DOCUMENT RESUME

ED 135 862 TM 006 093

AUTHOR Love, Judith A.; And Others
TITLE An Instrument to Describe College Teaching: A Reliability Study.
PUB DATE [Apr 77]
NOTE 21p.; Paper presented at the Annual Meeting of the American Educational Research Association (61st, New York, New York, April 4-8, 1977)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Classroom Observation Techniques; *College Instruction; Correlation; *Effective Teaching; Factor Analysis; Higher Education; *Lecture; *Reliability; Teacher Behavior; Test Reliability; *Video Tape Recordings
IDENTIFIERS Formal Lecture Observation Assessment Technique

ABSTRACT

Perhaps more than ever before, college teaching is being studied and evaluated. This paper describes the development of a simple descriptive instrument used to focus observers' classifications and ratings of college teachers' instructional behaviors as recorded on video tape. The need for such an instrument is reviewed, the methodology for testing the reliabilities of its measures is described and possible applications are outlined. The statistical findings (Spearman-Brown estimates of reliability ranged from .68 to .95 and averaged .81; four discrete factors emerged through factor analysis of the intercorrelations among grouped observers' mean ratings of instruction) and projected modifications are also presented. Further research is indicated and the plans for such research are summarized. (Author)

An Instrument to Describe College Teaching: A Reliability Study

Judith A. Love

Teaching Resources Center

Douglas L. Minnis

Jonathan H. Sandoval

George D. Yonge

Education Department

University of California, Davis

Paper presented at the Annual meeting of the
American Educational Research Association
New York, New York
April 4-8, 1977
Session 13.02

Running Head: Describing College Teaching

Abstract

Perhaps more than ever before, college teaching is being studied and evaluated. This paper describes the development of a simple descriptive instrument used to focus observers' classifications and ratings of college teachers' instructional behaviors as recorded on video tape. The need for such an instrument is reviewed, the methodology for testing the reliabilities of its measures is described and possible applications are outlined. The statistical findings (Spearman-Brown estimates of reliability ranged from .68 to .95 and averaged .81; four discrete factors emerged through factor analysis of the intercorrelations among grouped observers' mean ratings of instruction) and projected modifications are also presented. Further research is indicated and the plans for such research are summarized.

An Instrument to Describe College Teaching: A Reliability Study

Much attention is being accorded the evaluation of teaching at the higher education level. We frequently hear about student ratings and observations of instructors and other methods of identifying effective teaching. An important question not often asked and one which must be asked prior to perfecting methodology is: "What is effective teaching?" Is it that teaching which results in increased learning? Or in increased satisfaction with the learning process? Or in both? Does increased satisfaction facilitate learning? Certainly these sorts of relationships are implied when institutions that exist for the purpose of promoting learning reward teachers for "effective" teaching.

The literature on effective college teaching is not definitive. Dunkin and Biddle (1974) have suggested several reasons for the failure of educational researchers to define effective teaching despite over sixty years of study. They suggest the single most significant shortcoming is the failure to observe teaching activities. The present study is primarily concerned with developing a reliable observational measure of the lecture method of instruction. As a future step we intend to study other basic practices of college teaching and their possible relationship to learning. Ultimately, our aim is not to evaluate teachers, but to collect information to be used as feedback to the instructors observed with the goal of improving college teaching.

Background

Gage and Berliner (1976) describe the connections between models of teaching and theories of learning as tenuous. Investigations of learning styles, however, seem more likely to result in such connections (Gagne, 1976, Chap.2). It has been suggested (Joyce and Weil, 1972, Chap.17) that a framework for matching

learning styles and instructional strategies be created. Investigations purporting to measure teaching effectiveness are frequently criticized for their narrow scope. McKeachie and Kulik (1975, Chap. 6) cite the lack of concern for the instructor out-of-class time in models for evaluating teaching effectiveness. In the same vein, those who study teaching and the evaluation of its effectiveness are cogently remonstrated by Wilson, Gaff, Wood, Dienst and Bavry (1975) for their lack of consideration of the diversity of teaching context, the disparity between student and teacher goals and the range of institutional philosophical orientation. Hildebrand, Wilson and Dienst (1971) identified several types of teaching and, as reported by Wilson et al., concluded that no single model of effective teaching exists; "teachers who exhibit different teaching styles appeal to students who have different college goals" (p. 18). Meredith Gall (Note 1) has further cautioned researchers to consider specific teaching styles, as well as curriculum objectives and instructional settings in their assessments of teaching effectiveness. He suggests closing such inquiry with the administration of tests created to be responsive to the "idiosyncratic outcomes of the teaching skills and curriculum objectives." Similar concern for the measurement of student outcomes as criteria for teaching effectiveness has been registered by McKeachie, Lin and Mann (1971) and by Burnett (1974).

