

---

# JUDGMENT ANALYSIS: A METHODOLOGY FOR MERIT-BASED SALARY ALLOCATION

---

*Vern C. Vincent*

*DeWayne Hodges*

University of Texas–Pan American  
Edinburg ,Texas

Contrary to many predictions, the nationwide trend towards merit-based salary systems in higher education has not abated. In fact, the increase emphasis on accountability in higher education by the public has motivated a number of state legislative bodies to mandate salary increases solely on the basis of merit ([1], [3], [4]). Administrators and legislators claim that faculty union stances which traditionally oppose faculty merit review and performance appraisal systems have increased public pressure for accountability and merit-based salary systems [4]. Especially when funds are inadequate, some administrators and policy makers favor merit-based salary systems because they provide for rewarding the most productive faculty or, at least, provide a rational for administrative actions.

Critics of merit-based salary systems, however, cite a number of practical difficulties ([1], [4], [7]). Faculty point out that without cost-of-living increases during inflation cycles, merit-based salary systems result in a majority of faculty suffering real pay cuts. They further emphasize divergence in public and professional assessments of the relative importance of teaching, research, and service. Most important, perhaps, is the identification of valid criteria. The identification of variables which appropriately measure each of these functional service areas as well as the weight each of these variable should receive is always an issue [11].

Further complications arise since most institutions characterize themselves as teaching institutions and yet are unable to agree on how they wish to measure the teaching function. For example, many faculty claim that grades are exchanged for positive student evaluations. When other measures of teaching effectiveness are used, they generally are mistrusted or lack complete information [10].

Most of the articles in the literature do agree, however, that whatever methods of evaluation are used, the potential for success is greater when faculty input to the evaluation process is high ([1], [9], [11]). An evaluation system which incorporates a high degree of faculty input and can ultimately produce a consensus opinion as to how merit allocation can be obtained is based upon a statistical technique called Judgment Analysis.<sup>1</sup>

---

*Journal of Business Strategies*, Volume 7, Number 1 (Spring 1990)

<sup>1</sup>Editorial footnote: While the remainder of this article focuses on Judgment Analysis in a university environment, the principle outlines of the technique, its applications, and its uses are generalizable to a private sector, for-profit enterprise desiring to utilize a quantifiable method of merit-based salary adjustment. Clearly, this is a topic of special interest for compensation analysis.

### Judgment Analysis Technique

Judgment Analysis (JAN) is a statistical technique for combining multiple regression analysis and hierarchical grouping or clustering analysis in order to classify criteria in terms of the homogeneity of their prediction equations. Originally introduced by Ward [12] and later modified by Bottenberg and Christal [2], the technique had developed widespread popularity not only in business but also in education and psychology ([5], [6], [13]). Technical and computational aspects of JAN are available through cited references ([2], [13]).

The JAN procedure initially requires each of the decision-makers to rank stimuli on a set of predictor variables. The value assigned to each profile then serves as the criterion variable. In the first stage of JAN, a coefficient of determination ( $R^2$ ) is calculated for each individual judge's policy. Also, the  $R^2$  measures the judge's reliability in evaluating the profiles and serves as input for a comprehensive measure for all judges.

The second stage of JAN involves a clustering of the two judges having the most homogenous prediction equations. This procedure reduces the number of policies by one and generates a new  $R^2$  value. This new (unadjusted)  $R^2$  is, of course, lower than the original ones.

The grouping procedure continues. By examining one policy at a time and combining sequentially the most homogeneous policies, eventually all judgments will have been consolidated into one single policy with an overall  $R^2$  value. As will later be shown in Table 3, when significant drops in  $R^2$  occur between systems, then inter-rater agreement is confirmed and group policies are defined. The final result of JAN is that individual judgment policies as well as group policies can be examined, allowing for a better understanding of grouping dynamics in decision-making. A numerical example will help clarify these procedures.

### JAN Illustration

To apply JAN to faculty merit allocation, the first step is to establish faculty profiles containing the evaluation information on the service areas. For illustrative purposes, three faculty service areas to the university are defined as teaching, professionalism in the field, and service.

