DOCUMENT RESUME

ED 039 860 24 HE 001 603

AUTHOR	Hildebrand, Milton; Wilson, Robert C.
TITLE	Effective University Teaching and its Evaluation.
INSTITUTION	California Univ., Berkeley. Center for Research and Development in Higher Education.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-5-0248
PUB DATE	70
GRANT	OEG-6-10-106
NOTE	41p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.15
DESCRIPTORS	College Faculty, *Effective Teaching, *Faculty, *Higher Education, Student Opinion, Surveys, Teacher Characteristics, *Teacher Evaluation, *Teaching Quality

ABSTRACT

The major purpose of this study was to characterize effective teaching performance and provide a basis for the evaluation of teaching. Four surveys were conducted in which: (1) students described their best and worst teachers; (2) faculty described the teaching of colleagues they regarded as the best and worst instructors; (3) faculty described the ways they distributed their time among various academic pursuits; and (4) students independently described the teaching of instructors previously rated by other students and faculty. Results indicated: (1) agreement among students, and between faculty and students about the effectiveness of given teachers; (2) best and worst teachers engage in same professional activities, and allocate their time among academic pursuits the same way; (3) student rating of best teachers showed only negligible correlation with academic rank; and (4) a disproportionate number of best teachers taught seminars rather than lecture courses. Eighty-five items, which can be divided in 5 components of effective teaching, characterized best teachers as perceived by students; 54 items, also producing 5 scales, characterized best teachers as perceived by faculty. Nine types of effective teaching were identified using these scales. (AF)

Center for Research and Development in Higher Education

University of California, Berkeley

EFFECTIVE UNIVERSITY TEACHING AND ITS EVALUATION

Milton Hildebrand

Robert C. Wilson

1970

The research reported herein was supported by Grant No. OE-6-10-106, Project No. 5-0248-20-6, with the Office of Education, U. S. Department of Health, Education, and Welfare, under the provision of the Cooperative Research Program. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

EFFECTIVE UNIVERSITY TEACHING AND ITS EVALUATION

Report to the faculty on a study of teaching and the evaluation of teaching sponsored by the Academic Senate of the Davis Campus of the University of California (1, 2)

Milton Hildebrand and Robert C. Wilson
Codirectors (3, 4)

Advisory Committee: H. L. Alder, R. W. Hoermann, R. M. Johnson, M. P. Oettinger, M. Regan, G. D. Yonge, and P. E. Zinner

April, 1970

TABLE OF CONTENTS

SUMMARY

Our objective is to contribute to the improvement of teaching at the university by characterizing effective performance and providing a satisfactory basis for the evaluation of teaching. (For further description of the objective see p. 5 and related discussion starting on p. 19.)

Four surveys were conducted: 338 students described the teaching of teachers they regarded as their best and worst instructors; 119 members of the faculty described the teaching of colleagues they regarded as best and worst instructors; 162 members of the faculty described the ways they distributed their time among various academic pursuits; and 1015 students independently described the teaching of instructors previously rated by other students and/or faculty as to effectiveness of teaching. (For further description of the collection of data see p. 6.)

The principal results follow. For further description and justification, see the body of the report, pp. ' to 19.

1. There is excellent agreement among students, and between faculty and students, about the effectiveness of given teachers.

2. Best and worst teachers engage in the same professional activities and allocate their time among academic pursuits in about the same ways. The mere performance of activities associated with teaching does not assure that the instruction is effective.

3. Eighty-five items are listed that characterize best teachers as perceived by students (Appendix C), and 54 items are listed that characterize best teachers as perceived by colleagues (Appendix D). All items statistically discriminate best from worst teachers with a high level of significance.

4. Analysis of the items characterizing best teachers as perceived by students produced five scales, or components of effective performance (Table 1). Our conceptual interpretations of the scales are indicated by the headings assigned:

1) Analytic/Synthetic Approach
2) Organization/Clarity
3) Instructor-Group Interaction
4) Instructor-Individual Student Interaction
5) Dynamism/Enthusiasm

5. Analysis of the items characterizing best teachers as perceived by colleagues produced five scales (Table 2). They are described as follows:

1) Research Activity and Recognition
2) Intellectual Breadth
3) Participation in the Academic Community
4) Relations with Students
5) Concern for Teaching

We believe that the colleague scales are somewhat less useful and valid than the student scales as a basis for evaluating teachers (see discussion starting on p. 20).

6. The student scales were derived from a 1967 survey. A single summary description was phrased to express the nature of the component of effective teaching identified by the items composing each scale. Respondents to the 1968 survey rated their teachers on each of the five summary descriptions and also on each of the items from which the scales had been derived. Correlations of mean scores on the summary descriptions with mean scores on the full lists of respective items were very high. Thus, the five summary descriptions provide the basis for a short evaluation form demonstrated to be broad and highly discriminating.

7. In general, student ratings of best teachers showed only negligible correlations with academic rank of instructor, class level, number of courses previously taken in the same department, class size, required versus optional course, course in major or not, sex of respondent, class level of respondent, grade-point average, and expected grade in course.

8. A disproportionate number of best teachers were teaching seminar rather than lecture courses, and a wide range of excellence was revealed in the teaching of different subject areas.

9. Seventeen items describing the college goals of students were sorted into three scales (see Appendix H):

1) Upward Mobility/Security
2) Self-Knowledge/Humanism
3) Career/Subject Mastery

10. Thirteen items describing objectives of teaching as perceived by students were sorted into two scales (see Appendix I):

1) Contribution to General Development
2) Transmission of Fundamentals

11. Students evaluated the positive contributions made to their lives by best teachers in six areas: knowledge imparted, counsel given, objectives clarified, values developed, incentive elicited, and skills developed. Correlations of mean scores for these areas with mean scores for the components of effective teaching and with overall ratings of effectiveness of teaching are high (see Appendix J).

12. Nine types of effective teachers were identified by analyzing individual patterns of relatively high and low scores on the five components of effective teaching. Overall ratings of teachers having the various patterns correlate with certain course and student variables. The analysis is not presented in numerical form, but four examples are given (see p. 18).

13. Teachers rated as excellent by some observers and as poor by others are less even in their performance of the five components of effective teaching than are best teachers.

OBJECTIVE

Our general objective is to enable the Academic Senate ro provide guidance and incentive that can help to improve teaching.

One major aim is to identify and describe effective teaching so that instructors can be helped to improve, and so that the much neglected field of preparing graduate students for the teaching function of academic life may be benefited (5, 6). Articles allegedly describing good teaching are numerous, and many are sound, but most either largely represent the subjective judgment of individuals and committees, or are based on studies using small samples in restricted circumstances. Faculty members feel that there are different ways of teaching well (7, 8, 9, 10, 11), and believe that they disagree somewhat in the assessment of teaching. Hence, many instructors remain skeptical of advice offered. Reliable characterization of effective teaching is needed.

The other major aim is to find more valid, reliable, and effective means of incorporating the evaluation of teaching into advancement procedures. We believe this to be the most important single requirement for the improvement of university teaching. The incentive thereby provided will encourage instructors to devote the study, time, and effort necessary to do their best (12), and the stature of teaching will increase.

Eighty-seven percent of 303 members of the Davis faculty (and 92% of similar samples at six diverse colleges and universities) stated in 1968 that effectiveness of teaching <u>should be</u> "Quite important," or "Very important" as a criterion for advancement (13). Only 24% of the Davis sample (and 38% of the total sample) stated that effectiveness of teaching actually <u>is</u> "Quite important," or "Very important." Seventy-one percent of the Davis faculty (and 72% of the total sample) stated that the campus should have a formal procedure for evaluating teaching.

The University of California's system of review for advancement is as good as any we know (14), and since the revision, in Sept., 1969, of "Instructions to Appointment and Promotion Committees," greater efforts have been made to evaluate teaching. Nevertheless, procedures are largely unstandardized and untested, and hence do not adequately evaluate the teaching of the majority of the faculty. Consequently, here as elsewhere (15), "research and creative work" usually has outweighed quality of teaching as a criterion for advancement.

At Davis, as at most institutions (15), the dossier furnished by the department chairman to support promotion has been of the utmost importance, yet there are inherent weaknesses in a system that places great weight on evaluations of teaching as traditionally prepared by chairmen (or deans): A chairman may himself be doubtfully qualified as a judge of teaching. Opinions solicited from his staff may be biased, may not constitute an adequate sample, and often are in part secondhand. Most available measures

of involvement in teaching (e.g., number of courses taught, enrollments, number of advisees) do not necessarily correlate with quality of instruction. Classroom visitations are resisted or resented by most teachers, and hence are seldom utilized although considered by many administrators to be the most important element in evaluation (15). If a department is large, the chairman cannot visit any class more than once or twice, which is enough to judge certain elements of effective teaching but insufficient to make a comprehensive judgment. Classroom instruction is only part of the teaching function.

We believe that promotion letters cannot be improved sufficiently to achieve our objective unless new procedures assure that they include more thorough, more objective, and more comparable evaluations of teaching than have been usual in the past.

DEVELOPMENT OF TEACHER-DESCRIPTION SCALES

Collection of Data

Three questionnaires were distributed on the Davis campus in May, 1967, and one in May, 1968. A random sample of all students was asked to complete the first questionnaire. The 278 undergraduate and 60 graduate students who responded constituted 4% of the student body and 38% of those approached. Bias may have been introduced by self-selection. However, the sample was evenly divided between the sexes, it did not differ significantly from the population in regard to distribution by class level or major area (except perhaps for moderate over representation in the humanities), and the mean of the overall grade point averages of students in the sample was identical to the grade point average of the population for the spring, 1967 quarter. Respondents supplied biographical information and their academic backgrounds, and answered questions on their college goals, on objectives of teaching, and on the teaching of instructors identified by them as the best and worst they had had in the previous year (16). Assurance was given that the identity of teachers would be kept in strict confidence, even from Davis members of the research team.

The second questionnaire was returned by 119 of the faculty, which was 54% of the random sample approached and 21% of the resident teaching faculty. The respondent was asked to identify a best and a worst teacher among his colleagues and to answer, for each, questions about teaching activities observed outside the classroom, about in-class behavior, and about the presentation of talks and seminars.

The third questionnaire was returned by 162 members of the faculty who were not asked to complete the previous questionnaire. This was 80% of the random sample approached and 29% of the resident teaching faculty. Questions related to distribution of time among various academic pursuits.

Lastly, as a follow-up and validation study, a fourth questionnaire was distributed in 1968 to all students in 51 classes. These were selected

to include, in about equal numbers, those of instructors identified in 1967 as best teachers by three or more students or colleagues, those of instructors identified as worst teachers, and classes of another group not previously identified as either best or worst and presumed to include teachers of intermediate effectiveness. The 1015 respondents again provided biographical data, and answered questions on their college goals, on various objectives of teaching, and on the teaching of the given instructor. Overall ratings of the teachers were also secured.

Identification of Effective Teachers

It is important to know whether the various segments of the academic community concur in their identifications of the most effective and ineffective teachers.

