IPLE LINEAR REGRESSION VIEWPOINTS
)LUME 15, NUMBER 2, WINTER 1987

The Use of Judgement Analysis and A Modified Canonical JAN in Evaluation Methodology

Samuel R. Houston
University of Georgia

ABSTRACT

Judgment Analysis is presented as a technique for capturing and clustering unidimensional policies among a group of judges or evaluators. JAN utilizes a multiple linear regression model to represent each policy and then cluster evaluators together who are expressing similar policies. JAN is extended to a multidimensional situation in which a modified and simplified Canonical JAN (C-JAN) procedure for capturing policies on more than two criteria is described. Both unidimensional and multidimensional JAN procedures should be of general interest to the evaluation methodologist.

Teacher effectiveness is an area of great concern and the focus of much research in the educational community. The idea of teacher evaluation by students has been popular at the University of Northern Colorado campus for many years. The primary purpose of this paper is to present Judgment Analysis (JAN) as a technique for both capturing and clustering policies about what constitutes teacher effectiveness for individuals serving as evaluators.

Management personnel and evaluators often base decisions upon complex arrays of information. If these administrators could state explicitly how they used this information, these decision makers--and others--could replicate their judgments in subsequent sitations in which the same types of information are available.

By way of an example, consider a situation in which an organization is in the process of recruiting personnel for particular jobs at a specific point in time. The evaluation of prospective applicants for each position is often determined by the judgment of one or more administrators, judges or decision (policy) makers. Frequently the actual rating for each applicant is obtained by combining several different types of informatin into a weighted composite to produce a numerical indicator of the decision maker's judgment or value rating. One method of weighting is to have the decision maker provide the numerical weights to be used with the different types of information (variables) to form composite explicit-weighting evaluations. While explicit-weighting procedures are satisfactory in some situations, it is usually quite difficult to choose the proper multiplier values to form the composite evaluation of the applicant for the position in question that adequtely indicate the value of a person on a job. The problem of determining the appropriate numerical weights to be used can be illustrated in the following example. In Table 1 are presented three test scores in statistics for two students. The instructor desires that each test be weighted equally in the determination of the course grade. Both students obtained the same point total of 120 points. Yet, if the instructor wants each test to carry the same weight, he must <u>not</u> add the three scores together! While each test had the same mean score, the variances for the three tests are quite different. This variation actually influences any explicit-weighting approach which might be applied. As a result of these differences, <u>different</u> weights must be applied to each test score if each test is to carry the same weight in the evaluation process.

The difficulties encountered with explicit-weighting strategies in general have led to a second method--policy-capturing--which involves implicit determination of the numerical weights to be applied.

1. <u>JUDGMENT ANALYSIS</u>

A technique for determing implicitly the set of numerical weights to be applied in a decision-making situation was developed by J. H. Ward, Jr.

Table 1

ASSIGNING WEIGHTS TO THREE TESTS IN STATISTICS[1]

	Test Points			
	Test 1	Test 2	Test 3	Total Points
Student:				
Mary	30	40	50	120
Joe	50	40	30	120
	Z-Score			
	Test 1	Test 2	Test 3	Average Z-Score
Student:				
Mary	0.00	1.25	1.67	0.97
Joe	5.00	1.25	0.00	2.08
	Percentile Rank			
	Test 1	Test 2	Test 3	Average Rank
Student:				
Mary	50	89	95	83
Joe	99	89	50	98

[1]Assume Test 1 Scores $\sim N(30, 16)$, Test 2 Scores $\sim N(30, 64)$ and Test 3 Scores $\sim N(30, 144)$.

[2]Determined for the Z-Scores.

It is called Judgment Analysis (JAN) and it involves a hierarchial grouping of data using an iterative procedure (Ward 1961, 1963; Ward and Hook 1963). While this was a cluster analysis technique, Bottenberg and Christal (1968) used this idea of hierarchial grouping to combine regression equations, using minimal loss of predictive efficiency as the grouping criterion.

Originally, JAN was developed to solve problems faced by the Personnel Department of the Air Force (Christal 1968a; Bottenberg and Christal 1968).

2. POLICY-SPECIFYING AND POLICY-DEVELOPMENT WEIGHTS IN JAN

Weights

Policy-capturing requires a set of judgments (Y values) associated with n decision situations to obtain the implicit weights. However, in the policy-specifying process, the weights are determined without empirically obtained judgments (Y values) by stating desired properties of and relations among the predicted values in sufficient detail that the numerical weights become known.

Specifically let

b_j = the unknown weights to be determined by policy-specifying (corresponding to a_j in policy-capturing above). $j = 1,\ldots,k$

b_o = an unknown constant (corresponding to a_o)

x_j = variables corresponding to the predictor vectors above. These are not vectors of data but are variables which when given a set of weights b_j and b_o and a set of values for x_j will yield a composite value y.

Then we have the starting function

$$y = b_o + b_1x_1 + b_2x_2 + \ldots + b_jx_j + \ldots + b_kx_k$$

Prior to the policy-specifying process, the range of values for x_1, $x_2,\ldots x_k$ are known but the b_j and b_o values are not known. Policy-specifying proceeds by stating restrictive relations among the predicted values for various values of x_j. These policy statements result in restrictions on the values of b_j and b_o so that the numerical values of the weights can be determined. Specification is completed when k + 1 independent restrictions are imposed. Once the values of b_j and b_o are known, then predicted values, y, can be calculated for any values x_j.

Policy-capturing and policy-specifying can be combined to form a general process of policy-development. A particular decision maker may start by specifying several properties about relations among the predicted values. Whereas policy-specifying resulted in k + 1 restrictions on the k + 1 weights, b_j and b_o, the expressin of desired properties may result in only r k + 1 restructions on the b_j and b_o values.

Then imposing these r restrictions on the starting model results in a restricted model

$$y_r = c_o + c_1z_1 + c_2z_2 + \ldots + c_jz_j + \ldots + c_{k-r}z_{k-r}$$

where

z_i = new variables resulting from imposing the r restrictions.

Each z_i variable is a linear combination of the x_i variables. Now since there are still k + 1 - r unknown weights c_j and c_o to be computed it would be possible to use policy-capturing to find the c_j values. The decision maker could provide, for each of the n[n (k + 1 - r)] decision situations, y_i (i = 1,...,n) values associated with various profiles of information about the different situations. Then the least squares values of c_j can be computed for the model.

$$Y = c_oU + c_1Z^{(1)} + c_2Z^{(2)} + \ldots + c_jZ^{(j)} + \ldots + c_{k-r}Z^{(k-r)} + E^{(2)}$$

where

Y = a vector of judged values of dimension n.

$Z^{(j)}$ = the j^{th} predictor vector, of dimension n formed as linear combinations of the predictor vectors $X^{(j)}$ generated from information associated with the decision situations.

Having computed the least squares values for c_j and c_o the weighting system now produces values that both reflect the policy restrictions imposed by the policy-specifying process and the best fit to the empirical judgments.

3. GENERAL APPLICATIONS OF JAN

JAN has been used in several studies conducted by the U.S. Air Force for job evaluations and to stimulate officer promotion boards with a high degree of efficiency. Equations have also been designed to simulate career counselors in making initial assignments of airmen graduating from basic training (Dudycha, 1970).

The JAN technique has been applied in a prediction study of success in graduate education. In a study by Houston (1967) two variations of JAN were investigated--Normative JAN and Ipsative JAN. The purpose of the Normative JAN study was to determine the extent to which a policy regarding graduate admission standards existed among selected graduate faculty members at Colorado State College (now University of Northern Colorado). Basically, three sets of independent profile variables were used: (1) biographical dat (2) test data, and (3) major subject field data. Results from the Normative Jan study indicated essentially one policy was present in the group of judge

The Ipsative JAN study used for its dependent variable the rankings submitted by the judges who were requested to rank, without access to the three sets of independent profile variables used in the Normative JAN study, the doctoral graduates on a basis of personal knowledge. It was the intent this phase that the ratings or rankings be loaded with personality factors r readily available in the Normative JAN study. Results of this phase were tatistically significant, though weak from the predictive standpoint. The ractical significance of the Ipsative JAN study was in the suggestion of new lirections for subsequent research.

Williams, Gab, and Linden (1969) replicated Houston's Normative study at the University of North Dakota and sought to determine the policy of a university doctoral admissions board. Twelve members of the graduate faculty evaluated each graduate student's profile and place it into one of seven criterion categories (Q-sort). Each rater's policy was assessed or captured and the raters were grouped into appropriate clusters by the JAN process. The investigators found that at least two separate judgmental systems were present.

