

A Note on a Statistical Paradigm for the Evaluation of Cognitive Structure in Physics Instruction

Gail S. Gliner

University of California at Santa Barbara

Data from Shavelson's (1972) study investigating the change in cognitive structure due to instruction in Newtonian mechanics were reanalyzed using the quadratic assignment (QA) approach. This application of the QA technique involves a nonparametric confirmatory procedure to evaluate whether a hypothesized structure is present in a proximity matrix representing cognitive structure. The proximity matrices in Shavelson's study were obtained from multiple-response word association tests at a pretest and at the end of each day of a 5-day instructional sequence. The Euclidean distance measure Shavelson used to evaluate change toward a representation of the instructional content, however, did not measure *structural* change in the proximity matrices for cognitive structure. The present reanalysis using the QA paradigm showed that cognitive structure after instruction was similar to content structure before instruction, and word association tests did not measure any subtle changes toward greater similarity to content structure. However, the QA results provided some evidence for Shavelson's contention that the experimental group's cognitive structure changed toward greater homogeneity later in instruction.

Judging from the recent literature in educational psychology, it is apparent that simple test performance is no longer given the privileged status of being the best indicator of student learning. The *structure* of knowledge is now

viewed as crucial, both in terms of how the process of learning proceeds and how students organize concepts in memory after instruction. With this increased emphasis on the importance of developing and measuring the structure of knowledge in a student's memory, educators have attempted to develop procedures for evaluating cognitive structure and for comparing the student's cognitive structure to a representation of the content structure of the text material. For example, one technique for collecting proximity measures for concepts in memory involves multiple-response word association tests (P. E. Johnson, 1967; Preece, 1976a; Shavelson, 1972). Key concepts that share word (concept) responses in common are assumed to be "close" in memory. Thus, as instruction progresses, it is expected that related key concepts of the instructional unit will become more interrelated and that, in particular, concepts that are related by equations or by other constraints from the subject matter will become more tightly linked or clustered.

Much of the methodology in evaluating cognitive structure has involved exploratory procedures, such as multidimensional scaling (Johnson, Cox, & Curran, 1970; Shavelson, 1972) and cluster analysis (Shavelson, 1972) to find a revealing representation of proximity data. For example, a proximity measure for a pair of concepts might be the amount of overlap in the free

association response for each of a pair of words. The data for each pair of concepts are then organized in a matrix prior to deriving an overall representation of the proximity of the concepts in memory, possibly through scaling, cluster analysis, or some related variant. Based on data from Shavelson (1972), the present study illustrates a confirmatory nonparametric technique for comparing the pattern of relationship among concepts in cognitive structure and the content of an instructional text.

Shavelson Study

Shavelson's (1972) study evaluated the change in the cognitive structure of students studying Newtonian mechanics over a 5-day period and was based on the premise that, as instruction progressed, the structure of the main concepts in student's memories would be closer to the structure of the concepts as represented by the instructional material. Using syntactical rules (Shavelson & Geeslin, 1975) for mapping prose onto a digraph (e.g., Harary, Norman, & Cartwright, 1965), an appropriate proximity matrix for content was defined by the number of links between pairs of 14 key concepts in the digraph. This proximity matrix was interpreted as a representation of the structure of concepts in the subject matter. Since proximity measures for cognitive structure are assumed symmetric, the original digraph distance matrix was folded along the main diagonal and the corresponding cell entries were added to form a symmetric matrix. For convenience, a final transformation of each cell entry of the content matrix was carried out to force all entries to be less than 1.