We identified the instructors receiving three or more nominations as best teachers, and those receiving three or more nominations as worst teachers by the student respondents to our 1967 survey. In an independent study at Davis by Regan and Yonge (17), 57 of the same teachers were named by students as being particularly excellent or poor. Appendix A shows the very high degree of agreement between the two surveys: the chi square value indicates a level of significance of $p < .0005$ (that is, fewer than 5 chances in 10,000 that the observed result is fortuitous). This result indicates that the two groups of students probably used closely similar criteria of judgment. Since the Yonge-Regan study had a 90% return, we consider this to be indirect evidence that self-selection by our respondents did not introduce significant bias into the designation by them of their best and worst teachers.

Further, in our 1968 survey, all students of 15 instructors named in 1967 by three or more students as best teachers (18), all students of 18 instructors similarly named previously as worst teachers, and all students of 18 instructors not previously nominated as best or worst teachers, rated the excellence of their instructors. Ratings were along a seven-point continuum from "Among the very worst" to "Among the very best." Differences among the mean scores for the best, not nominated, and worst teachers of the previous year were all significant well below the .01 level (19).

Finally, each of 119 faculty respondents identified colleagues they considered outstanding and poor teachers. Of those named, 66 were common to the 1967 student sample. Appendix B shows the very high agreement between the two groups; again, $p < .0005$.

That there are any teachers nominated by some observers as best and by others as worst is of concern. This circumstance may result in part from a differing exposure of the respondents to individual faculty: the worst teacher observed by one student in the previous year might indeed be the best teacher observed by another student in the same period, even if the criteria of judgment were the same for the two students. There is, however, another interpretation for these few split nominations, to which we shall return at the end of this report.

Having learned that there is excellent agreement among students, and between faculty and students, about the effectiveness of given teachers, we proceed to the characterization of effective teaching.

Characterization by Students

Of the members of the Davis faculty who believed in 1968 that the campus should have a formal procedure for evaluating teaching, 86% believed that students should participate in the evaluations (13).

The student respondents to our 1967 survey stated whether each of 158 descriptions of aspects of teaching (or "items") was characteristic for the instructors they named as their best and their worst teachers of the year (20). Answers were "Yes," "No," and "Does not apply or don't know." The respondents to our 1968 survey stated whether most of the same items (and some new ones) were descriptive of their teachers, this time using a 4-point scale ranging from "Not at all descriptive" to "Very descriptive." We deemed it important that our surveys range widely over all general aspects of teaching. Items were drawn from the experience of the research staff and faculty advisory committee, and from studies by twelve other investigators (21).

Appendix C lists 85 of the 158 items to which at least 75% of respondents could answer "Yes" or "No," and which discriminate between best and worst teachers with the very high significance level of $p < .001$. For tabulation here, many of the items are somewhat condensed.

This table goes far to provide a description of fine teaching--our first major objective The included items are not equally useful, however, for making comparative evaluations of teaching. The general level of competence of instruction at Davis appears to be good, and students and colleagues both tend to rate instructors generously (22, 23, 24). Accordingly, items that discriminate at the top are particularly useful. When teachers in general are rated on selected items, it is desirable that the distributions of scores not be skewed so that there are many more high than low scores. Items 1 through 60 of Appendix C meet this requirement better than the remaining items. Asterisks and daggers mark the most discriminating items (see footnote to the appendix), with those marked by asterisks also providing the least skewed distributions of scores.

Some items (numbers 61 through 78 of Appendix C) are characteristic of a majority of <u>both</u> best and worst teachers (though sufficiently more typical of best teachers to discriminate at below the .001 level of significance). If teachers in general were rated on such items, one would expect the distributions to be markedly skewed: if an item were not descriptive of a given teacher, his teaching would probably not be effective in that regard, but if the item <u>were</u> descriptive, his teaching might still be <u>relatively</u> ineffective (on this campus at least). (Examination of the items gives confidence that even our "worst" teachers are competent in many respects.) To use such items for evaluation is

equivalent to giving an easy quiz to a class of variable but generally high excellence: all students earn 100% scores except the few already known to be at the foot of the class. A department chairman who wished to write, in a promotion letter, nice things about a relatively mediocre teacher could probably select several such items.

A smaller category comprises items (not included in Appendix C) characteristic of only a minority of best and worst teachers, yet less typical of best teachers to the extent that $p < .001$. Examples are: Has distracting mannerisms. Emphasizes grades. Gives ambiguous examinations.

Nondiscriminating items should be excluded from evaluation forms (though they may be useful for other purposes, such as the selection of teachers by students). Noteworthy among items found not to distinguish best from worst teachers, even at the comparatively low .05 level of significance, are these: Gives difficult examinations. Gives difficult assignments. Spends much of his time on research or projects other than teaching. Grades leniently. Grades subjectively. It is important to note that these responses, and many items listed in Appendix C (e.g., numbers 5, 20, 39, 64, 66, 67, 71, 80, 83, and particularly 81) give confidence that students do not equate "best" teachers with "easy" teachers.

Questions to which many students are unable to reply are of limited value for evaluating teachers, particularly when classes are small. The following are representative of items that discriminate best from worst teachers, but to which at least 25% of our respondents could not reply: Is always in his office during scheduled office hours. Puts me at ease when I visit him. Is involved in campus activities that affect students. Learns students' names promptly. Is well known in his field. Spends extra time with students having difficulty.

Some items discriminate best from worst teachers if ratings are by undergraduate students but not if ratings are by graduate students (25). We believe that the difference results largely from the nature of graduate instruction and the greater professional orientation and self-motivation on the part of graduate students.

Characterization by Colleagues

Of the members of the Davis faculty who believed in 1968 that the campus should have a formal procedure for evaluating teaching, 85% believed that colleagues should participate in the evaluations (13). Despite the importance attached to colleague judgments, we found no research that explored in depth the considerations which enter into such judgments (26).

Of our faculty respondents, 119 stated, for colleagues named as the most and least effective teachers known to them, whether each of 103 descriptions of aspects of teaching and other academic activities was

characteristic. Answers were "Yes," "No," and "Does not apply or don't know." Appendix D lists 54 items to which at least 66% of respondents could answer "Yes" or "No," and which discriminate between best and worst teachers with a significance level of $p < .001$. Many of the items are condensed somewhat for tabulation here. This appendix supplements Appendix C in characterizing the behavior of fine teachers.

The item Publishes frequently, is discriminating for best teachers at the .05 significance level. Noteworthy among items found not to be discriminating are these: Spends much of his time on research or projects other than teaching. Attends faculty social functions. Expresses concern about pressures to publish.

Of the numerous items to which more than a third of our colleague respondents replied "Does not apply or don't know," most related to instructor-student interaction.

As another part of the study, a random sample of 162 of the faculty was asked to state how often various functions of teaching, research, university and community service, consultation, and related academic pursuits had been performed in stated time periods Of all respondents, 38 had been named as best teachers and 32 as worst teachers by students or colleagues on the independent surveys already described. When the self-descriptions of the best and worst teachers were compared, remarkably little difference was found. Only two of the 143 items (Met informally with students outside of class or office and Talked with a colleague about my research) discriminate between effective and ineffective teachers below the .05 level of significance. None of the other results was of statistical significance (27). The following are examples of nondiscriminating items: Reviewed lecture notes. Revised a lecture. Prepared demonstration material for a class. Did background reading for a course. Graded examination papers. Helped students with individual projects (28). It appears that within our limits of discrimination, the more and the less effective teachers at Davis do the same general things with their time. Involvement with teaching on the part of candidates for promotion is a proper consideration in a recommendation report, but the mere performance of activities associated with teaching does not of itself assure that the instruction is effective.

Together, the items in Appendixes C and D give a picture of fine teaching as defined by students and colleagues. At the same time, the list of items is long, is miscellaneous in character, and does not fully characterize effective teaching in a conceptual manner. Further analysis is necessary.

Components of Effective Teaching

Many researchers (29) have sorted individual items describing aspects of effective teaching into related groups, thus identifying basic components,

dimensions, or scales of such teaching. Teacher-rating forms developed by students commonly do the same. Scales have been variously determined by subjective examination of a list of items or by factor analysis, which establishes mathematically the tendency of responses to the various items to associate in clusters. The number of scales developed in reports we have seen ranges from 4 to 13. Nevertheless, 4 to 5 particular scales (i.e., knowledge, presentation, relation with students, enthusiasm) appear rather consistently, even though the terminology differs. Our scales are generally consistent with those of previous studies.

Scales for student characterization of effective teaching were established by factor analysis of 91 items (30) describing the teaching of 338 best teachers as identified by respondents to our 1967 survey. (Our method (31) was a principal-components analysis with a varimax rotation (32).) After several analyses, a five-factor solution was selected as giving the maximum number of distinct and interpretable components of effective teaching. Items having factor coefficients (which show the tendency of an item to be associated with a particular scale) greater than .40 were retained and analyzed further (by pre-set cluster analysis (33)) to determine the consistency and reliability of the scales and their intercorrelations (34). The items were then re-analyzed with data from our 1968 validation survey. The five scales held together very well; the alpha reliabilities (showing internal consistency) range from .80 to .89 (35).

Table 1 presents the five scales and the included items. The factor coefficients from the 1968 survey are listed (36). Coefficients of .40 and higher are generally considered good; our cutoff value is .43. No item appears in more than one scale. Our conceptual interpretations of the scales are as follows:

Scale 1, Analytic/Synthetic Approach, is scholarship, with emphasis on breadth, analytic ability, and conceptual understanding.

Scale 2, Organization/Clarity, is skill at presentation, but is subject-related, not student-related, and is not merely rhetorical skill.

Scale 3, Instructor-Group Interaction, is rapport with the class as a whole, sensitivity to class response, and skill at securing active class participation.

Scale 4, Instructor-Individual Student Interaction, is mutual respect and rapport between the instructor and the individual student.

Scale 5, Dynamism/Enthusiasm, is the flare and infectious enthusiasm that comes with confidence, excitement for the subject, and pleasure in teaching.

Responses describing the performance of worst teachers were also subjected to factor analysis, but the results did not provide readily interpretable scales. The items showed less consistent relationships

than they did for best teachers. Thus, ineffective teachers were characterized best by their lack of attributes associated with effective teaching, rather than by possession of attributes of poor teaching.

Scales for the characterization of effective teachers by colleagues were prepared by factor analysis of 67 items describing the behavior of 84 best teachers identified by 119 members of the faculty. Items requiring attendance of the respondent at classroom instruction and seminars of the identified teacher (numbers 30 through 45 of Appendix D) were not factored, because many colleagues (51 and 17%, respectively) had not observed those activities (37).

Table 2 presents the five scales, which were established by the same method of factor analysis as for the student data. The factor coefficients of the included items are listed, the cutoff value being .37. Alpha reliabilities range from .65 to .86. Intercorrelations among the scales are low or negligible (38). Our conceptual interpretations of the scales are indicated by the headings assigned:

Scale 1. Research Activity and Recognition

Scale 2. Intellectual Breadth

Scale 3. Participation in the Academic Community

Scale 4. Relations with Students

Scale 5. Concern for Teaching

Utility of the Scales

The scales derived from the characterization of effective teaching by students provide conceptual understanding of the components of such teaching. Having been developed from items to which most students of a large random sample could respond, the student scales are applicable to most kinds of university-level teaching. Attention to the scales helps to assure that the major components of effective performance are considered when teaching or evaluating teaching. Many of the rating forms in our files (assembled from various campuses) fail in this regard.

