DOCUMENT RESUME

ED 041 045 | 24 | TM 000 011

AUTHOR Wick, John W.
TITLE The Initial Development of a Technique for Deriving Additional Information from Test Performance. Final Report.
INSTITUTION Northwestern Univ., Evanston, Ill.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-8-E-087
PUB DATE 31 Aug 69
GRANT OEG-0-8-080087-3716
NOTE 28p.

EDRS PRICE EDRS Price MF-$0.25 HC-$1.50
DESCRIPTORS *Cluster Grouping, *Evaluation Techniques, Measurement Techniques, *Profile Evaluation, *Psychometrics, Statistical Analysis, *Testing
IDENTIFIERS Similar Response Analysis (SRA)

ABSTRACT

A pattern-analytical technique, Similar Response Analysis (SRA), was developed, validated with contrived data, verified using previously reported data based on other pattern-analytical methods, and used successfully with "real" data. This technique orders subjects on the basis of the similarity of responses of adjacent individuals, not on the basis of the number of times each individual marks the answer keyed as "correct." The technique is offered as an alternative method for ordering a list of students on the basis of test scores. SRA is based on the theory of minimum inversions; i.e., if the number of matched responses decreases as subjects are further removed from one another in the list, no inversions will occur. However, when subjects A and B, adjacent to one another in the list, have fewer matched responses than A and C who are not adjacent, then an inversion occurs. The technique is designed to order the subjects so that only a minimum number of inversions can occur. The theory upon which SRA is based is discussed and a computer program for the technique is presented. (Author/ES)

Final Report

Project No. 8-E-087
Grant No. OEG-0-8-080087-3716

THE INITIAL DEVELOPMENT OF A TECHNIQUE FOR DERIVING ADDITIONAL INFORMATION FROM TEST PERFORMANCE

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Final Report

Project No. 8-E-087
Grant No. OEG-0-8-080087-3716

THE INITIAL DEVELOPMENT OF A TECHNIQUE FOR DERIVING ADDITIONAL INFORMATION FROM TEST PERFORMANCE

John W. Wick

Northwestern University

Evanston, Illinois

August 31, 1969

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

Table of Contents

List of Tables

SUMMARY

A pattern-analytical technique, entitled Similar Response Analysis, was developed, validated with contrived data, verified using previously reported data based on other pattern analytical methods, and used successfully with "real" data in this project. Pattern analysis procedures differ from "total score" procedures in that a list of subjects is ordered on the basis of the similarity of responses of adjacent individuals, and not on the basis of the number of times each individual marks the answer keyed as "correct" on an a priori basis.

The theory upon which Similar Response Analysis is based is quite simple. A minimum number of inversions in the basic matched response matrix is envisioned. If the number of matched responses decreases as subjects are further removed from one another in the list, no inversions will occur. When Subjects A and B, adjacent to one another on the list, have less matched responses than A and C, who are not adjacent, then an inversion occurs. The purpose of the technique is to order a list of subjects so that a minimum number of inversions occurs.

A computer program was written, based on the theory of minimum inversions. The program is included in this report. To make sure that the program operated as predicted, extensive empirical investigations were carried out, and these were successful. The program was then tried out on results which had been previously reported in the literature using other pattern analysis techniques, and the program also operated successfully in this instance. In fact, this technique gave not only the clusters of the other techniques, but ordered the clusters as well.

Finally, the program was used in five experiments with "real" data. The subjects included children from elementary school, junior high school, and some children with no schooling. The independent measures included achievement tests, word association lists, and two attitude measures (toward school and toward the law). With the independent measure a fifty item word association list, and the dependent measure the amount of schooling (with chronological age held constant), the technique was able to separate the schooled from unschooled children with a high level of reliability. This may be a very interesting way to relate word association lists to various dependent variables.

The second experiment used an attitude-toward-school measure as independent variable, and the teacher in the classroom as dependent variable. Again with high reliability the technique separated those students of one teacher from the students of a second, based on the patterns of the students' responses. Apparently the teacher in the room affects the response patterns of students, when they respond to an instrument of this type. The third experiment also indicated that the response patterns of students are fairly unique when they have had the same teacher. This time, however, a standardized achievement test was used. Apparently the teacher in the room also has a unique impact on the pattern of responses the students give in a standardized achievement testing battery.

I. Introduction and Theory

The purpose of this project was to develop a new pattern analytical technique. The technique, called Similar Response Analysis, was then to be used to obtain additional information from test scores. That is, test scores are generally interpreted on the basis of the "total score" obtained by the student. The items in the test are "keyed" to answers deemed "correct" on an *a priori* basis, and the number of times the student responds with the keyed response determines his total score. His performance is then interpreted relative to others who have also taken the test, or absolutely (to a fixed criterion) if it is a "mastery" type test.

One of the primary outcomes of "total score" interpretation is a ranking of students in some manner, from "best" to "worst." An example of such ranking, or ordering, occurs when students have been given a scholastic aptitude test, and are ordered from "best" to "worst" on the basis of their score on the exam. Sometimes the ranking is not that exact, and the students are first assigned to a smaller number of discrete groups. These groups are then ordered from "best" to "worst." The assumption often is made that the individual groups are internally homogeneous. An example of such grouping occurs whenever an instructor groups his students into five groups, and assigns a grade of "A" to the top group, "B" to the second, and so forth.

The purpose of the project was to develop an alternative method for ordering a list of students - alternative to the total score technique. An illustration at this point might be instructive, to show the distinction between "total score" and pattern-analytical techniques. In Table 1, five subjects have taken a four item test. The item is scored "1" if their answer matched the "keyed" response; "0" otherwise.