The teaching function comprises two components – student evaluations and an overall measure of teaching quality elements. How are these two measured? Student evaluations derive from a student opinion survey while teaching quality (an indication of a faculty member's teaching ability) derives from a faculty peer group.

Teaching quality is an indication of any activity on the part of the instructor to manage the teaching function in such a manner that promotes the learning and achievement of students. Typical sample activities would include curriculum innovation and development, special tutoring, attendance at conferences or workshops devoted to improving teaching quality, or documentation of output measures of student success before or after graduation.

Professionalism in the field comprises faculty development activities and scholarly research. Development activities are measured by a faculty member's accomplishments in remaining current in the discipline. Sample activities would include additional graduate course work and attendance at institutes, short-courses, and professional meetings. Scholarly research is measured by professional publications, such as books, refereed articles, program presentations, and non-refereed professional publications.

Service comprises contributions to the university and to the community. University service refers to documented activities at the department, school, or university level. Typical activities include committee assignments, student advising, sponsorships, and other special assigned tasks. Community service refers to work outside the university which effectively represents or promotes the university in the community-at-large. Community service activities include public workshops, speeches, consultantships and public-service assignments.

Six different measures, two per function area, were used to form the basis for policy determination in awarding merit. Each variable was measured based upon a five point scale ranging from unsatisfactory (0), satisfactory (1), above average (2), outstanding (3), to exceptional (4).

Fifty-nine profiles of university faculty members based on the above describes six variables and their measurement scale were generated by the use of a random number table. In addition to maximizing both mathematical and administrative soundness, this procedure also produces minimum correlation and maximum variability among the six profile variables. The net result of the procedure is to reduce multicollinearity among the profile variables and, consequently, to simplify policy determination. The 59 developed profiles are presented in Table 1.

A total of ten business school faculty and administrators were contacted to serve as judges and were requested to assign an overall rating on a five point scale from unsatisfactory (0) to exceptional (4) on each of the 59 hypothetical faculty profiles presented in Table 1. By using these profiles, the judges were required to determine merit ratings based solely upon academic credentials rather than allowing for the possibility of personalities to influence the ratings.

In the next step of the procedure, the JAN technique was applied to the judges' ratings in order to determine the merit policy present among the various administrator and faculty judges. JAN validate neither the variables nor the methods for obtaining the data information used in determining variable ratings. The variables and their measurement process are assumed to be reliable and valid or at the very least their identification and measurement have been agreed upon by all faculty and administrators involved in the merit process.

### Judges' Reliability and Merit Policies Results

Table 2 presents the  $R^2$  values for the ten faculty and administrators who serve as judges. All the  $R^2$  values are significantly different from zero at the .01 level with the exception of the  $R^2$  value associated with faculty judge number 5. The low  $R^2$

**Table 1**  
**Instrument**

Please review the six levels of performance for each hypothetical faculty member. Circle the overall merit that you believe is appropriate. The overall merit rating is to be measured as follows: (E) Exceptional, (O) Outstanding, (A) Above Average, (S) Satisfactory, and (U) Unsatisfactory.

Faculty No.	Teaching		Professionalism		Service		Merit
	Student Evaluation	Teaching Quality	Development	Research	University	Community	
1	E	E	U	A	A	S	E O A S U
2	O	A	O	O	U	A	E O A S U
3	S	O	A	A	A	S	E O A S U
4	O	U	A	U	S	U	E O A S U
⋮							
59	E	U	E	S	E	S	E O A S U

value for judge 5 indicates inconsistent ratings, suggesting that this judge either has no policy for awarding merit or perhaps is opposed to the concept of merit based salary allocation. Some investigators eliminate judges whose  $R^2$  values are less than .50 [14]; however, before making this decision an interview with that judge would be appropriate to assess what policy is being addressed.

**Table 2**  
**Judge Consistency**

Judge	$R^2$	Judge	$R^2$
1	0.7617*	6	0.8424*
2	0.8089*	7	0.8697*
3	0.9934*	8	0.9066*
4	0.9478*	9	0.8248*
5	0.1387	10	0.6186*

\*  $p < .01$

The ten stages of the JAN clustering procedure for the ten judges with their corresponding drops in  $R^2$  for each stage are presented in Table 3. In stage two, judges 6 and 10 are identified as having the most similar policies in determining merit awards. The drop in  $R^2$  between stages one and two, however, is only .0044, which is not statistically significant at .01 level. The JAN procedure continues combining the similar policies until a significant drop in  $R^2$  occurs between stages.