To measure cognitive structure, a multiple-response word association test for the 14 key concepts was given at pretest and after each day of instruction. Each of the six proximity matrices for the experimental group contained median relatedness coefficients (Garskof & Houston, 1963) that define the amount of overlap in responses for pairs of words in a multiple-response

word association test. Six proximity matrices were available for the control group, which took the word association tests but did not read the instructional text. Shavelson hypothesized that the matrices of median relatedness coefficients representing the instructional group's cognitive structure would become closer to the content matrix and that the matrices of relatedness coefficients for the control group should remain stable across the 5 days of instruction. To compare distances between pairs of instruction group and control matrices over instructional days, Shavelson computed a Euclidean distance measure that was designed to represent the absolute distance between two matrices. Specifically, the Euclidean distance measure was computed by squaring the difference score between corresponding cells in the two matrices, summing, dividing by the number of cells, and taking the square root. Based on this index, pairs of matrix distances for the control group changed little over the 5 days (Shavelson, 1972).

Shavelson concluded from the Euclidean distance matrix for the correspondence between the instruction and control group matrices that relatedness coefficient matrices for the experimental group earlier in instruction were more similar to the six control group matrices than to the experimental group matrices later in instruction. Moreover, based on the Euclidean distance measure comparing the six experimental group matrices and the six control group matrices to the transformed digraph distance matrix, he concluded that the distances to the content matrix decreased for the experimental group but remained fairly stable for the control group (Shavelson, 1972).

In response to Shavelson's paper, Preece (1976b) questioned the Euclidean distance measure as a measure of the distance between two matrices and noted that the trend in Shavelson's (1972) Table 6 of decreasing distances to the content structure for the experimental group is a result of the digraph transformation and of the increase in relatedness coefficients through in-

struction. Consequently, the observed trend does not necessarily represent an increasing correspondence between content structure and cognitive structure. Preece (1976b) argued that if a small distance between two matrices means that both contain a similar "pattern" of entries, then the Euclidean distance measure is inappropriate as a measure of the correspondence between cognitive structure and content structure. (This observation is very reminiscent of the discussion by Cronbach and Gleser, 1953. They noted that Euclidean distance used as a measure of profile similarity depends on the level of the entries, not just on the pattern of entries.) As an illustration of this problem, Preece (1976b) constructed an example of two half-matrices that had a similar pattern of high and low entries, and a third matrix with the same numbers as one of the matrices but a different pattern of entries. Between the matrices with the same absolute values, the Euclidean distance was smaller but there was a different pattern than between the two matrices with a similar pattern of entries. In conclusion, Preece suggested P. E. Johnson's (1967) use of the correlation of the entries between two matrices as an appropriate measure of their structural similarity.

The approach presented here is similar to Johnson's, but the use of a t test for the significance test of a correlation coefficient is bypassed. The usual t test for a correlation in the matrix comparison context involves comparing a test statistic to a sampling distribution that is assumed to be distributed as a t . In the present approach, a test index is compared to a sampling distribution that involves no assumptions about an underlying distribution. Also, the t test involves an approximation based on a reordering of all the cells in the matrix. A test for a significant correlation between two matrices should involve looking at whether the same structural pattern is present in both matrices, where different patterns are based on reorderings of whole rows and corresponding columns (Hubert & Schultz, 1976).

Method

As an alternative to P. E. Johnson's (1967) use of the correlation between two matrices and keeping Cronbach and Gleser's (1953) and Preece's (1976b) criticisms in mind, a quadratic assignment (QA) approach is proposed to analyze Shavelson's (1972) data. As discussed by Hubert and Schultz (1976), the QA paradigm provides a hypothesis testing procedure for confirming whether the pattern in each of the proximity matrices mirrored the structure of the content matrix representation of the concepts, and what pattern, if any, was present in the proximity data matrices for each day of instruction. The first question involved testing Shavelson's (1972) contention that cognitive structure for the experimental group changed in the direction of content structure.

