DOCUMENT RESUME

ED 054 392 AA 000 727

AUTHOR Cook, J. Marvin; Neville, Richard F.
TITLE The Faculty as Teachers: A Perspective on Evaluation.
INSTITUTION ERIC Clearinghouse on Higher Education, Washington, D.C.
PUB DATE Sep 71
NOTE 17p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Academic Performance; Achievement; *Effective Teaching; *Faculty Evaluation; *Higher Education; *Student Evaluation; Teacher Characteristics; Teaching Quality; *Teaching Skills
IDENTIFIERS Maryland University Baltimore Count

ABSTRACT

This paper examines the problem of measuring and evaluation teacher performance. Evaluation methods currently in use are reviewed, including the use of student questionnaires, and a recommendation for a more accurate measurement of teacher effectiveness is made. The authors specifically consider the relative merits of measurement based on student performance (direct measurement) and measurement based on teaching activities (indirect measurement) as they relate to the evaluation of faculty. This paper is based on a study performed by J. Marvin Cook for the Faculty Senate, University of Maryland, Baltimore County. (WVM)

The Faculty as Teachers:

A Perspective On Evaluation

ERIC CLEARINGHOUSE ON HIGHER EDUCATION • The George Washington University, One Dupont Circle, Suite 630, Washington, D.C. 20036 • September 1971

THE FACULTY AS TEACHERS: A PERSPECTIVE ON EVALUATION

J. MARVIN COOK
AND
RICHARD F. NEVILLE

Report 13

ERIC Clearinghouse on Higher Education
The George Washington University
1 Dupont Circle, Suite 630
Washington, D.C. 20036

September 1971

FOREWORD

This paper examines the problem of measuring and evaluating teacher performance. Evaluation methods currently in use are reviewed and a recommendation is made for the implementation of an approach to evaluation of teacher effectiveness. The authors, J. Marvin Cook and Richard F. Neville, define the concept of teacher effectiveness as the study of teaching outcomes. In that context they analyze the relative merits of measurement based on student performance (direct measurement) and measurement based on teaching activities (indirect measurement) as they relate to the evaluation of faculty. J. Marvin Cook is an associate professor and Richard F. Neville is division chairman and professor at the University of Maryland, Baltimore County, Division of Education. This paper is based on a study performed by Dr. Cook for the faculty senate at the Baltimore County campus.

This paper is the thirteenth in a series of reports on various aspects of higher education. These reports are based on recent significant documents found both in and outside the ERIC collection. Additionally, current research on higher education is abstracted and indexed for publication in the U.S. Office of Education's monthly volume. *Research in Education.* If you wish to order ERIC documents cited in the bibliography, send your request to ERIC Document Reproduction Service, Post Office Box Drawer 0, Bethesda, Maryland 20014. The ERIC document (ED) number must be given when ordering. Payment for microfiche (MF) or xerox/hard copy (HC) must accompany orders of less than $10.00. Documents are $0.65 for microfiche and $3.29 per hundred pages for hard copy. All orders must be in writing.

Carl J. Lange, *Director*
ERIC Clearinghouse on Higher Education
September 1971

TABLE OF CONTENTS

This publication was prepared pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official Office of Education position or policy.

I. INTRODUCTION

Now, more than ever before, the public at large and the young in particular pose penetrating questions about the quality of teaching in higher education. Russell Cooper (1970) notes "teachers are called upon to be accountable for the effectiveness of their performance as teachers," to defend their instructional objectives, and to develop new styles and methods suitable to teach large numbers of students of disparate backgrounds and expectations. In this context, this paper focuses on the evaluation of the university professor in his role as teacher: other areas of faculty performance—research, publications, and service—are not emphasized. The paper is comprised of five parts, including part I, the introduction. Part II assays the problem of measuring teacher performance; part III gives a summary of evaluation methods in use at colleges and universities across the country; part IV recommends specific models to evaluate teacher effectiveness; and part V contains a selective bibliography.

II. THE PROBLEM OF MEASURING TEACHER PERFORMANCE

Teacher performance evaluation has motivated literally thousands of studies. Gage (1960) found that literature about teacher effectiveness was so prolific that bibliographies on the subject bordered on the unmanageable. Another researcher, Tyler (1960) stated that:

> In the American college community the evaluation of teaching is as common as the judgment of the quality of dormitory food– and often as subjective. 'Professor Smith is a wonderful teacher!' 'We are a select college, proud of our tradition of excellence in teaching.' Comments like these are part of the typical pattern of college conversations. Yet we know they are not highly valid, objective and impartial appraisals. Sound and systematic evaluation of college teaching is exceedingly rare and yet it is highly essential to the improvement of college instruction.

Identifying where the teacher is successful and where he has found difficulties offers a starting point for directing efforts toward improvement. The Committee on College Teaching of the American Council on Education (1961) surveyed seven types of higher education institutions to examine their policies and practices in faculty evaluation. The Committee found "without exception, all seven types of institutions said that classroom teaching was the most important factor in evaluation" of faculty members. Faculty members frequently comment on the fact that while the university "states" a commitment to effective teaching, what really counts is research productivity and publication. It would appear that without a sound basis for studying teacher effectiveness, professors will direct their efforts toward those activities that they believe will be rewarded by their institution.

It is generally conceded that to become an effective teacher something more must be considered than the results of a trial-and-error approach or fortuitous successes—hypotheses about effective or ineffective teaching can become principles only when the reasons for effectiveness are carefully examined. There are at least three salient reasons for establishing teacher effectiveness guidelines:

- Every teacher would benefit from a systematic appraisal of his or her efforts.
- Teaching—along with research and service—is one of the three basic activities of a college or university professor and must be effectively performed.
- Basic criteria for teacher performance need to be developed to ensure that faculty are properly evaluated.

Developing a methodology

If a list of consistent predictors of teacher effectiveness were developed, qualified observers could enter a classroom, note a teacher's performance, and measure the way he promotes learning. From the data obtained, it would be possible to predict what traits or skills a teacher needed to assure that his students learned what he wished to teach them; in other words, it would then be possible to predict the teacher's *degree of effectiveness*. However, in spite of prodigious research in the area, Biddle (1964) reports that "few if any teacher traits, skills, or methodology have been identified by research as good predictors of teacher effectiveness."

To establish specific traits, skills, or methodology as predictors of teacher effectiveness in classroom situations, an investigator needs to gather data by direct observation and to record teaching performance by categories of teacher behavior. The resulting profile could then be related to student performance in specific areas of instruction. Unfortunately, as Biddle (1964) emphasized, observation of teacher behavior is a costly and tedious

process, with many variables affecting both student and teacher. One such varial is that a teacher's classroom performance is separated from student achievement by whatever span of time it takes to measure that achievement. A further complication is that most persons doing research on teacher effectiveness have not dealt with behavioral observation. The lack of data in this area is mentioned by Rosenshine (1970): "...among hundreds of research and evaluation reports at the ERIC Clearinghouse on Early Childhood Education, [one researcher] ... found only ten observational studies reported since 1960"; furthermore, Rosenshine notes that of 46 science projects on the higher educational level, only one included an observational analysis. In the same study, Rosenshine also points to the difficulty of meaningfully correlating data between teacher performance in the classroom and student achievement:

> Compared to the large number of descriptive studies, there have been relatively few studies of the relationship between measures obtained by the use of observational systems and measure of class achievement adjusted for initial aptitude or ability. Approximately forty studies were completed in this area. These studies are difficult to synthesize for at least three reasons: they varied widely in subject area, grade, and observational instruments; some studies used statistical procedures which were not appropriate; and in many studies the number of classrooms observed was less than twenty. Two attempts have been made to review some of the research relating classroom behavior to student achievement...but because of the above difficulties it is too early to identify relationships that can be stated with any confidence.