A further illustration of the versatility of the technique is provided in a study by Stock (1969) who sought to determine if systematic differences existed in the placement policies for special education students among special education personnel (teachers, administrators, and the members of the special education screening committee) responsible for placing the students in the public schools of Cheyenne, Wyoming. Colvert (1970) used JAN techniques in the identification and analysis of the consultant ratings of elementary student teachers at the University of Northern Colorado. Using JAN procedures, Chang (1970) designed a study to determine whether individuals serving in different official capacities in the State of Colorado had differing attitudes toward selection criteria for awarding college financial grants. Keelan et al. (1973) captured the leadership policies of selected fireman in the State of Colorado with the use of JAN.

The question of what is pornographic was investigated by J. Houston and S. Houston (1974) who used JAN as a methodology by testing this technique with three groups concerned wit this issue. These groups included doctoral students majoring in Psychology, Counseling and Guidance at the University of Northern Colorado, lawyers and police officers from the city of Greeley, Colorado. The JAN technique proved to be surprisingly effective in capturing and clustering the policies (specific and complex) of the judges from the three groups identified. As expected, many policies were present.

The problem of evaluating curriculum packages was explored by Torgunrud (1971) in a doctoral dissertation completed at the University of California at Lost Angeles under the direction of Dean John I. Goodlad. Torgunrud identified from the educational literature the following independent variables as important dimensions of any curriculum package or set of materials which are under consideration for possible adoption. These include: (1) valid and significant content, (2) significant elements of organization, (3) sequence providing a cumulative effect, (4) integration providing horizontal relationships, (5) value position clearly stated, (6) specificity providing direction, (7) flexibility providing alternatives, (8) accommodation for student participation, and (11) provision for measurement of achievement. After defining the variables, Torgunrud generated a sample of 100 profiles, each described on the 11 variables, by using techniques described by Naylor and Wheery (1965) for simulating stimuli with specified factor structure.

In another evaluation at the University of California at Los Angeles, Duff (1969) utilized JAN techniques to capture both the teacher-hiring policies (Ex Ante) of selected administrators and the administrators' evaluation policies (Ex Post) of teachers' on-the-job performance after their first year of paid teaching experience. Both types of policies (hiring and job performance) were analyzed for elements of predictive validity by the investigator.

The effectiveness of JAN in capturing and clustering raters' policies was investigated by Dudycha (1970) in a Monte Carlo evaluation of JAN as a methodology. Dudycha's outcomes show that the grouping process begins to break down when there are fewer than 200 stimuli being evaluated or 100 if ten or more stimulus dimensions are used. Consequently, the researcher using JAN must be concerned with the number of stimulus dimensions used in a relationships to the stimuli being evaluated. It is the present recommendation of the writer that a minimum of 100 stimuli be available for each judge on a maximum of 10 stimulus dimensions.

Other examples using Ipsative JAN are Christal (1968b) in which the researchers had to use their own knowledge to discover the variables being used by the single judge, and Holmes and Zedeck (1973) in which the judges were asked to judge paintings and also to relate qualities which the paintings exhibited. These qualities were then used to develop characteristics used as the predictors in the linear mathematical policy model. A Normative study using these characteristics followed.

The type of JAN used in a study can be further specified. Type A JAN would be used if the judges were dealing with the same subjects or profiles. Type B JAN designates a situation in which the judges each are making judgments on a different set of subjects or profiles.

Traditionally, JAN problems have involved predictors having a continuous distribution and have had dependent variables which were either ranked or categorical. It was demonstrated by Houston and Bolding (1974) that JAN is a special case of the general linear model. Because of this, any type of variable which could be used in a linear model could be used in JAN. Sets of non-redundant, dummy variables, for instance, can be used for the categories (Suits 1957). An example of this can be found in Christal (1968b) in which some of the variables were categorical.

Certain issues associated with the use of JAN have been debated (Houston 1974b). It has been suggested that a distribution be specified a priori for the judges to use. A second issue raised by statisticians was how many predictors (independent variables) should be used. Statistical studies have shown that ten should be the minimum. Practical considerations have suggested between five and seven. A third issue was the number of Ss to be given to each judge. Statistical studies employing Monte Carlo techniques have shown that a minimum of 200 should be used. Practical considerations indicate that between 30 and 60 profiles should be used in a policy-capturing situation. Another issue debated is whether a test of significance or a practical test should be used. Regression is a large sample procedure. Tests of significance useful in JAN (t and F) are designed to be powerful when samples are small with increasing power as the sample size increases. With a large sample size even the smallest decrease in predictability can be significant. Ward and Hook (1963) recommended looking for a break in the pattern of R^2 (RSQ) value decreases between stages in the analysis. Houston and Gilpin (1971) suggested a modification of this technique. They recommended establishing *a priori* the maximum decrease in predictability which the researcher would allow before considering the decrease to be meaningful. They suggested a .05 level as a general "rule of thumb".

JAN has been widely used as a policy-capturing procedure in the military. Some examples of military policy-capturing applications have been described in the following publications: Black (1973); Christal (1968a, 1968b); Gott (1974); Gooch (1972); Jones, Mannis, Martin, Summers, and Wagner (1976); Koplyay (1970); Koplyay, Albert, and Black (1976); Mullins and Usdin (1970); Ward and Davis (1963).

4. STUDENT POLICIES OF TEACHER EFFECTIVENESS

The student judgmental policies of teacher effectiveness were analyzed in study completed by Houston and Gilpin (1971).

Procedures. The primary problem of the investigation was to analyze the results of a teacher description study and to identify judgmental policies of selected subsets of students at the University of Northern Colorado. The subjects for which profile and judgment scores were generated were faculty members of the University of Northern Colorado.

The judges. Students rated the teachers using the criteria represented in Instrument One. For purposes of this study, the students were grouped into selected subsets. The first grouping was made by schools or colleges within the university and resulted in seven subsets or groups of students. The researcher treated each of the individual groups as a judge in the first JAN investigation. The second grouping of students was determined by grade level and allowed for five subsets of students ranging from freshman through graduate level. Each of these distinct groups was treated as an individual judge in the second JAN analysis. Therefore, in the JAN analyses, a slight innovation was used. In the usual JAN a judge is an individual; however, in this study the individuals were grouped into subsets and each subset, consisting of numerous individuals, was considered a judge.

The instrument. The student raters were requested to rank teachers on the first 9 items and to provide biographical information asked for in item 10 of the following instrument:

Teacher Description Instrument (Instrument One)

Please rate only this teacher in this particular course in accordance with this rating scale. 1) Poor 2) Fair 3) Average 4) Good 5) Excellent

1.	Teacher's interest and enthusiasm for course	1	2	3	4	5
2.	Ability to adequately answer questions	1	2	3	4	5
3	Ability to communicate the subject matter effectively	1	2	3	4	5
4.	Ability to interest and motivate students	1	2	3	4	5
5.	Fairness in testing and grading	1	2	3	4	5
6.	Personal interest and adaptation to student's needs	1	2	3	4	5
7.	Course objectives are clearly stated	1	2	3	4	5
8.	Course objectives are met	1	2	3	4	5
9.	Everything considered, including strengths and weaknesses, I would rate the instructor	1	2	3	4	5
10.	1) Freshman 2) Sophomore 3) Junior 4) Senior 5) Grad					

The first eight items of Instrument One were considered independent variables while item nine was treated as the dependent variable in multiple linear regression analyses. Responses to the first eight variables were also used as profile scores, and responses to item nine as judgments in the two JAN analyses.

JAN techniques. The JAN technique starts with the assumption that each judge has an individual policy. It gives and R^2 (multiple R coefficient squared) for each individual judge and an overall R^2 for the initial stage consisting of all the judges, and each one treated as an individual system. Two policies are selected and combined on the basis of having the most homogeneous prediction equations, therefore resulting in the least possible loss in predictive efficiency. This selection reduces the number of original policies by one and gives a new R^2 for this stage. The loss in predictive efficiency can be measured by finding the drop in R^2 between the two stages. The grouping procedure continues, reducing the number of policies by one at each stage, until finally all of the judges have been clustered into a single group.

Investigators examined the collective drop in R^2 from that of the original stage in each of the two JAN analyses. A determination of whether one or more policies were present among the judges was made on the basis of the sequential drop in R^2. A slippage greater than .05 was considered *a priori* to represent too great a loss in predictability.

Findings

The first JAN analysis considered the students grouped into the seven schools and/or colleges of the University of Northern Colorado. Each group was treated in the analysis as an individual judge. A listing and abbreviation of the variables for this study are found in Table 2.

Stages of the JAN procedure for judges by school and/or colleges. The R^2s for each of the seven initial systems are reported in Table 3. Note that the magnitudes of R^2 are restricted in range. The highest value is .8309 for judge four and lowest is .7443 for judge seven. These high values of R^2 for all judges indicated that the judges were consistent in their individual decision-making policies.

Table 4 reports the seven stages of the JAN clustering procedure for the seven judges and the corresponding R^2 for each stage. In stage 2, judges two and three have been combined to form one group while all other judges are treated individually. The drop in R^2 between stages 1 and 2 is only .004. Continuing this clustering procedure, stage 3 combined judges five and six resulting in a model consisting of five policies or systems. The resulting drop in R^2 from stage 1 is .0009.