Using QA, a data matrix, \mathbf{Q} , representing each of the 12 matrices of median relatedness coefficients (one for each of the 6 days for experimental and control groups) was compared to a structure matrix, \mathbf{C} , the digraph distance matrix, to assess whether the pattern represented in the structure matrix was also present in each data matrix. A measure of relationship, Γ , is used to assess the degree of correspondence, defined as the sum of the products of the corresponding elements of \mathbf{Q} and \mathbf{C} ,

$$\Gamma = \sum_{i < j} \sum q(o_i, o_j) c(o_i, o_j). \quad [1]$$

The measure, Γ , is the cross-product term from the numerator of a correlation coefficient between \mathbf{C} and \mathbf{Q} and is statistically equivalent to the correlation in terms of the inference model. The structure in \mathbf{C} is considered mirrored in \mathbf{Q} (i.e., the hypothesis of a random assignment of entries in \mathbf{C} is rejected) if the probability of obtaining the particular value of Γ that occurred from the data is sufficiently small. This latter probability is found by comparing the observed index Γ to the possible values of Γ based on all equally likely permutations of the rows and corresponding columns of \mathbf{Q} . This type of nonparametric significance test involves no underlying

assumptions of normality for the data population, as in a t test for a significant correlation, and could just as well be carried out on a correlation coefficient. In particular, since Γ is the only quantity that varies in the expression for a Pearson correlation between two matrices when the rows and columns are reordered, either the correlation itself or the less cumbersome Γ can be used in defining an index of correspondence and a reference distribution for a significance test. The index Γ is a measure of the correspondence in the pattern of entries in the two matrices C and Q . It does not depend on the level of entries, since the index is compared to the expected value of Γ in the sampling distribution for all possible reorderings of the rows and columns of one of the matrices.

Results

For the Shavelson (1972) data, the expectation and variance of Γ was found for each of the 12 data matrices compared to the digraph distance matrix, and a test using the z statistic was performed. The correlations between each of the matrices of median relatedness coefficients for concept pairs with the digraph distance matrix are represented in Table 1. Since, at pretest, $r = -.34$ and $r = -.35$ for the correlations for the experimental and control groups, respectively, both groups show cognitive structures that were related to a representation of the content structure. (Correlations are negative, since cognitive structure is represented by a similarity matrix

and content structure by a dissimilarity, or distance, matrix.) If the variations across time for the correlations of the control group represent only random change, then the change in the experimental group does not appear very large, even though it is systematic in the predicted direction. Shavelson (1972) has noted that multidimensional scaling solutions for pretest cognitive structure and for content structure indicated that students had a good understanding of the relationships among the key concepts before instruction, since Newtonian concepts such as acceleration, velocity, and mass are used in ordinary language. This was also found in the present analysis of the data.

The index Γ (or its normalized analogue, r), does not change greatly for the 12 data matrices compared to the digraph distance matrix, and all index values are represented by z values close to -3.0 , showing fairly strong agreement between the structure matrix and all the data matrices. Although it is tempting to rely on a normal distribution for the sampling distribution of the index Γ , not enough is known about an adequate size for n at the present time (Hubert & Levin, 1976). In fact, Mielke (1979) has noted that asymptotic normality may not even hold for the index Γ . Using the conservative Cantelli's inequality, however, a z value of -3.0 would have a p value of at most .1 (i.e., $1/(z^2 + 1)$) for a one-tailed test.

Preece suggested that the increase in the relatedness coefficients for the experimental group

Table 1
Correlations of Content Structure and
Cognitive Structure Matrices Across Time

Day	Experimental group		Control group	
	r	z value	r	z value
1	-.34	-3.09	-.35	-3.43
2	-.27	-1.85	-.31	-2.98
3	-.30	-2.34	-.26	-2.00
4	-.40	-3.97	-.38	-3.15
5	-.43	-3.92	-.33	-3.69
6	-.43	-3.29	-.27	-2.27

in Shavelson's study merely indicates a "strengthening of the links between concepts in cognitive structure rather than a reorganization of the structure" (Preece, 1978, p. 556). Preece's contention suggests that the patterns in the median relatedness coefficient matrices may not be different for the experimental and control groups at different times in the instruction. To evaluate this notion in detail using QA, an inter-correlation matrix, representing the correlations of each pair of data matrices, was generated (Table 2) to assess the pattern of entries for all pairs of matrices. Three comparisons are suggested from the design of Shavelson's study:

1. The degree of correspondence between the cognitive structure for the two groups during each day of instruction (Table 2, above diagonal);
2. The degree of correspondence between the six experimental group matrices and six control group matrices (Table 2, below diagonal);
3. The degree of correspondence among the cognitive structure of the experimental group matrices during the last part of instruction.