Defining relationships—a first step

There are at least two reasons for the difficulty of measuring teacher effectiveness: first, the inconsistent definition of terms; and second, an inability to relate definitions, once agreed upon, to a methodology for evaluating teachers. In the past, the phrase "teacher effectiveness" has too often been used as a synonym for: traits of teachers; classroom skills exhibited by teachers; student ratings received by teachers; classroom interaction or participation produced by teachers; classroom climate established by teachers; as well as student achievement effectuated by teachers. The same factors have been called "competence," "criteria for competence," "ability to teach," "value of a teacher," and others. In this paper, a distinction is made between teaching traits and skills, "effectiveness," "rating," and "evaluation." The following definitions incorporate these discriminations:

- *Teaching functions* refer to the categories of the role of the teacher as a facilitator of learning. These categories are established on the basis of the different ways a teacher relates to the student's learning process. For example. There might be three teaching functions which the teacher fulfills as a facilitator of learning: diagnostician, prescriber, and enabler.
- *Measurement of teaching traits and skills* (including methodology and academic competence) is the act of determining what and how different traits and skills (methodology) are used by a teacher for the different teaching functions.
- *Teaching effectiveness* is the ability of a teacher to facilitate the acquisition by his students of those ways of thinking, feeling, and acting that have been previously defined as his instructional objectives.
- *Measurement of teaching effectiveness* is the act of measuring the extent to which students have learned what the teacher is trying to teach.
- *Inference* refers to the process intervening between the objective data seen or heard by an observer and the coding of those data on a classroom observational instrument.
- *Rating a teacher* is a procedure by which the rater observes the teacher's performance in a classroom and then infers from his perception of a series of such performances whether or not the teacher is "enthusiastic," "communicates clearly," "has rapport with students," etc. Such ratings yield general impressions which lack specificity.
- *Delineating a set of values* involves the act of identifying the set of justifiable values that is to be superimposed upon a set of collected data.
- *Evaluating a teacher* is the act of weighting the measured data obtained about a teacher by superimposing a set of values onto the data.

These definitions, by distinguishing between the various acts in the teacher evaluation process, should remove a primary roadblock to solving the problem of teacher evaluation. Figure 1 displays these definitions hierarchically to clarify their functional interrelationships. Referring to the illustration, teacher evaluation can be seen as consisting of two major components: a data source of teacher effectiveness, and the set of values used to judge the data. When the measurement data has been obtained on a teacher's effectiveness, and a set of values has been decided upon by a promotion or tenure committee, the process of evaluation consists of weighting the measured data by superimposing a set of values predetermined by the committee onto the data known about a teacher.

For example, suppose the data collected showed that a teacher's students learned what he intended them to learn; and suppose what he intended them to learn were listed as instructional objectives A, B, C and D. If the set of values of the promotion and tenure committee were such that objectives A, B and C were not acceptable instructional objectives, then the superimposition of these values onto the data collected about the teacher would give the teacher credit only for having taught objective D. Of course, the outcome in this example could have been

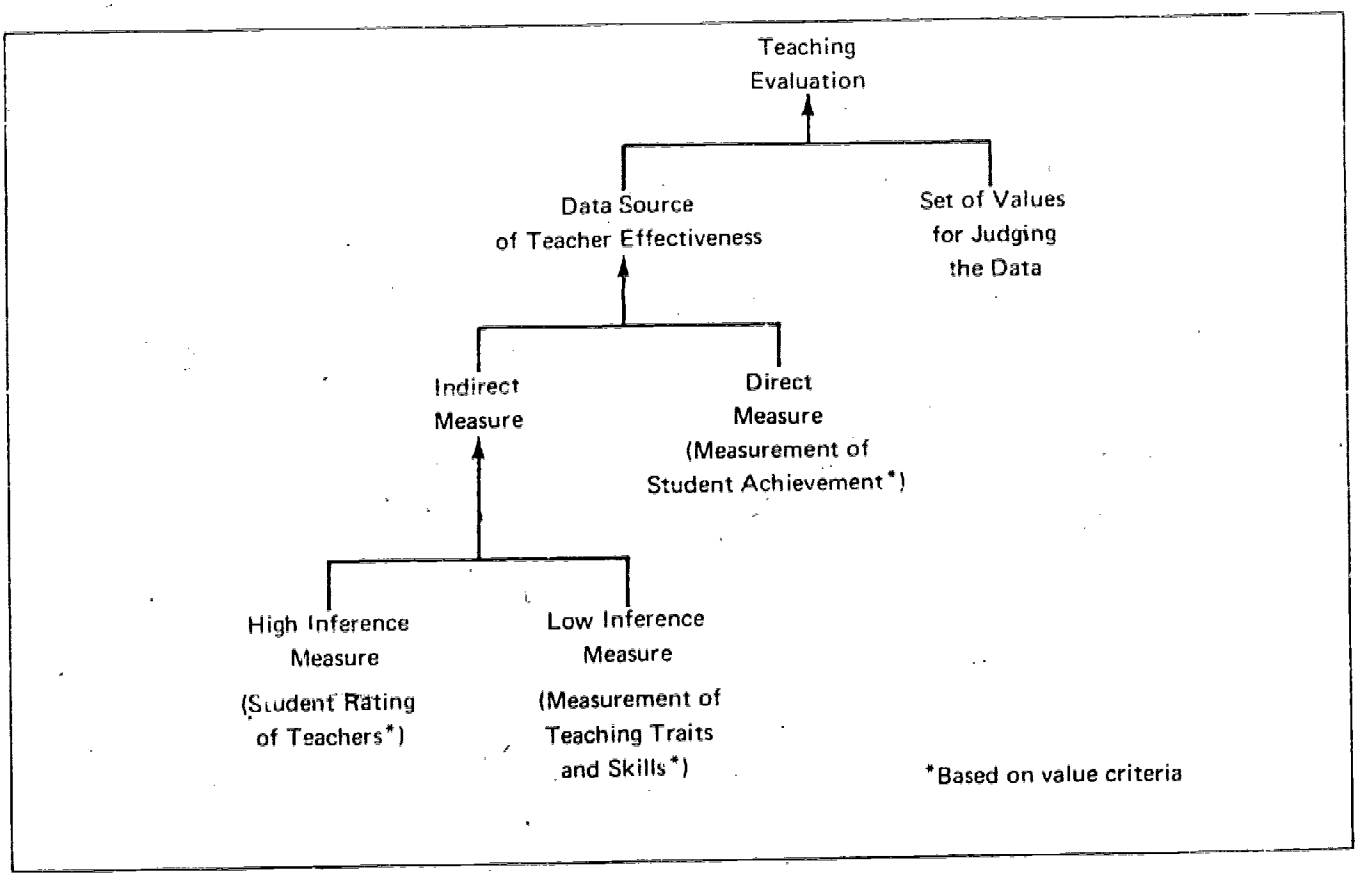

Figure 1. The relationship of the different components in the act of evaluating professors in the teaching role at a university.

prevented by making clear to the teacher what instructional objectives were to be included in the evaluation.