Stage 7 combined all seven judges into one cluster and resulted in a collective drop in R^2 of only .0248. The a priori criterion for permissible slippage in R^2 was .05. Since the collective drop of .0248 is well within this tolerance level, stage 7 was accepted as the appropriate grouping of judges. Therefore, the investigators concluded that only one policy was present among the seven judges.

Policy of the seven judges. Interpretation of the JAN procedure determined that only one policy existed among the seven judges representing the schools and/or colleges. Regression analysis was then employed in an effort to explain that policy.

The investigators were interested in determining the unique contribution of proper subsets of the predictor variables, 1 through 8, to the prediction of the criterion, GenP. The contribution of a set of variables to prediction may be measured by the difference between the R^2 for the full model (FM) and the R^2 for a restricted model (RM). The RM differs from the FM in that the proper subset of variables, for which the unique contribution to predictability is desired, have been deleted. The difference between the two R^2s may be tested for statistical significance through use of an F test or else an a priori acceptable drop can be established. The investigators chose the latter alternative and set a drop tolerance of .05. That is, if $R^2_F - R^2_F$ .05, the investigators concluded that the subset under consideration was making a unique contribution to prediction of the criterion.

A subjective hierarchy of the variables is presented in Table 5. This grouping was used in the regression analysis of the different policies.

Figure 1 presents a schematic to guide the sequence of tests from the FM through the various restricted models. The accompanying R^2 for each of these models is found in the appropriate block. For example, the information in block 1 indicates that the independent variables 1 through 8 were used as the predictors in the FM and that the R^2 for this model was .8123.

Block 2 displays FM - (5,6,7,8), indicating that variables (5,6,7,8) have been deleted from the full model. This also implies that variables 1, 2, 3, and 4 are used as the predictor variables in the RM. By dropping out variables (5,6,7,8), the unique contribution to prediction of these variables can be determined. The measure of this unique contribution was found by the difference between the R^2 = .8123 for the FM and the R^2 = .7742 for this RM. The difference .8123 - .7742 = .0381 was less than .05 and therefore indicated that these variables were making little or no contribution to prediction that could not be explained by the other four predictor variables. Since the drop in R^2 for this set was not significant, no further tests of subsets of these variables were necessary. The broken line in the chart indicates that further testing of subsets of variables (5,6,7,8) was terminated.

The expression in block 3, FM - (1,2,3,4), indicates that variables (1,2,3,4) were eliminated from the FM. These predictors were grouped on the subjective basis that they were related and measured a general hypothetical category called methodology. The drop .8123 - .6673 = .1450 was greater than .05 and therefore resulted in too great a loss in predictive efficiency. Therefore, further analysis of subsets of these variables was undertaken. However the R^2 for the model FM - (1,4) was .7768. Since the drop of .0335 was less than .05, variables (1,4) made no significant contribution to prediction of the criterion. An examination of the subset represented by the model FM - (2,3) showed that the drop in R^2 was equal to .0376. Again the drop was less than .05, and it was concluded that variables (2,3) made an insufficient unique contribution to the prediction of the criterion. Multicollinearity of the variables (1,2,3,4) accounted for the fact that no significant drop in R^2 was detected when further analysis of the branchings from this set were examined. That is, the variables in this set are highly intercorrelated, and when two of them are eliminated, the presence of the other two in the FM hold up the value of R^2. The broken line again indicates that further examination of subsets of these variables was not needed.

In summary, the eight predictor variables were very efficient in predicting the criterion since the R^2 was reported to be .8123. The model FM - (5,6,7,8) also had high prediction efficiency with an R^2 = .7742. Therefore, all of the judges who were clustered into the only policy-making system were attending to variables 1, 2, 3 and 4 when they were rating teachers in the general overall category.

As reported, the grouping of subsets of the eight predictor variables was a completely subjective determination. The investigators were interested in analyzing Table 6, the intercorrelations of predictors and the validities, to determine if a different hierarchy of variables would result. Perhaps a smaller subset of variables making a unique contribution to prediction could be found if the subsets were grouped differently.

The validities were comparatively high, ranging from .604 to a high of .804. The investigators grouped the predictors into a hierarchy base upon the correlations. This grouping is presented in Table 7.

The schematic sequence of tests is presented in Figure 2. The branching eading from block 2 was terminated in view of the resulting R^2 = .7848 for he model FM - (1,5,7,8). This represented a drop of only .0275, well within he .05 level. Of considerable interest was the alternate branching leading o and from block 3. The model FM - (2,3,4,6) yielded a significant drop in ^{2} of .8123 - .6758 = .1365. This prompted further investigation of subsets f this model. The model FM - (2,6) accounted for a drop of only .8123 - .7939 = .0184, and hence further investigation of subsequent branching was nded. However, the model FM - (3,4,6) was of extreme interest in view of the ignificant drop in R^2 of .8123 - .7248 = .0875. Consequently further ranching from this model was investigated. The model FM - (3,4) was also found to make a unique contribution since the drop of .8123 - .7558 = .0565. Further analysis of the unique contribution of variables 3 and 4, treated individually, resulted in nonsignificant findings. The reason for this finding was that variables 3 and 4 were highly related $r_{3,4}$ = .75.

The regression analysis based on correlations (Table 7) allowed for a more refined interpretation than did the analysis based on subjectivity. The hierarchy suggested by the correlations led not only to a set of three variables (3,4,6) making a unique contribution, but also to a set of only two predictors (3,4) making a unique contribution to prediction.

An interesting question arose at this juncture. The two sets of variables (3,4,6) and (3,4) both make unique contributions, but what about their absolute or total prediction? This information is not available from the sequence of tests in Figure 2. The researchers investigated the predictive efficiency of the RM models consisting of the set of variables (3,4,6) and (3,4). The R^2 for the RM consisting of variables (3,4,6) was equal to .7678. The difference was .8123 - .7678 = .0445 which, by virtue of the .05 convention used in this study, implied that this RM predicted as well as did the FM. However, the RM consisting of variables 3 and 4 had an R^2 = .7340 which obviously was not as efficient as was the FM.

<u>JAN by grade level.</u> The second JAN analysis grouped students according to grade level. Each of the five levels was considered as a judge. Table 8 shows the R^2s associated with the prediction equation for each of the five judges. The R^2s ranged in value from .7988 for freshmen to .8344 for seniors. The high R^2s indicated efficient prediction for each of the respective regression or decision-making equations.

The five stages of the JAN grouping technique are presented in Table 9. As conjectured from observation of the preliminary statistics, the collective drop in R^2 from the original stage to stage 5 was somewhat less than the .05 limit.

Stage 2 combined the freshmen and sophomores, leaving the juniors, seniors and graduates as the three single-member systems. This combination resulted in an R^2 slippage of only .002. Stage 3 clustered the juniors and seniors leaving the graduate students as the only singleton set. The collective drop in R^2 at this stage was a nearly indiscernible .0005. Stage 4 combined the sets containing two judges each into a cluster of four, again leaving judge five as the only single-member system. At this stage the overall drop in R^2 was an inconsequential .0015. Stage 5 grouped all of the judges into one decision-making system and resulted in a total R^2 slippage of only .003. Certainly this drop in R^2 was well within the tolerance range of .05. These data suggest that only one judgmental policy was existent among the five judges.

TABLE 2
List of Variables and Abbreviations

Number	Variable	Abbr.
1.	Teacher's interest and enthusiasm for course	IEth
2.	Ability to adequately answer questions	AnsQ
3.	Ability to communicate subject matter effectively	CSub
4.	Ability to interest and motivate students	MoSt
5.	Fairness in testing and grading	TeCr
6.	Personal interest and adaptation to student's needs	SNds
7.	Course objectives are clearly stated	CObS
8.	Course objectives are met	CObM
9.	General rating (criterion)	GenR

TABLE 3
R^2 Values for All Judges from Regression Models

	Judge	R^2
1.	School of the Arts	.7869
2.	College of Arts and Sciences	.8126
3.	School of Business	.7764
4.	College of Education	.8309
5.	School of Health, Physical Education, and Recreation	.7992
6.	School of Music	.8075
7.	School of Nursing	.7443

