, the items should have been rather difficult to remember. As Nunnally says later in the same volume (1971:164):

People are simply not accustomed to making absolute judgements in daily life, since most judgements are inherently comparative. Thus, subjects can respond with a high degree of confidence when asked which of two lights is brighter or which of two tones is louder. Whereas people are notoriously inaccurate when judging the absolute magnitude of stimuli, e.g., the length of a line in inches, they are highly accurate in making comparative judgements.

The steps in metric multidimensional scaling data analysis concerning collective representations are as follows. First, one creates a matrix of the mean distances from one concept to another. This assumes that persons are replicates of one another because the operation controls out individual differences.

Second, because the distance of a concept to itself is zero and this value lies on the diagonal of the mean distance matrix, one cannot directly do operations (factor analyses) to find the underlying dimensions. This is because this matrix has no inverse. The problem is compounded by the fact that the true origin of the dimensions is unknown and an arbitrary origin must be chosen.

One solution to this problem is to place the origin at the centroid of all of the stimuli rather than at a particular point. This procedure gives a unique solution and tends to allow the errors in the individual points to cancel one

another. On the average, we would expect an origin at the centroid of all the points to be less in error than an origin at one of the points selected arbitrarily. (Torgerson, 1958: 256-257).

This new matrix, the centroid scalar products matrix, is factorable since it has the squared distance of the centroid to the concept (point) on the diagonals. The distance of each point to each other point through the centroid are in the appropriate cells, such that these distances equal $i_j \cos$, where i is the distance between the centroid and point i , j , the distance between j and the origin and \cos the angle between the two vectors i and j . This matrix is obtained by pre-multiplying the transpose of the adjusted matrix (adjusted to the centroid) by the adjusted matrix. The formulas for these manipulations are located in Torgerson (1958:255-258).

Finally, this matrix is then factored by a principal axis or a jacobi routine to obtain the projections of the stimuli on N orthogonal axes of the space. Principal Axis analysis was performed on the data in this study to obtain the spatial locations in Euclidean space. These spatial locations are the multidimensional scale values.

This "factor" analysis will yield negative eigen values from about the tenth dimension on to the nth factor. The reason for this is stated as follows:

All multidimensional scaling techniques share one assumption commonly--that all the stimuli to be scaled may be represented as points in space. When such analyses are attempted with actual physical objects, however, this assumption fails, since physical objects actually occupy regions in space. If all measurements are made from the respective centers of volume of the object, the results will be a space of three or fewer dimensions in which the location of the object is represented by a point. But if each distance is measured from the respective peripheries of each pair of objects, the result will be a distorted configuration partly projected into imaginary space, since each distance will be exactly the sum of the radii of the two objects too short. (Woelfel, 1972: 101)

This "imaginary space" is indicated by the negative eigen roots because the original distance matrix was not positive semi-definate. If all of the error were removed from the distance matrix, and the size specific to each concept added to the distance matrix; the imaginary space would become the size of the concept, the matrix would become positive semi-definate, and the problem of negative eigen values would be removed. Indeed, if the scaling technique become more reliable as the number of cases increases, the size of the negative roots would decrease to a point where the imaginary loadings would be attributable only to the size of the concept.

Finally, the configuration obtained from the significant factor loadings at time 2, is rotated to congruence on the original (time one) space. This is similar to a Procustes solution, rotating the loadings to a least-square best fit and moving the space until a common origin is shared.

It is the reliability of these points as projections in Euclidean space that we wish to measure. Reliability, defined as consistency of measure, can be estimated by a number of different methods. The first of these, the alternative-form method, must be rejected for estimating the reliability coefficients of MDS. The reason being that it is impossible to construct equivalent forms which are very similar because each unique word has its own connotative aspects, supports the notion of concepts with sizes; circles, spheres or hyperspheres. The advantage this method provides of enabling one to get an indication of error due to sampling of content is irrelevant to MDS because of this researcher's desire to look at the relations between specific concepts. The subdivided-test method is likewise not appropriate for the measurement of reliability of MDS, again because of the necessity to look at a number of specific concepts and the

irrelevancy of the sampling problem. The problem with using the method of internal consistency, coefficient alpha, to measure the reliability of a MDS instrument is that this method assumes unidimensionality of the items in the test. Items vary either along single dimensions or covary along the same dimension, but this approach to measuring reliability does not account for simultaneous variation over multiple dimensions.