Shavelson (1972) concluded that the Euclidean distance measure between pairs of matrices suggested that the matrices in the experimental group became more similar during the later days of instruction. Unfortunately, as Preece (1976b) has pointed out, Euclidean distance does not measure structural similarity; moreover, even if it did, there is no structural significance test for the conjecture.

The first two comparisons were evaluated using a specialization of the QA paradigm (Hubert & Levin, 1976; Hubert & Schultz, 1976) to represent weighted partitions of a data matrix **Q**. Specifically, an index, Δ , defined by the sum of the within-subset proximities, was used to index the homogeneity of the partition derived from each comparison. This index is used when comparing a data matrix to a structure matrix representing a particular partition of the data matrix. For example, when examining homogeneity within the groups by days, the matrix above the diagonal in Table 2 was compared to a structure matrix containing one's in the position for each of the two within-subject proximities and zero's everywhere else. For the above diagonal matrix in Table 2, $\Delta = .86(1) + .86(1)$

Table 2
Intercorrelations of Relatedness Coefficient Matrices.
Six Group Partition by Days Above Diagonal; Two Group
Partition by Experimental (E) and Control (C)
Below Diagonal

Group & Day	Day												Group & Day
	1		2		3		4		5		6		
	E	C	E	C	E	C	E	C	E	C	E	C	
E, Day 1		.86	.82	.86	.81	.84	.73	.81	.74	.92	.67	.86	E, Day 1
E, Day 2	.82		.74	.84	.74	.81	.66	.76	.72	.78	.63	.78	C, Day 1
E, Day 3	.81	.87		.84	.87	.89	.70	.79	.73	.85	.72	.86	E, Day 2
E, Day 4	.73	.70	.86		.89	.90	.80	.77	.80	.88	.74	.88	C, Day 2
E, Day 5	.74	.73	.87	.93		.91	.86	.83	.87	.87	.80	.85	E, Day 3
E, Day 6	.67	.72	.80	.84	.86		.77	.84	.80	.88	.74	.90	C, Day 3
C, Day 1	.86	.74	.74	.66	.72	.63		.73	.93	.78	.84	.72	E, Day 4
C, Day 2	.86	.84	.89	.80	.80	.74	.84		.75	.86	.71	.80	C, Day 4
C, Day 3	.84	.89	.91	.77	.80	.74	.81	.90		.78	.86	.77	E, Day 5
C, Day 4	.81	.79	.83	.73	.75	.71	.76	.77	.84		.73	.89	C, Day 5
C, Day 5	.92	.85	.87	.78	.78	.73	.76	.88	.88	.86		.71	E, Day 6
C, Day 6	.86	.86	.85	.72	.77	.71	.76	.88	.90	.80	.89		C, Day 6

Group											
E						C					
1	2	3	4	5	6	1	2	3	4	5	6

+ .84(1) + .84(1) + .91(1) + .73(1) + .73(1) + .78(1) + .71(1) + .71(1). Thus, when the cross-products of each of the corresponding cells in the data matrix and structure matrix are computed and added, the index Δ becomes the sum of the within-subject proximities.

The six-group split (partitioned by days; Table 2, above diagonal) is less salient than the two-group split (partitioned by experimental versus control group; Table 1, below diagonal) as evidence by the z values given below.