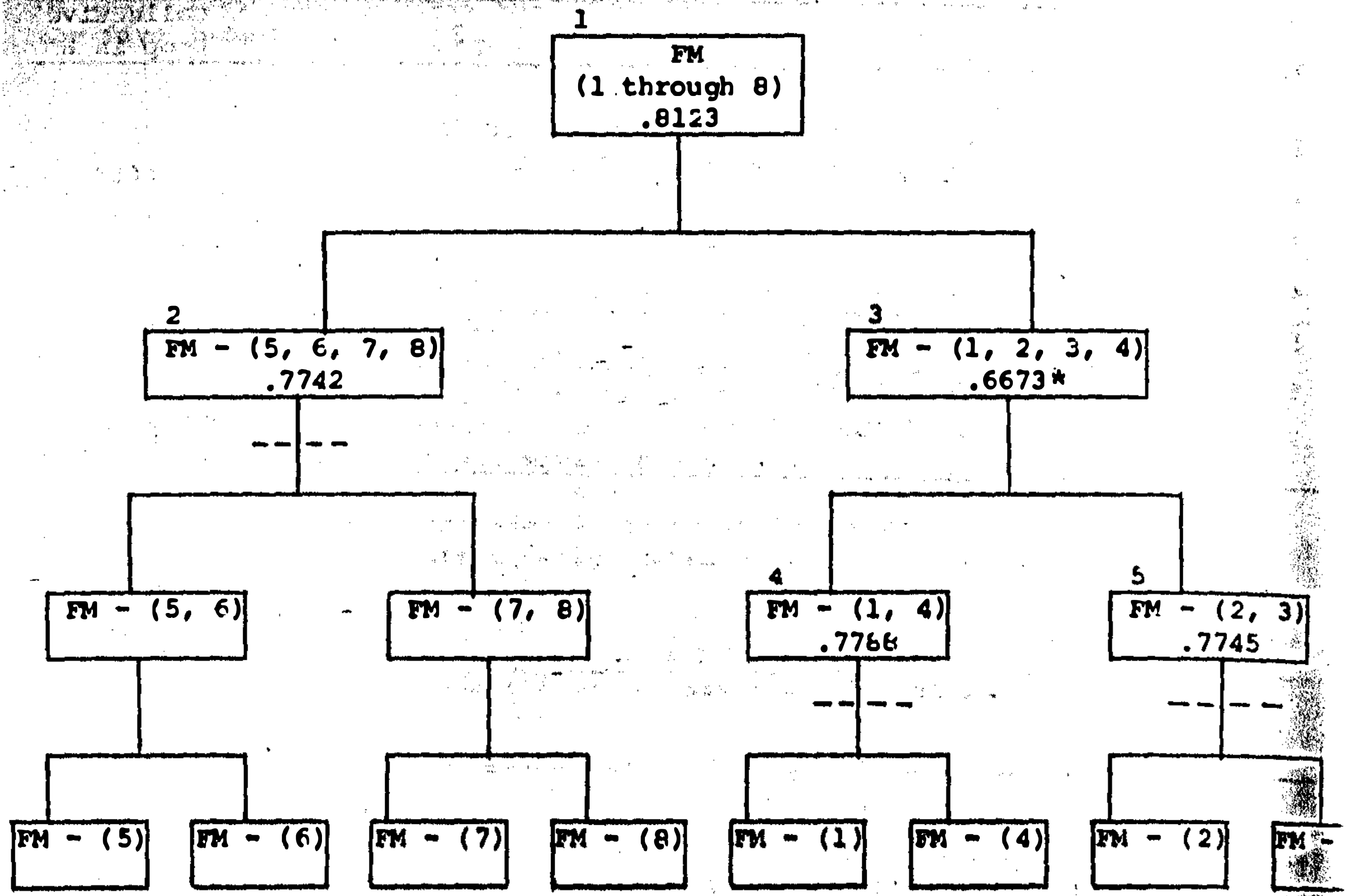
TABLE 4
Stages of the JAN Procedure for the Seven Judges

Stage	Judges	R^2	Collective Drop in R^2
1	1, 2, 3, 4, 5, 6, 7	.8141	
2	(2, 3), 1, 4, 5, 6, 7	.8137	.0004
3	(2, 3), (5, 6), 1, 4, 7	.8132	.0009
4	(1, 4), (2, 3), (5, 6), 7	.8121	.0019
5	(1, 4), (2, 3, 7), (5, 6)	.8099	.0042
6	(1, 4, 2, 3, 7), (5, 6)	.8064	.0077
7	(1, 4, 2, 3, 7, 5, 6)	.7893	.0248

TABLE 5
Subjective Hierarchy of Variables

Methodology:	
Teacher's interest and enthusiasm for course	(1)
Ability to interest and motivate students	(4)
Ability to adequately answer questions	(2)
Ability to communicate subject matter effectively	(3)
Humanistic:	
Fairness in testing and grading	(5)
Personal interest and adaptation to student's needs	(6)
Organizational:	
Course objectives are clearly stated	(7)
Course objectives are met	(8)

FIGURE 1
Seven Judged (Subjective Hierarchy)



*Significant drop in R^2.

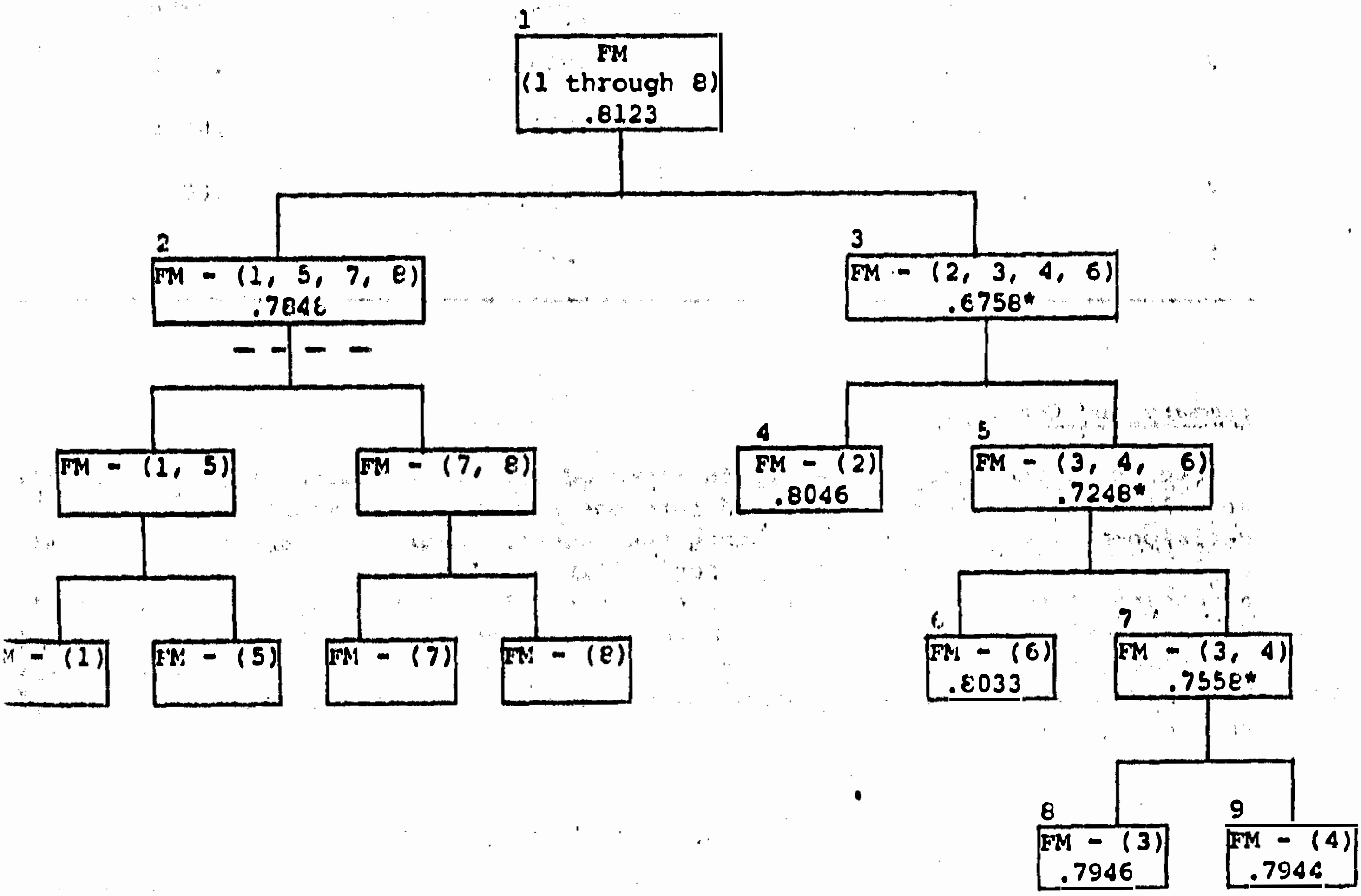
TABLE 6
Correlations of Predictor and Criterion Variables

Variable		1	2	3	4	5	6	7	8
1.	IEth								
2.	AnsQ	.580							
3.	CSub	.606	.696						
4.	MoSt	.646	.621	.746					
5.	ToGr	.426	.471	.492	.522				
6.	Snds	.558	.566	.613	.688	.582			
7.	CObS	.477	.507	.580	.550	.467	.532		
8.	CObM	.532	.564	.633	.618	.510	.578	.794	
9.	GenR	.688	.715	.716	.804	.604	.728	.623	.699

TABLE 7
Hierarchy of Variables Based on Correlations

Subset 1	
Sub-subsets:	
Ability to interest and motivate students	(4)
Ability to communicate subject matter effectively	(3)
Personal interest and adaptation to student's needs	(6)
Ability to adequately answer questions	(2)
Subset 2	
Sub-subsets:	
Course objectives are met	(8)
Teachers interest and enthusiasm for course	(1)
Course objectives are clearly stated	(7)
Fairness in testing and grading	(5)

FIGURE 2
Seven Judges (Hierarchy Based on Correlations)



*Significant drop in R^2.