This suggests that to get reliability estimates for Metric MDS instruments one must use a test-retest method where the concepts as points (or spatial locations) at time 1 are correlated with the positions of the concepts at time 2. The disadvantage of not being able to get some indication of the error attributed to sampling of content does not concern this researcher. The problem of an individual's memory of the answer at time 1 effecting the responses of time 2 has been discussed previously.

To measure the reliability of the metric MDS instrument, one takes the output of the rotation program and correlates the loadings of the significant factors of all the concepts at time 1 to the same points at time 2. The correlations of all the dimensions must be examined simultaneously, which would dictate a canonical correlation analysis. However, because one rotation is already made, nothing would be gained by another which would be performed in the process of finding the canonicals. Therefore, zero-order correlations are sufficient.

According to the law of large numbers, if one assumes a normal distribution about each mean distance, and then draws random samples from the above population; the larger the sample one selects, the smaller will be the variance in the sampling distribution. Thus, as the sample size increases the standard error of measure in the sample will decrease and the reliability coefficients will

increase positively because of the shrinkage in the variance about the mean in the population. It is this which will be examined in this study.

To see if these coefficients varied positively with the sample size, random samples from the matched pairs (individuals that were in the sample at time 1 and time 2) of varying size (25, 50, 75 and 100 cases) were selected. The number of matched pairs was 127 cases, which limited the variation in the number of the selections.

RESULTS:

The results of the principle axis analysis in all cases (sub-samples, as well as the entire waves) yielded three significant dimensions as determined by the method of the screen test (Tatusoka, 1971:147). The size of the last (seventeenth) negative eigen value for both waves decreased as the size of the random samples increased as indicated by table one.

TABLE ONE

SAMPLE SIZE	NEGATIVE EIGEN VALUE (FACTOR 17)	
	WAVE 1	WAVE 2
25	53.20	79.89
50	44.04	29.48
75	31.32	22.03
100	29.05	21.41
N	(390) 27.96	(410) 21.14

The size of the correlations of the first wave taken as the predictor variables and the second administration as the criterion also increases as the size of the random samples increase. These correlations are presented in table two.

Because of the difficulty of conceptualizing three relationships simultaneously, the results are also presented in graph one.

TABLE TWO
CORRELATION BETWEEN DIMENSIONS

SAMPLE SIZE	<u>DIMENSIONS</u>		
	1	2	3
25	.56	.53	.23
50	.70	.56	.53
75	.71	.73	.43
100	.80	.65	.57

It is clear that as sample size increases, the size of the last negative eigen value decreases and the magnitude of the correlations increase, indicating increased reliability.

Similar multidimensional scaling data was collected for three points in time during the spring of 1972. There were fifteen concepts comparing the spacial representation of environmental concepts with other social problems. The sample from which this data was collected was drawn from an introductory Sociology class, at the University of Illinois. One hundred ten (110) different students participated with 61 completing the questionnaire on all three occasions. The instructions were the same as for the previous study. There were 105 pairs in non-random order. That is 15 stimuli, whose order was random on the questionnaire.

The concepts were:

1. Over Population
2. The War
3. Public Transportation
4. Capitalism
5. The Most Serious National Problem
6. Automobiles

- | | |
|------------------------------|-------------------------|
| 7. Air Pollution | 12. Crime |
| 8. Socialism | 13. Water Pollution |
| 9. Noise Pollution | 14. Conservation |
| 10. Environmental Protection | 15. Recycling Resources |
| 11. Pesticides | |

The dates of administration were, April 1, April 28, and May 19. Thus, Nunnally's criterion for test-retest reliability was not adhered to strictly. Only four weeks occurred between time 1 and time 2, and only three weeks passed between wave 2 and wave 3. There were over 100 items.

Random samples of the matched pairs were drawn of 25, 50 and 61 cases, to further examine the relationship between sample size and reliability. The same operations were performed as with the other data; and, as in the other case, results showed only three significant dimensions.

The size of the last negative eigen value decreased as the sample size increased, although the pattern was not perfect. In this case, it was the fifteenth root. The results are presented in table three.

TABLE THREE

SAMPLE SIZE	NEGATIVE EIGEN VALUE (FACTOR 15)		
	WAVE 1	WAVE 2	WAVE 3
25	48.26	17.50	20.62
50	33.63	8.95	9.82
61	29.80	9.97	9.72
N	(85) 21.40	(108) 9.98	(110) 5.67

The zero-order correlations of the loadings at all three points in time are presented in the following matrices, in table four.