As shown in Figure 1, the data source of teacher effectiveness can be divided into two distinct types of measurements: direct and indirect. A *direct measurement* quantifies the effects of teaching and can be defined as the extent to which students have learned what the teacher is trying to teach as determined by the teacher, his department and other qualified judges. An *indirect measurement* is a measure of what the teacher does to facilitate learning; for example, the teacher selects instructional objectives, selects course materials, establishes a learning environment, prepares academically, organizes and presents materials, diagnoses students, interacts with students, etc.

Indirect measurement

Based on Figure 1, both student ratings of teachers and the measurement of teaching traits and skills are seen to be indirect measures of teacher effectiveness. To use such data a correlation must be established between what a teacher does to facilitate learning and what students do as a result of a teacher's activities. Correlating the activities of a teacher as a facilitator of student achievement is no easy task. It must be done if indirect measures are to be used to measure teaching effectiveness.

The indirect measure in Figure 1 is divided into high- and low-inference measures. The following comments by Rosenshine (1970) explain these two elements and highlight their advantages and disadvantages:

> Instruments for the observation of instruction are currently divided into *category systems* and *rating systems*. This division is based on the amount of *inference* required of the observer or of the person reading the research report. Inference here refers to the process intervening between the objective data seen or heard and to the coding of those data on an observational instrument. Category systems are classified as *low-inference* measures because the items focus upon specific, denotable, relatively objective behaviors such as "teacher repetition of student ideas," or "teacher asks evaluative question[,]" and because these events are recorded as frequency counts. Rating systems are classified as *high-inference* measures because they lack such specificity. Items on rating instruments such as "clarity of presentation," "enthusiasm," or "helpful toward students" require that an observer infer these constructs from a series of events. In addition, an observer must infer the frequency of such behavior in order to record whether it occurred "consistently," "sometimes," or "never," or whatever set of graduations are used in the scale of an observation instrument. To a reader, the statement that a teacher

repeated students answers 7% of the time is much more specific than the statement that a teacher was sometimes helpful toward students.

Category systems have become very popular in descriptive educational research and in teacher training because they offer greater low-inference specificity and because an "objective" *count* of a teacher's encouraging statements to students appears easier for a teacher to accept than a "subjective" *rating* of his warmth. The major disadvantages of category systems are the cost of using observers and the difficulty of specifying behaviors before they can be included in a category system.

Rating systems offer greater flexibility than category systems because they can include high-inference variables. Rating systems can also be less expensive if the students in the classrooms are used as observers. . . .The disadvantages of using rating systems are summarized by Mouly (1969); they include the halo effect, the error of central tendency, generosity or leniency error, and the lack of a common referent for scoring calibrations such as "excellent" or "seldom." Another disadvantage, noted by Gage (1969), is that high-inference items are difficult to translate into specific behaviors. This suggests that evaluative reports based on high-inference measures may offer few specific suggestions for improving an instructional program. An evaluative report which suggests that teachers need to improve their clarity and organization, without giving the low-inference correlates of such behaviors, may amount to little more than suggesting that the teachers be 'good and virtuous.'

It is a mistake to perceive specific acts of teaching and learning as being in a simple cause-and-effect relationship. Teaching and learning are a process comprised of many interactions between teacher and student. The teacher receives responses from the student and reacts to these responses in ways intended to induce learning. Concurrently, the student receives responses from and reacts to the teacher. This cycle of giving and receiving instruction is represented by Smith (1961) as follows:

$$\left\| P_t \rightarrow D_t \rightarrow R_t \rightarrow \left| P_p \rightarrow D_p \rightarrow R_p \right| \right\|$$

Where:

- P_t = the teacher's perception of pupil behavior;
- D_t = the teacher's diagnosis of the pupil's state of interest, readiness, knowledge, etc., made by inference from pupil behavior;
- R_t = the action taken by the teacher in light of the diagnosis;
- P_p = the pupil's perception of the teacher's behavior;
- D_p = the pupil's diagnosis of the teacher's interest, etc., as inferred from the teacher's behavior;
- R_p = the reaction of the pupil to the action of the teacher.

Note: Double vertical lines mark off instances of the complete cycle; single vertical lines divide the cycle into acts of teaching $\left| P_t \ldots R_t \right|$ or acts of taking instruction $\left| P_p \ldots R_p \right|$.

As illustrated above, learning new ways of thinking, feeling, and acting results from the student's participation in an exchange with the teacher. The tenuousness of this process suggests why it is difficult to obtain a positive correlation between specific traits or skills of teachers and specific learning by students. Other factors that increase this difficulty include the cumulative nature of learning described by Gagné (1965) and, as has been noted, the fact that a time lag often exists between the act of teaching and the measure of student achievement.

Various preconceptions among students about the teaching-learning situation must be taken into account as well. Each student brings to his environment a unique set of criteria, such as personal needs and beliefs, that act as intervening variables and add to the difficulty of identifying and using specific traits and skills as predictors of teacher effectiveness. Young (1970) observes that since each classroom situation is unique in itself, a unique set of conditions is necessary before learning can occur: in other words, the skills a teacher employs in one situation may not be suitable for other or even similar situations. Biddle (1964) comments on this problem:

> Some teachers may be inspirational leaders, others warm counselors, and still others walking encyclopedias. In certain contexts, each of these competencies [skills] may be highly effective, in others each might have little or negative effect.

For example, consider the teacher trait of warmth. It often has been assumed that "warm" teachers have a positive effect upon students, and that "cold" or "indifferent" teachers have a negative effect. But is this universally true? "Warmth" may have a positive effect upon first graders, but how about in terms of university students? Does "warmth" have the same value in respect to students with disparate backgrounds and expectations? Is "warmth" equally effective in the classroom, the laboratory, the office, etc.? Is "warmth" suitable to certain subject orientations—social science in contrast to a course in biology?

Because the problem of correlation is complex, the use of indirect measurement data obtained on a teacher's performance should be used only for descriptive purposes—and not as a measure of his effectiveness. Until positive correlations can be made between specific teacher activities and student achievement, ratings based upon observation by students, colleagues, or trained observers should not be considered adequate measures of teaching effectiveness.

Direct measure

The complexity of the task of evaluating teachers can be substantially simplified when a direct measure of teacher effectiveness is used. While many factors impinge on the behavior of the student, the teacher does have an extended opportunity to effect a change in the students' competence levels, and this change in competence becomes a relatively accurate barometer of a teacher's effectiveness. If a value is placed upon students acquiring

the instructional goals, then the task of evaluating a teacher, as noted by Tyler (1960), is that of determining "the extent to which the purposes of the [teaching] activity are actually realized."

At the National Conference on Teacher Evaluation, Herbert Hite (1968a) noted that the object of teaching is to bring about learning, and stated:

> . . .when [teachers] teach pupils, the pupils should demonstrate learning. . . .But I don't think all that's learned is taught. Inasmuch as we teach, we have purpose, and I think that without purpose behind it, you don't have teaching. The object, then, is for the teachers to define the evidence that they'll accept as proof that this learning has taken place; and then to arrange matters so that the individual learner does demonstrate this evidence.

Rosemary Pierrel (n.d.) of Brown University supports this attitude and expands upon it by suggesting that underlying any measure of effective teaching must be specification of the desired end behaviors, both in the classroom and, hopefully, well beyond the college years. These end behaviors should be defined in testable terms and would require observation of the students as well as the teacher. Ideally, Pierrel would require follow-up studies of alumni to determine long-term retention of learned behaviors. Her position is encapsulated by this remark: "Not only is what the student *hears* more important that what the professor says, but it is what the student *does* about it and continues to do about it, that demonstrates his professor's effectiveness."