There are some who express doubt that teaching technique has any relationship to learning (see Gagne and Rohwer, 1969) and others who claim teachers do influence learning but that there is no significant difference in the amount of learning based on teaching method (Dubin and Taveggia, 1968). On the other hand, McKeachie (1969) argues that teaching method or behavior does have something to do with student satisfaction and, though the correlation between satisfaction and learning is low, "we'd all prefer to have students leave class with warm feelings about their experience" (p. 210).

Joyce and Weil (1972) summarize seventeen models of teaching, claimed by their authors to be particularly effective for certain kinds of learning. Whether or not such relationships exist, Joyce and Weil conservatively, but convincingly, state "It seems reasonable to suppose that, as our technology for studying teaching and learning improves, people will discover regularities in the teaching-learning process that have not been apparent before. A few general methods may emerge as superior" (p. 4). The general methods or strategies resulting from such inquiry will employ certain behaviors which we will then correctly label "effective". In order, then, to define effective teaching our research must look at teaching and its relationship to learning. The development of a reliable observational measure of teaching behaviors is an important first step in the process. Because lecturing is one form of college teaching that is widespread across disciplines, our first efforts were directed at studying this mode of instruction. Future energies will be directed at other modes of instruction.

Method

Our initial step was to address the first question raised by this paper, "What is effective teaching?" First, we conducted a review of the literature on teaching which yielded a pool of variables assumed to contribute to better learning (see Rosenshine and Furst, 1971, Chap. 3; Nuthall, G. and Snook, I., 1973, Chap. 2). Next, we reviewed the literature on human learning which yielded a set of characteristics assumed to affect consistently the acquisition and retention of information. Finally, a review of the literature on course and instructor evaluation yielded some factors which seem to contribute to increased student satisfaction with the teaching-learning process. This tri-lateral review process showed a high degree of semantic and conceptual overlap. We listed the descriptors of teaching behaviors, then abstracted

commonalities, and ultimately created a set of potentially effective lecturing behaviors. The set is displayed in Table 1.

Based on the assumption that the items listed in Table 1 were reasonable descriptions of effective college teaching behavior during lectures, we proceeded to design an observational rating schedule and a study of the reliability of the measures which would result from the employment of the instrument. Having constructed the instrument as a group, the experimenters found it possible to arbitrate differences in their interpretations of teaching performances. We expected that similar arbitrations during an observer training period would result in a high degree of agreement in observer judgement. We thus expected our investigation would result in a reliable measure of lecturing behavior. During the practice observation sessions we described our rationale for the rating of each behavior in order to calibrate our judgement during the independent observation period. Noting some redundance in our expressed reactions, we decided we had not eliminated conceptual overlap in the wording of the descriptors. Accordingly, we expected several discrete factors to emerge during the analysis of the data.

Subjects

The 20 subjects were full-time faculty and teaching assistants from nine departments at the University of California at Davis. Most requested videotaping as part of the U.C. Davis Teaching Resources Center's instructional improvement program. When informed of the proposed concurrent investigation, they agreed to contribute three videotapes and to complete the required paperwork in return for detailed feedback from the Center's videotape analyst.

Observers and Observer Training

During a one-month period prior to the data-gathering stage, two of the authors met with two graduate students for a total of sixteen hours of training. The

training period commenced with detailed discussions of the nature and scope of the categories for the observation, and culminated with three practice observation-rating sessions of three training tapes. When the observers each expressed confidence in their ability to observe and rate the teaching behaviors described by the instrument, the taping and rating segment of the study began.

