PRL-TDR-63-3

JAN: A Technique for Analyzing Group Judgment

By

Raymond E. Christal

Technical Documentary Report PRL-TDR-63-3

February 1963

6570TH PERSONNEL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
Lackland Air Force Base, Texas

Project 7734

NOTICE

ASTIA release to OTS is not authorized.

Qualified requestors may obtain copies from ASTIA. Orders will be expedited if placed through the librarian or other person designated to request documents from ASTIA.

6570th Personnel Research Laboratory (AMD), Lackland AF Base, Tex. Rpt No. PRL-TDR-63-3. JAN: A TECHNIQUE FOR ANALYZING GROUP JUDGMENT. Feb 63, iii + 6 p. Unclassified Report

This paper indicates how a technique which clusters criteria in terms of the homogeneity of their prediction equations can be used to identify and describe the rating policies within a group or board of judges. The technique measures the consistency of ratings obtained from individual judges, and indicates the amount and nature of agreement between judges or groups of judges. A practical method is suggested for helping a board or committee to reach a consensus concerning how relevant factors shall be weighted in future situations so as to carry out its final joint policy. Examples are given to show how the technique can be used to determine the composition of a service school selection composite, to determine factor weights for a job evaluation plan, and to determine how relevant factors should be combined into a job-oriented criterion composite.

1 Psychometrics
2 Decision making
3 Group dynamics
4 Personnel
5 Selection
6 Statistical analysis
7 Data
8 Mathematical prediction
I AFSC Project 7734
II R.E. Christal
III In ASTIA collection

6570th Personnel Research Laboratory (AMD), Lackland AF Base, Tex. Rpt No. PRL-TDR-63-3. JAN: A TECHNIQUE FOR ANALYZING GROUP JUDGMENT. Feb 63, iii + 6 p. Unclassified Report

This paper indicates how a technique which clusters criteria in terms of the homogeneity of their prediction equations can be used to identify and describe the rating policies within a group or board of judges. The technique measures the consistency of ratings obtained from individual judges, and indicates the amount and nature of agreement between judges or groups of judges. A practical method is suggested for helping a board or committee to reach a consensus concerning how relevant factors shall be weighted in future situations so as to carry out its final joint policy. Examples are given to show how the technique can be used to determine the composition of a service school selection composite, to determine factor weights for a job evaluation plan, and to determine how relevant factors should be combined into a job-oriented criterion composite.

1 Psychometrics
2 Decision making
3 Group dynamics
4 Personnel
5 Selection
6 Statistical analysis
7 Data
8 Mathematical prediction
I AFSC Project 7734
II R.E. Christal
III In ASTIA collection

(5) N/A

PRL-TDR-63-3

(8) 7L (9) NA

(15) NA

(16) PRL 7734

(17) NA

(18) PRL

(19) TDR 63-3

(20) 7L

(6) JAN: A TECHNIQUE FOR ANALYZING GROUP JUDGMENT

(10) By
Raymond E. Christal

Technical Documentary Report PRL-TDR-63-3
(11) February 1963,
(12) 6p

(7) 450

6570TH PERSONNEL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
Lackland Air Force Base, Texas

Project 7734

ABSTRACT

This paper indicates how a technique which clusters criteria in terms of the homogeneity of their prediction equations can be used to identify and describe the rating policies within a group or board of judges. The technique measures the consistency of ratings obtained from individual judges, and indicates the amount and nature of agreement between judges or groups of judges. A practical method is suggested for helping a board or committee to reach a consensus concerning how relevant factors shall be weighted in future situations so as to carry out its final joint policy. Examples are given to show how the technique can be used to determine the composition of a service school selection composite, to determine factor weights for a job evaluation plan, and to determine how relevant factors should be combined into a job-oriented criterion composite.

This report has been reviewed and is approved.

Fred E. Holdrege, Col USAF
Commander

A. Carp
Technical Director

Hq 6570th Personnel Research Laboratory

JAN: A TECHNIQUE FOR ANALYZING GROUP JUDGMENT[1]

JAN (Judgment ANalysis) is a simple but powerful technique for identifying and describing the rating policies that exist within a board or committee of judges. It can be used to help the members of such boards or committees reach a consensus and will express their final joint policy in a precise manner.

JAN is a special application of a technique developed by Bottenberg & Christal (1961) which groups criteria in terms of the homogeneity of their prediction equations.[2] The criterion-grouping technique is a special application of a still more general hierarchical grouping model developed by Ward (1961). The reader is referred to these papers for computational expressions.

