Swanson, R. A., & Sisson, D. J. (1971). The development, evaluation and utilization of a university faculty appraisal system. Journal of Industrial Teacher Education, 9 (1), 64-79.

THE DEVELOPMENT, EVALUATION, AND UTILIZATION OF A DEPARTMENTAL FACULTY APPRAISAL SYSTEM

by Richard A. Swanson [1] & David J. Sisson [2]

DEVELOPING THE SYSTEM

Introduction

Evaluation of university faculty personnel is, in most instances, the responsibility of the department chairman. The chairman is charged to subjectively apply evaluative criteria such as: evident ability as a teacher, service to the academic community, engagement in scholarly research or creative work (Bowling Green State University, 1969, p. 27).

The evaluations by chairmen have many sources of error. Frames of reference among chairmen differ rather markedly. The chairman's ratings may be affected by faculty members who differ greatly in age, teaching field, sex, years of experience, abilities, and other factors within and between departments and universities (Stanley, 1962, p. 6). The chairmen's evaluations are at best unpredictable and in many cases without validity.

The chairman may not be best qualified to assess all dimensions of faculty performance. It is believed that, while the chairman's rating may constitute the best measure of performance in one or more dimensions of faculty performance, it cannot be considered the sole criterion (Domas, 1950). A university professor's duties include: teaching, scholarly productivity, and service (Bowling Green State University, 1969; Lauritis, 1967). A rating system should consider all the dimensions of faculty performance, as well as who might best assess performance in the dimensions.

Purpose of the Study

The purpose of this study was to develop and evaluate a university departmental faculty appraisal system (FAS). Specifically, it was the purpose of this study to (1) develop a theoretically sound faculty appraisal system, (2) select or develop appropriate instruments, and (3) to develop a scoring system.

The System

A theoretical model for the appraisal of university faculty should include appraisal of the three dimensions of performance that have been identified: teaching, scholarly productivity, and service (Table 1). The research shows that students are honest and reliable raters of teaching performance. Students are best qualified to rate the teaching performance of the faculty, but are not

[1] Dr. Swanson is Associate Professor, Department of Industrial Education and Technology, Bowling Green State University.
[2] Mr. Sisson is Instructor, Department of Industrial Technology, Kent State University.

1. What effects would feedback have on the performance of psychomotor tasks by students in grade levels other than those used in this study, either at a less advanced or a more advanced stage?
2. What effect would the different feedback techniques have if students were grouped according to psychomotor ability?
3. What effect would the treatments used in this study have on the performance of other psychomotor tasks found in industrial arts classes?
4. What effect would the feedback techniques used in this study have on the performance of psychomotor tasks which differ in complexity and degree of difficulty?
5. Would learning the correct performance of a psychomotor task under conditions where feedback is present effect retention?
6. Can other techniques of providing feedback be identified which are effective in raising the student's level of performance of a simple psychomotor task?

REFERENCES

Gagne, Robert M., and Edwin A. Fleishman. *Psychology and human performance.* New York: Hold-Dryden, 1959.

Hill, Winfred F. *Learning: A survey of psychological interpretations.* San Francisco: Chandler, 1963.

Sorenson, Herbert. *Psychology in education.* New York: McGraw-Hill, 1964.

While most evaluators would agree that "quality of the final product" is of great importance, they are likely to argue that some consideration should be given to the "process" by which the final product is obtained. They would evaluate the individual's care of the equipment he uses, his observance of safety rules, and his adherence to approved methods of work. They might also take into account the amount of material he wastes and the time he takes to do the job.

Boyd, J.L., Jr. and Shimberg, B. *Handbook of performance testing.* Princeton, N.J.: Educational Testing Service, 1971.

chairman have direct contact with the service performance of faculty members, are aware of their scholarly productivity, and somewhat knowledgeable of their teaching ability. The assessment of performance by the peer nomination technique was chosen as the most valid means of faculty evaluating their peers. The chairman simply ranks his faculty according to total performance with consideration for all three dimensions.