Apparatus

A Sony Portapak 3400 series videocamera and recorder were used to record the classroom instruction. A built-in timer which registers minutes and tenths of minutes permitted the observers to start and stop the tapes at very nearly the same points in recorded time. A Sony CVM-115 video monitor was used for playback purposes. The tapes were recorded in various classrooms and lecture halls on the Davis campus. Ratings were recorded on an observational schedule. Student and instructor evaluations of teaching were collected. Written classifications of instructor intent and reactions were also collected as were forms requesting demographic information.

Procedure

The twenty instructors permitted the first thirty minutes of three fifty-minute lectures to be videotaped. The four trained observers reviewed the sixty tapes in an essentially random order; the order of acquisition. The observers stopped the taped playback every five minutes and rated the instruction, using the twenty-one categories of the instrument to structure their observations. For each of the first five five-minute periods, the observers recorded ratings for items one through fifteen. Only after the sixth period were ratings recorded for all 21 variables. Ratings consisted of a four point Likert scale with the descriptors of "strong" or "slight disagreement" and "slight" or "strong agreement," encoded one through four, respectively. Variables which corresponded to behaviors not

present were encoded "non-applicable," with a value of zero, and deleted from calculations. When the sixty tapes had been observed and rated by the four observers, the data from those 240 rating schedules, incorporating 96 ratings each, were keypunched onto IBM cards and statistically analyzed.

Results and Discussion

The mean ratings for the combined observers and for each variable across all instructors and tapings were calculated and are presented in Table 2. The Pearson product-moment correlation program from the Statistical Package for the Social Sciences (SPSS) (Nie, Hull, Jenkins, Steinbrenner and Bent, 1975) was utilized and the resulting coefficients are also presented in Table 2. The coefficients are based on data from the sixth five-minute period across 58 of the 60 video tapes. Two video tapes of less than 30 minutes in length were deleted from the sample. In the third column of Table 2, estimated reliabilities for each variable are presented. The Spearman-Brown Prophecy formula:

$$r_{xx} = \frac{n\,\bar{r}}{1+(n-1)\bar{r}} \qquad (1)$$

Where n = number of raters and $\bar{r}$ = average correlation (based on z conversion)

was used to estimate the inter-rater reliability of composite ratings of the four observers. These reliability estimates range from .68 (for Item 13) to .95 (for Item 1) and are comparable to those obtained from most standardized tests. The mean of the estimated reliabilities among all the variables is .81. This value was obtained by averaging the z coefficients, converting to $\bar{r}$, and using the Spearman-Brown formula (see Equation 1).

Rowley, in an article on the reliability of observational measures, corrected the common misconception that reliability coefficients refer to the reliability of the instrument itself (1976). Rowley points out that such coefficients should only be considered in determining the relative reliability of each of the measures which an instrument is designed to calculate. The data presented in Table 2

clearly show the disparity in reliability among teacher behaviors the instrument was designed to describe. The infrequency of certain behaviors undoubtedly contribute to the lower reliability of some measures. Other low reliabilities may have been influenced by insufficient observer training, inconsistency of instructor behavior or difficult-to-rate interactions.

In any case, the lowest reliability coefficients obtained in this study are still sufficiently high to support the primary hypothesis, that the observational schedule is capable of generating several reliable observational measures of college teaching.

Strong intercorrelations among the means of the combined observer ratings of the 21 variables suggested an underlying factor structure that was explored and defined through factor analysis. Varimax Factor Analysis using SPSS's (Nie, et al, 1975) alpha procedure was utilized. To obtain the factor score coefficients in Table 3, the ratings of the four observers for each of the three tapings and for each of the six five-minute periods for each instructor and each variable were pooled. The 72 resulting ratings (4x3x6) were averaged and the intercorrelations among the means were factor analyzed. As shown in Table 3, four discrete factors emerged. Factor 1, which accounts for 74.4 percent of the rotated variance, seems to be a general measure of instructional competence; of Task-Orientation. Both cognitive and affective variables, as well as direct and indirect behaviors compose this cluster. Factor 2, explaining 11.5 percent of the rotated variance, is composed of those variables relating to Use of Resources. Allowing opportunity for student involvement, soliciting and receiving such involvement in a non-threatening fashion, using the student ideas that result, giving feedback, demonstrating prowess with questioning strategies and purposes, and giving the impression that one enjoys both content and process, contribute to a cluster that has face validity. The third factor, which explains 8.1 percent of the rotated variance,