The correlations between time 1 and time 2, along with the coefficients between time 2 and time 3 should be greater than the strength of the relationship between the first and third wave. By inspecting all three matrices, this is the case for the largest sample size. It is not true in both instances with fewer cases.

However, the coefficients do increase as the sample size increases with the most profound jump between 25 and 50 cases and a much smaller one between 50 and 61.

TABLE FOUR

N = 25				Time 1			Time 2			Time 3		
		1	2	3		1	2	3		1	2	3
Time 1	1	1.0										
	2		1.0									
	3			1.0								
Time 2	1	.74				1.0						
	2		.54				1.0					
	3			.52				1.0				
Time 3	1	.86				.74				1.0		
	2		.70				.60				1.0	
	3			.45				.13				1.0
N = 50				Time 1			Time 2			Time 3		
		1	2	3		1	2	3		1	2	3
Time 1	1	1.0										
	2		1.0									
	3			1.0								
Time 2	1	.96				1.0						
	2		.93				1.0					
	3			.84				1.0				
Time 3	1	.97				.95				1.0		
	2		.66				.56				1.0	
	3			.46				.44				1.0

N = 61				Time 1			Time 2			Time 3		
		1	2	3	1	2	3	1	2	3	1	
Time	1	1.0										
	2		1.0									
	3			1.0								
Time	1	.95				1.0						
	2		.90				1.0					
	3			.45				1.0				
Time	1	.95				.95			1.0			
	2		.78				.75			1.0		
	3			.49				.61			1.0	

This should be true for two reasons. One, is that the absolute number of cases is fewer between 50 and 1 (11) than between 25 and 50 (25). Two, as the sample size increases the curve between the number of cases and the coefficients appears to flatten out rapidly. This can be seen in the graph of the curve of two points in time.

The same MDS manipulations were performed on random data in order to gain a base line of the operation. Two sets of random three-digit numbers (000-999) were input as the mean distance of each pair, for 15 imaginary stimuli or 105 unreal pairs. Of interest in this data are the eigen roots from the principle axis factor analysis. It should be clear that there is no obvious set of underlying dimensions as with the real data. These roots are presented in Table 5.

TABLE FIVE

WAVE 1		WAVE 2	
PER CENT VARIANCE	CUMULATIVE VARIANCE	PER CENT VARIANCE	CUMULATIVE VARIANCE
45.86	45.86	44.22	44.22
35.00	80.87	38.10	82.32
25.88	106.75	31.11	113.42
22.77	129.52	29.91	138.33
18.53	148.06	22.66	160.99
16.48	164.53	16.69	177.68
11.17	175.71	12.44	190.12

Only the positive roots are given. The reason for the varying size of the roots is that the principal axis program extracts the largest root first (the vector that explains the most variance) and then the second largest and so on. The absolute size of the eigen roots here is purely random. It will be clearer that there exists no sub-set of underlying dimension less than all the positive roots if one examines the screen lines plotted on graph two.

For continuity, the first three dimensions were extracted and then the spaces rotated to congruence. Then the points were correlated in order to get a measure of reliability. The coefficients are, .21, .01, .38. These correlations are moderately low, but certainly not insignificant. It must be pointed out that unlike normal data these coefficients are not in descending order. The correlation of the two projections of the third dimension is higher than any of the other components. While there is no relation between the number two dimension, the correlation between the remaining pair is small. Thus, one should be cautious of data whose correlation between the points is below .4, because this could be attributed to random variance in the system.

DISCUSSION:

The two different sets of real data produced different levels of reliability. There are a number of reasons for this. First, there was a difference in the samples used. Although they were exclusively undergraduates from the University of Illinois, producing a very homogenous sample, the first study had students from all different levels of school, while the second set of data included only freshmen. Second, the first study was coded and keypunched by undergraduates in an introductory research method course, while the environmental data was pre-coded and the keypunching performed professionally.

Third, and of greater importance is that less time occurred between the administrations of the environmental piece than between the more general study. Three or four weeks as compared to five weeks. However, it must be noted that the reliability coefficients between the first and third waves were still higher than between the two waves of the more general study. The administration in the middle could have familiarized the subjects enough with the test items to produce higher coefficients in the final wave. However, as mentioned previously, subjects are not often required to make absolute decisions and there were over 100 items in both cases. Therefore, the potential effects of individual memory should be minimal.