Table 1a**

Original Data Matrix

	Item Number				Total
	1	2	3	4	Score
Subject 1	1	0	0	1	2
2	1	0	1	1	3
3	1	0	1	0	2
4	0	1	0	1	2
5	1	1	0	1	3

**Editor's note: Beginning with this illustration and through the completion of this "Introduction and Theory" section, the tables were taken from the article "Similar Response Analysis" by John W. Wick, accepted for publication by *Educational and Psychological Measurement*. The article will appear in the March 1970 issue of this journal.

Based on the concept of total score, subjects 2 and 5 are alike, having both obtained a score of 3, as are 1, 3, and 4, all of whom obtained a score of 2. Now suppose we count, for every possible pair of subjects, the number of times their item responses match. For example, the responses of subject 1 and 2 match on three of the items - namely items 1, 2, and 4. When all pairs are similarly compared, the results are summarized in a matched response matrix, and the entry for subjects 1 and 2 is 3. The entire matched response matrix is given in Table 2a.

Table 2a

Original Matched Response Matrix

	With Subject			
	2	3	4	5
Subject 1	3	2	2	3
2		3	1	2
3			0	1
4				3

Now suppose we reorder the information in Table 2a, so that in our final list people who have responded alike will be "near" each other in the list, and separate from people who responded in a very different manner. That is, people with a high number of matched responses should be near each other. The further two people are separated, in the final list, the smaller their number of matched responses should be. Table 2a can be reordered as follows:

Table 2b

Judiciously Reordered Matched Response Matrix

	With Subject			
	2	1	5	4
Subject 3	3	2	1	0
2		3	2	1
1			3	2
5				3

Because the number of subjects involved here was very low, and because the data were contrived, it was possible to locate the final order (based on similarity of response patterns) through a trial-and-error procedure. However, when the number of subjects increases, and data are "real" (and therefore "messy"), the procedure is more complicated. A reordering procedure was needed.

A study of Table 2b will illustrate the basic theory upon which the reordering program was based. Notice that when the subjects are placed in the list so that "similar responders" are near one another, and "dissimilar responders" are separated, the following two phenomena result: (a) the numbers along each row descend, and (b) the numbers down any column ascend. In general the program sought to reorder the original data matrix until this situation - descending numbers in the rows, ascending numbers down the columns - was maximally realized.

Suppose the first row of an unordered matched response matrix had these entries:

12 13 9 9 14 1

In the first two positions, 13 is larger than 12 - this is contrary to the theory - so we count one "inversion." Comparing 12 with each of the other entries, we also note that 14 is larger than 12, and the inversion count increases to 2. Now looking at 13, we note that only 14 is larger than 13, which is an inversion, and the count increases to 3. The next entry is 9, and again 14 is larger than 9, so the count increases to 4. The fourth entry is also 9, and 14 is larger than 9, so the count increases to 5, which is the final total for row 1. The "inversion counter" reads 5.

This procedure can be continued along each row (the rows become shorter each time as we proceed down the matrix, since only the upper triangular matrix is used) and as we proceed up each column. When the entire counting procedure is completed, the inversion counter has a final measure of disorder.

The program was written so that (a) an inversion count is obtained; (b) two people "change places" in the list; (c) a new inversion count is obtained. If the new inversion count has decreased (indicating that the goal of a minimum number of inversions is being approached) then the new list remains. If there has not been a decrease, the two subjects are shifted back to their original positions.

One complete cycle of the program allows each position to exchange places with each other position two times. An inversion count is given before and after each cycle. The program is allowed to run as long as there are reductions for each cycle. As soon as a cycle is completed, and there is no reduction, the "run" stops. The time for a cycle varies with the size of the original matrix, of course. When the matrix size increases from 10 x 10 to 20 x 20 (a factor of two on the side), the number of "moves" required increases by a factor of approximately four.

Three methods of counting inversions were developed. These have been called the "nonparametric," "nonparametric-ties," and "parametric" techniques. The nonparametric technique was illustrated above. For the nonparametric-ties procedure, we simply count 1/2 for each time a tie occurs. There is one tie in the row given as an example above (the two nines in the third and fourth positions) so the final count with the nonparametric-ties procedure would be 5 1/2.

The parametric technique takes into consideration not only the existence of an inversion, but also the magnitude of the inversion. An inversion exists at the first two positions (12 13) and the size of the difference is one. An inversion exists between the first and fifth positions (12 14), and the size is two. These differences are used as the inversion counters.

II. Computer Program Development

The basic program carries out the iterative procedures mentioned earlier. That is, it allows each position in the matched response matrix to change places with each other position, counting (and keeping track of) the number inversions before and after the change. The program has the decision procedure built in - that is, whether the previous change should remain, or be shifted back to their original position.

When the positions of two individuals in the list are exchanged, a whole series of elements in the matrix must be moved. Suppose subjects in the second and eighth positions were to be switched. Each of these has a matched response count for all the other subjects in the entire list. If there are a total of ten subjects in the list, sixteen elements must be moved - the matched response scores for subject 2 with subjects 1, 3, 4, 5, 6, 7, 9, and 10, and the matched response scores for subject 8 with these same eight other people.