is a measure of Verbal Presence. All the variables composing this cluster refer to verbalization or verbal pacing. The inclusion of variable 17, the utilizing of diverse resources, seems logical. An instructor who competently engages students in dialogue that contributes to the development of a topic would certainly be judged as one who uses diverse resources (the students and their contributions) to illustrate principles. Factor 4, accounting for 5.9 percent of the rotated variance, is composed of two variables, one measuring physical demeanor and the other, the use of the blackboard. This factor is a measure of Physical Presence. The actual correlation between variables 19 and 21, the two components of the Physical Presence cluster, is just .53, as presented in Table 4. It seems reasonable that some relationship would exist since those lecturers who are too podium-bound to move to the board certainly do not exemplify attractive physical demeanor. It also seems reasonable that the relationship would not be a particularly strong one since the opposite phenomena also occurs occasionally, i.e., the lecturer who is so absorbed by boardwork as to preclude non-verbal interaction with students. The absence of such interaction would lead observers to lower ratings on the physical demeanor variable. The inter-correlations presented in Table 4, on further examination, provide specific evidence of the relationships among the variables which compose the clusters described in Table 3.

Table 5 presents the correlations between average observer ratings in each category and average student ratings in each category for the 20 instructors. All but three are significant at the .05 level of confidence. Students were asked to rate the instructor's behavior at the time of the taping. Nevertheless, their ratings may be summative across all of their experience with their instructor. If this summation is taking place it would reduce the magnitude of the correlations. Such disparity is not, however, evidence that either students or observers ratings are more valid than the other. Cosin, Greenough and Menges (1971) did find, in their review of validity studies related to student ratings of teaching, that

summative ratings consistently and strongly correlate with early-in-the-course ratings. A more reasonable explanation for a reduction in magnitude of the correlations between student and observer ratings is one which emphasizes the basic differences in rating the teaching process alone versus rating the process as it explicates the content.

The three non-significant correlations were "integrates, relates and categorizes" (OS2), "attractive physical demeanor" (SP2), and "charts, diagrams and boardwork" (VI2). The first disagreement may have resulted from the students' focus on note-taking in contrast to the observers' attention to the verbal behavior of the instructor. The second disagreement (on the physical demeanor of the instructor) seems to have resulted from the refusal of many students to rate this trait. A number of students wrote on the rating sheet that they felt this category was unfair, arbitrary and inappropriate. The final disagreement (on the use of charts and diagrams) was probably a result of the fact that this behavior was relatively infrequent in most and totally absent in three instructors.

The other correlations range from the low .40's to one correlation of .88. Nearly all of the components of factor 1 (c.f. Table 3) were items with high agreement between students and observers. The highest correlation ("speaks audibly and clearly") indicates the importance of this behavior in raters minds; but it should be noted that this category had the smallest variance of all the ratings, with most raters assigning it a "strongly agree."

Plans for Future Research

The four factors identified through the factor analysis of grouped means (Table 3) will now compose the redefined observational instrument. Future investigations will center on an initial reliability study of the redefined instrument and cross-referenced validity studies where the ratings of students and the self-ratings of the instructors are factor-analyzed and correlated with the factor analyses of trained observers' ratings. Whitely and Doyle (1976) have

suggested that implicit relationships among facets of teaching behavior exist in the minds of those who rate such behavior. The authors of this study expect to address, in future examinations of the instrument, the hypothesis that implicit factors do exist.

Increasing attention is being paid to of generalizability of the measures which an instrument provid els and Dempsey, 1976). Since our data were collected in an ANOVA design, they will be re-analyzed from a generalizability standpoint at a later date.

Most importantly, the authors plan to design studies which will investigate the relationships between particular instructional styles positive student outcomes. It is hoped that such investigations will contribute to the identification of effective college teaching behaviors and provide impetus for instructional change programs.

Reference Note

1. Gall, M. D. The Importance of Context Variables in Research on Teaching Skills. In B. Ward (Chair), "What does recent research tell us about the study of teaching skills?" Symposium presented at the meeting of the American Educational Research Association, San Francisco, 1976.