TABLE 8

R^2 Values for All Judges from Regression Models

	Judges	R^2
1.	Freshmen	.7988
2.	Sophomores	.7954
3.	Juniors	.8165
4.	Seniors	.8344
5.	Graduates	.8276

TABLE 9

Stages of the JAN Procedure for the Five Judges

Stage	Judges	R^2	Collective Drop in R^2
1	1, 2, 3, 4, 5	.8136	.0000
2	(1, 2), 3, 4, 5	.8134	.0002
3	(1, 2), (3, 4), 5	.8131	.0005
4	(1, 2, 3, 4), 5	.8121	.0015
5	(1, 2, 3, 4, 5)	.8106	.0030

Summary and Conclusions

Results of the first JAN analysis revealed the seven judges, representing the schools and/or colleges, clustered into one system. This meant that only one decision-making policy existed among the judges. Regression analysis was used to explain this single judgmental policy and it was found that the judges were attending primarily to variables 3, 4, and 6. An interesting finding was that the RM using only variables 3, 4, and 6 resulted in predictive efficiency significantly equivalent to that of the FM. Judges representing the five grade levels were also clustered into one system as a result of the hierarchical grouping procedure of the second JAN analysis.

5. EVALUATING THE EVALUATORS VIA JAN

What is now presented is an application of JAN to indicate how it might be used to evaluate evaluators.

The League of Cooperating Schools (LCS) was launched in May 1966, as a 5-year project to study and promote planned change in American education. It was sponsored by a partnership of the University of California at Los Angeles, the Institute of Development of Educational Activities, Inc., and eighteen independent school districts in Southern California. Each school district contributed one League school and these districts ranged in size from the massive Los Angeles City system to a small district of only three schools. The districts and schools were selected in such a way as to represent, hopefully, a true microcosm of American elementary schools. It was the aim of this joint enterprise to develop a cohesive program of research, development, innovation, and dissemination of information in order to narrow the chasm between current educational theory and practice.

In order to effect educational change, a rationale was needed that would serve as a basis for research design while at the same time serving the interests of the cooperating schools. The result was the creation of a new social system in which principals and teachers in the LCS were to be challenged by I/D/E/A to fashion new norms, roles, supports and rewards for themselves.

Four members of the Intervention Staff were requested to score on a 5-point scale each of eighteen schools on eight characteristics deemed essential for effective schools. A list of these characteristics with explanations appears in Table 10 (variables 1-8). In addition, the Intervention Staff members were asked to rank the eighteen schools in terms of overall effectiveness. The rankings were used as the criterion variable in the JAN process. This procedure represents a slight modification of the usual JAN procedure in that the judges generated their own profiles by the scores they gave on variables 1-8.

In Table 11 appears the intercorrelations between all the variables. The means and standard deviations are presented in Table 12. A multiple linear regression equation was developed for each Intervention Staff member who served as judge. Table 13 contains the correlations of each predictor variable and the criterion variable (school rank). Also included for each rater is his multiple correlation coefficient.

Table 14 summarizes intercorrelations of judgmental policies. It appears that judges 3 and 4 have the most homogeneous policy as the correlation coefficient rating their rankings of effective schools is 0.90. This is borne out in Table 15 which gives the stage values for the JAN technique. In Stage 2, two groups have been formed and judges 3 and 4 have been first to be grouped. The investigators conclude that there are essentially two policies present. The justification for this stems from the fact that the collective drop in r^2 from Stage 1 to Stage 3 is just 0.0361 while the drop from Stage 3 to Stage 4 results in a loss of 0.0678 making the collective drop 0.1060. From Table 15 one can see in Stage 3 that judges 1 and 2 comprise one policy group while judges 3 and 4 form the second policy group.

In analyzing the policies one might wish to refer to Table 13 which reports the correlations between the school characteristics and judges. However, one finds a distressing situation in that all the intercorrelations are high. This means that the judges may have been guilty of the "halo effect" as they generated their profile scores for the eighteen schools.

The investigators were interested in determining the unique contribution of proper subsets of the predictor variables, 1-8, to the prediction of the criterion, JANCr, in both policies to compensate for multicollinearity.

For an explanation of the two judgmental policies, the investigators first made a subjective analysis of the predictors and conjectured that they formed a hierarchical pattern as displayed in Table 16.

Presented in Table 17 is a schematic to guide the sequence of tests associated with the single policy of Judges 1 and 2.

In summary the eight predictor variables were very efficient in predicting the criterion since the R^2 was reported to be 0.8672. Policy I as expressed by Judges 1 and 2 could basically be explained as a concern for the competence of the professional team (variables 1, 2, and 3).

In Table 18 appears a schematic which illustrates the second policy, namely the of judges 3 and 4. From blocks 2, 3, and 4, it can be seen that each of the three subsets in the subjective hierarchy was making a significant unique contribution to predicting the criterion.

TABLE 10
List of Variables

Number	Variable	Abbr.
1.	Extent professional team (principal and teachers) shows enthusiasm about their school program	PEnt
2.	Extent professional team is action-oriented; i.e., they put their ideas into practice	PAct
3.	Extent professional team is inquiring and searching intullecutally and self-critical	PInq
4.	Extent children are involved in educational activity (can observe and talk to children)	CInv
5.	Extent teacher concerns are with each child as an individual. (One can gain information from children, teachers, or parents.)	TChC
6.	Extent the district supports and shows pride in the school program	DSup
7.	Extent of community support (the program is supported by participation in school life, publicity, etc.)	CSup
8.	The quality of the educational program vis-a-vis individualization of instruction is evident (alternatives, conferences, different grouping procedures, etc.)	QEdPr
9.	JAN criterion--rank of school	JANCr

TABLE 11
Intercorrelations

Variable		1	2	3	4	5	6	7	8
PEnt	1								
PAct	2	.83							
PInq	3	.56	.79						
CInv	4	.66	.71	.71					
TChC	5	.70	.74	.72	.74				
DSup	6	.80	.60	.64	.73	.60			
CSup	7	.74	.76	.84	.77	.77	.67		
ÇEdPr	8	.58	.66	.65	.79	.73	.46	.67	
JANCr	9	.57	.74	.82	.75	.71	.56	.59	.71

TABLE 12
Means and Standard Deviations (N = 18)

Variable		Mean	Standard Deviation
1	PEnt	2.333	.594
2	PAct	1.944	.872
3	PInq	1.722	.826
4	GInv	1.388	.698
5	TChC	1.833	.707
6	DSup	1.777	.878
7	CSup	1.611	.850
8	ÇEdPr	1.666	.686
9	JANCr	9.500	5.338

TABLE 13
Correlations Between Judges and School Characteristics

Judge	School Characteristics								
	PEnt	PAct	PInq	GInv	TChC	DSup	CSup	CEdPr	R
1	0.56	0.74	0.82	0.75	0.71	0.56	0.59	0.71	0.95
2	0.57	0.59	0.62	0.77	0.69	0.63	0.63	0.59	0.81
3	0.87	0.88	0.69	0.77	0.83	0.66	0.76	0.63	0.94
4	0.85	0.85	0.71	0.73	0.80	0.69	0.82	0.69	0.93

TABLE 14
Intercorrelations of Judges

Judge	1	2	3	4
1	1.00	0.68	0.71	0.63
2	0.68	1.00	0.69	0.66
3	0.71	0.69	1.00	0.90
4	0.63	0.66	0.90	1.00