Partition by Days

$$\Delta = 9.67$$

$$E(\Delta) = 9.64$$

$$V(\Delta) = .06$$

$$Z = .11$$

Partition by
Experimental-Control

$$\Delta = 49.1$$

$$E(\Delta) = 48.2$$

$$V(\Delta) = .19$$

$$Z = 1.98$$

Based on a significance test for a random sample of 100 from the monte carlo sampling distribution of Δ , the index value obtained for the experimental-control partition could be declared significant at an approximate .07 level.

These results, for the partition by days and by experimental versus control group, are consistent with Shavelson's (1972) work. If instruction, not time, has an effect, then cognitive structure should not be highly correlated within each day of instruction between the experimental and control group. However, the more salient experimental versus control partition suggests that there is some homogeneity within one or both of the groups, which may not be due to instruction or to the temporal stability of the control group matrices.

The results for the second partition suggested further investigation of the homogeneity within the experimental and control groups. A variation of the Δ index was used to test the significance of a homogeneous subset of the intercorrelations of the relatedness coefficient matrix. In this case, the index only includes the within-

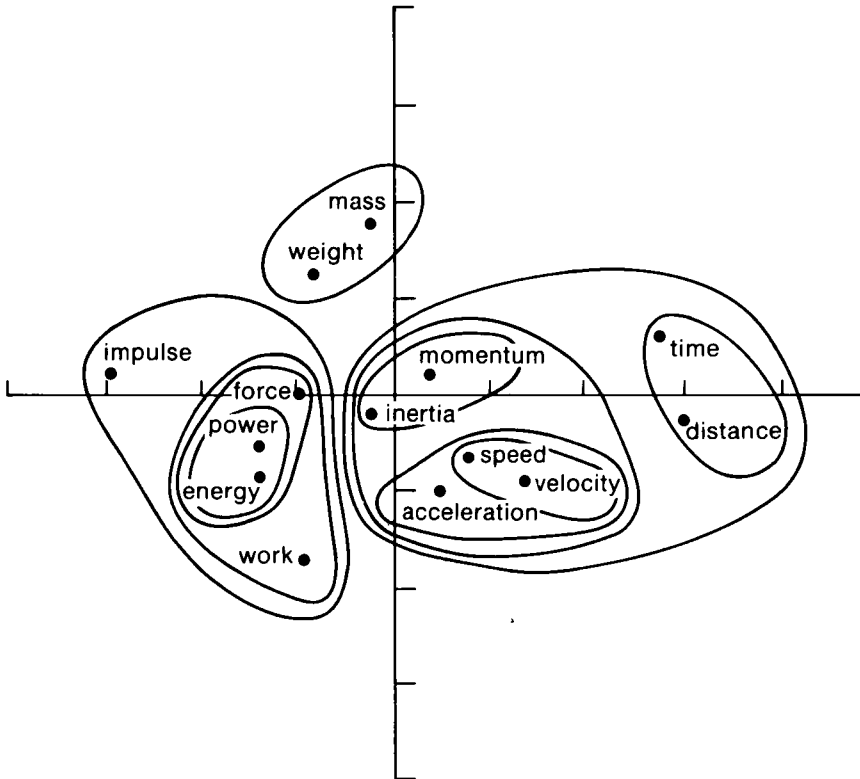
group proximities for a single subset. Two subsets were examined first: a homogeneous grouping of the control group for the first 3 days of instruction and for the last 3 days. The z values obtained were 1.03 and 1.06 for the first and second halves of instruction, respectively. These z values reflect day-to-day variability in the control group.

Using a similar analysis in the experimental group, the three matrices from the first half of instruction produced $z = .65$, suggesting that the cognitive structure of the experimental group early in instruction was not very similar. However, the cognitive structure for the experimental group during the last part of the instruction became more similar, as evidenced by the larger z value, $z = 1.58$. Although its significance cannot be asserted, apparently there is some change toward greater similarity in the cognitive structure of the experimental group from the beginning to the end of instruction, represented by an increase from $z = .65$ to $z = 1.58$.