Table 3a

Matched Response Matrix With Item Designations

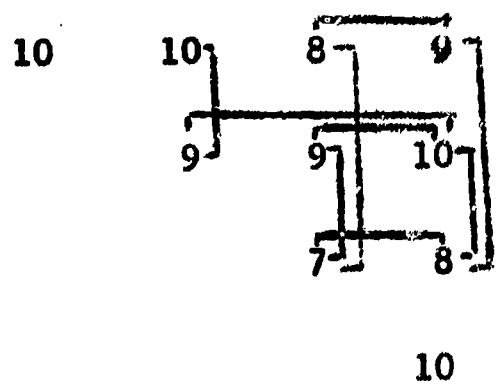
10	10	8	9
1,2	1,3	1,4	1,5
	9	9	10
	2,3	2,4	2,5
		7	8
		3,4	3,5
			10
			4,5

To illustrate the process whereby the position of subjects I and J is reversed in the matched response matrix, it is convenient to re-display the matrix as a vector. In this case the matrix presented in Table 3a appears as follows:

Computer position	1,2	1,3	1,4	1,5	2,3	2,4	2,5	3,4	3,5	4,5
Stored at this posn.	10 1,2	10 1,3	8 1,4	9 1,5	9 2,3	9 2,4	10 2,5	7 3,4	8 3,5	10 4,5

Table 3b

Inversions in the Matched Response Matrix



Now an I-storage vector, with N-2 positions (N = number of subjects) is defined, and similarly a vector of N-2 positions is defined for the J-term. For illustrative purposes, assume I = 1 and J = 2; that is, the row positions of subjects 1 and 2 in the matrix are to be exchanged. The program involves these steps.

1. Beginning at the left side of the vector above and in the "computer position" row, find the first pair which contains an "I" but not a "J." The values stored at this first location are placed in the first spot of the I-storage vector. Following along the row to the right, the I-storage vector is filled in order with the information stored at the computer storage positions which contain an I. The I-storage vector from the above example will be filled as follows:

10	8	9
1,3	1,4	1,5

and in a similar manner the J-storage vector will become

9	9	10
2,3	2,4	2,5

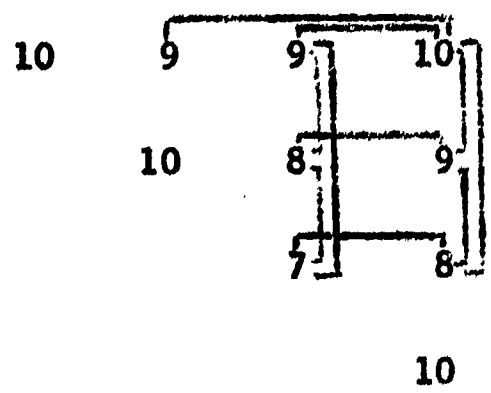
In a like manner, the J storage vector entries are inserted in the positions formerly occupied by the I-storage entries; and similarly the I-storage vector entries are inserted in the voids created by removing the J-storage entries. When this has been completed, the original vector will appear as follows:

Computer Position	1,2	1,3	1,4	1,5	2,3	2,4	2,5	3,4	3,5	4,5
Stored at this posn.	10 1,2	9 2,3	9 2,4	10 2,5	10 1,3	8 1,4	9 1,5	7 3,4	8 3,5	10 4,5

With the positions of subjects 1 and 2 exchanged, the matched response matrix assumes this configuration of Table 3c:

Table 3c

Matched Response Matrix with 1, 2 Exchanged



Using the parametric counting sequence, the count is 12. This is an increase from the original parametric count of 11, so we have moved away from minimum count. Thus we should re-exchange the positions of subject 1 and 2, following the same procedure. For the example, by the end of two complete iterations, a count of zero is obtained. The final order is subject 3 in the first position, followed by 1, 2, 5, and 4.