Research Questions and Procedures

To evaluate the theoretical system, the following research questions were developed:

1. Is the Faculty Appraisal System (FAS) reliable?
2. Is the Faculty Appraisal System valid?
3. Is it feasible to establish quantitative ratings for individual faculty members using the Faculty Appraisal System?

Data for this study were obtained from the Faculty of the Department of Industrial Education, Bowling Green State University. Full-time instructors, assistant professors, associate professors, and professors who taught courses during the Winter and Spring quarters 1969-1970 were evaluated through the use of the Faculty Appraisal System (FAS). There were twelve professors in all.

The Illinois Course Evaluation Questionnaire (CEQ) was administered to all classes in the department at the final exam or the last meeting. Students were given verbal directions in addition to the written directions of the CEQ. Proctors were used to administer the questionnaire to insure uniformity. The CEQ was designed to be machine scored to insure rapid processing and equal treatment for all instructors and classes (Spencer, 1969, p. 4). It was assumed that an instructor's CEQ rating was somewhat constant and therefore if there were several CEQ scores for an instructor the scores were averaged for the quarter. Total scores from the CEQ are the mean responses over all questionnaire items (Spencer, 1968, p. 7). The total scores were used to rank the teaching performance of the faculty. The rank order from the total CEQ scores was weighted one-third of the system rank.

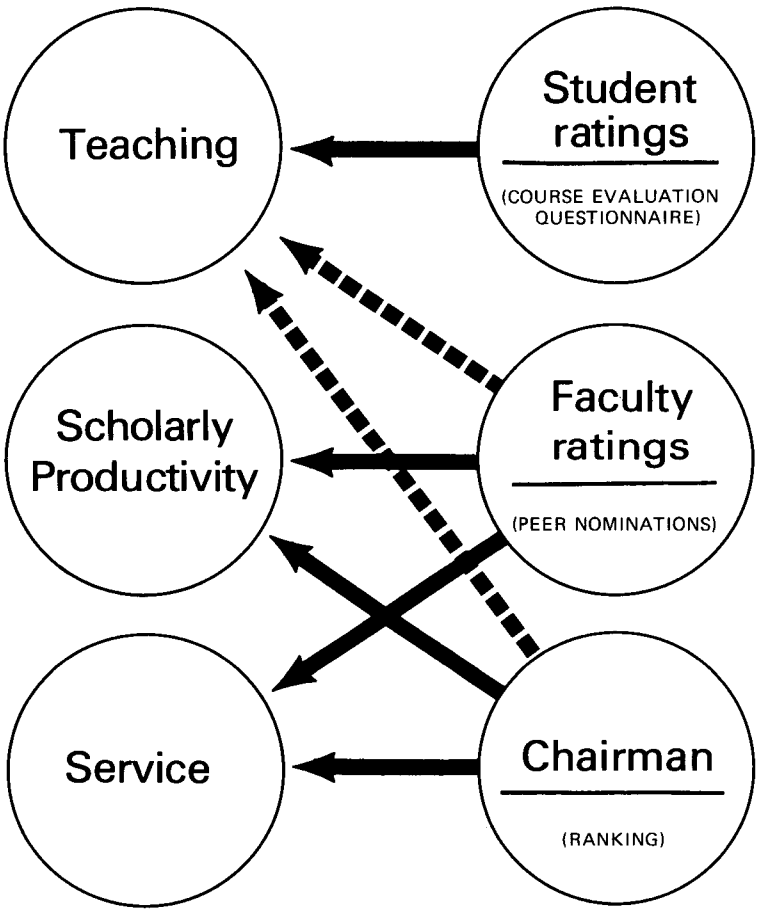
A simple roster (Table 2) of faculty members was used for the peer nomination technique. Nominations were made by writing the word "most" in the blank next to the four faculty members who most highly met the ideal of teaching, research, and service. Current faculty vitas circulated to every faculty member assist them in their decisions. Scoring consisted of a rank order determined by the number of times each member was nominated. The rank order from the peer nomination was weighted one-third of the system rank.