In order to learn if we could develop an effective short evaluation form, we phrased a summary description for each of the student scales derived from the 1967 survey. Each description was intended to express the component of effective teaching identified by the items composing that scale. The 1968 survey asked respondents to rate their teachers on each of the five summary descriptions. It also repeated the full set of original items from which the scales had been established. Correlations of mean scores on the summary descriptions with mean scores on the full list of respective items (N = 51) were very high (coefficients ranging

TABLE 1. COMPONENTS OF EFFECTIVE TEACHING AS PERCEIVED BY STUDENTS

		Factor coefficient
SCALE 1. ANALYTIC/SYNTHETIC APPROACH		
1.	Discusses points of view other than his own	.70
2.	Contrasts implications of various theories	.66
3.	Discusses recent developments in the field	.64
4.	Presents origins of ideas and concepts	.60
5.	Gives references for more interesting and involved points	.53
6.	Presents facts and concepts from related fields	.53
7.	Emphasizes conceptual understanding	.46
SCALE 2. ORGANIZATION/CLARITY		
8.	Explains clearly	.78
9.	Is well prepared	.63
10.	Gives lectures that are easy to outline	.62
11.	Is careful and precise in answering questions	.61
12.	Summarizes major points	.51
13.	States objectives for each class session	.50
14.	Identifies what he considers important	.47
SCALE 3. INSTRUCTOR-GROUP INTERACTION		
15.	Encourages class discussion	.70
16.	Invites students to share their knowledge and experiences	.65
17.	Clarifies thinking by identifying reasons for questions	.64
18.	Invites criticism of his own ideas	.62
19.	Knows if the class is understanding him or not	.58
20.	Knows when students are bored or confused	.57
21.	Has interest and concern in the quality of his teaching	.48
22.	Has students apply concepts to demonstrate understanding	.43
SCALE 4. INSTRUCTOR-INDIVIDUAL STUDENT INTERACTION		
23.	Has a genuine interest in students	.74
24.	Is friendly toward students	.71
25.	Relates to students as individuals	.69
26.	Recognizes and greets students out of class	.68
27.	Is accessible to students out of class	.65
28.	Is valued for advice not directly related to the course	.64
29.	Respects students as persons	.60
SCALE 5. DYNAMISM/ENTHUSIASM		
30.	Is a dynamic and energetic person	.80
31.	Has an interesting style of presentation	.76
32.	Seems to enjoy teaching	.74
33.	Is enthusiastic about his subject	.65
34.	Seems to have self-confidence	.64
35.	Varies the speed and tone of his voice	.63
36.	Has a sense of humor	.53

Based on the 1968 survey. N = 1015

TABLE 2. COMPONENTS OF THE ACTIVITIES OF EFFECTIVE TEACHERS AS PERCEIVED BY COLLEAGUES

		Factor coefficient
	SCALE 1. RESEARCH ACTIVITY AND RECOGNITION	
1.	Does work that receives serious attention from others	.69
2.	Corresponds with others about his research	.69
3.	Does original and creative work	.64
4.	Expresses interest in the research of his colleagues	.55
5.	Gives many papers at conferences	.55
6.	Keeps current with developments in his field	.49
7.	Has done work to which I refer in teaching	.48
8.	Has talked with me about his research	.38
	SCALE 2. INTELLECTUAL BREADTH	
9.	Seems well read beyond the subject he teaches	.66
10.	Is sought by others for advice on research	.60
11.	Can suggest reading in any area of his general field	.59
12.	Knows about developments in fields other than his own	.51
13.	Is sought by colleagues for advice on academic matters	.43
	SCALE 3. PARTICIPATION IN THE ACADEMIC COMMUNITY	
14.	Encourages students to talk with him on matters of concern	.60
15.	Is involved in campus activities that affect students	.58
16.	Attends many lectures and other events on campus	.47
17.	Has a congenial relationship with colleagues	.39
	SCALE 4. RELATIONS WITH STUDENTS	
18.	Meets with students informally out of class	.58
19.	Is conscientious about keeping appointments with students	.57
20.	Meets with students out of regular office hours	.57
21.	Encourages students to talk with him on matters of concern	.55
22.	Recognizes and greets students out of class	.37
	SCALE 5. CONCERN FOR TEACHING	
23.	Seeks advice from others about the courses he teaches	.70
24.	Discusses teaching in general with colleagues	.60
25.	Does not seek close friendships with colleagues (Negative)	-.47
26.	Is someone with whom I have discussed my teaching	.45
27.	Is interested in and informed about the work of colleagues	.44
28.	Expresses interest and concern about the quality of his teaching	.40

N = 119

from .88 to .96). Thus, a short-form rating instrument is established that is quickly answered, yet is objectively known to be broad, balanced, and highly discriminating between effective and ineffective teachers.

The five summary descriptions that we recommend follow (39). We suggest that if these scales are used for evaluations, the respondent be asked to use a seven-point continuum (40) ranging from "Low score" to "High score."

1. Has command of the subject, presents material in an analytic way, contrasts various points of view, discusses current developments, and relates topics to other areas of knowledge.

2. Makes himself clear, states objectives, summarizes major points, presents material in an organized manner, and provides emphasis.

3. Is sensitive to the response of the class, encourages student participation, and welcomes questions and discussion.

4. Is available to and friendly toward students, is interested in students as individuals, is himself respected as a person, and is valued for advice not directly related to the course.

5. Enjoys teaching, is enthusiastic about his subject, makes the course exciting, and has self-confidence.

For reasons explained later in this report, we believe that colleague Scale 5, Concern for Teaching, is the most useful of the colleague scales. This is a summary description of that scale: Expresses interest and concern in the quality of his teaching, discusses teaching in general with his colleagues, seeks advice regarding his teaching, and is sought by others for counsel on their teaching.

Respondents to the 1968 student survey made a single overall rating of the effectiveness of their teachers on a continuum of 1 to 7. Appendix E shows the correlations between the overall rating of effectiveness and the five separate summary descriptions. Scale 5, Dynamism/Enthusiasm, is the most highly related with the teachers named as best, and Scale 2, Organization/Clarity, is in second place. For all the correlations, $p < .001$.

The utility of the five scales for discriminating best from worst teachers is shown in another way. Each teacher named in the 1967 student survey was given a score for each scale based on the total number of contained items stated to be descriptive of his performance. The scores for each scale were then converted so that the mean score for all teachers is 50 and the standard deviation is 10. Appendix F shows frequency distributions for the converted scores of best and of worst teachers.

Similarly, Appendix G presents the percentages of best and of worst teachers that fall within each range of the converted scores. These percentages can be interpreted as the probabilities that any teacher with a given score would be nominated by students as a best or a worst teacher.

The scales are stressed because they have greater utility and conceptual value than do the individual items. Even so, they do not include all of the useful data; some discriminating items do not cluster sufficiently with others to fall in any scale. Even a short evaluation form might well supplement the five summary descriptions with selections from items of this kind (i.e., items from Appendix C that do not also appear in Table 1).

RELATION OF RATINGS OF TEACHERS TO THE COURSE AND THE STUDENT

Course and Student Characteristics

It is important to know what variables significantly affect teaching and student ratings of teachers. Our overall ratings of effectiveness of teaching from the 1968 survey were correlated with: 1) academic rank of teacher; 2) course level; 3) number of courses previously taken in the same department; 4) class size; 5) required versus optional course; and 6) course in the major or not. The highest correlation of any of these six variables with rated quality of teaching is .06, which is negligible. However, since our samples are large (41), statistical significance is achieved with a very small correlation: correlations bordering on the .05 level of significance were found for the last two variables listed. These data confirm results of Solomon (42) in regard to class size and of Guthrie (43) in regard to academic rank. They are partly in disagreement with a survey at the University of Illinois noted by Cohen and Brawer (44) in regard to class size.

Although the variables listed above are seen not to significantly bias overall ratings of effectiveness of teaching, they might be expected to influence the characteristics of teaching. The six variables were, therefore, correlated with the scores assigned to teachers for each of the five student description scales of components of effective teaching. Of the 30 elements of the matrix, only 5 coefficients are high enough ($\pm$ .20 to $\pm$ .30) to establish definite but small correlation: Scale 4 (Instructor-Individual Student Interaction) correlates positively with higher level of course, smaller class size, and course in the major; Scale 1 (Analytic/Synthetic Approach) correlates positively with higher level of course; and Scale 3 (Instructor-group Interaction) correlates positively with smaller class size (45). For 18 elements of the matrix, $p < .01$.

Turning to variables related more directly to the student, we correlated the 1015 overall ratings of teachers with 1) sex of student; 2) class level of student; 3) grade-point average; and 4) expected grade in course. All correlations are negligible (highest coefficient .09),

though female sex and high expected grade in course correlate positively with high rating at just below and above the .01 level of significance. Cohen and Brawer (44) report similar results. Other studies have reported that there is a relation between expected grade and rating of teacher (23, 46), a relation only at lower class levels (47), and no relation (22, 48). These contradictions seem consistent with the presence of a definite but trifling correlation.

The four variables listed above were also correlated with scores for each of the five student description scales of effective teaching. Of the 20 elements of the matrix, only one coefficient is high enough (.24) that the correlation can be considered definite though small: Scale 4 (Instructor-Individual Student Interaction) correlates positively with higher class level of student. Half of the correlations are significant at the .01 level or better. The matrix indicates that high achievers and advanced students are slightly less dependent than their counterparts on organization and motivation supplied by the instructor, and also that female students respond slightly more than males to personal and group interaction with their (predominantly male) instructors. Other investigators have related grade-point average to the needs, responses, and motivation of students (49). The effects of authoritarianism, personality, and sex-related needs have also been studied (8, 9, 50, 51).

These results show that in general, the ten course and student characteristics listed do not significantly bias student ratings of teachers. Measuring is usually not needed for these variables; they might well be omitted from short evaulation forms. However, ratings of teachers having particular attributes may be somewhat influenced by certain of these variables (e.g., the personality of a particular teacher might tend to antagonize students of one sex more than the other). Analysis of the influence of course and student characteristics on teacher ratings may, therefore, help individual instructors to adapt to local circumstance. (See section below on matching students with teachers.)

Two other correlations proved to be more marked. When number of nominations for most and for least effective teachers ($N = 676$) were compared by subject areas (allowances being made for the sizes of the areas), differences were found which are significant at the .01 and .001 levels (52). Corresponding correlations by type of course presentation revealed proportionately more best teachers in seminar courses than in lecture courses ($p < .001$), with lecture-with laboratory courses being intermediate.

Goals of Students

Effective teaching cannot be fully studied without attention to the goals, perceptions, and values of students. We approached this subject in several ways. First, our 1967 student survey included 24 items on reasons for going to college (53). Responses were subjected to factor analysis and the results validated in 1968 following the procedures described above in the section on components of effective teaching. A three-scale solution

was selected having alpha reliabilities of .80, .81, and .81. Appendix H presents the scales and the 17 contained items having acceptable factor coefficients. Our interpretations of the scales are indicated by the headings selected: Scale 1, Upward Mobility/Security; Scale 2, Self-Knowledge/Humanism; and Scale 3, Career/Subject Mastery. Items that did not appear in the scales tend to relate to social pressure or apathy (54). Scale 1 has a low correlation with Scale 3; the other intercorrelations are negligible. Female sex has a low positive correlation with Scale 2.