Biddle (1964) concurs with Hite and Pierrel: "teacher effectiveness [is] the ability of a teacher to produce agreed-upon educational effects in a given situation or context." These educational effects are those desired by the teacher and may be exhibited by the students during the teaching process as well as at the end. This direct approach to measuring teacher effectiveness is well expressed by Tyler (1960):

> What is involved in evaluating teaching? In essence, the task is similar to that of evaluating any other purposeful activity. It is one of finding out the extent to which the purposes of the activity are actually realized. This seems simple enough [.] but as we examine the purposes of teaching, we find certain complications. The purposes of teaching are to facilitate various kinds of desirable learning on the part of the students. Hence, *the evaluation of teaching involves appraising the learning of students.* But learning is the acquisition by the student of ways of behaving, that is, ways of thinking, feeling and acting, which he has not previously followed. Thus a college student may develop an understanding of the physical structure of the atom, an understanding which he did not possess before, or he may acquire the ability to analyze a problem in plant ecology which he could not do before, or he may develop greater skill in reading literary works than he had before, or he may acquire intellectual or aesthetic interests which he did not have previously. These are a few random illustrations of kinds of learning which college teachers may seek to bring about in their students.
>
> Of course not all things students learned in college are desirable, so that the evaluation of college teaching is not simply finding out whether students have learned, but *whether they have learned the things which the instructors were trying to teach.* Students may learn bad habits as well as good ones, they may acquire misinformation as well as sound understanding, they may acquire a distaste as well as an interest in various intellectual and aesthetic fields, and these, too, are cases of learning, but not cases of the learning which college teachers are trying to bring about. Furthermore, since the college years are relatively short—far too short to learn all that we wish could be learned—teachers must make a fairly rigorous selection of those kinds of behavior, that is, of those ways of thinking, feeling and acting, that they consider most important for students to acquire. This, then, makes the evaluation of teaching a matter of finding out how far the students are acquiring the important ways of thinking, feeling and acting on which the teachers are focusing their attention.

III. CURRENT APPROACHES

In the previous section, two basic methods to obtain data for the evaluation of university teachers are noted. One method is a general assessment of the teacher's ability to motivate learning—the *indirect* method. This method would rate teachers by colleagues' and students' judgment of the teacher's performance and by a trained observer using a low-inference category instrument. This indirect evaluation method has validity as a measure of what a teacher does, but is not valid as a measure of the results of teaching.

The second method – *direct* method – seeks to measure a teacher's effectiveness by whether students have learned ways of thinking, feeling, and acting that the instructor was trying to teach. In contrast to the frequency of the student-colleague rating procedure used on other campuses, this method is not known to be used at this time at the college or university level.

The indirect method—evaluating teaching ability on the basis of a general assessment by student or colleague rating procedures—is used most frequently on campuses across the country. Most often, the rating is done by students based upon their perception of the teacher. There appears to be considerable agreement among students, faculty, and administrators at the point of rating teachers. If students rate a teacher high or low on a particular teaching skill (e.g., clarity of presentation), it is highly probable that the teacher's colleagues as well as administrators will rate him similarly.

The use of student evaluations of teachers' performance has greatly increased during the past decade; however, the use of student evaluations is not new. Harvard's annual *Confidential Guide* to courses was first published in 1924, and student evaluation began at the University of Washington in the same year. A large collection of current

questionnaires is maintained by the National Student Association (NSA) at its Washington headquarters. An extensive appendix of different kinds of questionnaires in use at a wide range of colleges and universities is included in Philip Werdell's book for NSA, *Course and Teacher Evaluation* (1967). The availability of such questionnaires is a definite asset when a university considers establishing a student rating system for evaluating teachers.

A report on the evaluation of teaching sponsored by the Academic Senate of the Davis Campus, University of California, yielded five factors that are common to most rating forms used by students and faculties. These factors, epitomizing desirable teaching traits or skills, are listed below:

- *Analytic/Synthetic Approach*—scholarship, with emphasis on breadth, analytic ability, and conceptual understanding.
- *Organization/Clarity*—skill at presentation, but is subject related, not student related, and is not merely rhetorical skill.
- *Instructor-Group Interaction*—rapport with the class as a whole, sensitivity to class response, and skill at securing active class participation.
- *Instructor-Individual Student Interaction*—mutual respect and rapport between the instructor and the individual student.
- *Dynamism/Enthusiasm*—the flare and infectious enthusiasm that comes with confidence, excitement for the subject, and pleasure in teaching.

Universities wishing to create a student rating form could well begin with this list of definitive categories.

Jointly sponsored by the American Association of University Professors (AAUP) and the Association of American Colleges, who are supported by a grant from the Carnegie Corporation, a 2-year program entitled "Project to Improve College Teaching" began its formal existence September 1, 1969, from offices at the AAUP national headquarters, Washington, D.C. The director of the project, Dr. Kenneth E. Eble, traveled to a number of colleges and universities to learn first-hand about promising programs that attempt to measure competency in teaching functions. The following account of approaches on various campuses is from that report (Eble (1970)):

Princeton

> In the first stage of a Pilot Study in evaluation at Princeton, four questionnaires were given to a sampling of courses and students in Biology, Politics, Electrical Engineering, and French. The questionnaires consisted of a course evaluation, a departmental evaluation, an evaluation of the total undergraduate experience, and an evaluation of the first two years.
>
> The results from the four departments were so promising that in the second stage of the Pilot Study the questionnaires were given to all undergraduates in all courses in the university. Some of the conclusions resulting from this experience are worth quoting:
>
> '1. Our experience has set at rest our misgivings about student interest in such questionnaires and about the seriousness with which they would be treated.
>
> '2. Though we have taken no poll, our experience with the comments and letters of the faculty show that those who have reported to us have found the reports of their courses helpful.
>
> . . . we are convinced that the administration of questionnaires of the kind we have tried are important in giving students a sense of greater participation in their education.'
>
> On the basis of this extensive study, Princeton has incorporated the four questionnaires into the regular educational process. The course evaluation at the present time is a modified open system with the machine print-outs for individual courses available to the instructor and his department chairman, and to a central file available to students and other members of the university.

The University of Washington

> Students evaluation of teaching has been going on at the University of Wasnington since 1924. An office of student ratings is supported by the university and directed by a faculty member expert in the area of testing and counseling. The office not only responds to faculty requests for evaluation but *conducts continuing research on the evaluation process.*
>
> Faculty participation in the program has been voluntary; results were made available only to the individual teacher, but he might have them forwarded to his department chairman or dean if he chose. The ten items on the current questionnaire were factored out of forty-one items most frequently mentioned in the extensive literature in the field.
>
> This year, the Faculty Senate recommended the adoption of additional procedures that will insure that every undergraduate class is evaluated at least once in three years and perhaps oftener. These evaluations will be reported to department chairmen and deans and made available to department and college review committees.

The University of Texas

> University of Texas students publish one of the largest and most detailed of course evaluation booklets. The booklet is part of an extensive effort in the college of arts and sciences toward 'a more systematic effort in teaching evaluation.' It includes not only the statistical data from a questionnaire, but an extensive sampling of representative written comments from the students and a self-appraisal of the course by the teacher.
>
> Dean John Silber is inviting each faculty member with tenure to participate in the evaluation at least once every three years and those without tenure at least once a year. Teaching assistants and associates are invited to participate at least once and possibly twice a year. Participants will have the option of releasing results to the department chairman, dean, or the student committee responsible for publishing the results.