The department chairman was charged to subjectively apply the criteria of the university faculty charter to each faculty member. A rank order of the faculty was determined by the chairman using personal observations and interaction with the faculty along with a review of the information supplied by the faculty vitas.

These measures were made at the end of the Winter and Spring quarters.

TABLE I

EVALUATING THE THREE DIMENSIONS OF FACULTY PERFORMANCE THROUGH THREE POINTS OF VIEW



likely to be aware of or able to assess the research and service dimensions of performance. Of the available instruments, the Illinois Course Evaluation Questionnaire (Spencer, 1966) was selected as the most acceptable. Peers and

(Spearman-Brown) .92, and Kuder-Richardson 21 reliability of .932. The content validity was tediously established (Spencer, 1969).

Wolins (1956) reported, in a factor analytic study, that peer ratings made by 974 subjects had reliabilities in the high .80's. The technique has shown results that correlate highly with criterion such as success in leadership ability (Hollander, 1956; Wherry, 1945).

The reliability of chairmen ratings was never thought to be a problem. The area in question was the ability of the chairman to make a valid assessment of total faculty performance. This weakness was part of the rationale behind the development of the FAS. In any case, it is assumed that a degree of faith must be placed in the opinion of the academic leader, and thus is included.

The validity of the FAS is enhanced as the independence of the measures was supported. Intercorrelations of .08 to .88 between the measures established the fact that the three groups rated faculty on independent criterias. Tables 3 and 4 show the Spearman-rho correlation coefficients for the measures during the Winter and Spring quarters. One can conclude that the independence of the measures was partially supported and that the three measures of the FAS individually contributed to the valid assessment of total faculty performance.

TABLE 3

CORRELATION COEFFICIENTS AMONG STUDENT, PEER, AND CHAIRMAN RATINGS OF FACULTY - WINTER QUARTER

Groups	N	Correlation
Peer rating to Chairman rating	12	.82
Peer rating to Student rating	12	.70
Student rating to Chairman rating	12	.50

*Spearman — rho

TABLE 4

CORRELATION COEFFICIENTS AMONG STUDENT, PEER, AND CHAIRMAN RATINGS OF FACULTY - WINTER QUARTER

Groups	N	Correlation
Peer rating to Chairman rating	12	.88
Peer rating to Student rating	12	.28
Student rating to Chairman rating	12	.08

*Spearman — rho

Rank data from each of the measures for both quarters were averaged into a mean rank score for each faculty member. This mean rank score was then used as the index of faculty performance.

TABLE II

PEER NOMINATION FORM
FACULTY APPRAISAL SYSTEM

Instructions

The list of names[1] below includes all the faculty members in the Department of Industrial Education. Based upon the professional charge of quality teaching, research, and service expected of university faculty members, select the four members that you feel most highly meet the ideal by placing a "MOST ' in front of their names.

Do not choose more than four men. You may choose yourself if you wish. Do not sign this sheet or identify yourself in any way. Place this sheet in the enclosed campus mail envelope, seal, and return to the department chairman as soon as possible.

---------- SMITH
---------- JONES
---------- GREEN
---------- JOHNSTON
---------- BROWN
---------- DOE
---------- HOWARD
---------- LONG
-- -------- FISCHER
---------- ADAMS
---------- MILLER
---------- HOOD

RANDOM ORDER OF FACULTY NAMES WILL CHANGE EACH QUARTER

[1] The names presented here are fictitious.

Reliability and Validity

The reliability and validity of the Faculty Appraisal System has, in effect, been established through the reliability and or validity of the three component measures reported by other researchers.