TABLE 15
Stages of the JAN Procedure

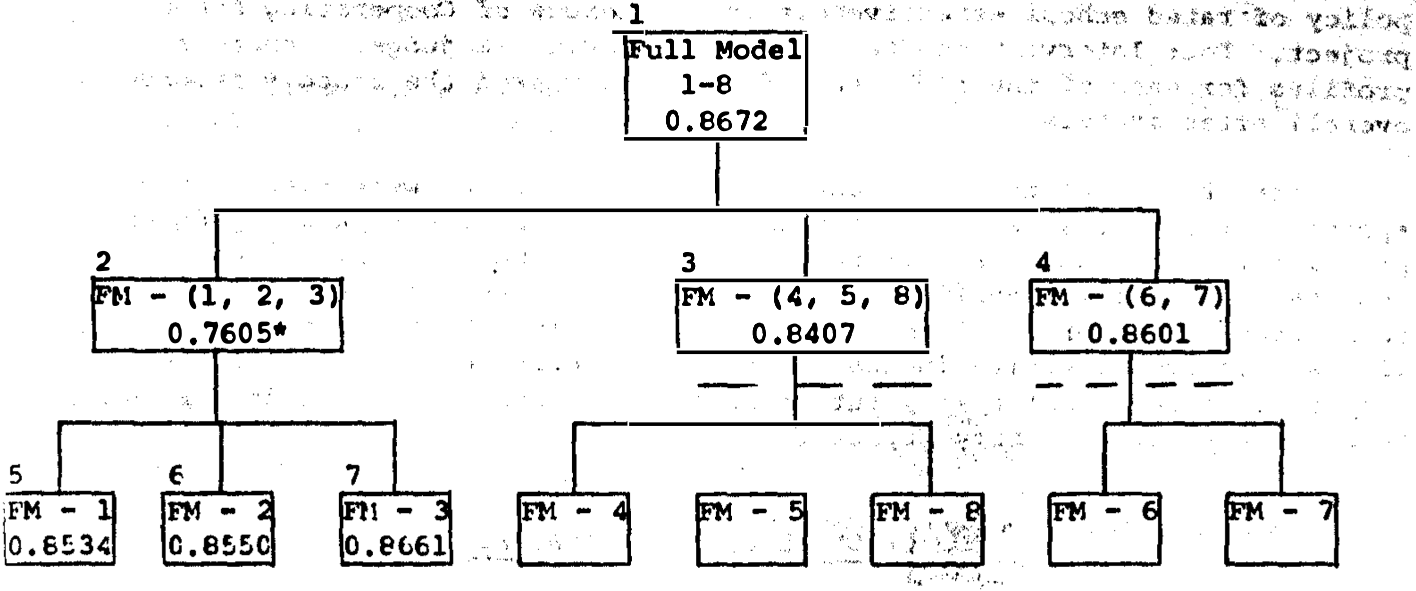
Stage	Judges	R^2	Collectiv Drop in
1	1,2,3,4	.8302	
2	(3,4), 1,2	.8222	.0080
3	(3,4), (1,2)	.7921	.0381
4	(1,2,3,4)	.7242	.1060

TABLE 16
Subjective Hierarchy of Variables

Professional staff competence:	Extent professional team is enthusiastic	(1)
	Extent professional team is action-oriented	(2)
	Extent professional team is inquiring and self-critical	(3)
Concern for children:	Extent children are involved in educational activity	(4)
	Extent teacher concerns are with child as individual	(5)
	Extent of individualized instruction	(8)
Outside support:	Extent of district support	(6)
	Extent of community support	(7)

TABLE 17

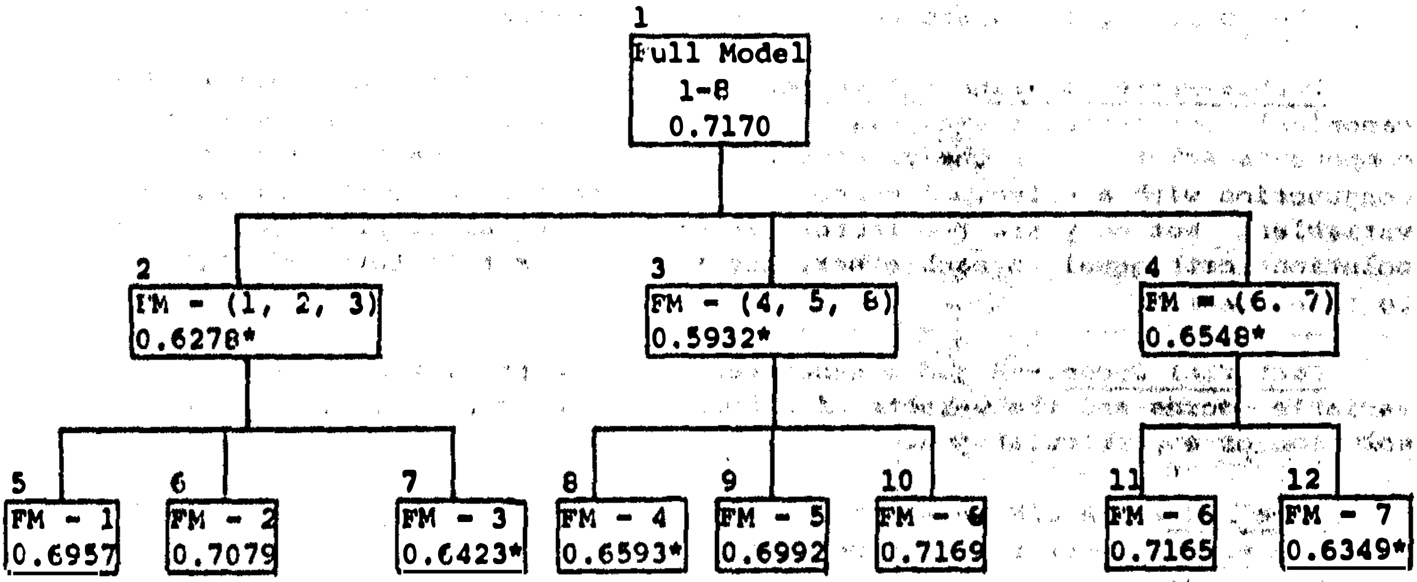
Flowchart of Regression Analysis of Policy I (Judges 1 and 2)



*Significant drop in R^2.

TABLE 18

Flowchart of Regression Analysis of Policy II (Judges 3 and 4)



*Significant drop in R^2.

In summary, the eight predictor variables were efficient in predicting the criterion for judges 3 and 4, though not as efficient as in Policy I. Policy II differed from Policy I in that each of the three hypothetical subsets made a significant unique contribution.

Summary.--In this study, an attempt was made to demonstrate the feasibility of utilizing a modified form of JAN as a vehicle for identifying a policy of rated school effectiveness in the League of Cooperating Schools project. Four Intervention Staff members, serving as judges, generated profiles for each of the eighteen LCS and then ranked the schools in order of overall effectiveness.

With the use of the JAN technique, the four judges were placed into appropriate clusters, and it was found that at least two separate judgmental policies were present. A regression analysis of the two policies was undertaken. Policy I could be explained basically as a concern for the competence of the professional team in the schools. On the other hand, Policy II was more comprehensive in that it not only reflected a concern for a competent professional staff, but it included a concern for children as well as a concern for community support.

6. CANONIAL JUDGMENT ANALYSIS

What is now proposed is a strategy in which the JAN technique can be extended to include the ratings of judges on two or more criterion variables or dimensions. The technique is identified as Canonical Judgment Analysis or C-JAN. The C-JAN technique was successfully used by Johnson and King (1973) in a team doctoral dissertation at the University of Northern Colorado.

Definition of Terms

The following terms are defined in the development of C-JAN:

Double-Barreled Principal Components Solution.--A factor solution for a canonical correlational analysis. In this type of factor solution a principal components solution for the predictor (profile) variables is given in conjunction with a principal components solution for the criterion (judgment) variables. Not only are the factors in each of the above principal component solutions orthogonal to each other, but the cross-set factors are orthogonal to each other.

Factorial Judge.--A judge generated from the predictor and criterion variable scores and the weights of a double-barreled principal components solution of a particular judge.

Type A JAN.--A JAN in which all the judges give ratings on the same subjects with respect to the same criterion variable and predictor variables.

Type B JAN.--A JAN in which the judges do not rate the same subjects with respect to the same criterion and predictor variables.

teps in C-JAN Process

tep 1

For each judge run a canonical correlation analysis using Veldman's 1967) CANONA program. Let the judges be J_k for $k = 1,\ldots,m$

tep 2

For each judge, J_k, determine the number of factorial judges, $_{K,F1}, J_{K,F2}, \ldots J_{k,Fn_F}$.

is is where J_k, would be the ith factorial judge generated from the ith ctor for the kth judge. Also, n_F = the number of significant factors.

1. Let $\underline{Z}_{Fi}$ be the canonical predictor factor score vector for the ith factor for the kth judge..

2. Let $\underline{U}_{Fi}$ be the canonical criterion factor score vector associated with $\underline{Z}_{Fi}$ for the kth judge.