Fourth, but of questionable validity is the problem of variable test length. The general study had 136 items. The environmental study was 82% as long. There were only 105 pairs to make distance estimates upon. The shorter study was more reliable. Thus, the test length could have produced higher coefficients because there was less fatigue in the respondents.

Finally, and most important to the author is the problem of selection of concepts. The levels of explained variance in the first piece are much smaller than in the second. One reason is that the concepts chosen for the subjects to analyze are more homogeneous in the later study. They all deal with social problems. There were no disparate concepts as in the first, where suicide was present. Because there were no extremely different ideas included all the points are more stable. As with correlation, MDS is sensitive to extreme values. This problem becomes emphasized when the remaining points form a fairly tight cluster. The first dimension, like the regression line, will go from the disperant point through the cluster. Any small change in the location of that point will be amplified because of its relationship to the others. Just as the

regression line and the correlation coefficient change dramatically with extreme values, so will the loadings on specific factors. For example, the mean distance between driving a car and being anti-materialistic was changed from 153 to 1,000 and then the rest of the analysis completed on 100 cases. The reliability coefficients dropped to, (1) .71, (2) .60, (3) .46. This is only about as strong as a relationship as with half as many cases. These altered results are also plotted on graph 1.

The above results would seem to dictate small space analysis, metric MDS on a specific dependent concept and others which are in close proximity to it. The results should be more reliable and the points more sensitive to change with respect to the other concepts in the smaller space, while more stable in the larger space.

The population from which both these sets of data were gathered was very homogeneous. They were all undergraduates at a large university. The sample size at which these coefficients were obtained is not suggestive of the number of cases needed in all instances when using a metric multidimensional scaling instrument. In a "society" that is more highly integrated (as indicated by cultural patterns) than a group of college students, fewer cases might be necessary to obtain this level of reliability. Likewise, in a population of divergent cultural patterns a sample of greater quantity would be necessary to accurately measure the spatial location of cultural concepts in the social space. Gillham reported reliability coefficients above .90 with only 29 cases. The reason was that his sample was from a group, well-defined by its culture and information flows.

Thus, the results do indicate that the reliability does increase with the size of the sample and it should be noted that the curve of the coefficients

does flatten out between 75 and 100 cases in the first study and between 50 and 61 cases in the second. This change in the curve gives rise to the notion that the reliability is dependent on the population size. The author will not suggest the specific number of cases necessary to perform these operations but recommends that one look at the homogeneity of the population and the concepts he wishes to measure before choosing a sample size. Certainly, for a nationwide study, more than 100 respondents would be necessary to obtain acceptable levels of reliability.

The sources of instrument unreliability apparent at this time are; one, errors in scoring, notably keypunching and two, rounding error in the computer analysis. The third source of unreliability is the actual change in the spatial locations over time. Also, the instrument appears to be extremely sensitive to extreme values. Therefore, in an attempt to increase the reliability of the instrument, one could bound the range or likewise attempt to structure the format of the items as has traditionally been done in psychometrics, with the goal to be to control for very extreme values. This would violate the metric qualities of the instrument. Deviate or extreme cases are taken into account in at least three points in the analysis. The first being that taking the mean reduces the effect of the deviate case. The second is the centroid scalar products manipulations for reasons mentioned above. The third, small space analysis, the selection of concepts has been discussed previously.