The basic reordering program is in two parts. The main program carries out the "shifting" in core; the subroutine counts the number of inversions before and after the shift. The program, as revised, follows:

```
      PROGRAM SRA2(INPUT,OUTPUT)
      DIMENSION MATCH(30,30),ID(30),FMT(12)
  200 READ 1, N
      IF(N-99) 201,300,201
    1 FORMAT(I2)
    2 FORMAT(24I2)
    3 FORMAT(1H1,*ORIGINAL REVERSAL COUNT =*,F10)
    4 FORMAT(12A6)
    5 FORMAT(2X,24I3)
    6 FORMAT(2X,24I3)
    9 FORMAT(1H1)
  201 READ 4,FMT
      NM1=N-1
      READ 2, (ID(I),I=1,N)
      DO 10 I=1,N
   10 READ FMT, (MATCH(I,J),J=1,N)
      COUNT1=COUNT(MATCH,N)
      SAVE=COUNT1
      PRINT 3, COUNT1
      COUNT2=9999999999.0
      ITER=0
      GO TO 85
   20 ITER=ITER+1
      COUNT2=SAVE
      IF(ITER-4) 25,100,100
   25 DO 80 I=1,NM1
      III=I+1
      DO 80 J=III,N
      DO 40 II=1,NM1
      IP1=II+1
      DO 40 JJ=IP1,N
      IF((I.NE.II.AND.I.NE.JJ).OR.(I.EQ.II.AND.J.EQ.JJ)) GO TO 40
      IF(I-II) 35,36,35
   35 TEMP=MATCH(II,JJ)
      MATCH(II,JJ)=MATCH(II,J)
      MATCH(II,J)=TEMP
      GO TO 40
   36 TEMP=MATCH(I,JJ)
      IF(JJ-J) 38,40,39
   38 MATCH(I,JJ)=MATCH(JJ,J)
      MATCH(JJ,J)=TEMP
      GO TO 40
   39 MATCH(I,JJ)=MATCH(J,JJ)
      MATCH(J,JJ)=TEMP
```

```
40 CONTINUE
   COUNT1=COUNT(MATCH,N)
   IF(SAVE-COUNT1) 70,70,50
50 PRINT 7, COUNT1
 7 FORMAT(1X,* REVERSAL COUNT =*F10)
   ID1=ID(J)
   ID2=ID(I)
   ID(I)=ID1
   ID(J)=ID2
   SAVE=COUNT1
   GO TO 80
70 DO 75 II=1,NM1
   IP1=II+1
   DO 75 JJ=IP1,N
   IF((I.NE.II.AND.I.NE.JJ).OR.(I.EQ.II.AND.J.EQ.JJ)) GO TO 75
   IF(I-II) 76,77,76
76 TEMP=MATCH(II,JJ)
   MATCH(II.JJ)=MATCH(II.J)
   MATCH(II,J)=TEMP
   GO TO 75
77 TEMP=MATCH(I,JJ)
   IF(JJ-J) 78,75,79
78 MATCH(I,JJ)=MATCH(JJ,J)
   MATCH(JJ,J)=TEMP
   GO TO 75
79 MATCH(I,JJ)=MATCH(J,JJ)
   MATCH(J,JJ)=TEMP
75 CONTINUE
80 CONTINUE
85 CONTINUE
   PRINT 5, (ID(I),I=1,N)
   PRINT 205
205 FORMAT(//)
   DO  90 I=1,N
 90 PRINT 6, (MATCH(I,J),J=1,N)
   PRINT 9
   IF((COUNT2*0.95)-SAVE) 100,100,20
100 CONTINUE
   PRINT 13, SAVE
 13 FORMAT(///1X,*FINAL REVERSAL COUNT =*,F10)
   GO TO 200
300 CONTINUE
   STOP
   END
```

```
      FUNCTION COUNT(MVAL,M)
      DIMENSION MVAL(30,30)
      REAL NR
      LIM=M-2 $ NR=0 $ JL=M-1
      DO 3 I=1,LIM
      JS=I+1
      DO 3 JP=JS,JL
      JT=JP+1
      DO 3 J=JT,M
      IF(MVAL(I,JP).LT.MVAL(I,J))NR=NR+MVAL(I,J)-MVAL(I,JP)
    3 CONTINUE
      DO 4 J=3,M
      JK=J-1
      JM=J-2
      DO 4 I=1,JM
      JN=I+1
      DO 4 JS=JN,JK
      IF(MVAL(I,J).GT.MVAL(JS,J))NR=NR+MVAL(I,J)-MVAL(JS,J)
    4 CONTINUE
      COUNT=NR
      RETURN
      END
```

III. Validation with contrived data

A theory is one thing; making it work is another. As a first step in verifying the technique, we contrived matrices of various sizes, where the correct result was known in advance. Then we treated these contrived matrices like they were "real" data, to find out if the techniques worked in the expected manner. The matrices were set up similar to the matrix in tables 2a and 2b, so that if the program did reorder the matched response matrix to the final minimum count, that count would be zero.

Matrices of sizes 10 x 10, 20 x 20, and 30 x 30 were used. For each size, the ordering of the subjects was randomly rearranged, in twenty different orders. The initial count, in each case, was a value larger than zero. Each of the twenty random orders was used with the three counting routines. A record was kept of the amount of time necessary to reach a zero count, the number of complete iterations required, and the initial and final inversion count. With three matrix sizes, twenty reorderings per size, and three counting procedures, it can be seen that 180 different runs were used in this stage.

The results are summarized in Table 4.

Table 4

Results with Contrived Data

Matrix Size	Counting Sequence	No. of times Count = 0 obtained	Avg. No. of iterations to count zero.	Avg. Time required on CDC 6400 computer
10 x 10	NP	20	2.6	<10 sec.
	NT	20	2.2	<10 sec.
	P	20	2.0	<10 sec.
20 x 20	NP	14	2.7	58 sec.
	NT	14	2.5	59 sec.
	P	20	2.7	61 sec.
30 x 30	NP	12	2.4	107 sec.
	NT	12	2.1	109 sec.
	P	20	2.8	117 sec.

Thus it can be seen that the program operates in a satisfactory manner. The Parametric sequence seems to be the most consistent, although this may be a function of these particular contrived data. When the sequencing does "stick" and stops before a zero count (with contrived data) it is probably due to the restriction of moving only two subjects at a time. This problem is reduced, in practice, by submitting each matched response matrix in many random orders, and using the result which reaches the lowest count.

IV. Similar Response Analysis application to reported results using other techniques

There is no absence of different pattern analysis technique available. A survey of these is found in Wick (1970, in press) cited earlier. As a second validation of Similar Response Analysis, the program was applied to data which had been previously reported, using other techniques. The philosophy behind such empirical verification as this one, and the previous one with contrived data, is that a technique ought to give back information known in advance. While this does not prove that it will work with real data, it is not unreasonable to expect a procedure which gives correct answers in such cases will also give correct answers with "real" data.

The data first presented by Zubin (1938) involves the use of agreement scores for twenty objects. McQuitty and his colleagues have twice discussed these data (1967, 1968) using different approaches. In the 1968 report, the authors introduce a technique called Iterative, Intercolumnar Cluster Analysis, a process involving the computation of the correlation coefficients for each column with all other columns. This is followed by an iterative technique which has as its primary purposes the identification of clusters (or types). The technique differs from most of the previous McQuitty techniques, which depended heavily on hierarchical groupings, in that it might be termed a "reverse hierarchical" technique. Two (or more) major clusters are first defined, and these are subsequently further divided.

The verification procedures were as follows:

(a) The original agreement matrix was first randomly rearranged in ten different orders.