The CEQ has a reported split-half reliability of .849, corrected for length

TABLE 6

MEAN FACULTY RANKS, NATURAL BREAKS, AND TENABLE GRADES - SPRING QUARTER

Professor	Mean Rank	Break	Tenable Grade
H	2.00		
		.67	A
D	2.67		
		1.83	— — — — —
C	4.50		
		.50	B
K	5.00		
		.83	
I	5.83		
		.84	— — — — —
J	6.67		
		.16	
L	6.83		C
		.67	
G	7.50		
		.50	
F	8.00		
		.67	— — — — —
M	8.67		
		.66	D
A	9.33		
		1.67	— — — — —
B	11.00		F

TABLE 7

MEAN FACULTY RANKS, NATURAL BREAKS, AND TENABLE GRADES - TOTAL SYSTEM RATING

Professor	Mean Rank	Break	Tenable Grade
H	1.75		
		.75	A
D	2.50		
— — — — — — — — — —		1.92	— — — — —
K	4.42		
		.16	
C	4.58		
		1.33	B

Quantitative Ratings

Distinct divisions of faculty performance occurred during the Winter and Spring quarters (Tables 5 and 6) and for the total system ranking (Table 7). The natural breaks among the total mean ratings produced additional validity of the system's ability to discriminate among faculty performance. These natural divisions also provide a method of assigning values to the various levels of performance.

TABLE 5

MEAN FACULTY RANKS, NATURAL BREAKS, AND TENABLE GRADES - WINTER QUARTER

Professor	Mean Rank	Break	Tenable Grade
H	1.50		
		.83	A
D	2.33		
		1.50	— — — — —
K	3.83		
		.84	
C	4.67		
		1.00	
A	5.67		B
		.33	
I	6.00		
		.00	
L	6.00		
		2.00	— — — — —
G	8.00		
		.33	C
J	8.33		
		1.34	— — — — —
M	9.67		
		1.00	
F	10.67		D
		.66	
B	11.33		

TABLE 9

STUDENT, PEER, CHAIRMAN, AND MEAN RANKS OF FACULTY SPRING QUARTER

Professor	Student Ranking	Peer Ranking	Chairman Ranking	Mean Rank
A	9.0	11.0	8.0	9.33
B	11.0	11.0	11.0	11.00
C	2.0	5.5	6.0	4.50
D	4.0	1.0	3.0	2.67
F	1.0	11.0	12.0	8.00
G	12.0	5.5	5.0	7.50
H	3.0	2.0	1.0	2.00
I	8.0	5.5	4.0	5.83
J	5.0	8.0	7.0	6.67
K	10.0	3.0	2.0	5.00
L	6.0	5.5	9.0	6.83
M	7.0	9.0	10.0	8.67

TABLE 10

Professor	Combined Quarter Rank by Students	Combined Quarter Rank by Peers	Combined Quarter Rank by Chairman	Mean Rank
A	7.0	7.50	8.0	7.50
B	11.5	11.00	11.0	11.17
C	2.5	5.25	6.0	4.58
D	4.0	1.00	2.5	2.50
F	5.0	11.00	12.0	9.33
G	11.0	7.25	5.0	7.75
H	2.0	2.25	1.0	1.75
I	7.5	6.25	4.0	5.91
J	8.0	7.50	7.0	7.50
K	8.0	2.75	2.5	4.42
L	4.0	6.25	9.0	6.42
M	7.5	10.00	10.0	9.17

As demonstrated by some movement of faculty ranks within the system from Winter to Spring quarter, the results of one quarter's system rank should be treated quite tentatively. Considering the apparent face validity and reliability

I	5.91			
		.51		
L	6.42			
	— — — — — — — — — —	1.08	— — — — —	
A	7.50			
		.00		C
J	7.50			
		.25		
G	7.75			
	— — — — — — — — — —	1.42	— — — — —	
M	9.17			
		.16		D
F	9.33			
	— — — — — — — — — —	1.84	— — — — —	
B	11.17			F

The relative student, peer, chairman and mean ranks of faculty for individual quarters can be seen in Tables 8 and 9. The results of the total system rank are presented in Table 10. The total system rank is a mean rank for all measures during both test periods.