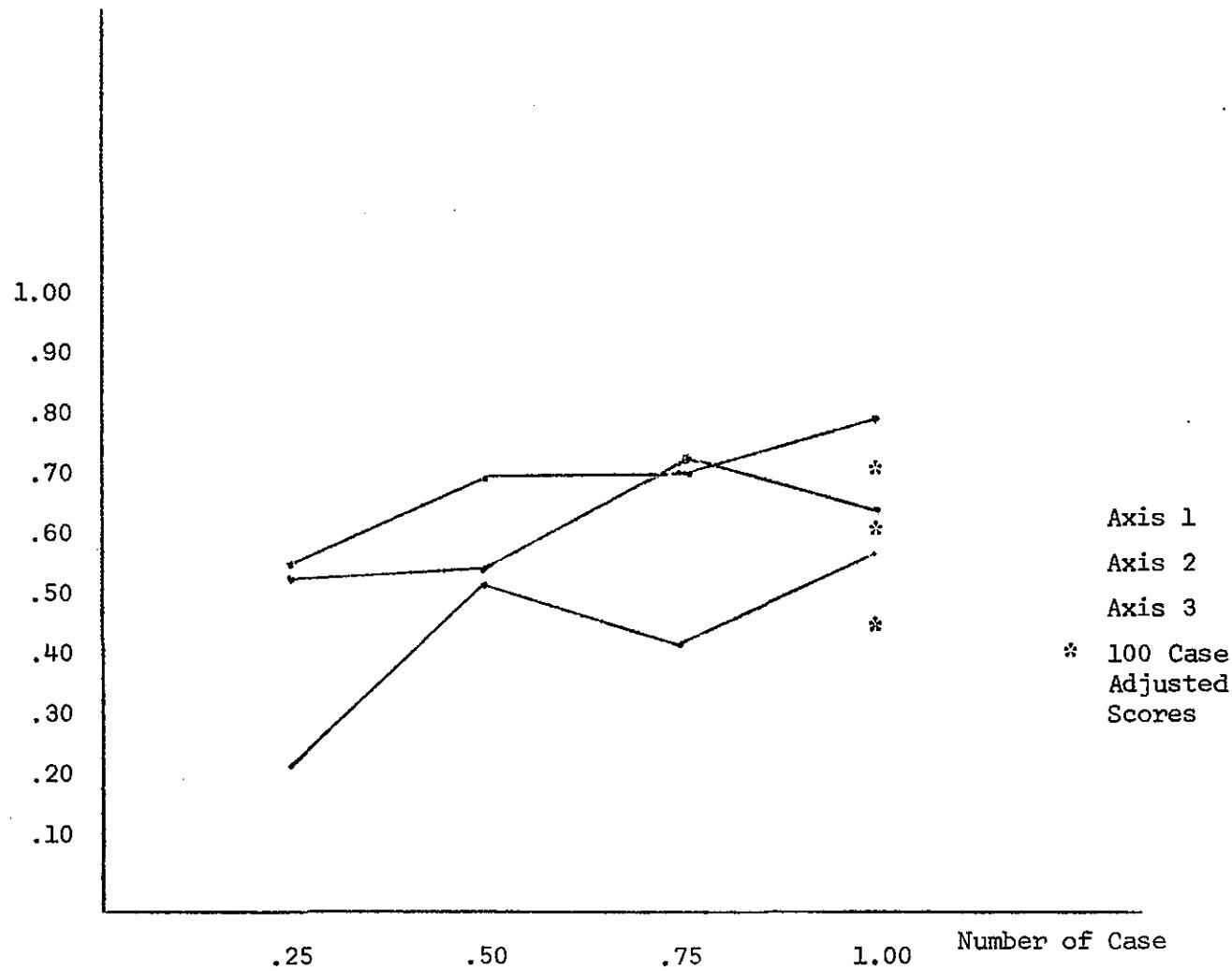
One analysis was performed with suicide, an extreme act in our society. It was compared to more stable activities like walking or eating. This produced extreme distances between pairs. Although suicide is positioned a good distance from the other behaviors in the space, it did not alter the metric and reasonably good levels of reliability were obtained. If limits were placed on pairs,

suicide would have moved closer to the other concepts and while this would increase the reliability, the subjects would be responding to experimenter bias and altering the metric qualities of the instrument.

In conclusion, metric multidimensional scaling instruments as conceived in this paper are potentially useful to explain group perceptions of concepts, despite the unstructured format of the items. This paper has indicated with two separate sets of data that the instrument is increasingly reliable as the size of the sample increases, depending on the specific population in the analysis. Thus, reliability becomes only a function of cost. Because of the continuous, unbounded nature of metric multidimensional scaling, its utilizing as a tool of the social sciences should be immediately apparent, especially for cross-national research or to measure long term changes over time where a structured instrument may impute experimental bias.