(b) Each of these ten new matrices was used with the three counting procedures.

The average initial count, minimum final count, and final resulting order are given in Table 5.

Table 5

First Results with Previously reported Data

Count Sequence	Average Ini. Count	Minimum Final Count	Final Order associated with min. count*
NP	1166	331	HLRDBJTGOQMAPCNIFSEK*
NT	1192	404	HLRDBJTGOQAMPCNIFSEK
P	7066	1117	RLDBGJTOQMANPCIFSKEH

*Letter designations taken from McQuitty articles.

The following comments are appropriate:

1. The final order, corresponding to the minimum count, for the NP and NT counting sequences are the same, except for the minor shift of M and A. There is excellent agreement of this final order with the results previously reported (McQuitty and Clark, 1968). Their results after five (or less) iterations, identified these clusters:

I: CFIMPANQ II: EKS III: JTRGB(O) IV. HDL

Object O was not added to cluster III until the seventh iteration. The final order for the NP and NT counting sequences, disagrees only at object R. To make this close agreement clearer to the reader, the final order from Table 5 is recopied below, arbitrarily grouped according to the clusters obtained by McQuitty and listed above.

H L R D B J T G O Q M A P C N I F S E K

IV III I II

2. The clusters emerge as reported previously. In addition, the order of the clusters is important. When the clusters were resubmitted with the Similar Response Analysis program in different orders, such as I, III, II, IV, the count in each case increased. Clusters II and IV appear to be the most diverse, and between clusters III and I, pairs BJ and IF are most dissimilar. In each final arrangement, these behave in a manner similar to "like poles" on magnets.

3. The final order corresponding to the minimum count using the P counting technique differs slightly. This order, along with the clusters reported previously, is shown below.

R L D B G J T O Q M A N P C I F S K E H

IV III I II

The agreement is excellent, except for the very erratic behavior of the H object. The P counting sequence places much emphasis on the magnitude of the agreement scores - much more than do the NP and NT sequences. The average agreement score for object H is 14.5, which is markedly below those reported for the other 19 objects. The range for the others is from 18.5 to 23.5, making the difference between object H's average and the next lowest average nearly as much as the range of average scores for the other 19 objects. This may help to explain erratic behavior of object H.

The second example again utilizes a McQuitty (1957) technique. Here ten response patterns are given, based on a series of questions designed to indicate the respondent's concept of what constitutes a "best" and "poorest" mechanic. Once again the ten subjects were rearranged in ten random orders. Table 6 summarizes the results.

Table 6

Results from the "Best" and "Poorest"
Mechanics Matrix

Count Sequence	Average Ini. Count	Minimum Final Count	Final Order (Groupings arbitrary)
NP	124	48	AB FEDC JIGH
NT	131	57	AB FEDC JIGH
P	491	116	AB FEDC JIGH

In each case, a substantial reduction in the original count is indicated. This time, the final order for the three counting sequences is the same. In each of the ten cases, that is, the ten rearranged input orders, the groups AB, FEDC, and JIGH always appeared adjacent to one another, even when the minimum count was not finally obtained. Apparently, of the ten patterns, these three can be called "clusters," and order is important. When the cluster JIGH was arbitrarily placed between AB and FEDC, a higher count occurs in each case, and the final output was AB HGJI FECD. Even within the two large clusters, the pairs HG and CD were rearranged as far as possible from one another, without leaving their grouping. These pairs are the most dissimilar in the two clusters.

V. Experiments with "real" data

After the theory had been worked out, the computer program written, and two types of empirical validation of the program completed, the next step was to see if the technique has any usefulness with "real" data. Especially of interest was the question of unique usefulness. The project was not inclusive enough to exhaust the many different kinds of investigations which would be necessary to completely resolve this question, but five experiments were carried out, with a broad variety of data inputs.

Experiment 1

This experiment dealt with word association data. From Kendler (1963) a word association test is one where

> "a person responds with the first word that occurs to him upon seeing or hearing a given stimulus word."

Usually word association results are reported in terms of frequency counts; that is, for the stimulus word "table" a certain percentage respond "chair," others may respond "sit," "seat," or "eat." If a large number of subjects are given a word association test, it is likely that a wide variety of responses will be given, even though some obviously will predominate. A list of frequency counts is useful in

the normative sense - that is, as a basis for later comparison of some special group or individual. The data will elude most "traditional" analysis procedures, since there is no "correct" or "incorrect" response, and since the number of different responses for a list of words will be a variable. With 100 children, only three different responses might be given for "cat" but as many as 25 are possible with "pretty."

The subjects for the study were Spanish-speaking natives of Colombia.[1] Two groups of children were defined. One group was called the "low education" group, since they had received less than 1 1/2 years of schooling. The other, called "high education" had received more than 1 1/2 years of schooling. Each group was given a word-association test, consisting of fifty stimulus words, and the responses were recorded. The task at hand was to assemble equal sized samples from the high- and low-education groups, matched, subject by subject, in terms of chronological age. Twelve subjects were selected from each group. The age range in the two groups was from 8 years, 7 months through 10 years, 10 months.

The underlying question was: Does formal schooling affect response patterns in such children, even when chronological age is held constant? To answer the question, a matched response matrix was computed for the 24 subjects, using their coded responses to the fifty stimulus words. Whenever two children gave the same response to a stimulus word, a "1" was added to their score in the matched response matrix. If they did not give exactly the same response, a zero was added.