As a second approach, 20 items on the perception by students of the objectives of teaching (53) were processed into two scales having alpha reliabilities of .83 and .84. Appendix I presents the scales, Contribution to General Development and Transmission of Fundamentals, and the 13 contained items. There is no interscale correlation. Female sex has a low positive correlation with the first scale named.

Relating the scales on college goals with those on objectives of teaching, Contribution to General Development has moderate correlation with Self-Knowledge/Humanism (coefficient .54). Transmission of Fundamentals has moderate correlation with Career/Subject Mastery and low correlation with Upward Mobility/Security (coefficients .47 and .34, respectively).

As a third approach to the goals, perceptions, and values of students, we asked respondents to the 1968 survey to rate their teachers on a seven-point continuum as to constructive contributions made to their lives in each of six areas. Appendix J shows correlations of the mean scores for these areas with mean scores for the components of effective teaching and with overall ratings of effectiveness of teaching.

Matching Students with Teachers

Correlations of both college goals and objectives of teaching with the components of effective teaching are low (55). This doubtless results in part from the fact that only ratings of best teachers were utilized in the calculations. These teachers rate sufficiently high on all components of effective teaching so that students having any goals and objectives find the attributes they admire. Nevertheless, we identified nine types of effective teachers by analyzing individual patterns of relatively high and low scores on the five components of effective teaching. Overall ratings of teachers having the various patterns were then correlated with course and student variables. The analysis is complicated by so many factors that we do not present results in numerical form lest the conclusions seem more exact than in fact they can be. The following two contrasting pairs of relationships are reported, however, to illustrate the concept of matching students with teachers.

Best teachers who were rated relatively high on Scale 4, Instructor-Individual Student Interaction, were particularly favored by female, upper-dividion, and graduate students having low Upward Mobility/Security, valuing Contribution to General Development, and majoring in the arts. The courses tended to be small lecture-with-laboratory classes. By

contrast, teachers who were rated relatively low on the same scale were particularly favored by female, lower division students having moderate Upward Mobility/Security, and valuing the Transmission of Fundamentals. The courses tended to be large lecture classes.

Best teachers who were rated relatively high on Scale 2, Organization/Clarity, were particularly favored by male, lower-division students having high Upward Mobility/Security, valuing the Transmission of Fundamentals, and majoring in the biological sciences. The courses tended to be large lecture or lecture-with-laboratory classes. By contrast, teachers who were rated relatively low on the same scale were particularly favored by female, senior students valuing Self-Knowledge/Humanism and Contribution to General Development, and majoring in the humanities. The classes were of various sizes and tended to be lecture courses.

It seemed probable that controversial teachers (rated as excellent by some observers and as poor by others) would be less even in their performance than best teachers: some students might accept relatively poor performance in a given component, whereas other students, having different goals and objectives, might not. To test this hypothesis, the within-individual variances between the converted (standardized) scores for each component of effective teaching and the mean converted score for all five components were calculated separately for 112 ratings of 32 best teachers and contrasted with those for 154 ratings of 48 controversial teachers. As predicted, the within-individual variances were greater for the latter group ($p < .01$). That is, controversial teachers have greater variation in the rated effectiveness of their performance of the five components of teaching than do best teachers. This explains, in part, their controversial status when rated by students having various goals, and indicates that it might be well for them to be matched with those students who are most inclined to value their assets. The procedures we recommend would help to identify those controversial teachers who are not well matched with students.

DISCUSSION

What is Effective Teaching?

Many persons consider teaching to be excellent in proportion to progress made by learners toward stated educational objectives (22, 50). We consider this concept to be generally sound. Except in restricted circumstances, however, the concept is difficult to apply to the characterization or evaluation of university teaching. For the present, there is insufficient agreement on the desired objectives, or on who should determine what objectives are desired. Even if specific objectives were accepted, it is unlikely that there could now be agreement on how to test progress toward the attainment of many of them. Information learned from teachers can be tested, but its value cannot; the contribution a teacher makes to spiritual or emotional maturation cannot easily be assessed.

Alternatively, teaching may be considered excellent in proportion to its constructive contribution to the life of the learner. The constructive contribution may be knowledge imparted, wisdom instilled, experience offered, counsel given, objectives clarified, human values developed, incentive and inspiration elicited, or skills developed. Effective teaching usually contributes to the life of the student in several ways according to the individual teacher-student relationship. The learner may not be able to fully assess the constructive contribution made to his life by a teacher, and his judgment may change with time. Nevertheless, the learner is often (or usually) the best judge of contributions made to his own life. In order for this concept of effective teaching to be generally applicable, different learners must tend to judge the same teachers as having made constructive contributions to them.

We included no definition of effective teaching in our questionnaires, leaving it to each respondent to select his best and worst teachers by his own criteria. Thus we have derived a descriptive definition of fine teaching as actually perceived by students and colleagues (Tables 1 and 2, and Appendixes C and D). The uniformity of judgment found in both the identification of best and worst teachers and in the characterication of best teaching, leads us to believe that this descriptive approach is not only practical, but also generally consistent with the views noted in the two preceding paragraphs. We question that those two views are as far apart in either application or theory as they may seem at first.

Other opinions, which we have not seriously considered, are that teaching should be judged primarily by students' increased ability to solve assigned problems (56); by out-of-classroom accomplishments (57), or by the academic prowess of former students (58).

Comparison of Evaluations by Students and Colleagues

Colleague Scales 1 (Research Activity and Recognition) and 2 (Intellectual Breadth) relate to scholarship as expressed in research. Results of this study support the widely recognized close relation between excellence in research and excellence in teaching. However, the former is questionably essential for establishing the latter provided that the teacher remains a scholar and applies his learning. Excellence in research is clearly not sufficient for establishing excellence in teaching, particularly at the undergraduate level. It is highly inappropriate, we believe, that at most institutions, research productivity is the primary consideration in evaluating teaching ability (5). We found that colleagues tend to rate full professors relatively high on Scale 1, doubtless because it takes time to establish a reputation for competence in research, even though professorial rank as such did not affect student or faculty ratings of teaching. Also a criterion for advancement at Davis, as at many universities, is "professional competence." Since measures of professional competence (e.g., positions held, honors received) are largely responses to reputation for research, not for teaching, beyond the home campus, research is, in effect,

counted another time. We believe that since excellence in research is considered separately as a criterion for advancement, it should specifically be eliminated in evaluating effectiveness of teaching. Hence, all of colleague Scale 1, and items 10 and 13 of Scale 2, should not be used for rating teaching. Student Scale 1 (Analytic/Synthetic Approach) is not equivalent to colleague Scales 1 and 2, but does also relate to scholarship; if this scale is used, scholarship will be considered as it is expressed in teaching.

Colleague Scale 3 (Participation in the Academic Community) appears to us to be relatively weak conceptually, although the items composing the scale are individually satisfactory.

Ratings of teachers made by the various members of the academic community are rarely completely independent: communication among students and between faculty and students influences judgments. This is particularly true, we believe, for the items in colleague Scale 4 (Relations with Students) which are usually known to the faculty primarily indirectly from students. Accordingly, colleague Scale 4 appears to us to be less direct, more superficial, and hence less valid than related student Scales 3 (Instructor-Group Interaction) and 4 (Instructor-Individual Student Interaction).

Items 30 through 45 of the colleague survey (Appendix D) relate to teaching observed in seminars and in the classroom. However, 17% of our faculty respondents had not attended a seminar given by their own selected best teacher, fully 51% had not observed classroom teaching of the best teacher, and a surprising 75% had not observed classroom teaching of the selected worst teacher. Further, most members of the faculty who had observed the teaching of the named colleague had done so only briefly or infrequently.

We conclude that ratings by colleagues should be used to supplement, though not to substitute for, ratings by students. Accordingly, our analysis stresses the student scales. However, colleague Scale 5 (Concern for Teaching) relates directly to teaching and is based on items that faculty, not students, can observe. This scale could profitably be represented in any evaluations of teaching made by colleagues.

Sample Size and Norms

It is of consequence that teacher evaluations be based on adequate samples of opinion. Teachers of small classes, and teachers regarded as excellent by some observers and as poor by others, should be rated by as many observers as possible. If a short rating form is used, items should be avoided that are descriptive of the majority of both best and worst teachers and that relatively few respondents can answer. Teachers of even small classes can be rated adequately if evaluations are accumulated from the classes taught between periods of eligibility for promotion.

Whether the teaching of individuals and departments should be evaluated on an absolute or relative basis is open to question. In practice, however,

academic advancement, and the selection by students of courses and curricula, are often based at least in part on comparisons of teacher with teacher and department with department. It is desirable that norms be calculated at the campus level for at least some elements of any evaluation form that is used in promotion procedures. The summary descriptions of the five principal components of effective teaching would be satisfactory for the calculation of such norms. Departments or subject areas might find it useful also to calculate their own norms, particularly if they have developed their own evaluation forms.

Finally, we suggest that norms be recalculated at frequent intervals to make the system of evaluation responsive to change.

A Potential Weakness in the Use of Student Evaluations

It is unlikely that our results could guide an instructor to elicit higher student ratings than he deserved. Scholarship, rapport, and enthusiasm are difficult to simulate, and students are not easily deceived. In various circumstances, however, a teacher may be better than his teaching. His ratings may be adversely affected because his work load is too heavy, his classes are too large, he is assigned to teach outside the area of his greatest competence, his course is new and untried, or he is experimenting with innovations. The student properly rates his teacher on how good he perceives the instruction to be, not on how good it could have been or will become. It would be unfortunate if rating procedures either penalized teachers for factors beyond their control or encouraged them to offer only "safe," familiar instruction. We believe that this danger can be minimized if it is recognized and appropriate steps are taken to bring any mitigating circumstances to the attention of the administration (see recommendation 6 on p. 24).

Alternative Student Evaluation Forms

Our results can be used in many ways, depending on objectives and facilities. We present and comment on three kinds of evaluations which are intended to be suggestive rather than limiting.

1) Long form. The 85 items of Appendix C provide the basis for a long evaluation form. The list might well be altered to better adapt it to the requirements of a particular teacher, department, or subject area. Such a form provides much information and thus is useful to teachers, whether new or established, who wish to improve. (Some instructors believe, however, that a single open-ended question such as Part D of Appendix K, elicits the most useful responses for this purpose.) A long form is relatively slow to complete. Results, being diverse, would be difficult to apply to advancement procedures. This being true, evaluations would probably be ignored by some teachers.

2) Short form. Appendix K is a model of a short evaluation form (59). The derivations of the items have been explained in the body of this report.

Such an instrument would be effective for evaluating teaching for use in advancement procedures. It is applicable to most university teaching and therefore would permit the calculation of departmental, college, and campus norms. A short form is less directly useful than a long form for helping teachers to improve their performance. However, though there is as yet little evidence to support us, we believe that if teaching were to become a more effective criterion for academic advancement, performance would be benefitted indirectly through improvement in the status of teaching.