Communication of the JAN concept can be simplified by describing how it might be applied in a hypothetical situation. Suppose one were faced with the problem of selecting the first class of 300 students for a newly established service school. Scores are available on 15 selection variables for each of 3000 applicants. A special board of 10 officers has been assembled and charged with the responsibility for determining how these variables shall be weighted together to compute the final selection composite. Since this is a new school, criterion data are unavailable. However, each board member is familiar with the curriculum to be offered and with the general characteristics of the selection variables.

In order to set the stage for a discussion of JAN, a portion of the proceedings will be fabricated. The letters MA, MB, and so forth, will be used to refer to the individual board members.

The board has been called to order. The chairman has recognized MC.

MC: MB and I would like to propose a set of weights for the board's consideration. If these weights are used as a beginning point for our discussion, perhaps we can proceed more rapidly.

Copies of the weights are passed out to all board members. Members study the proposed weights.

MA: What concerns me is the low weight you have assigned the High School Activities Index. After all, it's the only variable in the selection battery that measures leadership.

MC: But we gave it a weight of three, and that's pretty high. It's almost as much weight as we give to all of the verbal tests combined.

MA: Yes. But the highest Activities Index score is only 300, while the *lowest* scores on the verbal tests are usually higher than 300. I think we should weight the Activities Index by at least 20.

COMMENT: Notice that this is a very powerful argument, even though it is based on false premises. The board at this point does not recognize that the effect of an applied weight is influenced by the standard deviation of the scores to be weighted and is not necessarily related to the mean or size of the scores to be weighted.

The board compromises on a weight of 15 points for the Activities Index. The conference continues. Various members argue for increasing or decreasing the weights for certain variables. Compromises are made until, finally, a tentative set of weights is agreed upon. As a check, these weights are applied to the scores for a sample of the applicants. Certain peculiarities are observed, which are taken up during the next day's meeting.

---

[1] Reproduced with some minor editorial changes from a paper presented at the annual meeting of the Southwestern Psychological Association, Fort Worth, Texas, 6 April 1962.

[2] Dr. Joe H. Ward, Jr. is credited with suggesting use of a least-squares weighted regression formula to capture the policy of a single rater (also see Hoffman, 1960). The present paper indicates how the Bottenberg-Christal criterion-grouping technique can be applied to a series of such equations in order to analyze the rating behavior of a group.

ME: There's something wrong with our weighting system. You will observe that several of the applicants with the highest composite scores made less than 400 on mathematical aptitude. I don't think we should let in anyone who scores below 450 on the Mathematics Aptitude Test. They will never pass our engineering curriculum.

MF: I think the problem is that we are weighting the Activities Index too much. You will notice that all of those cases with low quantitative scores have high Activities Indexes.

MA: Well, I don't think we should lower the weight for the Activities Index. It's the only measure of leadership we have. Let's establish cutoffs on the verbal and quantitative tests; then we will refuse to accept any candidate who scores below our established cutoffs.

The board compromises by lowering the weight on the Activities Index to 12, and by establishing cutoffs of 450 on the verbal and quantitative tests. New composites are computed for the sample cases, and the board is reconvened.

ME: I don't like this multiple-cutoff technique. I found several applicants who were high on nearly all variables, but who were disqualified because of a single verbal or mathematical score of 440 or 445. We should make an exception and let these cases in. After all, how much difference is there between a score of 440 and one of 450?

MF: I still think we are weighting the Activities Index too much. For some reason it seems that an applicant with high scores on all variables except the Activities Index comes out with a low composite score.

After a prolonged discussion, the chairman decides that the weight for the Activities Index will be lowered to 8, and that the cutoff scores will be lowered to 400.

This conversation has been carried far enough to make clear certain points. First, the board was unable to predict accurately the effect of their proposed weights. Very early in the conference, a tentative set of weights was agreed upon. However, no one was satisfied with the results when these weights were applied to the applicant group. Recommendations from this point on were based primarily upon the individual board members' satisfaction or dissatisfaction with the rank position of applicants. If they felt that particular applicants were ranked too high or too low, recommendations were made for changing weights in a manner predicted to rectify the problem. Often as not, the proposed changes either failed to resolve the problem, or created new problems.