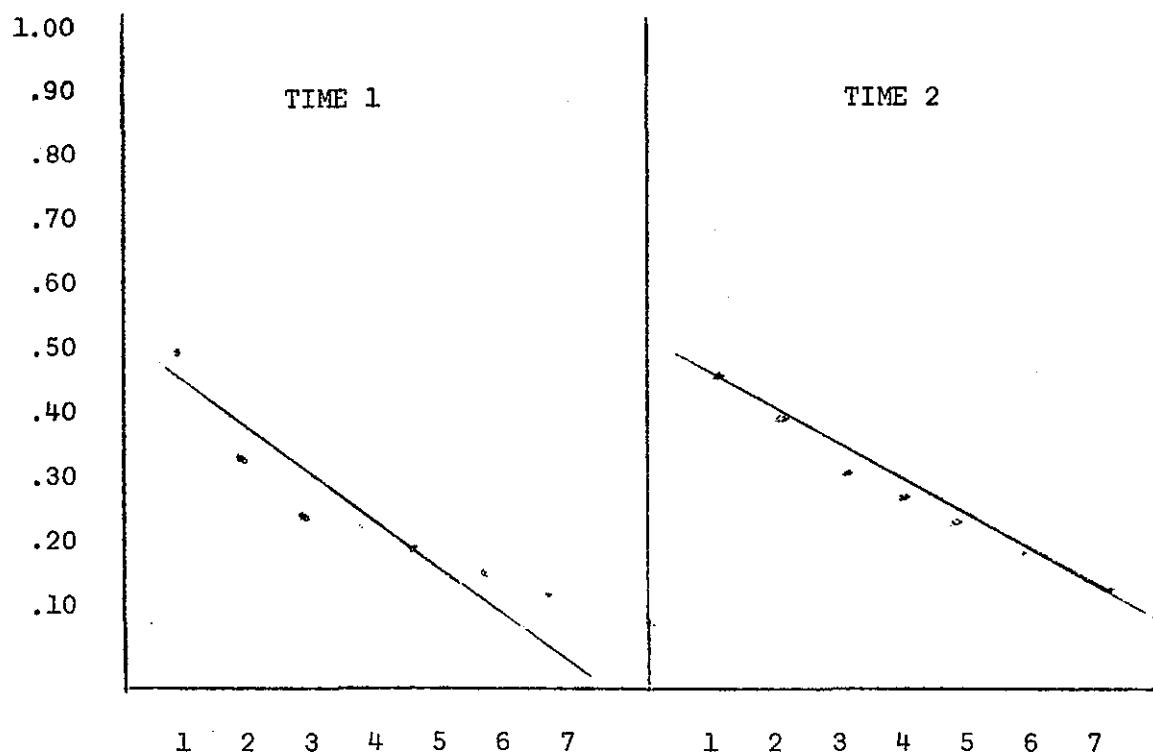
GRAPH 1

RELIABILITY COEFFICIENTS BY NUMBER OF CASES



GRAPH 2

GRAPH OF SCREE LINES OF TWO SETS OF RANDOM DATA



REFERENCES

- Abelson, Robert P.
1954 A Technique and a Model for Multi-dimensional Attitude Scaling,
Public Opinion Quarterly
- Carmone, Frank J. and Paul E. Green
1970 Multidimensional Scaling and Related Techniques in Marketing
Analysis
Allyn and Bacon, Boston
- Ekman, Gosta
1963 A Direct Method for Multidimensional Ratio Scaling
Psychometrika
- Gillham, James R.
1972 The Aggregation of Shared Information in a Sociology Department
University of Illinois
- Helm, Carl E.
1959 A Multidimensional Ratio Scaling Analysis of Color Relations
Educational Testing Service, Princeton
- Helm, Carl E., Samuel Messick and Ledyard Tucker
1959 Psychophysical Law and Scaling Models
Educational Testing Service, Princeton
- Jackson, Douglas N., Samuel Messick and Charles M. Solley
1957 A Multidimensional Scaling Approach to the Perception of Personality
The Journal of Psychology
- Klingberg, F.L.
1941 Studies in Measurement of the Relations Among Sovereign States
Psychometrika
- Lord, Frederick and Melvin Novak
1968 Statistical Theories of Mental Test Scores
Addison-Wesley, Reading, Mass.
- Messick, Samuel J.
1956a An Empirical Evaluation of Multidimensional Successive Intervals
Psychometrika
1956b The Perception of Social Attitudes
Journal of Abnormal Social Psychology
- Tatsuoka, Maurice, M.
1971 Multivariate Analysis Techniques for Educational and Psychological
Research
John Wiley and Sons, New York
- Torgerson, Warren G.
1958 Theory and Methods of Scaling
John Wiley and Sons, New York
- Woelfel, Joseph
1972 Sociology and Science (mimeo)