TABLE 8

STUDENT, PEER, CHAIRMAN, AND MEAN RANKS OF FACULTY WINTER QUARTER

Professor	Student Ranking	Peer Ranking	Chairman Ranking	Mean Rank
A	5.0	4.0	8.0	5.67
B	12.0	11.0	11.0	11.33
C	3.0	5.0	6.0	4.67
D	4.0	1.0	2.0	2.33
F	9.0	11.0	12.0	10.67
G	10.0	9.0	5.0	8.00
H	1.0	2.5	1.0	1.50
I	7.0	7.0	4.0	6.00
J	11.0	7.0	7.0	8.33
K	6.0	2.5	3.0	3.83
L	2.0	7.0	9.0	6.00
M	8.0	11.0	10.0	9.67

Interpreting the Data

It is assumed that the preceeding portions of this paper are not intended as an academic exercise, but an exercise that yields valid information for making the "real" decisions of retention, promotion, tenure, and salary in the academic world. It is at the full realization of this point that the "alarm factor" is ignited in many professors. It is necessary to remind the reader that decisions of retention, promotion, tenure, and salary have, are, and will continue to be made. This is nothing new. The question is, will they be made via a single input (chairman rating only) or will they be based on multiple inputs? Even the most cautious reader must admit that the thousands of bits of information generated for the assessment of performance levels by the FAS are more valid than the present chairmen ratings typically being used. Obviously, these writers consider the FAS much more than a "better than" alternative.

Establishing a Retention, Promotion, Tenure, and Salary Committee

There are several decisions that must be made before the FAS data can be utilized for retention, promotion, tenure, and salary considerations. It seems appropriate that a standing departmental Retention, Promotion, Tenure, and Salary Committee (RPTS) be established to handle these decisions. Its membership might consist of one tenured faculty, one untenured faculty, one graduate student, and one undergraduate student. Once the committee has collected and digested the data into mean faculty ranks it must determine the range of performance within the department. It is conceivable that even though faculty are ranked from highest performer to lowest performer that they all could fall into one level of performance (e.g. all "A's" even though the scale runs A, B, C, D, F). Departmental faculties will vary in their range of performance. Some will run the gamut while others may span a shorter range. The committee must make a value judgement by carefully studying all the data input of both the top man and the bottom man and then assign these men a grade (A, B, D, D, or F). Once the range is established, the natural breaks between the ranking can serve as divisions of performance between the extremes (see Table 10 for tenable example).

Criteria for Retention, Promotion, and Tenure

The next obvious extention to the utilization of the FAS is the comparison of faculty levels of performance to minimum criteria established by the committee. An example of one such set of criteria is as follows:

Instructor—Retention and Promotion (Minimum qualifications)

A. "C" rating or better overall

B. Must perform in the upper half in at least one dimension

Assistant Professor—Retention and Promotion (minimum qualifications)

A. "B" rating or better overall

B. Must perform in the upper half in at least one dimension

of the FAS, consistently low ranks validated over two or more quarters would seem to indicate problem areas in the performance of acceptable faculty duties. In practice, data should be collected on at least two or three occasions (during the academic year) determined by whether institution is on the quarter or semester plan.

The theoretical elements of the faculty appraisal system and proposed independence of the three adopted measures were partially supported. The following rank-difference correlations for these three measures during the two ten week quarters were: .50 for student to chairman rating; .70 for student to peer rating; .82 for peer to chairman rating for the winter quarter. For data collected the spring quarter the correlations were: .28 for student to peer rating; .08 for student to chairman rating; and .88 for peer to chairman rating.

Additional validity of the system was evidenced by the emergence of distinct divisions in faculty performance. With the previously validated instruments discrimination occurred among faculty in the form of spread in ratings and in the emergence of natural breaks.

UTILIZING THE SYSTEM

Administrative Concerns

The weighting of the instruments within the proposed system were equal. Each of the three measures was weighted one-third of the total. If adopted by other institutions or departments, the nature of the institution may dictate a different method of weighting. For example, if a greater emphasis is placed on research rather than on teaching, the instruments that assess that domain could carry greater weight.