When one carefully analyzes the board's actions, he comes to realize that the board should not have been asked to propose weights in the first place. Its main concern was that, whatever the weights, the final selection composite would rank applicants in a manner which is acceptable to the board members. Rather than argue over the relative weights to be applied to the selection variables, the board should have restricted its attention to the relative acceptability of applicants with given profiles of scores on the selection tests. The impact of applied weights on the final rank-ordering of applicants is a highly complex matter which is influenced by several things, including the standard deviations and intercorrelations among the selection variables. A sophisticated measurement psychologist cannot solve for a 15-predictor regression equation in his head. Yet this is, in effect, what was required of the board members when they were asked to determine the weights to be applied to the selection tests.

Nor should the board members have attempted to establish cutoff scores. It may be true that applicants below given score levels on certain variables tend to become unacceptable, regardless of their scores on other variables. However, there is no clear basis for deciding *where* these cutoff points should be established. As ME put it, "After all, how much difference is there between a score of 440 and one of 450?" What the board really wants is for these variables to be weighted so that applicants obtaining scores below certain points will have a rapidly decreasing probability of receiving a total composite score high enough to be accepted into the school. The desired result could be obtained by introducing appropriately weighted squared test scores into the selection composite, but it is highly improbable that the board would come up with such a suggestion. Again, one arrives at the conclusion that the board should reach agreement on the appropriate rank position of applicants, and leave the selection test weights to be determined by analytical methods.

Now that the proper groundwork has been laid, the time has come to describe how the JAN technique can help such a board reach closure. The first step is to prepare a card for each of approximately 150 men selected at random from the total applicant group.[3] The card for each man lists his 15 selection test scores and his identification number—nothing more. Irrelevant information which might bias the rater, such as the applicant's name, religion, or home town, is omitted.

Next, every member of the board is given a set of these cards. Board members are asked to study the test scores for each of the 150 applicants, and then rank them from 1 to 150. That is, a board member places the card on top which, in his judgment, describes the applicant most deserving to be accepted—and so on. The card he places on the bottom is assigned a rank of 150, and describes the applicant who, in his judgment, is least qualified.[4]

Now let's consider this information for one of the raters. After carefully studying the selection score information, MF has rank-ordered the sample of 150 applicants. *MF's rating policy can be captured by using the selection test variables as predictors and his ranking of the 150 applicants as a criterion variable.* This simply requires computation of a 15-predictor regression equation. The magnitude of the obtained squared multiple correlation coefficient (R^2) indicates how well MF's policy can be replicated (in this sample) using a least-squares-weighted composite.

Observe that MF had to render his judgments solely on the basis of selection test score information. Nothing else was provided to him. Under these conditions, he could not have been biased by information in the applicants' folders which the board has judged to be irrelevant. Perhaps of more importance, however, is the fact that one should be able to do a good job of accounting for MF's evaluations using the information which he studied. In fact, there are only two reasons why one should not be able to predict MF's judgments perfectly using these data.

First, it may be that MF considered certain interactions among the predictors, or that he perceived the relationship between certain of the factors and applicant acceptability as being nonlinear. Under these conditions, it would not be possible to account for MF's assigned evaluation scores unless the appropriate product or power terms were introduced into the equation. However, it is a simple matter to introduce any such terms which are hypothesized. For example, if it were hypothesized that MF considers applicants who scored below certain values on the verbal and quantitative tests as being unacceptable, regardless of how they scored on other tests, then squared terms are introduced into the equation. If introduction of squared terms causes a significant increase in predictive efficiency, then the formula for expressing MF's policy has been improved. Notice that this is a much better approach than simply arbitrarily establishing cutoff scores.[5]

A second and perhaps more likely explanation for an inability to account for MF's judgments is that he was inconsistent in his rating behavior. Such intrarater instability will lead to a reduced value of the obtained R^2. To the extent that MF is consistent in his rating behavior, it should be possible to produce a formula which will capture his policy.

But MF is only one of the 10 members serving on our board. The goal is to obtain a board policy on selection—not one man's opinion. In order to guide the board to closure, the first step is to compute 10 separate regression equations. Each of these 10 equations expresses the policy of one of the 10 judges—and each equation could be used to rank the entire 5000 applicants in a manner which would please at least one of the 10 board members.

---

[3] The number of cases to be selected depends on the number of selection variables being used. A goal is to avoid "overfitting" when these variables are weighted into a least squares regression equation.

[4] It would be permissible to have these cases rated instead of ranked.

[5] The combination of a properly-weighted squared and first order term has the ability to fit a regression line to any set of points which displays a single bend or change in direction.