3. Let $(a_{1}, Fi)i=1$ be the weight vector for the jth predictor factor for the kth judge.

4. Let $(b_{i,Fj})i=1^{t}$ be the weight vector for the jth criterion factor for the kth judge.

5. Let the following model be used in the JAN process for the factoral judge $J_{k,Fi}$ for $i=1, \ldots, n_F$.

   The criterion vector: $(\underline{Z}'_{Fi}, \underline{U}'_{Fi})'$

   The profile matrix:

   $a_{1,Fi}{}^{*}X_1 \quad a_{2,Fi}{}^{*}\lambda_2 \ \ldots \ a_{s,Fi}{}^{*}\lambda_s \quad b_{1,Fi}{}^{*}Y_1 \ \ldots \ b_{t,Fi}{}^{*}Y_T$

xx	xx	...	...	xx					
xx	xx	...	...	xx		O_{Nxt}			
xx	xx	...	...	xx					
					xx	...	...	xx	
		O_{Nxs}			xx	...	...	xx	
					xx	...	...	xx	

N = number of subjects for J_k.

Step 3

Determine the judges who should be retained. Judges who identify at least one significant canonical factor should be retained in the analysis. Any judge who is unable to identify at least one significant factor should be eliminated as he is failing to relate any predictor variable set to any criterion variable set. After eliminating inconsistent judges, a Type A or Type B (JAN) should be completed on all of the factorial judges identified in the study.

Step 4

For every policy captured in Step 3 form a matrix in which each column represents the respective factorial judge's original factor loadings. These loadings will be obtained from the CANONA printout for the judge from which the factorial judge was generated. Include along with this matrix the corresponding vector of canonical correlations for the original CANONA printout.

Step 5

At this point aided with the data presented in Step 4, the researcher should make an intuitive analysis of each of the captured factorial policies in order to determine relationships between predictor variable sets and criterion variable sets.

A limitation in this approach to C-JAN is that a single judge may be allowed to express more than one policy as more than one canonical correlation associated with his judgments may be significant. Unfortunately this full C-JAN technique is so complex that it has rarely been used.

Instead we propose a simplified C-JAN methodology which may be suitable for use in many practical situations and avoids much of the complexity of the full C-JAN methodology. Essentially, the canonical analysis will only be used as a data reduction technique to reduce the multiple criterion variables to a single criterion variable. This then allows use of the standard JAN analysis. This approach would be suitable for the case in which judge's rankings on the multiple criterion variables display a degree of redundancy. The basic steps are as follows:

1. Give a set of N profiles to the K judges and have them rank the profiles on the specified criterion variables.

2. Use canonical correlation analysis to produce a set of canonical functions for each judge using the judge's rankings as one canonical set and the profile variables as the second canonical set.

3. Check the canonical correlation between the first and second two canonical functions for each judge. To continue with the simplified C-JAN procedure, it would be necessary for the first canonical functions to be of practical significance and the second and further

possible canonical functions to be of little or no practical significance. If even the first canonical R is of no significance for a particular judge, the judge should not be used in further analysis. If more than the first canonical functions are highly important, the more complex C-JAN procedure must be used.

4. Use the first criterion canonical function to produce a new canonical variate for each judge. Substitute the new canonial variate for the original set of criterion ranking variables for each judge. Substitute the new canonical variate for the original set of criterion ranking variables for each judge.

5. Proceed with the standard JAN analysis as described in the previous section.

6. If multicollinearity of the profile variable set is not a problem, then regression analysis can be used to capture the judgment policies as usual. If multicollinearity is a problem, then canonical correlation analysis may be used to help determine the judgmental policies.

The logic behind this procedure is quite straightforward. The first ınonical criterion function is the linear combination of the criterion ıriables which extracts the maximum possible variance of the criterion ıriables and has the maximum covariance with the first canonical function of ıe profile variables. We are attempting to maximize the simplicity of ıbsequent data analysis while minimizing the loss of information.

plication Example

Many institutions of higher education have internal funds which are used ɔ support the beginning stages of research which may lead to outside funding nd publishable journal articles. It is typical for such funds to be llocated by committee decision. Several interesting questions might be aised about such decisions:

1. Given a set of profile descriptors of a research proposal, how many different judgmental policies exist among the committee members in determining the quality of the research proposals?

2. Which descriptors do the differing judgmental policy groups emphasize in determining proposal quality?

The following example illustrated the C-JAN approach in answering the stated questions. We first constructed a set of 32 hypothetical descriptions of proposals by use of simulation techniques. A sample profile appears in Table 19.

TABLE 19
Sample Research Proposal Profile

Profile Variable ID Numbers and Descriptors	Weak 1	2	3	Average 4	5	6	7	Strong 8	9	10
1. Need			3							
2. Feasibility								8		
3. Cost benefit				4						
4. Quality of writing		2								
5. Originality						6				
Judges' Overall Rating (repeated rankings not allowed)	Rank Profile from 1st (strongest) to 32nd (weakest)									
Possibility of generating outside funding										
Possibility of leading to publishable journal research										

The set of 32 profiles was then submitted to each of four members of a hypothetical proposal funding committee. The judges were required to independently rank their set of profiled from strongest (1st) to weakest (32nd) based on the profile descriptor values. This ranking had to be accomplished first for the possibility that the proposed research would lead to outside funding, and secondly, for the possibility thd proposed research would generate journal publication. The rankings for each of the criterion variables should be carried out at separate times in order to minimize halo effect. Tied rankings were not allowed for any particular criterion variable.

Tables 20 and 21 show means, standard deviations and intercorrelations of the five simulated profile variables. The simulated profiles appear to be quite good with consistent means, standard deviations, and low intercorrelations between the profile variables.

TABLE 20
Means and Standard Deviations (N = 32)

Variable	Mean	Standard Deviation
1	6.25	2.54
2	5.69	2.76
3	5.34	2.73
4	5.72	3.15
5	5.25	2.80

TABLE 21
Intercorrelations of the Profile Variables

	Research Proposal Profile Variables				
	1	2	3	4	5
1	1.00	-.28	-.23	-.24	.23
2	-.28	1.00	-.03	-.19	-.13
3	-.23	-.03	1.00	.09	-.06
4	-.24	-.19	.09	1.00	.01
5	.23	-.13	-.06	.01	1.00

The set of two criterion variable rankings and the five profile variables were then subjected to canonical correlation analysis for each judge. The canonical correlations for this analysis are displayed in Table 22.

TABLE 22

Canonical Correlations Between the Ranking and Profile Variable Sets by Judge

Judge Number	Canonical R 1st	2nd
1	.959	.272
2	.899	.541
3	.916	.367
4	.915	.329

In each case the first canonical correlation is very strong while the second is comparatively weak. We therefore proceeded with the simplified C-JAN procedure. The first canonical function for the criterion variable set was used to produce a single canonical variable for each judge. The original set of two criterion variable rankings was replaced by the single canonical variable.

The modified data were then analyzed by means of the JAN procedure which computes a regression equation for each judge and then hierarchically clusters the judges based on the homogeneity of their prediction equations. A general idea of which judges will cluster together can be determined by looking at Table 23 which shows the intercorrelations of the judges.

TABLE 23

Intercorrelations of Judge's Ratings

Judge	1	2	3	4
1	1.00	.46	.39	.49
2	.46	1.00	.95	.94
3	.39	.95	1.00	.95
4	.49	.94	.95	1.00

stages of the JAN process are displayed in Table 24.

TABLE 24
Stages of the JAN Procedure for the Four Judges

ge	Judges	System R^2	Total System R^2 Drop
1	1, 2, 3, 4	.8507	
2	(2, 4), 1, 3	.8497	.0011
3	(2, 3, 4), 1	.8472	.0035
4	(1, 2, 3, 4)	.6864	.1643

ing an *a priori* criterion of an R^2 drop of .05 or more as indicating a parture from linearity, the clustering of judges is easily determined. The op in overall system R^2 for stages one through three are of little nsequence. Judges which cluster together are indicated by parentheses. The drop from stage 3 to 4 is considerably larger than the .05 criterion and dicates a substantial loss of predictive efficiency. We therefore conclude at two policies were present in the committee. Judge 1 has Policy I while dges 2, 3 and 4 have Policy II.

To explain the two policies, all possible subsets regression was used. A ugh idea of the profile variables the judges were attending to while making eir ranking can be gained from Table 25.

TABLE 25
Correlations Between Judges and
Research Proposal Profile Variables

Judge	Research Proposal Variables				
	1	2	3	4	5
1	-.46	.27	-.11	-.60	-.46
2	.08	-.13	-.75	-.31	-.26
3	-.13	-.24	-.75	-.26	-.26
4	.04	-.17	-.72	-.33	-.29

To explain Policy I, the use of Table 26 is required. Table 26 indicates all possible combinations of profile variables ordered by their R^2 values for predicting the canonical variables of Judge 1.