Table 3  
JAN Stages with Drops in  $R^2$

Stage	Judges	$R^2$	Successive Drops in $R^2$	Collective Drops in $R^2$
1	1,2,3,4,5,6,7,8,9,10	.7481	—	—
2	(6,10),1,2,3,4,5,7,8,9	.7437	.0044	.0044
3	(1,6,10),2,3,4,5,7,8,9	.7359	.0078	.0122
4	(1,6,10),(4,8),2,3,5,7,9	.7278	.0080	.0203
5	(1,6,10,7),(4,8),2,3,5,9	.7187	.0092	.0294
6	(1,6,10,7,9),(4,8),2,3,5	.7001	.0186	.0480
7	(1,6,10,7,9,5),(4,8),2,3	.6675	.0325**	.0806 <sup>a</sup>
8	(1,6,10,7,9,5),(2,4,8),3	.6247	.0428	.1234
9	(1,6,10,7,9,5,3),(2,4,8)	.5317	.0930	.2163
10	(1,6,10,7,9,5,3,2),(4,8)	.4304	.1013	.3177

<sup>a</sup> First collective drop which satisfies the *a priori* criterion of a .05 drop in  $R^2$ .

\*\* First statistically significant drop in  $R^2$  at .01 level.

NOTE: Parentheses indicates judges grouped by the JAN procedure.

Although the most common procedure for determining how many policies are present in a group of judges is based on tests of statistical significance, researchers familiar with JAN procedures caution that  $R^2$  values calculated in JAN procedures, as with many other statistical tests, are a function of the number of cases to be judged and can often artificially generate statistical significance while not being practically significant [13]. Ward and Hook [13] recommend looking for a break in the objectives function as measured by a drop in  $R^2$  values between stages. The general rule recommended is an  $R^2$  drop of more than .05 from the initial stage  $R^2$  value [6]. Applying this rule to the present analysis suggests that a reasonable break occurs between stages six and seven which results in a .0806 drop in  $R^2$  from stage one.

Another reason that the break in policies should occur between steps six and seven has to do with the consistency of the judges' ratings. Recall that judge 5 was identified as having inconsistent ratings. At stage six and all previous stages, judge 5 was identified as a single-member system. It would be inappropriate to include judge 5 with a group of judges in stage seven who have established consistent policies.

Stage six indicates that five decision making policies are present among the ten faculty judges. Judges 1, 6, 10, 7, and 9 represent policy one, while judges 4 and 8 represent policy two. Judges 2, 3, and 5 are single-member systems and represent three additional policies. A clearer representation of the five policies captured by the JAN process is obtained by examining the validity coefficients, presented in Table 4. A particular judge's policy can be determined by examining the six validity coefficients

for each judge in question. Each validity coefficient is determined by correlating the overall rating of the 59 hypothetical faculty with each independent variable. Judge 2, for example, attends to only one variable – student evaluations of teaching.

**Table 4**  
**Validity Coefficients**

Judge	Student Evaluation	Teaching Quality	Development	Research	University Service	Community Service
1	0.33**	0.40**	0.07	0.54**	0.28	0.38**
2	0.89**	0.10	-0.27	0.06	0.07	-0.07
3	0.05	0.09	-0.07	0.99**	-0.02	0.25
4	0.79**	0.15	-0.28	0.60	0.15	0.18
5	-0.02	-0.12	0.01	0.26	-0.06	0.26
6	0.38**	0.46**	0.08	0.58**	0.34**	0.23
7	0.58**	0.43**	-0.29	0.57**	0.21	0.46**
8	0.66**	0.24	-0.19	0.70**	0.10	0.14
9	0.43**	0.45**	-0.04	0.56**	0.33**	0.34**
10	0.54**	0.46**	-0.15	0.40**	0.23	0.16

\*\*  $p < .01$

Of the five policies presented, policy one (judges 1, 6, 10, 7, and 9) primarily addresses both variables in the teaching function, scholarly research and to a minor extent the service variables. Policy two, judges 4 and 8, asserts that merit should be awarded on only two variables – student evaluations of teaching and scholarly research. Judge 2 attends to only one variable, student evaluations of teaching while judge 3 only considers scholarly research. Because of the lack of consistency in ratings performed by judge 5, there is some question as to what policy is exhibited or even if this judge should be included in any further analysis.