University of Hartford—University of Delaware

> Both of these universities have begun special efforts to improve teaching and to examine the contribution that evaluation of teaching may make. The departments of History, Political Science, and Speech and Drama at the

University of Hartford are working with the project in devising, administering, and evaluating the results of a student evaluation questionnaire.

A University of Delaware faculty committee has finished an extensive examination of teaching and ways of improving it. The Senate has accepted the report, and the next step will be to get many of its excellent recommendations implemented.

Hamline University

Hamline University reports a successful teaching improvement program which involves annual evaluation. Associate Dean Kenneth Janzen attributes its success to 'a faculty which has insisted on being involved in the process of evaluation, has been willing to take the responsibility for devising means for making that evaluation, and has been willing to assume the burden of making evaluations.'

All faculty members at Hamline are reviewed every year. As part of that review, each faculty member is required to submit student evaluations of three classes during the year or to be visited twice a year by each of two visitors from an evaluation team of eight experienced teachers. The faculty member, at his option, may also submit his own self-evaluation, and/or an evaluation of his work by a colleague. The most widely chosen option is the student evaluation, though a number of faculty use all the evaluation procedures.

There have been recent efforts on a number of other campuses (e.g., the University of Toledo, New Mexico State University, Northern Illinois University, Western Kentucky University, and University of California at Davis) to identify characteristics of teaching that predict teaching effectiveness. In each of these efforts teaching effectiveness *was not* measured in terms of what students had learned. Each of the studies focused on measuring teacher behavior in terms of traits and skills as such—divorced from the effect of those traits and skills and student achievement.

With reference to teaching traits and skills (not teacher effectiveness), the above research tended to provide support for the following:

- Faculty and students tend to agree in their rating of teachers in terms of specific teaching skills.
- The mere allocation of time among academic pursuits and the performance of professional activities associated with teaching does not assure that instruction is effective.
- Ratings of competency in teacher functions in the classroom by faculty colleagues should be used as a supplement, not as a substitute, for ratings by students.
- Student ratings of teaching assembled by faculty should be made available to the students; otherwise students can be expected to collect and disseminate their own data.
- Relatively short student rating forms (15 to 40 items) are preferable to longer forms.
- Student rating forms should be used to supplement, but not to substitute for, other kinds of evaluations.

There are many arguments for and against student evaluation. It may be helpful for future reference to note that the evidence supports the assertion that students are competent to measure the level at which the teaching skills are performed in the classroom. Do student evaluations have a positive and continuing effect upon improving teaching? The evidence is not conclusive. Eble notes in his Special Report (1970) that initially student evaluations call attention to teaching and aid in the dialogue that could lead to improved teaching. At the same time, evaluation may create defensive postures that harm good teaching and may even work specific harm on certain individuals. However, Eble notes that in the evaluations he observed, many of the harmful effects were anticipated and guarded against.

Eble also offers some opinions on the reliability of student evaluations and the advisability of using them. He states that although undergraduate student evaluations are limited in the information they can provide, undergraduates do have competence to judge in some important matters and are able to shed some light on such matters. In addition, while student evaluations are limited, they probably are less limited than other means such as classroom visitation by colleagues.

IV. A MODEL TO EVALUATE TEACHER EFFECTIVENESS

The study of teacher effectiveness is a search for the relationship between teacher behavior and the effects of this behavior on the knowledge, feeling, and performance of students. Biddle (1964) points out the relationship between teacher traits and teacher behavior and states it is unlikely that significant advances will be made about understanding teacher competence without a clearer picture of teacher behavior and its effects. Teaching effectiveness moves beyond the description of a teacher's traits or the methods (teaching decisions, strategies, etc.) used by a teacher, to the examination of these factors in relation to what the student learns; in other words, the study of teacher effectiveness considers the extent to which the purposes of teaching are realized. Therefore, teaching effectiveness means appraising the learning of students. A college student may develop an understanding about the structure of an atom; he may learn mathematical operations and proofs; or he may acquire new intellectual and aesthetic interests. These acquired understandings and interests are examples of cognitive and affective learnings that students attain at a university. Obviously, there are many other things learned that are beyond the prescribed curriculum. The university environment stimulates many opportunities for the informal cultivation of tastes,

knowledge and social roles. The question of teacher effectiveness, however, rests in the analysis of the progress students make in the acquisition of knowledge, and ways of thinking and acting that their instructors were trying to teach.

In summary, anyone who sets out to study teacher effectiveness and to construct an instrument for the purpose of collecting data on teacher effectiveness must recognize that measures of effectiveness are directly related to the performance of students. Therefore, procedures established to guide the collection of information on teacher effectiveness would appropriately include ways of determining the results of a teacher's efforts as expressed in some form through the demonstrated knowledge and abilities of his or her students.

Tyler (1960) has identified the generic tasks that are associated with obtaining a direct measure of teacher effectiveness. These tasks can be simply stated; however, a commitment to the tasks is considerably more involved. There are basically two generic tasks involved in obtaining a direct measure of teacher effectiveness: the first task is the clarification of the instructional objectives; and the second task is the measurement of the extent to which the instructional objectives have been achieved by the students. These tasks can and must be performed if the results of the efforts of teachers are to be measured.

Defining instructional objectives

The first difficulty faced by a university teacher, or a university faculty, is stating instructional objectives clearly enough in terms of behavior that the students are learning so that evaluators know what to assess when judging the degree to which students have acquired these behaviors. It is no secret that many statements appearing in college and university syllabi, course outlines, or catalogues are listings of content areas that the teacher will cover, not statements of the things students are expected to learn. Often students in different sections of the same course learn quite different things, even though the teachers of the different sections use the same list of course topics. One teacher may have the students memorize certain dates, formulae, facts, names, and definitions; another might want students to discriminate among concepts in analyzing situations and issues under these topics; a third might want students to select and apply the necessary principles to solve problems; and a fourth might want students to develop an attitude toward the subject that will be expressed in a continuing interest beyond the course. Simply listing topics does not indicate which of the above or other kinds of behavior are teaching goals for any particular course.

Tyler (1960) suggests that each teacher needs to answer such questions as:

> What things am I really trying to get students to understand? What kinds of abilities and skills in thinking, analysis, problem-solving, and the like am I really trying to develop? What kinds of competence in reading, or writing or mathematical operations am I actually seeking to help students to acquire? Am I trying to arouse certain intellectual and aesthetic interests, and, if so, what are they? Am I trying to develop some study habits or practices that will aid the student in his continuing development, and, if so, what ones? Am I seeking to instill certain attitudes, appropriate to understanding or appreciation of phenomena in this field, and, if so, what are they?

Until teachers can state in clear terms the different types of student behavior they wish students to acquire through their teaching, a sound evaluation of their teaching cannot be made because the evaluators will have no common criteria for judgment.