As Sawyer (1962, p. 268) pointed out, some faculty do not regard the evaluation of teaching performance as necessary. Accordingly, there was some resistance to the first application of the CEQ and the peer nomination technique. Possibly because of the diagnostic as well as evaluative nature of the instruments, resistance was nil on subsequent applications.

To insure uniformity for this study, the CEQ was administered to the classes by proctors. The questionnaire was designed to be administered by classroom teachers (Spencer, 1969, p. 4). In the ongoing application of the FAS, proctors may not be necessary. The CEQ could be administered in the classroom by the faculty members and the peer nomination forms could be distributed and collected through departmental mails.

The 24-item CEQ was selected for the FAS rather than the 50-item form to prevent rater boredom and indiscriminate marking. The 50-item form supplies print-out data additional to that of the 24-item form. The additional data are: fallability (lie) coefficient, total, and all university norms. As the cost of the forms is the same, the 50-item form may be more valuable.

TABLE II

CRITERIA AND METHOD FOR DETERMINING SALARY INCREASES

1. To be considered for a salary increase a faculty member must have a "C ' rating overall.

2. Salary increases should be independent of rank and longevity. Salary increases should be based on relative performance of the members of the department.

Performance level	Weighted salary increase	Faculty[1] performing at the various levels	People X units	Money available ÷ total people/units	Recommended salary increases
A	4	1. Hood	8		1. $1,2680.00
		2. Green			2. 1,2680.00
	3.5	3. Johnston	3.5		3. 1,110.00
		4. Doe		(e.g. $10,000.00 available)	4. 951.00
B	3	5. Adams	9		5. 951.00
		6. Smith		317	6. 951.00
	2.5	7. Howard		31.5 $\overline{)10,000}$	7. 792.00
		8. Jones	5		8. 792.00
		9. Wagner			9. 634.00
C	2	10. Long	6	Approximately $317.00 per unit	10. 634.00
		11. Brown			11. 634.00
D	0	12. Fischer			12. 00.00
		13. White	0		13. 00.00
F	0	14. Miller	0		14. 00.00
			Total 31.5		$10,000.00

[1] The names presented are fictitious.

CONCLUSIONS

The effective utilization of the FAS is primarily dependent upon the validity and relative independence of the component measures of the FAS. The research effort reported here supports both of these concepts as well as to outline the suggested utilization of the data. As with any measurement system, constant evaluation and revision is desirable. Considering the seriousness of the decisions being made from the data in this situation, it is felt to be mandatory.

REFERENCES

Adams, R.B., *The application of a special approach to teacher selection and appraisal.* New York: Vantage Press, 1967.

Barr, A.S., The measurement and prediction of teacher efficiency: Summary of investigations. *The Journal of Experimental Education,* 1948, 16:216.

_____., Wisconsin studies of the measurement and prediction of teacher effectiveness. *The Journal of Experimental Education,* 1961, 30:1-156.

Bowling Green State University, *Faculty charter.* Unpublished manuscript. 1969.

Associate Professor—Retention and Promotion (minimum qualifications)

A. "B" rating or better overall
B. Must perform in the upper half in at least one dimension

Full Professor—Retention and Promotion (minimum qualifications)

A. "B" rating or better overall
B. Must perform in the upper half of two dimensions

Tenure—(minimum qualifications)

A. "B" rating or better overall
B. Must perform in the upper half of one dimension

Exceptions: There are two categories of terminal exceptions that can be made if the above criteria are not met. First, if the department cannot find a suitable replacement a person may be retained until the replacement is hired. The second deals with major variables that might have effected performance of a faculty member on a terminal basis (e.g. health problems).

The perceptive reader can easily see that a non or low performer will never reach the tenure stage. A criticism may be lodged against the FAS in that it will encourage everyone to work until tenure and then they will stop. The answer to this is simple. It's better to get five years of productivity than none. Utilization of a correlated productivity to salary reinforcement system can act as an ongoing incentive to faculty after tenure to perform at their maximum.