### Merit Allocation Model

A logical extension to the JAN procedures outlined above would be to include all of a school's faculty in the evaluation of the hypothetical profiles and use the last stage of the JAN procedure as the model for merit-based salary allocation. The actual faculty profiles could be substituted into the regression equation generated in the last stage of the JAN procedure. Faculty merit awards could then be based upon the model. *Objectivity* in the synthesis of ratings would be ensured because faculty names never are presented to a peer committee or university administrators during the evaluation process. Also every faculty member would have input to the model and, consequently, have a determination as to how merit would be awarded. Consideration could also be given to elimination of faculty or administrators from the model who failed to demonstrate consistency on their evaluations.

In conclusion, JAN is a powerful statistical methodology which provides business school decision-makers an objective technique for identifying not only what policies exist in awarding merit-based salary allocations but also serves as a mechanism for setting or determining new salary allocation policies.

### References

1. Baker, M. A. "Merit-based Salary Systems: A Continuing Problem," *The Texas Association of College Teachers Bulletin*, Vol. 41, No. 1 (August, 1988), pp. 12, 24-25.
2. Bottenberg, R. A. and R. E. Christal. "An Iterative Technique for Clustering Criteria Which Retains Optimal Predictive Efficiency," *Journal of Experimental Education*, Vol. 36 (1968), pp. 28-34.
3. Brown, W. S. "Performance Review Instruments and Merit Pay Programs in an Academic Environment," *Journal of College and University Personnel Association*, Vol. 35 (1984), pp. 7-13.
4. Ehil, G. J. "Faculty Attitudes Toward Merit Pay in South Dakota's Public Colleges and Universities," *Journal of College and University Personnel Association*, Vol. 37 (1986), pp. 6-12.
5. Glorfield, L. and G. Fowler. "A Multivariate Methodology for Simultaneous Capturing and Clustering Judgment Policies," *Decision Sciences Journal*, Vol. 19, No. 3 (Summer 1988), pp. 504-20.
6. Houston, S. R. *Judgment Analysis: Tool for Decision Makers*. MSS Information Corporation (1974), pp. 69-73.
7. Kasten, K. L. "Tenure and Merit Pay as Rewards for Researching, Teaching, and Service at a Research University," *Journal of Higher Education*, Vol. 55 (1984), pp. 500-14.
8. Madden, J. M. "An Application to Job Evaluation for a Policy Capturing Model for Analyzing Individual and Group Judgment," *Journal of Industrial Psychology*, Vol. 2 (1964), pp. 36-42.
9. Magnusen, K. O. "Faculty Evaluation, Performance, and Pay: Application and Issues," *Journal of Higher Education*, Vol. 58 (1987), pp. 516-29.
10. Medley, D. M., H. Coker, and R. S. Soar. *Measurement-Based Evaluation of Teacher Performance*. Langman, NY: and Landau (1984).
11. Spaight, E. and E. Bridges. "Peer Evaluations of Salary Increases and Promotions Among College and University Faculty Members," *North Central Association Quarterly*, Vol. 60 (1986), pp. 403-10.

12. Ward, J. H. "Hierarchical Grouping to optimize an Objective Function," *Journal of the American Statistical Association*, Vol. 58 (1963), pp. 236-44.
13. Ward, J. H. and M. A. Hook. "Application of an Hierarchical Grouping Procedure to a Problem of Graying Profiles," *Education and Psychological Measurement*, Vol. 23 (1963), pp. 69-81.
14. Williams, J. D., D. Gab, and A. Lindem. "Judgment Analysis for Assessing Doctoral Admission Policies," *The Journal of Experimental Education*, Vol. 38, No. 2 (1969), pp. 92-96.