Table 1 - FLOAT*

A list of teaching behaviors employed in the college classroom and intended to structure the systematic observation and rating of college teaching.

Item	Variable	Description
1	$Clarity_1$	The instructor speaks audibly and clearly.
2	$Clarity_2$	Instructor is able to simplify concepts.
3	$Enthusiasm_1$	Instructor is energetic, stimulating, involved.
4	$Enthusiasm_2$	Instructor appears to enjoy both content and process.
5	Organization-Structuring$_1$	Instructor presents materials in an orderly, logical manner, making clear and logical transitions.
6	Organization-Structuring$_2$	Instructor integrates, relates and categorizes to aid both comprehension and note-taking.
7	Knowledge of Subject Matter$_1$	Instructor projects command of material.
8	Knowledge of Subject Matter$_2$	Instructor answers student questions and concerns or redirects them to appropriate resources.
9	$Interaction_1$	Instructor provides opportunity for expression of other opinions, assenting and dissenting.
10	$Interaction_2$	Instructor uses student ideas and comments.
11	$Support_1$	Instructor solicits and receives involvement in an accepting, non-threatening manner.
12	$Support_2$	Instructor gives feedback appropriately.
13	Questioning Tactics$_1$	Instructor displays facility with strategic questioning tactics.
14	Questioning Tactics$_2$	Instructor utilizes questions for many purposes.
15	Questioning Tactics$_3$	Instructor addresses questions to many individuals and groups.
16	$Variability_1$	Instructor displays a repertoire of skills, adapting his pace and techniques to the material being presented.
17	$Variability_2$	Instructor utilizes diverse resources to illustrate principles
18	$Stage\ Presence_1$	Instructor's verbal behavior is confident and effective.
19	$Stage\ Presence_2$	Instructor's physical demeanor is attractive and assured.
20	$Use\ of\ Visuals_1$	Media easy to follow and pertinent to lecture.
21	$Use\ of\ Visuals_2$	Charts, diagrams, boardwork are appropriate and presentation is effective.

*Formal Lecture Observation Assessment Technique

Table 2

Average z Coefficients, Converted Pearson $\bar{r}$ Coefficients and Estimated Reliability Among Four Observers' Ratings of College Teaching Behaviors

Item	Variables	Average z Coefficient among all observers	Converted Pearson $\bar{r}$ Coefficient	Estimated Reliability (corrected by Spearman-Brown Prophecy Formula)
1	Clarity$_1$	.82	.68	.95
2	Clarity$_2$	.58	.52	.85
3	Enthusiasm$_1$	.53	.48	.82
4	Enthusiasm$_2$	.53	.49	.82
5	Organization-Structuring$_1$	.43	.41	.76
6	Organization-Structuring$_2$	.45	.42	.77
7	Knowledge of Subject$_1$	.58	.53	.85
8	Knowledge of Subject$_2$	.38	.37	.71
9	Interaction$_1$	.52	.48	.81
10	Interaction$_2$	.41	.39	.74
11	Support$_1$	.56	.51	.84
12	Support$_2$	.35	.34	.68
13	Questioning$_1$	.35	.33	.68
14	Questioning$_2$	.39	.38	.72
15	Questioning$_3$	.36	.34	.69
16	Variability$_1$	.77	.65	.93
17	Variability$_2$	.52	.48	.81
18	Stage Presence$_1$	.56	.57	.84
19	Stage Presence$_2$	.67	.58	.89
20	Use of Visuals$_1$	No data	No data	No data
21	Use of Visuals$_2$	.44	.42	.76

Table 3

Varimax Rotated Factor Matrix by Grouped Observers

Factor 1 Task Orientation Item	Description	Coefficient
2	"...simplification of concepts..."	.87
3	"...energetic, stimulating, involved..."	.85
4	"...enjoys content and process..."	.84
5	"...orderly, logical presentation..."	.93
6	"...integrates, relates, categorizes..."	.92
7	"...projects command..."	.94
8	"...answers or redirects questions..."	.69
10	"...uses student ideas..."	.65
11	"...solicits involvement..."	.51
12	"...gives feedback..."	.56
16	"...repertoire of skills..."	.73
17	"...uses diverse resources..."	.74
18	"...confident, effective verbal behavior..."	.64
Eigenvalue		12.398
Percent of Rotated Variance		74.4