To gather additional evidence concerning the correspondence among the groups throughout instruction, multidimensional scaling using the KYST procedure (Kruskal, 1964) was used along with hierarchical cluster analysis with the diameter method (S. C. Johnson, 1967). Multidimensional scaling generated a pictorial representation of the relationship among the 14 key concepts based on the relatedness coefficient matrices. Three spaces were generated: one from the control group data, one from the experimental group during the first half of instruction, and the other from the experimental group during the second half of instruction.

The KYST solutions are given in Figures 1, 2, and 3 with the hierarchical clustering solutions up to the three-group partition embedded in the representation. The multidimensional scaling solutions are two-dimensional solutions. A general rule suggested by Kruskal and Wish (1978) is that the number of stimuli minus one should be at least four times the dimensionality. Since there were 14 stimuli, to insure a stable solution,

Figure 1
 KYST Solution of Control Group Data
 with Embedded Hierarchical Clustering Solution



only solutions up to three dimensions were considered. The following are the stress values for the three scaling solutions using Kruskal's Stress 1 Formula (Kruskal, 1964) for goodness of fit:

Dimension	Control	Experimental	
		Early	Late
1	.235	.322	.285
2	.084	.069	.118
3	.049	.045	.101

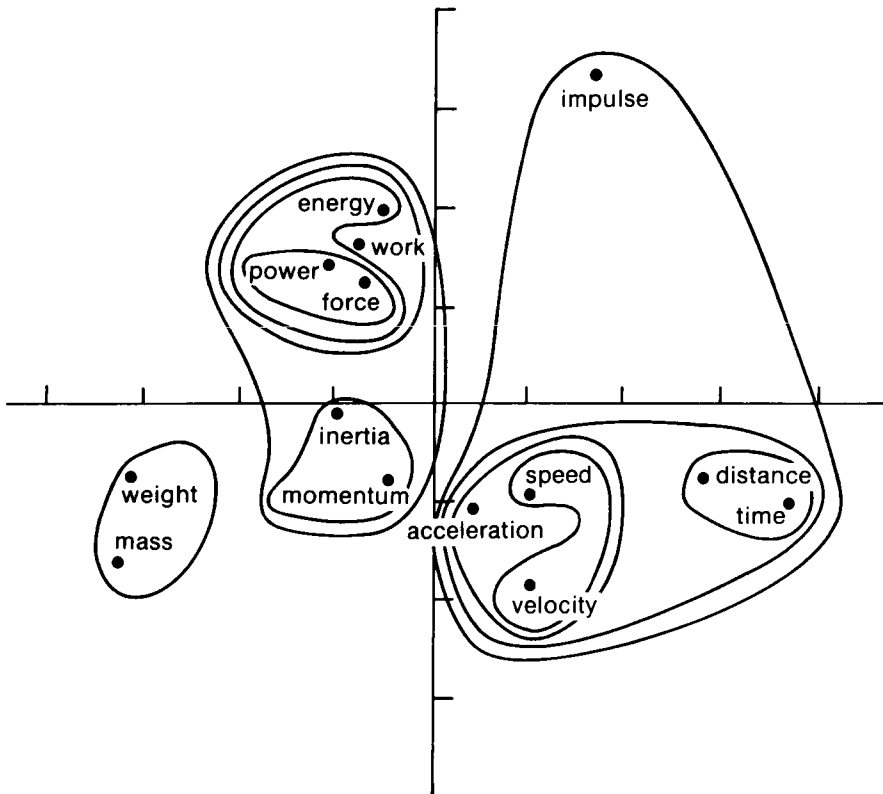
Since stress did not decrease appreciably between two and three dimensions, two-dimensional solutions were considered adequate. Shavelson (1972) applied KYST to the experimental and control group matrices for each day of instruction and found two-dimensional solutions provided an adequate fit to the data.