Evaluation requires not only a knowledge of students' behavior that teachers wish their students to acquire, but each kind of behavior must be clearly described so that it can be determined when the student has attained that objective. Teachers often say they are not trying merely to teach facts or simple skills, but are trying to develop "understanding" of the topic. However, many teachers are hard pressed when asked what they mean by "understanding" when they attempt to differentiate it from memorization of data. Other teachers have difficulty defining what is meant by such terms as "aware of," "appreciation for," etc. For example, some teachers believe that students have an "understanding" of certain concepts and principles if they can *describe* them in their own words; or if they can *distinguish* among examples of them; or if they *apply* them in analyzing and solving problems. Teachers might accept different student behavior for an expression of "understanding"; yet, if they are able to state their objectives as such, these statements can be used as a basis for evaluation, since it is then known what is being looked for. To determine how effective teaching has been in the development of "understanding," each instructional objective should be similarly stated.

It is important to keep in perspective the role instructional objectives play in the total educational experience of students. The objectives for a course must support the instructional objectives (goals) of the department. In similar manner, the department's objectives must support the objectives of the division. To assure that the students in different sections of a course acquire the necessary competence to continue work toward achieving objectives established by the department and reflected in other courses, it is necessary that the teachers of different sections agree upon instructional objectives for the course—*in terms of observable student behavior*. Only by such agreement will students in all sections have the prerequisite competencies to succeed in the next course that supports the objectives of that department. Thus, besides the need for course objectives to be stated in terms of observable student behavior, it is important that the objectives be agreed upon by colleagues who are teaching different sections of the same course. Obviously, if a department's objectives are stated in terms of what the students will be able to do as a result of the department's efforts, the task of colleagues reaching agreement on instructional objectives of a course is made easier.

After the individual instructor has formulated agreed-upon course objectives, the second step is to select ways of determining whether or not students have attained the competencies described in the instructional objectives. The means for gathering evidence of student achievement in large part will be determined by the objectives of the course; that is, the desired behavior of the students that has been described by the teachers as reflecting "understanding." Although evidence of student understanding in some disciplines can be obtained by means of paper and pencil tests; these means will not be satisfactory in other disciplines. However, in all disciplines it is highly important that the tests (term papers, recitals, lab reports, etc.) actually require the students to exhibit the type of competence that is described in the instructional objectives. If "understanding" in the instructional objectives is defined as the student being able to select and apply the necessary concepts and principles for solving problems, then the tests would not have performance agreement with the instructional objectives if they require the student to be able to select and name the necessary concepts and principles.

Tyler (1960) speaks to this point in the evaluation of teacher effectiveness:

> The evaluation of the results of teaching requires more than one appraisal of the students' behavior. Learning is the acquisition of new ways of behaving. It involves changes in the students' reactions. Hence, we shall need at least two appraisals, one toward the beginning of a course or the college experience and another toward the end in order to see what changes have taken place and thus to estimate the effectiveness of the teaching in stimulating and developing learning. Furthermore, many of us are interested in the permanency of learning and we shall, therefore, want to make a further appraisal of the students' behavior some time after the course is completed, or after they have graduated from college, to find out the extent to which the learning which was noted while the students were taking the course or were in college is still evident some time later or whether a considerable amount has gone.

Options for teacher evaluation

These are at least three options for initial action in proceeding to determine instrumentation and procedures to guide the study of teacher effectiveness at the university or college:

1. *Direct Measurement of Teacher Effectiveness.* This option requires that attention be given to the collection of data about class performance level. Performance measures such as formal tests, interpretive papers, creative efforts, etc., would be used as criteria. The review of these efforts would be made by the teacher's colleagues in accordance with procedures established by the department. Clearly, student efforts must be considered in terms of classroom goals, and the relative sophistication of the students with whom the instructor has been working should also be taken into account.

A plan that immediately moves to a direct measure teaching effectiveness, however, may not be practical. Before it can be applied, the faculty of each department must do a great deal of study and decisionmaking about the program. Agreement on total program and the objectives of courses within the programs are needed; also, the means by which initial information on student performance is obtained must be designed, along with the means for determining the competence of students at the end of the course. The extensiveness of the direct-measure method makes it prudent to suggest that this approach be seen as a long-range goal to be used initially in conjuction with other options.

2. *The Construction of a Student Rating Instrument.* The second option involves the use of a student rating form as the primary source of data. This option has most frequently been selected by universities in the past several years. Student rating forms and instrumentation, as described earlier, are used as indirect measures of teaching effectiveness. Such instruments are typically used to record student perceptions on a host of teacher traits, teaching methods, and class organization, as well as taking inventory of the attitude of feelings of the student(s) about the course. It has been stated that these concerns are not of themselves direct measures of effectiveness, but may be used as high-inference indirect measures. The research on the use of indirect measures as predictors of teacher effectiveness, is, at best, inconclusive. However, in cases where instruments were constructed in accordance with the purposes of the evaluation process, and where there was participation by both instructors and student representatives in this effort, the use of student rating forms turned out to be a constructive factor in teacher evaluation. Furthermore, instructors gain the benefit of student thinking, and students have been described as taking the process seriously. Beyond an inventory of student perceptions about individual instructors or courses, student ratings may offer a faculty important information about its progress in articulating its own predetermined program goals.

The analysis performed at the University of California, Davis, included a factorial analysis of rating forms, and concluded that student rating forms are usually comprised of items centering around five major factors, noted earlier: Analytic/Synthetic, Organization/Clarity, Instructor-Group Interaction, Instructor-Individual Student Interaction, and Dynamism/Enthusiasm. These categories can be used as an instrument for student rating of teachers by any college or university; or categories can be made up from a variety of sources to reflect the response modes that meet a particular institution's requirements.

Equally important is the development of procedures and guidelines for the use of the instrument and the purposes of such a measure. This step would allow student involvement at the level of an indirect measure of teaching effectiveness and would provide a description of teachers' traits and skills as well as a decription of the attitudes and feelings of the students about the courses.

3. *The Combination of Student Ratings and the Review of Student Products.* This third option includes

the design and construction of a student rating instrument, as well as the promotion of the student performance concept and explicit statements as to the proposed use of the findings. This option would be considered the first phase of a long-range effort to implement Option 1:

- Design of a student rating form – An Instrument should be designed to enable students and faculty to record student perceptions about their courses and teacher performance. This instrument would be constructed while an institution considered the more long-range implications of Option 1.
- Communication of the concept of direct measurement of teacher effectiveness – The concept of teacher effectiveness, defined as the study of the outcomes of teaching, would be fully examined and communicated in its intent and meaning to the academic divisions of an institution. Since a long-range study of teaching effectiveness would require depth analysis of curriculum by the departmental and program areas, faculty groups would be provided with a set of guidelines and a list of instructional resources to aid them in defining programs in terms of student learning behaviors. This option would include initial efforts by the faculty of each department to identify its desired student outcomes in reference to the division's goals. In addition, each department would determine its own methods of appropriating and examining student efforts, and these methods, in turn, would be communicated to the division chairman. A study of the range of approaches formulated by each division will aid the development of institutional guidelines for a direct, outcome-oriented approach to teacher effectiveness.
- Use of findings – In the best sense, student ratings of faculty and faculty measurement of the outcomes of teaching must provide the conditions for teaching improvement. Hence, this third option would include a policymaking body – consisting of faculty and students – to determine how to use these findings. A necessary, but not sufficient condition for establishing the criteria for teacher evaluation is the continual communication among the groups involved. If such communication is not maintained, any of the evaluation procedures could become a divisive factor. The need for a good start and the complexity of teacher evaluations suggest that Option No. 1 should be the primary goal, with Option No. 3 as the practical choice for an institution's first-year effort. When implementing Option No. 3, an institution should simultaneously initiate efforts toward Option No. 1. The following recommendations to implement a long-range program similar to Option 1 are submitted with the understanding that their implementation would require considerable faculty and student effort – a prerequisite to directly measure teacher effectiveness.