TABLE 26

Results from All Possible Subsets

Regression for the Single Judge Cluster (Judge 1)

Profile Variables in Equation	RSQ
1, 2, 3, 4, 5	.919
1, 3, 4, 5	.909
1, 2, 4, 5	.874
1, 4, 5	.868
1, 2, 3, 4	.817
1, 3, 4	.810
1, 2, 4	.775
1, 4	.771
2, 3, 4, 5	.584
2, 4, 5	.577
3, 4, 5	.574
4, 5	.567
1, 2, 3, 5	.420
1, 3, 5	.411
2, 3, 4	.390
2, 4	.387
1, 2, 5	.372
3, 4	.366
4	.362
1, 5	.358
1, 2, 3	.293
1, 3	.278
2, 3, 5	.272
2, 5	.255
1, 2	.248
3, 5	.229
1	.228
5	.211
2, 3	.082
2	.072
3	.012

We again look for a jump in R^2 using the a priori .05 criterion. This jump occurs when going from the equation with variables 1, 4 and 5 to the equation with variables 1, 2, 3, and 4. Judge 1 was attending to variables 1, 4 and 5. We can also see that major emphasis was placed on variable 4. In other words, the Policy I judge was primarily considering need, quality of writing, and originality while ranking the proposals and essentially ignoring feasibility and cost benefit.

Policy II can be explained in a similar manner using Table Table 27 shows the all possible subsets regression for Judges 2, 3 and 4 combined as a single data set.

TABLE 27
Results from All Possible Subsets
Regression for the Three Judge Cluster (Judges 2, 3, 4)

Profile Variables in Equation	RSQ
1, 2, 3, 4, 5	.824
2, 3, 4, 5	.790
1, 2, 3, 4	.729
1, 2, 3, 5	.718
2, 3, 5	.709
1, 3, 4, 5	.708
3, 4, 5	.702
2, 3, 4	.667
1, 3, 5	.649
3, 5	.648
1, 3, 4	.624
1, 2, 3	.612
3, 4	.603
2, 3	.588
1, 3	.554
3	.547
1, 2, 4, 5	.240
2, 4, 5	.239
1, 4, 5	.167
4, 5	.162
1, 2, 4	.155
2, 4	.149
1, 2, 5	.129
2, 5	.120
1, 5	.095
1, 4	.090
4	.090
5	.073
1, 2	.034
2	.033
1	.007

In this case we see that a major jump in R^2 occurs when going from variables 2, 3, 4, and 5 to 1, 2, 3 and 4. It is obvious that variable 3 was of major importance. That is, the Policy II judges were attending to feasibility, cost benefit, quality of writing and originality with a primary emphasis on cost benefit while ranking the proposal profiles. Need was not viewed as important. It is interesting to note that neither of the policy groups attended to all the profile variables.

Although JAN and C-JAN are useful and innovative procedures, they do have some general problems. As with any statistical procedure, it would oftentimes be advisable to validate the results by use of split sample techniques or replication. Since the JAN procedure is based on regression, it suffers from the same problems encountered with regression. For example, JAN must have a sufficient ratio of profiles to profile variables to avoid overfit which results in inflated and unstable R^2s. Since JAN clusters on the basis of homogeneity of prediction equations, multicollinearity of the profile variables is also a serious problem. High multicollinearity will lead to questionable clustering results and make the interpretation of the captured policies quite difficult. However, if utilized properly, JAN and C-Jan are promising tools for evaluation methodologists to be used as additional techniques in decision-making and policy-capturing situations.

BIBLIOGRAPHY

ck, D. E.
1973 Development of the E-2 weighted airman promotion system. AFHRL-TR-73-3, AD-767 195. Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

tenberg, Robert A. and Raymond Christal
1968 Grouping criteria - a method which retains maximum predictive efficiency. The Journal of Experimental Education 36, 4: 28-34.

ing, N-K.
1970 Hierarchical groupings of judges according to selected criteria for financial aid awards. Unpublished doctoral dissertation, University of Northern Colorado.

istal, Raymond E.
1968a JAN: a technique for analyzing group judgment. The Journal of Experimental Education 36, 4: 24-27.

1968b Selecting a harem - and other applications of the policy-capturing model. The Journal of Experimental Education 36, 4: 35-41.

dycha, A. L.
1970 A monte carlo evaluation of JAN: a technique for capturing and clustering raters' policies. Organizational Behavior and Human Performance, 5: 501-506.

ff, W. L.
1969 A quantitative study of teacher selection and evaluation of policies at a suburban elementary school district. Unpublished doctoral dissertation, University of California at Los Angeles.

och, L. L.
1972 Policy capturing with local models: the application of the AID technique in modeling judgment. Unpublished Ph.D. Dissertation, The University of Texas, Auston.

tt, C. D.
1974 Development of the weighted airman screening system for the air reserve forces. AFHRL-TR-74-18, AD-781 747. Lackland AFB, TX: Computational Sciences Division, Air Force Human Resources Laboratory.

uston, Judith A., Houston, Samuel R. and R. LaMonte Ohlson
1974 On determining pornographic material. The Journal of Psychology, 86: 277-287.

Houston, Samuel R.
1967 The Judgment Analysis regression technique applied to the admission variables for doctoral students at Colorado State College, 1963-1966. Unpublished Ph.D. Dissertation, University of Northern Colorado, Greeley.

1968 Generating a projected criterion of graduate school success via normative Judgment Analysis. The Journal of Experimental Education 37, 2: 53-58.

1974a Classification of Judgment Analysis. In: Judgment Analysis: tool for decision makers, edited by Samuel R. Houston, pp. 52-53. New York: MSS Information Corp.

1974b Issues associated with the use of Judgment Analysis. In: Judgment Analysis: tool for decision makers, edited by Samuel R. Houston, pp. 69-73. New York, MSS Information Corp.

1974c Faculty policies of teaching effectiveness. In Judgment Analysis: tool for decision makers, edited by Samuel R. Houston, pp. 140-147. New York, MSS Information Corp.

-----, and James T. Bolding, Jr.
1974 The general linear model and Judgment Analysis. In: Judgment Analysis: tool for decision makers, edited by Samuel R. Houston, pp. 54-60. New York, MSS Information Corp.

-----, and Joseph W. Gilpin
1971 Hierarchical groupings of students according to their policy of rated teacher effectiveness. SPATE 10, 2: 28-53.

-----, and Gary C. Stock
1973 Judgment Analysis (JAN): tool for education decision-makers. SRIS Quarterly 6, 2: 22-24.

Holmes, George P. and Sheldon Zedick
1973 Judgment analysis for assessing paintings. The Journal of Experimental Education 41, 4: 26-30.

Johnson, J. W., and King, F. S.
1973 Multiple criterion judgment analysis for the educational researcher. Unpublished team doctoral dissertation, University of Northern Colorado.

Jones, K. M., Mennis, L. S., Martin L. R., Summers, J. L., and G. R. Wagner
1976 Judgment modeling for effective policy and decision making. Research Report for Air Force Office of Scientific Research Grant No. AFOSR-74-2656, AD-A033 186.

Keelan, J. A., Houston, S. R., and Houston, S. F.
1973 Leadership policies as perceived by firemen. Colorado Journal of Education Research, 12: 20-23.

plyay, J. B.
1970 Extension of the weighted airman promotion system to grades E-8 and E-9. AFHPL-TR-70-2, AD-703 687. Lackland AFB, TX: Personnel Research Division, Air Force Human Resources Laboratory.

plyay, J. B., Albert, W. G., and D. E. Black
1976 Development of a senior NCO promotion system. AFHRL-TR-76-46, AD-A030 607. Lackland AFB, TX: Computational Sciences Division, Air Force Human Resources Laboratory.

illins, C. J. and E. Usdin
1970 Estimation of validity in the absence of a criterion. AFHPL-TR-70-36, AD-716 809. Lackland AFB, TX: Personnel Division, Air Force Human Resources Laboratory.

aylor, J. C., and Wherry, R. L.
1965 The use of simulated stimuli and the JAN technique to capture and cluster the policies of raters. Educational and Psychological Measurement, 25: 969-966.

tock, G. C.
1969 Judgment analysis for the educational researcher. Unpublished doctoral dissertation, Colorado State College.

uits, D. B.
1957 Use of dummy variables in regression equations. Journal of the American Statistical Association 52: 548-551.

orgunrud, E. A.
1971 Criteria guiding curriculum decisions of selected school superintendents. Unpublished doctoral dissertation, University of California at Los Angeles.

ard, Joe H., Jr.
1961 Hierarchical grouping to maximize payoff. Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division.

1963 Hierarchical grouping to optimize an objective function. Journal of the American Statistical Association 58: 236-244.

-----, and K. Davis
1963 Teaching a digital computer to assist in making decisions. PRL-TDR-63-16, AD-407 322. Lackland AFB, TX: 6570th Personnel Research Laboratory, Aerospace Medical Division.

-----, and Marion E. Hook
1963 Applications of an hierarchical grouping procedure to a problem of grouping profiles. Education and Psychological Measurement 23: 69-81.

Williams, J. D., Gab, D., Linden, A.
1969 Judgment analysis for assessing doctoral admission policies. Journal of Experimental Education, 38: 92-96.