3) Medium-length form. An evaluation form of medium length might provide a desirable compromise between the advantages and disadvantages of longer and shorter forms. The 36 items of Table 1 (p. 13) plus Part B of Appendix K would constitute such a form. Some demographic items might also be included (see p. 16).

General Recommendations

We recommend that:

1) The faculty, students, and administration at Davis support in principle the regular use of student evaluations of teaching for the benefit of individual instructors, students, and advancement procedures.

2) Ratings and campus norms (and departmental or subject area norms if appropriate) be made accessible to the entire academic community. Academic Freedom should not be interpreted as granting to a professor immunity from orderly and responsible appraisal of his competence in so major an area of his professional activity as teaching. Evaluation of a scholar's teaching, like his research and artistic and literary works, should, we believe, be accessible for use by those directly concerned. If student evaluations of teaching assembled by the faculty are denied to the students, then the students can be expected to assemble their own data. Such duplication of effort would jeopardize the success of all evaluation programs and would, we believe, constitute an unwarranted expression of distrust of the students by the faculty.

3) Policies and procedures relative to the collection and processing of student evaluations of teaching be established by the Committee on Teaching of the Academic Senate and the Student Academic Affairs Council, in consultation, subject to review by their respective parent bodies.

4) The Academic Senate request the Vice Chancellor for Academic Affairs to establish an Office for the Evaluation of Teaching (60), to be competently staffed and adequately funded to direct the distribution and processing of evaluation forms. We consider this job to be too big and important to be performed in an entirely satisfactory manner by committees of either the faculty or student body. Ultimately the Office might also contribute to the improvement of teaching in other ways (research on the evaluation of teaching, video tape service for self-evaluation, references, consultation, seminars, etc.).

5) First priority be given to the wide use of a short evaluation form (about 15 items); second priority be given to the occasional substitution, at the instructor's request, of a form of medium length (about 50 items); and third priority be given to a long evaluation form to be used, at the instructor's request, as a supplement to the other forms.

6) Student evaluations of teaching be used to supplement, but not to substitute for, other kinds of evaluations, and that all available valid evidence be used in making judgments about teaching relative to academic advancement. Departments should be encouraged to report any mitigating circumstances in relation to student evaluations (see p. 22).

REFERENCES AND NOTES

1. The study was recommended in 1966 by an ad hoc Committee on Teaching. The members were R. M. Cello, M. M. Green, R. W. Hoermann, T. J. Shank, and M. Hildebrand, chairman.

2. The $31,670 provided for the study by the president of the university and chancellor of the campus was supplemented by the University's Center for Research and Development in Higher Education under contract OE-6-10-106 of the U. S. Office of Education. The directors received no stipends.

3. Hildebrand (Department of Zoology, Davis) suggested the project and, with help from many colleagues, defined objectives, secured funds, coordinated, and brought to the study Wilson (Center for Research and Development in Higher Education, Berkeley), who, with Evelyn R. Dienst and Nancy Watson, designed the surveys and conducted the analyses.

4. We thank the Advisory Committee which reviewed the research plan, questionnaires, and drafts of this report. We also thank the many other persons, on this campus and elsewhere, who reviewed a draft of this report. Their suggestions led to many changes and substantial improvement.

5. A. W. Astin and C. B. T. Lee, The Educ. Record 47, 361 (1966).

6. P. Dressel,(in) Conference on the appraisal of teaching in large universities, Univ. of Michigan Press, Ann Arbor (1959).

7. H. H. Remmers,(in) ibid.

8. B. A. Doty, J. Educ. Research 60, 363 (1967).

9. I. Carpenter, E. Van Egmond, and J. Jochen, Sci. Educ. 49, 235 (1965).

10. R. E. Eckert and D. C. Neale, Rev. Educ. Research 35, 304 (1965).

11. W. J. McKeachie, Rev. Educ. Research 30, 351 (1960).

12. It is unlikely that prizes for fine teaching (commonly $500 or $1000) influence the instruction of many professors. Such awards are a pittance compared with prizes for fine research (in the form of fellowships and grants) or with the benefits of promotion. At Davis, prizes were discontinued for lack of confidence in methods of evaluation, and hence lack of prestige.

13. A current study of characteristics of the faculty by one of us (Wilson).

14. An appointment or promotion is initiated by letter from the department chairman and is supported by reprints and other documents. The dean of the college adds a letter to the file. A secret ad hoc committee of the faculty reviews the case, adhering to a uniform statement of criteria. It obtains additional evidence if the dossier is considered to be deficient, and reports to the select faculty Committee on the Budget and Interdepartmental Relations, which reviews all evaluations and recommends to the Chancellor.

15. J. W. Gustad, The Educ. Record 42, 194 (1961).

16. A one-year period was selected to standardize exposure and recall.

17. A long-range study of "Educational relevance: the meaning of college," by M. Regan and G. D. Yonge.

18. Or by a margin of three best over worst nominations if the teacher received both kinds of rating.

19. Mean scores for best, not nominated, and worst teachers were respectively 6.16 ($s = 1.02$, $N = 573$), 5.28 ($s = 1.39$, $N = 297$), and 4.58 ($s = 1.59$, $N = 283$). For the difference between best and worst, $p < .0005$, for the difference between best and not nominated, $p < .005$, and for the difference between not nominated and worst, $p < .01$. ($N > 1015$ because responses are included that were eliminated from subsequent analysis.)

20. The list had been shortened from 236 items following analysis of a pretest taken by 44 students.

21. D. J. Cosgrove, J. Educ. Psychol. 50, 200 (1959); C. W. Crannel, J. Psychol. 36, 417 (1953); C. A. Gibb, Educ. and Psychol. Measurement 15, 254 (1955); E. R. Guthrie, The evaluation of teaching: a progress report, Univ. of Washington, Seattle (lithoprint) (1954); R. B. Hayes, J. Educ. Research 57, 47 (1963); T. F. Hodgson, The general and primary factors in student evaluation of teaching ability, Univ. of Washington, Seattle (mimeograph) (1958); R. L. Isaacson et al., J. Educ. Psychol. 55, 344 (1964); A. A. Lacognata, J. Exptl. Educ. 33, 107 (1964); A. G. Rezler, J. Educ. Research 58, 282 (1965); D. G. Ryans, Characteristics of teachers, Am. Council on Educ., Washington (1960); D. Solomon, Am. Educ. Research J. 3, 35 (1966); D. Solomon et al., J. Expt. Educ. 33, 23 (1964).

22. L. Kent, (in) C. B. T. Lee, ed., Improving college teaching, Am. Council on Educ. (1967).

23. C. H. Weaver, J. Educ. Psychol. 51, 21 (1960).

24. D. B. Gowin and D. E. Payne, School Rev. 70, 207 (1962).

25. See items 4, 13, 14, 24, 30, 43, 48, 55, 63, 65, 76 and 85.

26. E. R. Guthrie (citation in note 21) reported results of 121 ratings of colleagues on an 8-item questionnaire.

27. Larger samples would doubtless reveal statistically significant but minor differences between the groups.

28. These items cast doubt on the validity of item 8 of Appendix D, which indicates that colleagues perceive these kinds of activities as discriminating between effective and ineffective teachers, but see again note 27.

29. A. W. Bendig, J. Exptl. Educ. 21, 333 (1953); W. E. Coffman, J. Educ. Psychol. 45, 277 (1954); D. J. Cosgrove, J. Educ. Psychol. 50, 200 (1959); C. W. Crannel, J. Psychol. 36, 417 (1953); H. A. Estrin, Improving Col. and Univ. Teaching 13, 137 (1965); G. J. French, College students' concept of effective teaching determined by an analysis of teacher ratings, Univ. of Washington, Seattle (mimeograph); C. M. Garverick and H. D. Carter, California J. Educ. Research 13, 218 (1962); C. A. Gibb, Educ. and Psychol. Measurement 15, 254 (1955); R. L. Isaacson et al., J. Educ. Psychol. 55, 344 (1964); H. H. Remmers and P. C. Baker, Manual of instructions for Purdue rating scale for instruction, Div. of Educ. Reference, Purdue Univ., Lafayette, Indiana (1952); D. Solomon, Am. Educ. Research J. 3, 35 (1966); D. Solomon et al., J. Exptl. Educ. 33, 23 (1964); R. J. Wherry, Control of bias in rating: instructor rating scales, Personnel Research Section, AGO, U.S. Dept. of the Army, Washington, D.C.

30. Items were eliminated from the original list of 158 if: a) not discriminating between best and worst teachers at the .001 level (the item was retained but the responses of graduate students discounted if they did not meet this standard as a group); b) 25% or more of respondents could not reply "Yes" or "No"; c) descriptive of virtually all best teachers; d) descriptive of few best or worst teachers; e) descriptive of most best and worst teachers; f) applicable only to small classes; or g) related to examinations and assignments.

31. We consulted Norman Cliff, Professor of Psychology at the Univ. of So. Calif., about methods for factor analysis.

32. H. F. Kaiser, Psychometrika 23, 187 (1958).

33. R. C. Tryon and D. E. Bailey, Multivariate Behavioral Research 1, 95 (1966).

34. The highest intercorrelations are: 3 with 4, .38; and 1 with 3, .32.

35. Alpha reliabilities for the data from the 1967 survey range from .58 to .76. The values are lower because only best teachers were included.

36. The 1967 values are similar; the 1968 values are shown because several new items had been added.

37. Items were also excluded if not discriminating at the $p < .001$ level, and if more than 33% of respondents replied "Does not apply or don't know."

38. The highest intercorrelations are: 1 with 2, .41; and 3 with 4, .39.

39. These descriptions have been modified somewhat from those used in the 1968 survey so as to emphasize the items found most discriminating and, for reasons already explained in text, to give less emphasis to items that, even though discriminating, are characteristic of both best and worst teachers.

40. Respondents tend to use the upper part of the rating scale; a seven-point continuum therefore provides more discrimination at the high end of the scale than does a five-point continuum.

41. N = 1015 for all variables except academic rank, course level, and class size, for which N = 51.

42. D. Solomon, Am. Educ. Research J. 3, 35 (1966).

43. E. R. Guthrie, The evaluation of teaching: a progress report, Univ. of Washington, Seattle (lithoprint) (1954).

44. A. H. Cohen and F. B. Brawer, Measuring faculty performance, Am. Assoc. of Junior Coll., Washington, D.C. (1969).

45. The correlations of class size with Scales 3 and 4 are higher (coefficients .32 and .51) when the five summary descriptions of the 1968 survey are substituted for the full scales of the 1967 survey.

46. C. T. Stewart and L. F. Malparr, J. Educ. Research 59, 347 (1966).

47. A. M. Anikeef, J. Applied Psychol. 37, 458 (1953).

48. V. Voeks and G. M. French, J. Higher Educ. 31, 334 (1960).

49. N. M. Downie, J. Higher Educ. 23, 495 (1952); E. Spaights, Improving Coll. and Univ. Training 15, 15 (1967).

50. W. J. McKeachie, (in) N. L. Gage, Handbook of research on teaching, Rand McNally & Co., Chicago (1963).

51. A. C. Maney, Sociology of Educ. 32, 226 (1959); A. G. Rezler, J. Educ. Research 58, 282 (1965); W. Haythorn et al., J. Abnorm. Soc. Psychol. 53, 210 (1956); M. F. Freehill, Improving Coll. and Univ. Teaching 15, 18 (1967).