Beginning with these equations, the criterion-grouping technique is employed to define the areas of agreement and disagreement among the board members. As the first step, a *single value* of R^2 is computed which indicates the overall predictive efficiency obtained when a separate least-squares-weighted regression equation is used for each rater. During the second step, each equation is compared with every other equation. In this way, the two raters are located who are in closest agreement concerning how the selection variables should be weighted to determine the relative acceptability of applicants. These two raters have the most homogeneous regression equations. The computer then determines the single equation that best represents the joint policy of these two raters. It also indicates the overall loss in predictive efficiency that results when the original N equations are replaced by the best set of N-1 equations. This is accomplished by comparing the magnitude of the overall R^2 in the two instances.[6]

In subsequent operations, the program systematically reduces the number of prediction equations (or rater clusters) by one at each step so as to minimize the loss in overall predictive efficiency. In each instance, the computer prints out an overall R^2, as well as the single equation for each cluster which best represents the joint policy of all raters in that cluster. At the last step, a single prediction equation is generated so as to express the joint policy of all raters with as little error as possible.

Examination of the loss in overall predictive efficiency at various stages of the grouping process will help to identify the number of rating policies existing among the board members. For example, if very little overall predictive efficiency is lost in going from ten to two equations, but a considerable amount of efficiency is lost in going from two to one equation, there is strong evidence that two different rating policies are operating. That is to say, some of the board members are using one weighting system and the rest a different weighting system. Of course the hope is that a single equation can be found which predicts the criterion values assigned by all raters. In this case the board could be dismissed without further ado, since their single agreed-upon policy would have been determined. However, should more than one rating policy be found to exist, the board is *not* asked to arbitrate the differences in weights. Rather, the several equations are used to rank the applicant sample, and the board is asked to arbitrate the differences in the rank position of cases resulting from application of these several equations. Once a single agreed-upon rank-ordering of the sample has been obtained, it is a simple matter to determine the appropriate set of weights for the final selection composite.

The JAN technique enables the investigator to make a complete analysis of interrater agreement. As a matter of fact, results of the first step alone indicate the amount of agreement between each rater and every other rater.

We have seen that a situation may exist where the level of interrater agreement among the entire sample of raters is low, but where the raters can be divided into two or more groups within each of which there is high agreement. Conventional analysis techniques for determining interrater agreement would not detect this situation. The JAN approach will not only identify the groups, but will at the same time provide equations expressing the different rating policies.

So far, only one example has been given illustrating how JAN might be used. But the Air Force is constantly calling boards together to determine how variables should be weighted for making decisions: Who should be promoted? What specialties should receive proficiency pay? Who should be given a regular commission? What are the grade requirements for jobs? Which officers should be

[6] What the program actually does during this first step is to evaluate each of the $\frac{N(N-1)}{2}$ alternate ways of reducing the number of equations from N to N-1. Any one of these ways involves predicting the criterion observations for a particular pair of board members with a single equation, while retaining separate least squares-weighted equations for each of the remaining N-2 board members. The end result is identical to that which would be obtained by the procedure outlined in the text, under the conditions set forth in this paper.

given responsibility pay? The solution to all such problems involves either the implicit or explicit weighting together of factors judged to be relevant for servicing the agreed-upon goals. If these factors are made explicit, then JAN can be used to help such boards reach a consensus. Moreover, the JAN approach results in a precise statement of the factors and weights to be used in carrying out the board's recommendations.

Very often a group's policy can be captured by having it rate a series of simulated situations. For example, The Ohio State University Research Foundation is conducting a job-oriented criterion-development contract in which scores on hypothesized criterion-relevant factors are simply ascribed to a sample of several hundred pseudo-incumbents presumably working in a particular Air Force specialty.[7] Supervisors in this specialty will be asked to study the resulting score profiles and make a global evaluation of the worth or effectiveness of each simulated incumbent. Then the JAN technique will be applied to determine which factors should be weighted into an acceptable criterion composite for use in the specialty. This approach to criterion development has the advantage of not requiring expensive field testing. It also makes possible the evaluation of factors which at the present time cannot be measured easily in the field. For purposes of the study, it is only necessary that each factor be definable in terms that present unambiguous concepts to the supervisors who must make the ratings.

Job evaluation is another area to which the JAN technique has been applied (Madden, 1963). Work requirement factors were developed as likely candidates for inclusion in a revised Air Force officer job-evaluation system. The factors cover such requirements as formal education, special training and work experience, management responsibilities, and working conditions. A sample of Air Force jobs was rated on these factors, and a score profile for each job was generated by computing its mean rating on each factor. Senior officers at the Air Force Command and Staff College were asked to study these profiles and rate each of the corresponding jobs on the basis of deserved compensation. In subsequent analyses, the JAN technique was able in most instances to capture between 80 and [illegible] percent of the variance in the ratings made by these individual officers. Furthermore, a great deal of similarity was found in the raters' policies concerning which factors should determine the pay level for Air Force jobs.