Next, the list of twenty-four subjects was rearranged in a completely random order, so that there was no initial separation in the list based on group membership. If the "low-education" group tended to make responses which had high intra-group similarity; the "high-education" group tended to make responses with high intra-group similarity; and where the intergroup similarity was low; then it follows that a list with members of the "low-education" group at one end, and the "high-education" group at the other should have the least number of inversions.

The subjects were numbered 1 through 24. The first 12 numbers were assigned to the "high-education" group, and the last twelve numbers (13 through 24) to the "low-education" group. Three random orders were submitted, as follows:

---

[1]The data were made available by Associate Professor James W. Hall of Northwestern University, and are from a project he carried out in Colombia, South America. Author is deeply indebted to Professor Hall for his cooperation in this experiment.

Table 7

Original Orders, Experiment 1

Trial	Original Order
1	13 1 4 8 21 20 6 14 17 7 9 10 3 12 22 15 16 11 19 23 2 18 5 24
2	1 5 4 15 14 16 17 7 23 8 18 12 22 13 3 9 2 21 20 19 10 6 24 11
3	7 24 18 4 22 19 16 9 3 10 12 1 13 15 20 11 8 2 6 17 5 14 21 23

In these lists, it can be seen that 7, 6, and 7 of the first twelve positions, respectively, are occupied by members of the "high-education" group (numbers 1 through 12). The lists are apparently "random" orders.

The results of the three runs using the Similar Response Analysis technique, are summarized below:

Table 8

Results from Experiment 1

Trial	Original Inversion Count	Final Inversion Count	Percent Reduction Dec./orig.	Number of "high-education" in 1st 12 posns.	Chance Probability of event
1	8583	2232	74%	8	0.220
2	13835	2151	84%	9	0.039
3	9768	2037	79%	10	0.003

The probability computations are based on a hypergeometric distribution (Kraft and Van Eeden, 1968). Assume there are 24 slips of paper in a bowl, and 12 are marked "H" and 12 marked "L." We wish to draw a sample of 12 slips of paper from the bowl (that is, fill the first twelve positions) at random. If each "slip" has an equal probability of selection, the first position has a 50-50 chance of an "H" or an "L." If an "H" fills this position, however, the next position has these probabilities: 11/23 for another H (.478) and 12/23 for an L (.521). In the long run, the L has the better chance. If two H's are drawn for the first two positions, then the probabilities for the third are 10/22 for the H, and 12/22 for the L. As disproportionately more H's are drawn than L's, the chance of even further imbalances decreases. A table of values for the case of a sample size of 12 from a population of 24 is as follows:

Table 9

Hypergeometric Probabilities for these Experiments

X = no. of high- or low-education students in first 12 positions	Probability of this single event	Cumulative Probability	Double cum. probability
X = 0	0.00000	0.00000	0.00000
1	0.00005	0.00005	0.00010
2	0.00161	0.00166	0.00332
3	0.01790	0.01956	0.03912
4	0.09061	0.11017	0.22034
5	0.23196	0.34213	0.68426
6	0.31573	0.65786	
7	0.23196	0.88982	
8	0.09061	0.98043	
9	0.01790	0.99833	
10	0.00161	0.99994	
11	0.00005	0.99999	
12	0.00000	0.99999	

The probability of getting three or less H's in the first twelve positions is 0.01956; and this is also the probability of getting three L's. Of interest in our experiment was the probability of obtaining a small number of H's or L's; thus we need to use the "double cumulative frequency" column.

The results shown in Table 8 are very interesting. Note first that the original inversion counts are reduced by a considerable amount - averaging approximately an 80% reduction. Something systematic is happening - there must be an underlying "orderliness" in these data.

But also note that as the final inversion count goes down (2232 to 2151 to 2037) so does the number of H's in the first 12 positions go up (8 to 9 to 10). We could have gotten the same amount of reduction in original inversion count, but not the second part - the separation of the predefined groups. Not only is there an orderliness in the data, but that orderliness apparently is related to the predicted factor - amount of schooling.

The implications, in terms of the use of this technique and word association data, are twofold. First, a variety of dependent variables can be defined, such as sex, socioeconomic status, ethnic group, geographic area, or classroom unit. Then matched samples can be obtained, and word association lists administered. If the similar response analysis separates the two groups in the list associated with minimum inversion count, then the predetermined dependent separating variable is very likely associated with the responses to the word association task.

However, similar response analysis also has the advantage of ordering - it is not limited to defining groups. When a group of subjects have an 80% reduction in inversion counts, there is very likely to be something in those data which is underlying the increased orderliness. When the final list is obtained, corresponding with a minimum inversion count, this list can be correlated with other information on the subjects. High correlations between the final order and measures of scholastic aptitude would suggest what that "something" underlying the increased orderliness is. The final list could be correlated with the students socio-economic status, attitude toward school, or age - any number of measures are possible.

Experiment 2

The present study treats the responses of the students of four different teachers in elementary classrooms. By the end of a school year, the students in a self-contained elementary room have had a considerable amount of contact with their teacher. It is reasonable to hypothesize that many teachers make a unique type of impact on their students, and that this impact might be reflected in the students' responses to items. If this is true, and the test items are properly selected, the students of one class may have somewhat homogeneous response patterns, different from those of a similar group of students who were taught by a different teacher. The present work is an investigation of this notion, using Similar Response Analysis.

Subjects: Twelve students were randomly chosen from each of four teachers. The teachers all taught at the same grade level (upper elementary) in the same school building. Presumably, the students were originally randomly assigned to the four teachers. The testing was done in the spring of the school year.[1]

The test: The instrument used for this study was the Describe Your School test (Hoyt, 1964). This is a list of fifty statements about the school, which the student responds to in a "yes-no" manner.