52. When this study was proposed to the Academic Senate it was pledged that individual and departmental ratings would not be revealed.

53. Areas to cover and specific items were taken in part from A. G. Cohen and G. M. Guthrie, Educ. and Psychol. Measurement 26, 89 (1966); W. McKeachie, A report on student evaluation forms (Term 1 (1966-67)), Univ. of Michigan, Ann Arbor (ditto); and Student report on introductory psychology, Univ. of Washington, Ann Arbor (mimeograph).

54. Examples: Because my parents wanted me to. Because my friends were going. For the fun and excitement of college. Because I had no other plans.

55. Coefficients of the 25 elements of the matrix range from -.19 to .22. N = 338.

56. G. J. Beichl, Science, 17 Feb. 1967.

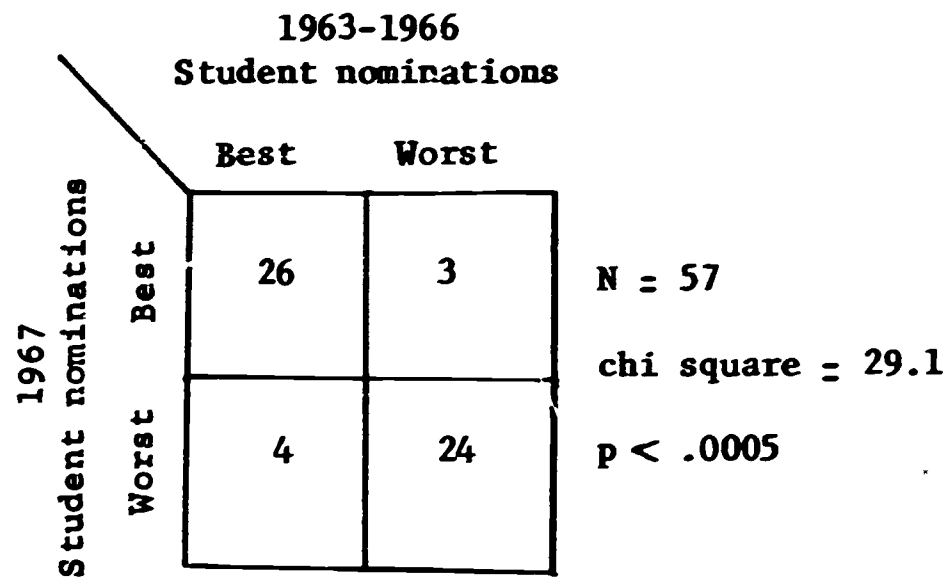
57. R. Brandis, The Educ. Record 45, 56 (1964).

58. At Davis, candidates for advancement must report the current positions held by former graduate students, a practice which we consider to be undesirable.

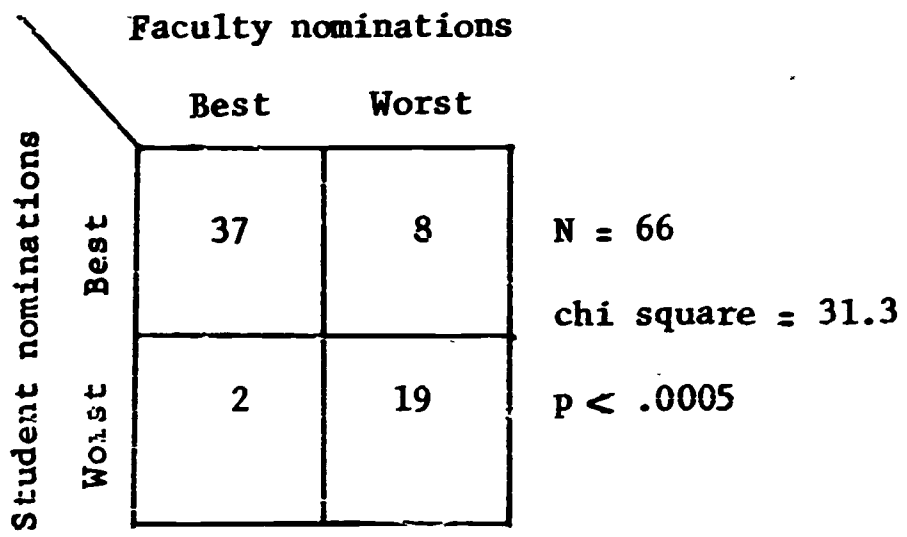
59. As indicated by the heading and instructions, this form is clearly evaluative. This strategy is favored by the Advisory Committee because most items clearly have "right" and "wrong" answers. An alternative strategy, favored by Wilson, is to present the form as a description of teaching, reserving evaluation for a subsequent action.

60. The University of Washington has long had a comparable Office of Student Ratings.

APPENDIX A. AGREEMENT BETWEEN NOMINATIONS FOR BEST AND WORST TEACHERS BY OUR 1967 STUDENT SAMPLE AND SIMILAR NOMINATIONS BY THE 1963-1966 REGAN/YONGE STUDENT SAMPLE.



APPENDIX B. AGREEMENT BETWEEN NOMINATIONS FOR BEST AND WORST TEACHERS BY OUR 1967 STUDENT SAMPLE AND SIMILAR NOMINATIONS BY A SAMPLE OF THE FACULTY.



APPENDIX C. CHARACTERIZATION BY STUDENTS OF EFFECTIVE TEACHERS

CHARACTERISTICS OF A MAJORITY OF BEST TEACHERS AND OF A MINORITY OF WORST

Course content and presentation

†* 1. Contrasts implications of various theories
2. Presents origins of ideas and concepts
* 3. Presents facts and concepts from related fields
4. Talks about research he has done himself
5. Emphasizes ways of solving problems rather than solutions
6. Discusses practical applications
7. Explains his actions, decisions, and selection of topics
† 8. Seems well read beyond the subject he teaches
* 9. Is an excellent public speaker
† 10. Speaks clearly
*11. Explains clearly
12. Gives lectures that are easy to outline
13. Reads his lectures or stays close to his notes (Negative)
14. Assigns text as background, but lectures include other topics
*15. Makes difficult topics easy to understand
16. Summarizes major points
17. States objectives for each class session
18. Identifies what he considers important
*19. Shows interest and concern in quality of his teaching
20. Gives examinations requiring creative, original thinking
21. Gives examinations having instructional value
22. Gives examinations requiring chiefly recall of facts (Negative)
23. Gives interesting and stimulating assignments
24. Stresses the aesthetic and emotional value of the subject
*25. Is a dynamic and energetic person
†*26. Seems to enjoy teaching
† 27. Is enthusiastic about his subject
† 28. Seems to have self-confidence
29. Varies the speed and tone of his voice
30. Has a sense of humor

Relations with students

31. Is careful and precise in answering questions
† 32. Explains his own criticisms
33. Encourages class discussion
*34. Invites students to share their knowledge and experiences
*35. Clarifies thinking by identifying reasons for questions
*36. Invites criticism of his own ideas
†*37. Knows if the class is understanding him or not
38. Knows when students are bored or confused
39. Has students apply concepts to demonstrate understanding
†*40. Keeps well informed about progress of class
41. Anticipates difficulties and prepares students beforehand
42. Has definite plan, yet uses material introduced by students
43. Provides time for discussion and questions
*44. Is sensitive to student's desire to ask a question

45. Encourages students to speak out in lecture or discussion
†46. Quickly grasps what a student is asking or telling him
47. Restates questions or comments to clarify for entire class
48. Asks others to comment on one student's contribution
49. Compliments students for raising good points
50. Doesn't fully answer questions (Negative)
51. Determines if one student's problem is common to others
52. Reminds students to see him if having difficulty
53. Informs students of coming campus events related to course
54. Encourages students to express feelings and opinions
55. Relates class topics to students' lives and experiences
†56. Has a genuine interest in students
57. Relates to students as individuals
58. Recognizes and greets students out of class
*59. Is valued for advice not directly related to the course
60. Treats students as his equals

CHARACTERISTICS OF A MAJORITY OF BEST AND WORST TEACHERS, BUT MORE TYPICAL OF BEST

61. Discusses points of view other than his own
62. Discusses recent developments in the field
63. Gives references for the more interesting and involved points
64. Emphasizes conceptual understanding
65. Disagrees with some ideas in textbook and other readings
66. Stresses rational and intellectual aspects of the subject
67. Stresses general concepts and ideas
68. Seems to have a serious commitment to his field
69. Is well prepared
70. Gives examinations stressing conceptual understanding
71. Gives examinations requiring synthesis of various parts of course
72. Gives examinations permitting students to show understanding
73. Is friendly toward students
74. Is accessible to students out of class
75. Respects students as persons
76. Is always courteous to students
77. Gives personal help to students having difficulty with course
78. Has an interesting style of presentation

RESULTS TYPICAL OF TAKING A COURSE FROM A BEST TEACHER AND NOT FROM A WORST

†*79. Have developed increased appreciation for the subject
*80. Have learned new ways to evaluate problems
81. Have worked harder than in most other courses
82. Know how to find more information on the subject
83. Have studied a topic from the course on own initiative
84. Plan to take more courses on the subject
85. Have gained self-knowledge

---

* Descriptive of 75% or more of best teachers and 25% or less of worst teachers

† Descriptive of 95% or more of best teachers and 45% or less of worst teachers

Items are not listed in rank order

APPENDIX D. CHARACTERICATION BY COLLEAGUES OF EFFECTIVE TEACHERS

CHARACTERISTICS OF A MAJORITY OF BEST TEACHERS AND OF A MINORITY OF WORST

1. Does original and creative work
2. Expresses interest in the research of his colleagues
3. Gives many papers at conferences
4. Has done work to which I refer in teaching
5. Has been consulted by me about my research
6. Has been consulted by me about problems in his field
7. Discusses students' work with colleagues
†8. Spends much time planning and preparing for his teaching
9. Seems well read beyond the subject he teaches
10. Is sought by others for advice on research
†11. Can suggest reading in any area of his general field
12. Is sought by colleagues for advice on academic matters
13. Encourages students to talk with him on matters of concern
14. Is involved in campus activities that affect students
15. Attends many lectures and other events on campus
16. Enjoys controversy in discussion and may provoke opposing views
†17. Comes to departmental or committee meetings well prepared
18. Meets with students informally out of class
19. Meets with students out of regular office hours
20. Encourages students to talk with him on matters of concern
†21. Seems to have a congenial relationship with students
†22. Seems to have a genuine interest in his students
* 23. Seeks advice from others about the courses he teaches
†24. Discusses teaching in general with colleagues
25. Does not seek close friendships with colleagues (Negative)
26. Is someone with whom I have discussed my teaching
27. Is interested in, and informed about, the work of colleagues
28. Expresses interest and concern about the quality of his teaching
†29. Seems to enjoy teaching