The three KYST solutions in Figures 1, 2, and 3 do not appear very different from each other. One difference between the early and late experimental groups is that IMPULSE has moved into a cluster with ACCELERATION. Other clusters, such as WEIGHT and MASS, TIME and DISTANCE, MOMENTUM and INERTIA, and SPEED and VELOCITY are fairly constant across the three solutions for control, early experimental, and late experimental.

Discussion

The results of the present analysis provide mixed evidence for Shavelson's (1972) contention that the relatedness coefficient data may provide a measure of the change in cognitive

Figure 2
 KYST Solution of Early Experimental Group Data
 with Embedded Hierarchical Clustering Solution



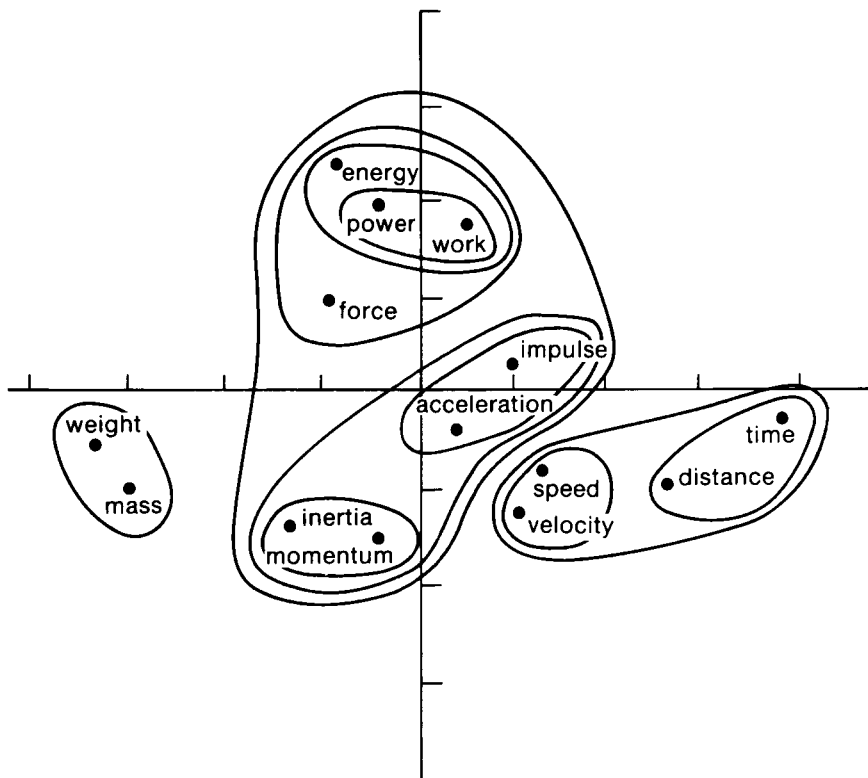
structure before the change is measured on a post-test. Although the effect is weak and non-significant, the experimental matrices did become somewhat more similar to each other later in the instruction. Shavelson (1972) noted that post-test word association responses increased and produced more common associates for concepts that define each other, e.g., force, mass, acceleration. However, this change may reflect the learners' increased awareness of "links" among concepts and not a change in the overall structure, a conclusion proposed by Preece (1978).

Unfortunately, the QA analysis did not suggest a significant change for the experimental group toward greater similarity to the content structure. At pretest the correlation between the

structure of the concepts in the instructional material was already similar to the students' cognitive structure ($r = -.34$ for the experimental group and $r = -.35$ for the control group), so that it may be difficult to measure subtle changes in cognitive structure due to instruction using word association tests and the relatedness coefficient measure. Also, the control, the early experimental, and the late experimental group data generated similar multidimensional scaling solutions and hierarchical cluster analyses from relatedness coefficient matrices.