Analysis: Each of the four teachers was compared with the others, meaning that the following pairs were studied:

A,B A,C A,D B,C B,D C,D

For the first pair, twelve students were selected from teacher A, and 12 from teacher B. Using a random number table, these 24 were randomly arranged in a list, and a matched response matrix computed for this random order. The original number of inversion in this randomly ordered matrix was 11,678. After four complete

---

[1]The author wishes to extend his appreciation to Professor Norman Bowers of Northwestern University and Professor Frank Vogel of Northeastern Illinois State College, who collected these data.

iterations, this count had been reduced to 1777, a reduction of about 85%. The original (random) order and the final order (associated with the minimum number of inversions) follow:

Original Order

Subject	7	3	18	6	14	24	8	4	13	12	1	22	21	19	10	2	16	11	15	5	9	17	20	23
Teacher	A	A	B	A	B	B	A	A	B	A	A	B	B	B	A	A	B	A	B	A	A	B	B	B

Final Order

Subject	24	12	10	14	8	2	5	18	3	6	7	9	16	11	15	13	17	22	4	19	21	20	23	1
Teacher	B	A	A	B	A	A	A	B	A	A	A	A	B	A	B	B	B	B	A	B	B	B	B	A

Results and Discussion

Even a perfunctory review of the original and final orders given above will indicate that a considerable amount of shifting around has taken place. Of primary interest, however, is the fact that in the final order, 9 of the first 12 positions are filled by the students of teacher A. The probability of 9 (or more) A-types in the first 12 positions by chance alone is less than 0.02. It appears that the response patterns of the students of teacher A are relatively homogeneous, and are somewhat different from the response patterns of Teacher B's students.

The following table summarizes the results for all six combinations of teachers:

Table 10

Final results for the six teacher combinations

Teacher Pair	Final Order	Inversion Count Original	Final	% red.	No. A's in 1st 12 pos.	Prob. by chance
A,B	BAABAAABAAAABABBBBABBBBA	11,678	1777	84.8	9	.039
A,C	AACAAAAAAAACCCCCACCCCACC	11,992	2083	82.6	10	.003
A,D	ADAAAAADAAADDDADDDDADADD	16,289	3161	80.6	9	.039
B,C	CCBCCBBBBBBBBCBCCCCBCCCB	11,165	3764	66.3	8	.220
B,D	BBBBBBDDBDBBDDBDBBDDDDDD	11,965	2060	82.8	9	.039
C,D	CCDDDDCCCCCCDDDDCCDDCDCD	9,517	2411	74.7	8	.220
			Avg.	78.6		

The data suggests that there does seem to be a tendency for the pupils of a given teacher, near the end of the school year, to have similar response patterns. The students who have been with one teacher for an entire school year are relatively close together in the final list.

Future Research

This small investigation indicates that there may be some merit in a pattern analytical investigation of teacher impact. These preliminary results indicate that the teacher apparently does influence the response patterns of her students.

It is quite possible that lower final inversion counts might have been reached if some other random orders had been used for each of the six cases. Previous empirical work by the author has shown that with 24 x 24 matrices, the program sometimes "sticks" before reaching its minimum count. Further work with the A,D and B,C teacher combinations seems especially desirable, since these final counts were quite a bit higher than the other four.

This study covered only upper elementary students, using an attitude inventory. Future studies are being designed to use the responses of elementary school students through high school students. At the same time, the use of achievement test responses will be considered. With these items, it may be necessary to put greater weight on a matching incorrect response than on a matching correct response.

Finally, no value judgments regarding the impact have been made at this point. It is conceivable that students might have similar response patterns due to their common disgust with their teacher. Investigations of the characteristics of the classrooms which tend to be grouped together must also be carried out.

Experiments 3, 4, 5

Design. For this test of Similar Response Analysis (SRA) with "real," non-contrived, data to evaluate the effectiveness of the method in ordering response patterns to different types of sets of questions by several groups of students, data were drawn from three populations. For each of these populations, identified as A, B, and C, the responses to a set of questions were used to calculate a matched response matrix with the answers given by selected samples of students. Similar response analysis was then applied to each of these samples. The manipulations performed for the data from each sample were identical. The variations among the populations were subject and task characteristics.

Subjects. The subjects for this study were elementary school and junior high students. Population A was basically middle class third grade pupils. The B and C students were from inner city schools. Population B subjects were fourth graders, while C pupils were in the seventh grade.

Data. The data which were utilized in the similar response analysis were the responses to different types of questions. For population A, the questions were the paragraph meaning, language skills, and arithmetic computation segments of a standardized achievement test. There were 47 items coded on a scale from 0 to 4 used for the raw scores for these subjects. One hundred questions from a reading test, with the responses coded from 0 to 4, were the data for population B. The reactions of population C students to 25 questions evaluating their attitudes toward the law were scored on a scale from 1 to 5.

Procedure. For population A three classes were arbitrarily selected to be the basis for the original samples of 12 students. In the determination of the three classes for populations B, and C, the selection occurred by drawing three identification numbers, corresponding to the classes, from those possible for that particular population. From each of these three classes 12 students were obtained at random by drawing slips of paper with their identification numbers from the identification numbers existing for that class. Thus, for population A there were three, 12 student samples labeled A1, A2, A3. The three samples from each of the other populations were obtained in the same manner. After these samples were determined, those within a particular population were combined to form three larger samples of 24 students each. Within a population these larger samples were the merging of the twelve students from the class arbitrarily labeled 1 with those from class 2, class 1 with class 3, and class 2 with class 3. Twelve random orders were created by drawing slips of paper numbered from 1 to 24 and recording the order in which the numbers were selected. Each random order was then assigned to one of the combined samples.