Factor 2 Use of Resources Item	Description	Coefficient
4	"...enjoys content and process..."	.46
9	"...opportunity for opinion..."	.48
10	"...uses ideas..."	.47
11	"...solicits involvement..."	.71
12	"...gives feedback..."	.64
13	"...strategic question..."	.91
14	"...many purposes for questions..."	.71
15	"...questions many individuals..."	.90
Eigenvalue		1.922
Per cent of Rotated Variance		11.5

Factor 3 Verbal Presence Item	Description	Coefficient
1	"...speaks audibly..."	.77
8	"...answers or redirects questions..."	.38
9	"...opportunity for opinions..."	.40
16	"...repertoire of skills..."	.43
17	"...utilizes diverse resources..."	.51
18	"...confident, effective verbal behavior..."	.58
Eigenvalue		1.356
Per cent of Rotated Variance		8.1

Factor 4 Physical Presence Item	Description	Coefficient
19	"...attractive physical demeanor.."	.80
21	"...use of blackboard..."	.87
Eigenvalue		0.978
Per cent of Rotated Variance		5.9

Table 4

The Intercorrelations among 21 Teaching Behaviors as Rated by Four Observers

Item 1 - Clarity$_1$	Item 2 - Clarity$_2$	Item 3 - Enthusiasm$_1$
Item 12 .51 Item 19 .58	Item 3 .59 Item 7 .74	Item 2 .59 Item 11 .51
	4 .54 8 .70	4 .82 12 .55
	5 .81 16 .54	5 .58 14 .52
	6 .77 19 .50	6 .53 16 .69
		7 .56 17 .57
		8 .70 18 .57
		10 .50 19 .51

Item 4 - Enthusiasm$_2$	Item 5 - Organization-Structuring$_1$	Item 6 - Organization-Structuring$_2$
Item 2 .54 Item 11 .62	Item 2 .81 Item 8 .75	Item 2 .77 Item 5 .76
3 .82 12 .61	3 .58 10 .53	3 .53 8 .81
5 .67 15 .56	4 .67 16 .57	4 .60 16 .59
6 .60 16 .66	6 .93 17 .50	5 .93 19 .56
7 .61 17 .60	7 .80 19 .50	
8 .65 18 .60		
10 .59 19 .60		

Item 7 - Knowledge of Subject Matter$_1$	Item 8 - Knowledge of Subject Matter$_2$	Item 9 - Interaction$_1$
Item 2 .74 Item 5 .80	Item 2 .70 Item 7 .69	Item 10 .58 Item 13 .54
3 .56 6 .76	3 .70 13 .52	
4 .61 8 .69	4 .65 16 .70	
Item 12 .56	5 .75 18 .57	
	6 .81 19 .73	

Item 10 - Interaction$_2$	Item 11 - Support$_1$	Item 12 - Support$_2$
Item 3 .50 Item 5 .53	Item 3 .51 Item 10 .67	Item 1 .51 Item 7 .56
4 .59 9 .58	4 .62 12 .65	3 .55 11 .65
Item 11 .67	Item 15 .70	4 .61 15 .51
		Item 17 .52

Item 13 - Questioning Tactics$_1$	Item 14 - Questioning Tactics$_2$	Item 15 - Questioning Tactics$_3$
Item 8 .52 Item 9 .54	Item 3 5.2	Item 4 .56 Item 11 .71
Item 16 .601		Item 12 .51

Item 16 - Variability$_1$	Item 17 - Variability$_2$	Item 18 - Stage Presence$_1$
Item 2 .54 Item 8 .70	Item 3 .57 Item 5 .50	Item 1 .58 Item 8 .57
3 .69 13 .60	4 .60 12 .52	3 .57 16 .62
4 .66 17 .64	Item 16 .64	4 .60 19 .53
5 .57 18 .62		
6 .59 19 .61		

Item 19 - Stage Presence$_2$	Item 20 - Use of Visuals$_1$	Item 21 - Use of Visuals$_2$
Item 3 .51 Item 6 .56	- No Data -	Item 19 .53
4 .60 8 .73		
5 .51 16 .61		
Item 21 .531		

Note: Only correlations $\geq$.50 are cited in this table. The inverses of correlations are accumulated as the listing progresses.