Recommendations for long-range direct measurement of teaching effectiveness

Five steps are recommended for the long-range study of teaching effectiveness. These steps are summarized as follows:

- Each discipline or program area develop a comprehensive statement of its program.
- The faculty in the various disciplines devise a clear statement about the range of performance expectations held for students in the faculty's component of the program.
- With the clear statement of objectives as a reference point, the faculty of a discipline establishes consensus as to general and specific definitions of their program.
- Obtain baseline data via a premeasure when students enter a course.
- Measure the growth of students in a given course by the objectives of the course. Use this measure of student achievement as a direct measure of teacher effectiveness.

The implementation of these steps would take the following form:

1. *Comprehensive statement of program.* Each discipline or program area would develop a comprehensive statement of its program. This includes both a general affirmation of departmental and program goals, as well as a more detailed definition of the contribution to be made by each component to the total program. This step should provide a clear understanding of the aims of the discipline, as well as the expected contribution of courses, field work, seminars, and independent study arrangements. This information would be available for the analysis of faculty, students, and administration.

2. *Clearly stated objectives for each course.* The general definition of a discipline's program is an important step, but is not complete enough to enable the evaluation of teaching in terms of teacher effectiveness to proceed. Consequently, the faculty in the various disciplines and programs should continue to work on more specific interpretations of each course or program component for which faculty members are responsible. This step would include the itemization of course content in a more exact form than that usually distributed to students in course outlines. This step also includes descriptions of what students will be able to do as they complete parts of or the complete program component. In this way, faculty and students will have a common basis for understanding the inclusion of a particular course within a program. Both

students and faculty will benefit from a clear statement about the range of performance expectations held for students as defined for that component (course, etc.) by the faculty.

This procedure may seem overly complex; even worse, it may seem an abridgement of the instructor's freedom in the development of his teaching. From another perspective, however, the completion of program definition, including specific course elements and performance expectations for students, may actually increase opportunities for creative teaching. All instructors will not, nor should they be expected to teach in the same way. Given rather specific common agreement among faculty as to the purpose and composition of a course, individual faculty may devise widely divergent approaches to instruction. Certain students, for example, will demonstrate great competence in a course in the early stages of instruction. The recognition of gifted students as well as those who need help, when devising the basic elements of a course, will enable the teacher to make better decisions for the class both as a group and as individuals. Precise program definition and planning is not a deterrent to creative teaching and is clearly not inimical to student learning. The general definition of a departmental program as well as a more specific course or program component structure is never a completed process. The stated program is a guide and should be viewed as a tentative definition of things; that is, as experience is gained with a program, and as additional students and faculty are added, it becomes necessary to constantly assay the meaning and relevance of the departmental program. This reassessment should proceed in terms of overall goals (e.g., to prepare students for the medical professions, research in industry, or graduate work, etc.), as well as in terms of the component courses that comprise the program. Program definition is not immutable, but a base which is subject to change. It does, though, become the stated intention of the departmental faculty, and it is against these stated intentions that faculty effectiveness might be reasonably studied.

3. *Faculty consensus about program definition.* The faculty of a discipline must arrive at a consensus about the general and specific definitions of their program. Most disciplines or program areas have curricular committees that serve the purpose of stimulating curricular change, revising course proposals, and many other related matters. Instructors should contribute to and be informed about the program in effect and should be made aware of the procedures to be followed if they decide to change the focus or structure of a course. This kind of communication is important when judgments on teaching effectiveness are related to a teacher's ability to nurture certain kinds and levels of student performance. Again, the definition of "program" should be flexible: in a college or university there should be ample opportunity for innovative approaches. Where a basic and substantive change is to be acted upon, however, it is reasonable that such a change should be discussed in the hope that consensus can be reached about the implications of the change vis-á-vis the department's total program. Consensus among the faculty regarding its program provides the teacher with information about departmental expectations in his area of instruction.

4. *Obtaining baseline data.* Since there will be a range of ability and background represented among students in different classes, an assessment should be made of the range of student competence in a given teacher's classes. It is expected that in time this variable would even out, so that classes would reflect the general range of ability present in the institution's student population. If the teacher's effectiveness is viewed in regard to the intellectual growth of students, then there must be a system designed to gain information about the level of competence at which students begin a particular course. In a sense, the mechanism designed to elicit this kind of information becomes a premeasure; and the nature of the premeasure, whether it be a pencil and paper exercise(s) or a demonstration of some type, is best determined by an examination of the goals of a course and the nature of the discipline of which it is a part. The purposes of the premeasure are: (1) to provide the teacher with information on the background of his students; (2) to help the teacher make instructional decisions that are appropriate to his students; and (3) to establish baseline data in the measurement of student progress in the course. If a group of teachers are teaching different sections of the same course, the same premeasure could be employed. The information collected on the premeasure would provide helpful data on the range of student abilities, as well as indicate the starting point from which a class could begin its study. The design of the premeasure should be determined by the instructors who teach that part of the program; a variety of opinions could be gathered to increase the likelihood that the premeasure would correspond to all major course aims.

5. *Measurement of student achievement.* An assessment of the change in student competence should be obtained either at the end or at pivotal points during a given course. Using the baseline data obtained by premeasures, the effectiveness of instruction can be measured in terms of the extent to which student performance has changed or in terms of the extent to which students have learned what a teacher intended to teach. A procedure for measurement of such change has been detailed by Cook (1971) in *Direct Measurement of Teaching Effectiveness: A Challenge*. Agreement about performance criteria using the objectives originally established as goals for the course and the assessment measure (i.e., postmeasure) is critical when measuring teacher effectiveness. Group efforts similar to those related to the premeasure are recommended. If colleagues teaching the same course agree upon the objectives of the course, their agreement on the assessment items would be significant. Once the data describing

the intellectual growth of students is available, the teacher's effectiveness may be established—taking into consideration the variables pertaining to the data.

Implementation of these and other complementary steps requires a commitment on the part of the university community—students, faculty, and administration—of significant dimensions; however, the result of such a commitment could be rewarding beyond the dreams of present teacher evaluation committees.

V. SELECTED BIBLIOGRAPHY

Banathy, Bela H. *Instructional Systems.* Palo Alto: Fearon Publishers, 1968.

Bentley, G.E., *et al. Report of the President's Committee for a Pilot Study In Student Evaluation at Princeton.* Princeton: Princeton University Press, 1968.

Biddle, Bruce J. "The Integration of Teacher Effectiveness Research." *Contemporary Research on Teacher Effectiveness,* ed. Biddle, B. J. and Ellena, W. J. New York: Holt, Rinehart, and Winston, 1964. Contains extensive bibliography.

Bloom, B. S., ed. *Taxonomy of Education Objectives, Handbook I: Cognitive Domain.* New York: David McKay Co., Inc., 1956.

Brickman, W. W. "The Professor and his Salary." *School and Society* 85:44.

Brock, E. H. "Rating Teaching Personnel for Advancement Purposes." *Journal of Engineering Education* 44:593-99.

Burkhart, Robert C., ed. *The Assessment Revolution: New Viewpoints for Teacher Evaluation.* Buffalo: State University College at Buffalo, 1968. Contains an extensive bibliography.

Cohen, A. M. *Objectives for College Courses.* Beverly Hills: Glencoe Press, 1970.