The applications of JAN suggested in this paper have related to Air Force problems, but it should be apparent that the technique is generally applicable to any situation where a committee or board wishes to express its policy concerning how relevant factors shall be combined for making subsequent decisions.

The technique also can be used to analyze and describe the similarities and differences among ratings collected from individuals who are separated by time or space, and who in no way can be considered as being members of a board or committee. Perhaps it could be used to analyze customers' preferences for combinations of commodity characteristics or to analyze voters' preferences for certain characteristics in real or simulated political candidates.

Finally, it should be recognized that application of the JAN technique to large problems is not feasible without use of an electronic data processing machine. However, most investigators now have access to computers, and studies of the type discussed are easily accomplished. If a board of 10 judges each made 150 evaluations based on 15 factors, it is estimated that a complete analysis of the data using the JAN technique would take less than five minutes using a modern computer comparable to the IBM 7090.[8]

---

[7] As one investigator stated in jest: "We're tired of being unable to predict the behavior of real-live people, so we decided to create some of our own."

[8] Qualified requesters can obtain a set of Fortran computing statements for the criterion-grouping technique by writing to the 6570th Personnel Research Laboratory. This Fortran program was written by Mr. Daniel Rigney under the general direction of Dr. Joe H. Ward, Jr.

REFERENCES

Bottenberg, R.A. & Christal, R.E. *An iterative technique for clustering criteria which retains optimum predictive efficiency.* Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, March 1961. (WADD-TN-61-30, ASTIA Document AD-261 615)

Hoffman, P.J. The paramorphic representation of clinical judgments. *Psychol. Bull.*, 1960, 57, 116-131.

Madden, J.M. *An application to job evaluation of a policy-capturing model for analyzing individual and group judgment.* Lackland Air Force Base, Texas: 6570th Personnel Research Laboratory, Aerospace Medical Division, 1963, in press.

Ward, J.H., Jr. *Hierarchical grouping to maximize payoff.* Lackland Air Force Base, Texas: Personnel Laboratory, Wright Air Development Division, March 1961. (WADD-TN-61-29, ASTIA Document AD-261 750)

6570th Personnel Research Laboratory (AMD), Lackland AF Base, Tex. Rpt No. PRL-TDR-63-3. JAN: A TECHNIQUE FOR ANALYZING GROUP JUDGMENT. Feb 63, iii + 6 p. Unclassified Report

This paper indicates how a technique which clusters criteria in terms of the homogeneity of their prediction equations can be used to identify and describe the rating policies within a group or board of judges. The technique measures the consistency of ratings obtained from individual judges, and indicates the amount and nature of agreement between judges or groups of judges. A practical method is suggested for helping a board or committee to reach a consensus concerning how relevant factors shall be weighted in future situations so as to carry out its final joint policy. Examples are given to show how the technique can be used to determine the composition of a service school selection composite, to determine factor weights for a job evaluation plan, and to determine how relevant factors should be combined into a job-oriented criterion composite.

1 Psychometrics
2 Decision making
3 Group dynamics
4 Personnel
5 Selection
6 Statistical analysis
7 Data
8 Mathematical prediction
I AFSC Project 7734
II R.E. Christal
III In ASTIA collection

6570th Personnel Research Laboratory (AMD), Lackland AF Base, Tex. Rpt No. PRL-TDR-63-3. JAN: A TECHNIQUE FOR ANALYZING GROUP JUDGMENT. Feb 63, iii + 6 p. Unclassified Report

This paper indicates how a technique which clusters criteria in terms of the homogeneity of their prediction equations can be used to identify and describe the rating policies within a group or board of judges. The technique measures the consistency of ratings obtained from individual judges, and indicates the amount and nature of agreement between judges or groups of judges. A practical method is suggested for helping a board or committee to reach a consensus concerning how relevant factors shall be weighted in future situations so as to carry out its final joint policy. Examples are given to show how the technique can be used to determine the composition of a service school selection composite, to determine factor weights for a job evaluation plan, and to determine how relevant factors should be combined into a job-oriented criterion composite.

1 Psychometrics
2 Decision making
3 Group dynamics
4 Personnel
5 Selection
6 Statistical analysis
7 Data
8 Mathematical prediction
I AFSC Project 7734
II R.E. Christal
III In ASTIA collection