Table 5

Correlations between Grouped Students and Grouped Trained Observers' Mean Ratings of Lecturing Behavior across Three Samples of Twenty Instructors' Lectures for each of Twenty-one Variables.

Item	Variables	Correlations	Item	Variables	Correlations
1	Clarity$_1$	.88**	11	Support$_1$	.57**
2	Clarity$_2$	.59**	12	Support$_2$	.51**
3	Enthusiasm$_1$	.67**	13	Questioning$_1$	.48*
4	Enthusiasm$_2$	.59**	14	Questioning$_2$	.41*
5	Organization-Structuring$_1$	.43*	15	Questioning$_3$	.52**
6	Organization Structuring$_2$	.33	16	Variability$_1$	.43*
7	Knowledge of Subject$_1$	.44*	17	Variability$_2$	.52.**
8	Knowledge of Subject$_2$	.49**	18	Stage Presence$_1$	.49**
9	Interaction$_1$	.43*	19	Stage Presence$_2$	.02
10	Interaction$_2$	.47*	20	Use of Visuals$_1$	.49**
			21	Use of Visuals$_2$	.37

**p ≤ .01
*p ≤ .05

References

Burnett, C. W. Relating to learners. Improving College and University Teaching, 1974, 22(1), 49-52.

Costin, F., Greenough, W.T. and Menges, R.J. Student ratings of college teaching: Reliability, validity and usefulness. Review of Educational Research, 1971, 41(5), 511-535.

Dubin, R. and Taveggia, T. C. The teaching-learning paradox: A comparative analysis of college teaching methods. Eugene: Center for the Advanced Study of Educational Administration, University of Oregon, 1968.

Dunkin, M. J. and Biddle, B. J. The study of teaching. New York: Holt, Rinehart and Winston, Inc., 1974.

Gage, N. L. and Berliner, D. C. Educational psychology. Chicago: Rand McNally College Publishing Company, 1976.

Gagne, R. M. The learning basis of teaching methods. In N. L. Gage (Ed.), The psychology of teaching methods. Chicago: University of Chicago Press, 1976.

Gagne, R. M. and Rohwer, W. D. Instructional psychology. In P. H. Mussen and W. E. Rosenzweig (Eds.) Annual Review of Psychology. Palo Alto: Annual Reviews Inc., 1969.

Hildebrand, M., Wilson, R. C. and Dienst, E. R. Evaluating university teaching. Berkeley: Center for Research and Development in Higher Education, University of California, 1971.

Joyce, B. and Weil, M. Models of teaching. Englewood Cliffs: Prentice-Hall, Inc., 1972.

McKeachie, W. J. Teaching tips - A guidebook for the beginning college teacher. Lexington: D. C. Heath and Co., 1969.

McKeachie, W. J. and Kulik, J. A. Effective college teaching. In F. N. Kerlinger (ed.), Review of research in education (Vol. 3). Itasca, Ill: F. E. Peacock Publishers, Inc., 1975.

McKeachie, W. J., Lin, Y. G. and Mann, W. Ratings of teacher effectiveness: Validity studies. American Educational Research Journal, 1971, 8(3), 435-45.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., and Bent, D. H. Statistical package for the social sciences. New York: McGraw-Hill Book Company, 1975

Nuthall, G. and Snook, I. Contemporary models of teaching. In R. M. W. Travers (Ed.), Second handbook of research on teaching. Chicago; Rand McNally Publishing Co., Inc., 1973.

Rosenshine, B. and Furst, N. Research on teacher performance criteria. In B. O. Smith (Ed.) Research in teacher education: A symposium. Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Rowley, G. L. The reliability of observational measures. American Educational Research Journal, 1976, 13(1), 51-59.

Shavelson, R. and Dempsey, N. Generalizability of measures of teaching behavior. Review of Educational Research, 1976, 46(4), 553-611.

Whitely, S.E. and Doyle, K. O. Implicit theories in student ratings. American Educational Research Journal, 1976, 13(4), 241-253.

Wilson, R. C., Gaff, J. G., Dienst, E. R., Wood, L. and Bavry, J. L. College professors and their impact on students. New York: Wiley-Interscience Publication, 1975.