Criteria for Salary

The resulting levels of performance via the FAS can also be used as a basis for awarding annual salary increases. Assuming that the department receives a lump sum of dollars the money could be distributed according to the model presented in Table II. A minimum level of performance should be established. In this example a "C" performance is the minimum acceptance level. Any tenured faculty performing lower than the "C" level receives no salary increase. In the criterias presented in this paper any non-tenured instructor performing below the "C" level would be dropped from the faculty. Column two of the model again requires a valid judgement from the RPTS committee. In this case it is the relative worth of a "C" man as compared to a "B" man and to an "A" man. In the proposed weighting the "C" man is equivalent to 50% of an "A" man. The number of people within the levels of performance are multiplied by the weighting factor. The sum of the people times the weighting factors is then divided into the sum of money available to the department. The result is a dollars per unit figure that can then be converted into individual salary increments. This method of rewarding annual salary increases puts the instructor and the full professor in the same boat. The salary awards have nothing to do with rank, base salaries, or tenure status.

Ryans, D.G., Assesment of teacher behavior and instruction. *Review of Educational Research,* 1963, 33: 4-15-441.

_____., Teacher behavior can be evaluated. *The Evaluation of Teaching,* Washington, D.C.: Pi Lambda Theta, 1967.

_____., *Characteristics of teachers.* Washington, D.C.: American Council on Education, 1960. a

_____., Prediction of teacher effectiveness. *Encyclopedia of Educational Research,* Macmillan Company, 1960, 1486-1491.b

Sawyer, B.E., Teacher evaluation in the liberal arts college. *Journal of Teacher Education,* 1962, 13: 268-272.

Shao, O.H., Faculty evaluation plan in a liberal arts college. *Liberal Education,* 1968, 54: 413-417.

Sisson, D.J., *Development and evaluation of a university faculty appraisal system.* Unpublished master's thesis, Bowling Green State University, 1970.

Spencer, R.E., *The course evaluation questionnaire: Manual of Interpretation.* Unpublished manuscript. Report No. 200 of the Measurement and Research Division, Office of Instructional Resources, University of Illinois, Urbana, Illinois 1965. Mimeo.

_____., *The course evaluation questionnaire: Manual of interpretation (Revised).* Unpublished manuscript. Report No. 270 of the Measurement and Research Division, Office of Instructional Resources, University of Illinois, Urbana, Illinois 1968. Mimeo.

_____., A history of the development of the Illinois course evaluation questionnaire. Unpublished manuscript. Report No. 306 of the Measurement and Research Division, Office of Instructional Resources, University of Illinois, Urbana, Illinois 1969. Mimeo.

Stander, N.E., The assessment of trade and technical education courses: The use of instructor and peer ratings as intermediate criteria of curricula effectiveness, *The development of achievement measures for trade and technical education.* Raleigh, North Carolina: Office of Education grant number OEG-2-6-000517-0585, August, 1966.

Stanley, J.C. , and Weiley, D.E., *Development and analysis of experimental designs for ratings.* Washington, D.C.: Cooperative Research Report No. 789. United States Office of Education, 1962.

Stallings, W. M., and Singhal, S., *Some observations on the relationships between research productivity and student evaluations of courses and teaching.* A paper presented at the annual meeting of the American Educational Research Association, February, 1969.

Stillings, F.S., The n-th attempt—A criterion for promotion. *Contemporary Education,* 1968, 40: 90-91.

Swanson, R.A., *Auditory automotive mechanics diagnostic achievement test.* Unpublished doctor's dissertation, University of Illinois, 1968.

Bryan, R.C., Student rating of teachers. *Improving College and University Teaching,* 1968, 16:200-202.

Cook, W.W., Leeds, C.H., and Callis, R., *Minnesota teacher attitude inventory.* New York: The psychological Corp., 1957.

deJung, J.E., Effects of rater frames of reference on peer ratings. *Journal of Experimental Education,* 1964, 33(2): 121-131.

Domas, S.J., *Report of an exploratory study of teacher competance.* Cambridge, Mass.: New England School Development Council, 1950.

Druckers, A.J., and Remmers, H.H., Do alumni and students differ in their attitudes toward instructors? *Journal of Educational Psychology,* 1951, 42: 129-143.