Further characterization if speech or seminar was attended

†30. Gives a well organized presentation
* 31. Is an excellent public speaker
32. Summarizes major points at the end of a presentation
* 33. Uses wit and humor effectively
†34. Uses well chosen examples to clarify points
†35. Communicates self-confidence

Further characterization if classroom teaching was attended

36. Encourages students to express feelings and opinions
* 37. Clarifies thinking by identifying reasons for questions
38. Presents facts and concepts from related fields
* 39. Anticipates difficulties and prepares students beforehand
† 40. Quickly grasps what a student is asking or telling him
† 41. Is careful and precise in answering questions
42. Presents origins of ideas and concepts
† 43. Emphasizes ways of solving problems rather than solutions

CHARACTERISTICS OF A MAJORITY OF BEST AND WORST TEACHERS, BUT MORE TYPICAL OF BEST

44. Invites discussion of points he raises
45. Is careful and previse in answering questions
46. Keeps current with developments in his field
47. Has talked with me about his research
48. Knows about developments in fields other than his own
49. Has a congenial relationship with colleagues
50. Is conscientious about keeping appointment with students
51. Recognizes and greets students out of class
52. Is enthusiastic about his subject
53. Does work that receives serious attention from others
54. Corresponds with others about his research

---

* Descriptive of 75% or more of best teachers and of 25% or less of worst teachers

† Descriptive of 95% or more of best teachers and of 45% or less of worst teachers

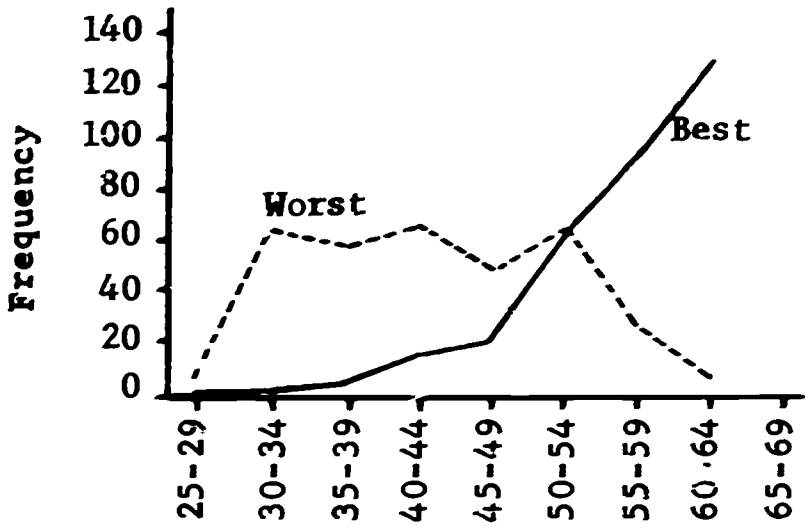
Items are not listed in rank order

APPENDIX E. CORRELATIONS BETWEEN STUDENT RATINGS OF THE OVERALL EFFECTIVENESS OF 51 TEACHERS AND RATINGS OF THE FIVE PRINCIPAL COMPONENTS OF THEIR TEACHING

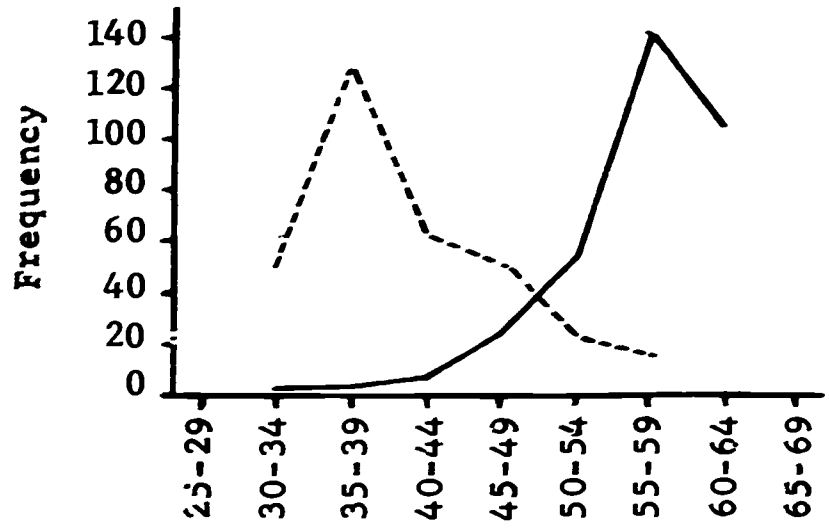
Component	Correlation with overall rating
1. Analytic/Synthetic approach	.60
2. Organization/clarity	.74
3. Instructor-group interaction	.59
4. Instructor-individual student interaction	.63
5. Dynamism/enthusiasm	.83

Correlations > .70 = high (underlining); .70 to .40 = moderate. N = 51

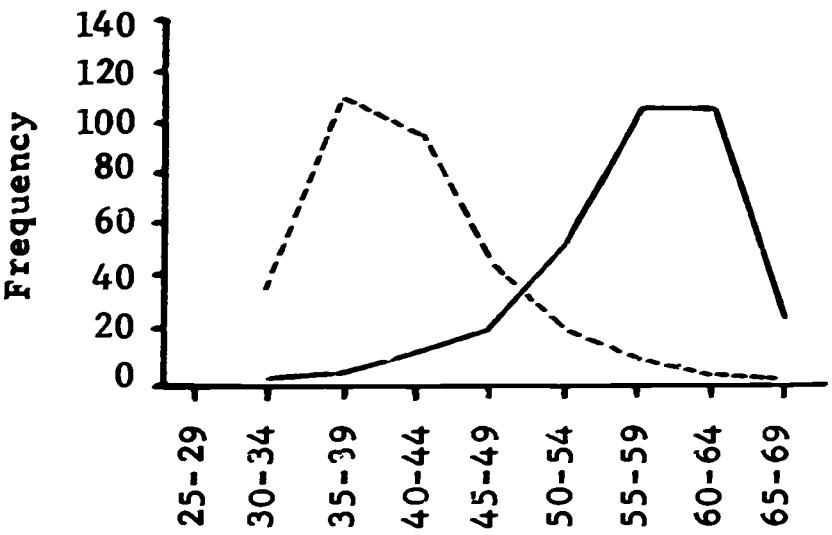
APPENDIX F. FREQUENCY DISTRIBUTIONS OF THE CONVERTED SCORES ($\bar{x}$ = 50, s = 10) OF 338 BEST AND 338 WORST TEACHERS FOR EACH OF FIVE SCALES OF EFFECTIVE TEACHING.



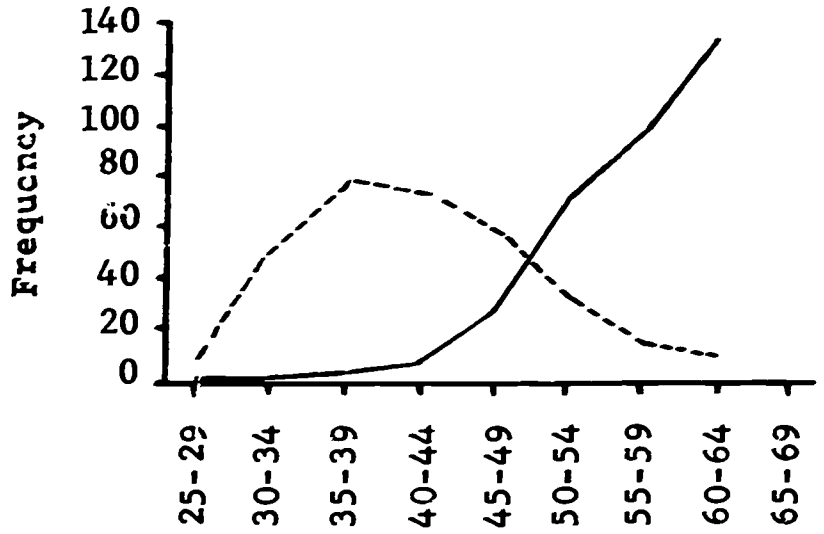
Scale 1. Analytic/Synthetic Approach



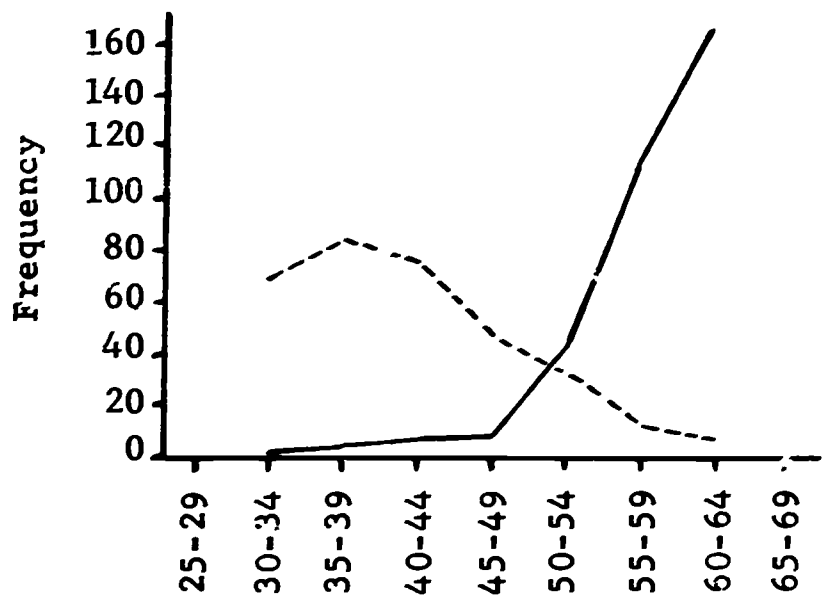
Scale 2. Organization/Clarity



Scale 3. Instructor-Group Interaction

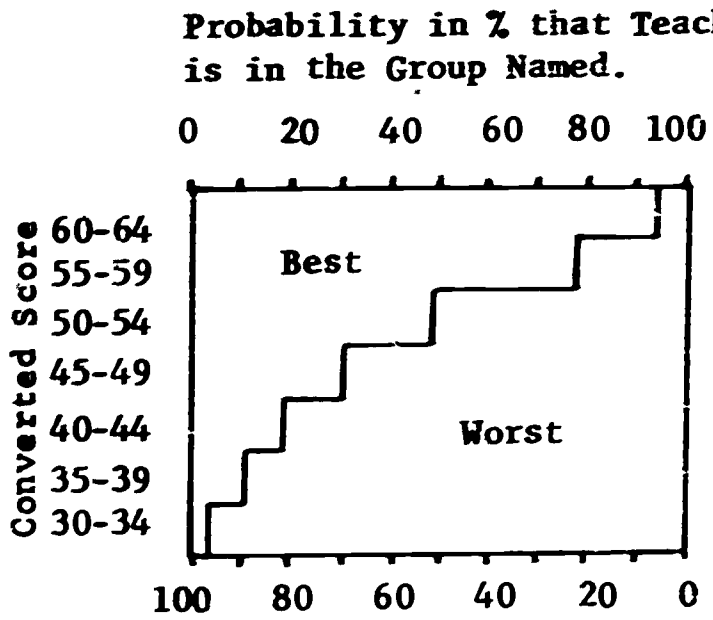


Scale 4. Instr.-Indiv. Student Interac.