The lack of an increase in correlation across time suggests that subjects were familiar with many of the key concepts before instruction, and differences in cognitive structure were not ap-

Figure 3
KYST Solution of Late Experimental Group Data
with Embedded Hierarchical Clustering Solution



parent later in instruction. However, it may be that each student possessed an incomplete knowledge of the structure before instruction, and the aggregation of responses in forming median relatedness coefficients for each group may have produced a more complete structure. The same aggregate structure may be present after instruction, and changes for individual students may not be apparent. Perhaps the presentation of an instructional unit with which subjects are clearly not familiar may more easily demonstrate a change in cognitive structure toward the content structure of an instruction text.

The use of the QA paradigm provides a flexible procedure to confirm the presence of a particular structural pattern in proximity matrices. Exploratory procedures in the multivariate framework, such as multidimensional scaling

and cluster analysis, do not lend themselves easily to hypothesis testing. The reader is referred to Hubert and Levin (1976) and Hubert and Baker (1978) for other applications of the QA paradigm in psychology and sociology.

References

- Cronbach, L. J., & Gleser, G. C. Assessing similarity between profiles. *Psychological Bulletin*, 1953, 50, 456-473.
- Garskof, B. E., & Houston, J. P. Measurement of verbal relatedness: An idiographic approach. *Psychological Review*, 1963, 70, 277-288.
- Harary, F., Norman, R. Z., & Cartwright, D. *Structural models: An introduction to the theory of directed graphs*. New York: Wiley and Sons, 1965.
- Hubert, L. J., & Baker, F. B. Evaluating the conformity of sociometric measurements. *Psychometrika*, 1978, 43, 31-42.

- Hubert, L. J., & Levin, J. Evaluating object set partitions; freesort analysis and some generalizations. *Journal of Verbal Learning and Verbal Behavior*, 1976, 15, 459-470.
- Hubert, L. J., & Schultz, J. V. Quadratic assignment as a general data analysis strategy. *British Journal of Mathematical and Statistical Psychology*, 1976, 29, 190-241.
- Johnson, P. E. Some psychological aspects of subject-matter structure. *Journal of Educational Psychology*, 1967, 58, 75-83.
- Johnson, P. E., Cox, D. L., & Curran, T. E. Psychological reality of physical concepts. *Psychonomic Science*, 1970, 19, 245-247.
- Johnson, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, 32, 241-254.
- Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, 29, 1-27.
- Kruskal, J. B., & Wish, M. *Multidimensional scaling*. Beverly Hills CA.: Sage Publications, 1978.
- Mielke, P. W. On asymptotic non-normality of null distributions of MRPP statistics. *Communications in Statistics, Theory and Methods*, 1979, 15, 1541-1550.
- Perfetti, C. A. Psychosemantics: Some cognitive aspects of structural meaning. *Psychological Bulletin*, 1972, 78, 241-259.
- Preece, P. F. W. Associative structure of science concepts. *British Journal of Educational Psychology*, 1976, 46, 174-183. (a)
- Preece, P. F. W. A note on the comparison of cognitive structure and subject matter structure. *Journal of Research in Science Teaching*, 1976, 13, 353-354. (b)
- Preece, P. F. W. Exploration of semantic space: Review of research on the organization of scientific concepts in semantic memory. *Science Education*, 1978, 62, 547-562.
- Shavelson, R. J. Some aspects of the correspondence between content structure and cognitive structure in physics instruction. *Journal of Educational Psychology*, 1972, 63, 225-234.
- Shavelson, R. J., & Geeslin, W. E. A method for examining subject matter structure in instructional material. *Journal of Structural Learning*, 1975, 4, 199-218.

Acknowledgments

The author expresses special thanks to Lawrence Hubert for his help during the preparation of this manuscript and to Richard Shavelson for providing use of his data.

Author's Address

Send requests for reprints or further information to Gail Gliner, Ergonomics Department, University of California at Santa Barbara, Santa Barbara CA 93106.