Treatment of data. For each combined sample, a matched response matrix was calculated. The response matrix was then ordered to correspond with the random order which had been paired with the subjects for that sample. This ordered matrix was reordered by the similar response analysis to obtain a smaller reversal count, through comparison of two adjacent locations in the matrix. The process of comparing, and subsequent testing to evaluate whether or not a reduction in the number of reversals was achieved by the reordering, was executed three times for each matrix to create the final reversal count.

Results

The hypergeometric distribution for $N = 12$, $n = 12$, and $0 \leq x \leq 12$ was used to determine the probabilities for the distribution of the subjects from the two smalle samples in the combined samples for the random orders and the final similar response analysis orders.

Table 11

Probabilities for class distributions for the random and similar response analysis orders

Sample	Probability of random order distribution	Probability of final SRA order distribution
A1&A2	.22034	.00332
A1&A3	>.22034	.03912
A2&A3	>.22034	.22034
B1&B2	>.22034	>.22034
B1&B3	>.22034	>.22034
B2&B3	.22034	>.22034
C1&C2	>.22034	>.22034
C1&C3	>.22034	>.22034
C2&C3	>.22034	>.22034

As can be seen in Table 11, only two of the distributions were significant at the .05 level. These were in the final SRA orders for A1&A2 and A1&A3. The probabilities for the remainder of the orders were equal to or greater than .22034 indicating that these distributions were highly probable by chance.

There was considerable variation among the samples in the original reversal counts (ORCs) and the effectiveness of the SRA program in reducing the number and size of these inversions.

Table 12

Original reversal counts and percentages of reductions in reversal counts

Sample	Original reversal count	Percentage of reduction
A1&A2	10,228	73.6
A1&A3	9,879	69.9
A2&A3	7,563	73.9
Average	9,233.3	72.5
B1&B2	25,397	48.9
B1&B3	24,443	74.3
B2&B3	18,625	72.1
Average	22,818.3	65.1
C1&C2	6,051	67.7
C1&C3	5,779	56.2
C2&C3	6,296	49.4
Average	6,042.0	57.7

The calculations in Table 12 indicate that on the average population C had the smallest amount of original inversions and the greatest average percentage in the reductions of these reversals. Population B, while having the largest average original reversal count, had the second largest average percentage in inversion reductions.

VI. Discussion of the Experiments

Five experiments with "real" data were performed. One dealt with a word-association list, and amount of schooling was the dependent measure. The other four had "teacher" as the dependent measure, and various independent measures. The following general comments are appropriate.

1. In most cases, the reduction in inversion count was quite large. The reductions were 79, 79, 73, 65, and 58 percents, respectively. There was clearly a good deal of reordering going on as the program progressed.

2. In the first and second experiments, the results are very supportive of our choice of dependent variable. In the first experiment, the variable was "amount of schooling." In the second experiment, the variable was "teacher in the classroom." The second dependent variable, an attitude survey into the attitude of the students toward school, apparently picks up different patterns of responses for students of different teachers.

3. The third experiment again used "teacher in the classroom" as the dependent variable, but this time used a standardized achievement test as the independent measure. All three trials indicated that the dependent variable chosen was associated with the reduction in inversions. The chance probability levels for two of the three were below the .05 level, and the third trial was as the .22 level. The results suggest that, if an achievement test is given to students from two different classes near the end of the school year, the response patterns for one teacher's students will differ from the response patterns of the second teacher.

4. The fourth and fifth experiments are difficult to understand. The fourth experiment also dealt with an achievement test, and different teachers as the dependent variable, but the results did not replicate those seen in Experiment 3. In a conversation with the head of the project from which these data were obtained, two facts came to light which may help to explain the results. First, the children did not have a single teacher - they had many different teachers. Second, in this particular area of the city, the turnover in students is very high. Possibly no teacher ever really had the opportunity to make enough of an impact so that the students' response patterns were affected.

The last experiment is not difficult to understand. Through a series of delays the units upon which the independent measures were based did not reach the classrooms until two or three weeks before the end of the school year. It is most likely that two or three weeks is not enough time for any teacher to make enough of an impact on these students' attitudes toward the law (the independent measure) so that the response patterns would be affected.

References

Hoyt, Cyril J. Describe Your School. Minnesota Test Publishers, 1964.

Kendler, Howard H. Basic Psychology. (New York: Appleton-Century-Crofts, 1963), pp. 350-2.

Kraft, C. H. and VanEeden, Constance. A Nonparametric Introduction to Statistics. (New York: Macmillan Co., 1968).

McQuitty, L. L., Price, L., and Clark, J. A. The Problem of Ties in a Pattern Analytic Method. Educational and Psychological Measurement, XXVII (1967), 787-796.

McQuitty, L. L. Improving the Validity of Crucial Decisions in Pattern Analytic Methods. Educational and Psychological Measurement, XXVIII (1968), 9-22.

McQuitty, L. L. Clusters from Iterative, Intercolumnar Correlational Analysis. Educational and Psychological Measurement, XXVIII (1968), 211-238.

McQuitty, L. L. A Pattern Analysis of Descriptions of "Best" and "Poorest" Mechanics Compared with Factor-Analytic Results. Psychological Monographs: General and Applied, 71, No. 446 (1957), 1-24.

Wick, John W. "Similar Response Analysis." Educational and Psychological Measurement (1970, in press).

Zubin, J. A. A Technique for Measuring Like-Mindedness. Journal of Abnormal and Social Psychology, 33 (1938), 508-516.