Fishman, J.A., Cross-cultural perspective on the evaluation of guided behavior change. *The Evaluation of Teaching,* Washington, D.C.: Pi Lambda Theta, 1967.

Gibb, C.A., Classroom behavior of the college teacher. *Educational and Psychological Measurement,* 1955, 15:254-263.

Goodhartz, A.S., Student attitudes and opinions relating to teaching at Brooklyn College. *School and Society,* 1948, 68:345-49.

Gustad, J.W., *Policies and practices in faculty evaluation.* Washington, D.C.: American Council on Education, 1961.

Hayes, R.B., A way to measure classroom teaching effectiveness. *Journal of Teacher Education,* 1963, 14:168-176.

Hollander, E.P., The friendship factor in peer nomination. *Personnel Psychology,* 1956, 9: 435-447.

Hudelson, E., *Class size at the college level.* Minneapolis, Minnesota: The University of Minnesota Press, 1928.

Lauritis, J., Thoughts on the evaluation of teaching. *The Evaluation of Teaching,* Washington, D.C.: Pi Lambda Theta, 1967.

Marshall, M.S., Grading the teacher. *Association of American Colleges Bulletin,* 1952: 257-267.

McKeachie, W.J., Student ratings of faculty: A research review. *Improving College and University Teaching,* 1957, 5: 4-8.

Mitzel, H.E., Teaching effectiveness. *Encyclopedia of Educational Research,* Macmillan Company, 1960.

Mueller, F.J., Trends in student ratings of faculty. *Bulletin of American Association of University Professors,* 1951, 37:319-324.

Phillips, B.N., The individual and the classroom group as frames of reference in determining teacher effectiveness. *The Journal of Educational Research,* 1964. 58: 128-131.

Remmers, H.H., and Wykoff, G.S., Student rating of college teaching A reply. *School and Society,* 1929, 30: 232-234.

Riley, Ryan, and Lifshitz, *The student looks at his teacher.* New Brunswick, New Jersey: Rutgers University Press, 1960.

Vielhaber, D.P., and Gottheil, E., First impressions and subsequent ratings of performance. *Psychological Reports,* 1965, 17(3): 9-16.
Webb, W.B., The problem of obtaining negative nominations in peer ratings. *Personnel Psychology,* 1955, 8:61-63.
Wherry, R.J., and Fryer, D.H., Buddy ratings: popularity contest or leadership criteria? *Personnel Psychology,* 1945, 2:147-59.
Wolins, L., *The validity of peer ratings as a function of rater characteristics and methods of scoring.* Unpublished doctor's dissertation, Ohio State University, 1956.

One of the most useful nonperformance "performance" tests is the identification test. It may involve the use of actual objects or it may rely on photographs. In either case, there may be great variation in the degree to which the identification task approaches the reality of a job-sample measure.

Boyd, J.L., Jr. and Shimberg, B. *Handbook of performance testing.* Princeton, N.J.: Educational Testing Service, 1971.

An examinee may be given a set of problems requiring him to order the parts needed to make repairs on a given make and model of a car; or he might be required to select the equipment that would meet certain heating or cooling specifications and conform to space limitations indicated on an accompanying blueprint. Such tasks are highly realistic and job-oriented. The fact that the response is in a paper-and-pencil format in no way diminishes the validity of the performance.

Boyd, J.L., Jr. and Shimberg, B. *Handbook of performance testing.* Princeton, N.J.: Educational Testing Service, 1971.

Readers who are familiar with the sets of rules and principles for writing multiple-choice test items may feel uncomfortable because the exact "right" and "wrong" ways of preparing performance tests have not been prescribed. Human performance covers such a wide range of behaviors that a variety of techniques must be employed to test them. No set of rules or cookbook approach can be adequate to define the boundaries of acceptable performance.

Boyd, J.L., Jr. and Shimberg, B. *Handbook of performance testing.* Princeton, N.J.: Educational Testing Service, 1971.