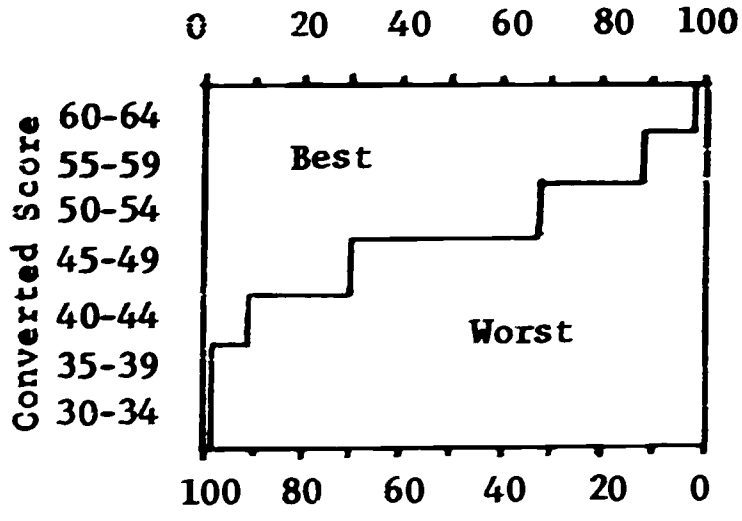


Scale 5. Enthusiasm/Dynamism

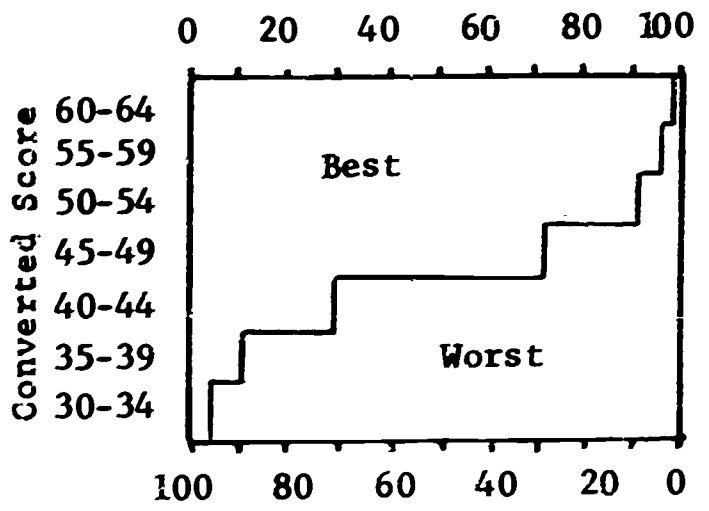
APPENDIX G. PROBABILITY CHARTS OF CONVERTED SCORES ($\bar{x}$ = 50, s = 10) OF 338 BEST AND 338 WORST TEACHERS FOR EACH OF FIVE SCALES OF EFFECTIVE TEACHING.



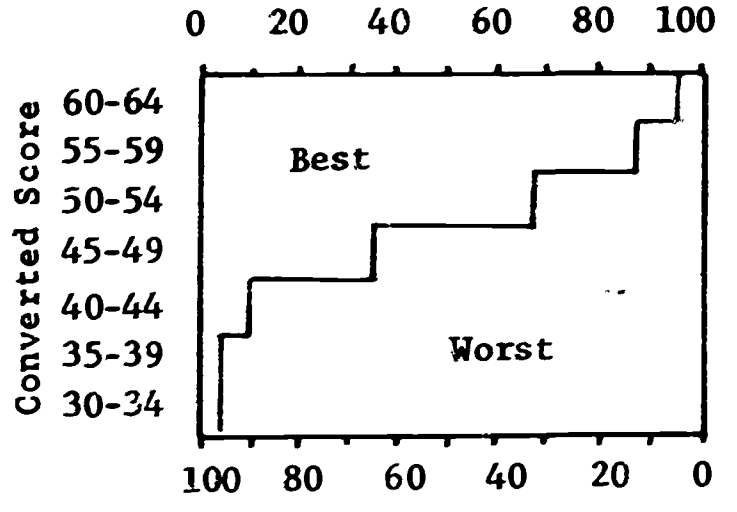
Scale 1. Analytic/Synthetic Approach



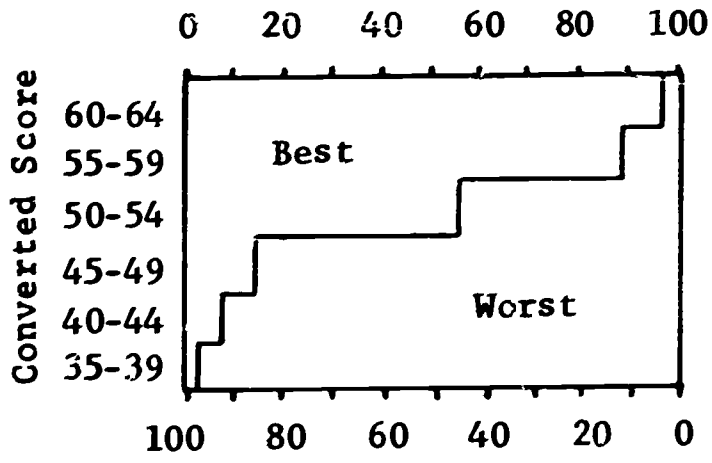
Scale 2. Organization/Clarity



Scale 3. Instructor-Group Interaction



Scale 4. Instr.-Individ. Stud. Interac.



Scale 5. Enthusiasm/Dynamism

APPENDIX H. COLLEGE GOALS OF STUDENTS. N = 1015

SCALE 1. UPWARD MOBILITY/SECURITY

		Factor coefficient
1.	To get the respect a college education brings	.72
2.	To prepare for a better-paying job	.67
3.	To earn a living more easily	.66
4.	To gain greater security	.63
5.	To have a better life than my parents	.50
6.	To become a better citizen	.50
7.	To associate with the preferred kind of people	.49

SCALE 2. SELF-KNOWLEDGE/HUMANISM

8.	To meet and learn from interesting people	.78
9.	To learn more about myself and others	.75
10.	To become more creative	.68
11.	To broaden my overall viewpoint	.66
12.	To be able to lead an interesting life	.45

SCALE 3. CAREER/SUBJECT MASTERY

13.	To get the training needed for success	.83
14.	To learn the skills needed for my career	.77
15.	To gain mastery of my field	.76
16.	To earn the degree needed for my work	.60
17.	To prepare for graduate school	.45

APPENDIX I. OBJECTIVES OF TEACHING FAVORED BY STUDENTS. N = 1015

SCALE 1. CONTRIBUTION TO GENERAL DEVELOPMENT

		Factor Coefficient
1.	To help students mature	.73
2.	To help students understand themselves	.68
3.	To help students understand other people	.68
4.	To help students develop their creative abilities	.66
5.	To help students discover and develop their abilities	.65
6.	To help students analyze their opinions and actions	.64
7.	To teach students to communicate	.55

SCALE 2. TRANSMISSION OF FUNDAMENTALS

8.	To teach facts	.79
9.	To teach fundamental principles	.74
10.	To explain technical terms	.69
11.	To transmit information	.65
12.	To summarize important concepts	.60
13.	To train students in the skills needed for their careers	.52

APPENDIX J. CORRELATIONS OF RATINGS BY 1015 STUDENTS OF DEGREE OF CONSTRUCTIVE CONTRIBUTION MADE TO LIFE IN EACH OF SIX AREAS BY 51 TEACHERS, WITH MEAN SCORES ON SCALES OF COMPONENTS OF EFFECTIVE TEACHING AND OVERALL RATING OF EFFECTIVENESS OF TEACHING

	Constructive contribution to life					
Components of effective teaching	Knowledge imparted	Counsel given	Objectives clarified	Values developed	Incentive elicited	Skills developed
1. Analytic/synthetic approach	.54	.28	.59	.50	.68	.14
2. Organization/clarity	.71	.32	.49	.38	.58	.42
3. Instructor-group interaction	.44	.66	.49	.48	.66	.53
4. Instructor-individual student interaction	.39	.80	.42	.46	.63	.45
5. Dynamism/enthusiasm	.61	.38	.48	.72	.74	.33
Overall rating of effectiveness	.79	.55	.64	.62	.84	.55

Correlations > .70 = high (underlining); .70 to .40 = moderate; .40 to .20 = low. N = 51.

STUDENT EVALUATION OF TEACHING – UNIVERSITY OF CALIFORNIA AT DAVIS

Instructor _____ Department _____ Course number _____

Year _____ Quarter _____ Teaching observed (give approximate number of each): Lectures _____ Labs _____

Conference/Discussion/Seminar _____ Other (specify) _____

A. Each of these statements describes a basic component of teaching. Give the instructor an overall rating for each component, reserving the highest scores for unusually effective performance.

	Low Score						High Score
1. Has command of the subject, presents material in an analytic way, contrasts points of view, discusses current developments, and relates topics to other areas of knowledge.	1	2	3	4	5	6	7
2. Makes himself clear, states objectives, summarizes major points, presents material in an organized manner, and provides emphasis.	1	2	3	4	5	6	7
3. Is sensitive to the response of the class, encourages student participation, and welcomes questions and discussion.	1	2	3	4	5	6	7
4. Is available to and friendly toward students, is interested in students as individuals, is himself respected as a person and is valued for advice not directly related to the course.	1	2	3	4	5	6	7
5. Enjoys teaching, is enthusiastic about his subject, makes the course exciting, and has self-confidence.	1	2	3	4	5	6	7

B. These items are not covered in the statements above and thus extend the evaluation.

	Low Score				High Score	Doesn't apply or don't know
6. Has increased my appreciation for the subject.	1	2	3	4	5	☐
7. Keeps well informed about the progress of the class.	1	2	3	4	5	☐
8. Anticipates problems and makes difficult topics easy to understand.	1	2	3	4	5	☐
9. Is an excellent speaker.	1	2	3	4	5	☐
10. Quickly grasps what a student is asking or telling him.	1	2	3	4	5	☐

(over)

	Low Score				High Score	Doesn't apply or don't know
11. Presents the aesthetic and emotional values of the subject.	1	2	3	4	5	☐
12. Relates class topics to students' lives and experience.	1	2	3	4	5	☐
13. Gives interesting and stimulating assignments.	1	2	3	4	5	☐
14. Gives examinations that require creative, original thinking.	1	2	3	4	5	☐
15. Gives examinations that have instructional value.	1	2	3	4	5	☐

C. Additional items may be presented by the instructor and/or department.

16.	1	2	3	4	5	☐
17.	1	2	3	4	5	☐
18.	1	2	3	4	5	☐
19.	1	2	3	4	5	☐
20.	1	2	3	4	5	☐
21.	1	2	3	4	5	☐
22.	1	2	3	4	5	☐
23.	1	2	3	4	5	☐
24.	1	2	3	4	5	☐
25.	1	2	3	4	5	☐

D. You are invited to comment further on the course and/or effectiveness of the instruction.