Commentary for Teachers of Science – A Process Approach. New York: Xerox Corporation, 1970.

Cook, J. Marvin. "Behavioral Objectives and Rate of Forgetting." Mimeographed. A paper presented at the American Educational Research Association's 1970 Annual Meeting, March 1970.

——. "Direct Measurement of Teaching Effectiveness: A Challenge." Mimeographed. Baltimore: University of Maryland, Baltimore County, Division of Education, 1971.

Cooper, Russell. "The College Teacher Today." In *Effective College Teaching,* edited by William H. Morris, Washington, D.C.: American Council on Education, 1970.

Davis, A. J. and T. A. Derbigny. "Toward Objectivity in Faculty Ranking." *School and Society* 81:137-39.

Deterline, W. A. "The Secrets We Keep from Students." *Educational Technology,* February 15, 1968.

Domino, George, "Interactive Effects of Achievement Orientation and Teaching Style on Academic Achievement." Available from: American College Testing Program, P.O. Box 168, Iowa City, Iowa 52240. ED 046 353. 9 pp.

Dressel, Paul L. *Evaluation in Higher Education.* Boston: Houghton-Mifflin Co., 1961.

Eble, Kenneth E. *Special Report on Evaluation of College Teaching.* Washington, D. C.: American Association of University Professors, 1970.

Eckert, R. E. and R. J. Keller. *A University Looks at its Program: 1942-1952.* Minneapolis: University of Minnesota Press, 1954. See especially chapters entitled "Faculty Promotion: Policies and Practices" and "Student Ratings of College Teaching."

Eiss, A. F. and M. B. Harbeck. *Behavioral Objectives in the Affective Domain,* Jackson. Michigan: National Science Supervisory Association, 1969.

Evans, T. H. "Evaluation of Engineering Teaching Ability." *Journal of Engineering Education* 45:708-12; and reply by F. W. Roys, 46:288-89.

Finn, Stanley R. ed., Higher Education Research Abstracts No. 8, Characteristics of the Effective College Teacher. Hempstead, New York: Hofstra University Center for the Study of Higher Education, September 1969, ED 032 841. 9 pp.

Gage, N. L. Address appearing in "Proceedings." *Research Résumé.* Burlingame, California: California Teachers Association, 1960.

Gagné, R. M. and N. E. Paradise. "Abilities and Learning Sets in Knowledge Acquisition." *Psychological Monographs* 75(1961):1-18.

——, *et al.* "Factors in Acquiring Knowledge of a Mathematical Task." *Psychological Monographs* 76(1962):1-19.

——, Robert M. *The Conditions of Learning.* New York: Holt, Rinehart and Winston, 1965.

——, *The Conditions of Learning.* 2d rev. ed, New York: Holt, Rinehart and Winston, Inc., 1970.

——, ed. *Learning and Individual Differences.* Columbus, Ohio: Charles E. Merrill Books, Inc., 1968.

Gowenlock, B.G. *et al.* "Assessment of Objectives in Chemistry Teaching." *Chemistry in Britain* 6:341-43.

Science—A Process Approach: A Guide for Inservice

Instruction. Washington, D.C.: American Association for the Advancement of Science, Commission on Science Education. Miscellaneous Publication 67-9.

Gustad, John W. "Policies and Practices in Faculty Evaluation." *The Educational Record* 42:194-211.

Hildebrand, Milton and Robert C. Wilson, *Effective University Teaching and Its Evaluation.* Davis: University of California, 1970. ED 039 860. 41 pp.

Hite, Herbert. "A Model for Performance Certification." *The Assessment Revolution: New Viewpoints for Teacher Evaluation,* edited by Robert C. Burkhart. Buffalo, N.Y.: State University College at Buffalo, 1968a.

——, "A Systematic Approach to the Analysis of a Non-Systematic Process," *The Assessment Revolution: New Viewpoints for Teacher Evaluation,* edited by Robert C. Burkhart. Buffalo, N.Y.: State University College at Buffalo, 1968b.

Hoetker, James. "The Limitations and Advantages of Behavioral Objectives in the Arts and Humanities: A Guest Editorial." *Foreign Language Annals,* May 1970.

Hoyt, Donald P., *Identifying Effective Teaching Behaviors.* Manhattan, Kansas: Kansas State University, December 1969. ED 039 197. 136 pp.

Jones, R. V., "Lap over Lap: A Learning Activity Package," Mimeographed. Turlock, California: Stanislaus State College, 1968.

Keller, F. S. "A Programmed System of Instruction," Paper presented to the Autumn Conference of the Pacific Northwest Association for College Physics, Seattle, Washington, 1968.

Kluss, Jill. "Performance Objectives and Learning Hierarchies: A Practical Approach." Mimeographed. Baltimore: University of Maryland, Baltimore County, Summer 1970.

Krathwohl, D. R., *et al. Taxonomy of Education Objectives, Handbook II: Affective Domain.* New York: David McKay Co., Inc., 1964.

Kurtz, E. B. "Help Stamp Out Non-Behavioral Objectives," *The Science Teacher* 32, no. 1.

——. and D. M. Bosanko. *Biology – An Individualized Approach.* Emporia, Kansas: Teachers College Press, 1970.

Langen, Thomas D. F. "Student Assessment of Teaching Effectiveness." In *Improving College and University Teaching.* Winter, 1966.

LeFever, H. M. and J. E. Taylor, *Genetics – An Individualized Approach,* Emporia, Kansas: Teachers College Press, 1971.

Leigh, Terry, *A Selected and Annotated Bibliography on Evaluating Performance of College Faculty Members.* Lexington, Kentucky: Kentucky University, 1969. ED 035 376. 30 pp.

Mager, R. F. *Preparing Instructional Objectives.* Palo Alto: Fearon Publishers, 1962.

——, *Developing Attitude Toward Learning.* Palo Alto: Fearon Publishers, 1968.

Mayhew, Louis B. "Current Practices in the Evaluation and Training of College Teachers." *Improving College and University Teaching* 15:5-10.

McKeachie, Wilbert J. *The Appraisal of Teaching in Large Universities.* Ann Arbor: University of Michigan Press, 1959.

——. *Research on College Teaching: A Review.* Washington, D.C.: ERIC Clearinghouse on Higher Education, November 1970. ED 043 789. 20 pp.

——, "Student Ratings of Faculty," *AAUP Bulletin* 55:439-44.

Nicholas, I. C. "Salary Schedules Are Based on Effectiveness of Teaching." *Nation's Schools* 57:52-56.

Pierrel, Rosemary. "The Evaluation of College Teaching: The View of an Administrator-Psychologist." Mimeographed. Providence, R. I.: Brown University, n.d. [1966].

Perry, Richard. "Criteria of Effective Teaching in an Institution of Higher Education." Mimeographed. Toledo, Ohio: University of Toledo, 1968.

Popham, W. J. and E. L. Baker. *Instructional Objectives.* Chicago: Rand McNally and Co., 1969. AERA Monograph Series on Curriculum Evaluation No. 3.

——, *et al. Systematic Instruction.* Englewood Cliffs, N.J.: Prentice-Hall, Inc., 1970.

——, *et al. Planning an Instructional Sequence.* Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1970.

Powell, J. P. "Experimentation and Teaching in Higher Education: Measuring Teaching Performance," *Educational Research* 6:188-89.

Preservice Science Education of Elementary School Teachers: Preliminary Report. Washington, D. C.: American Association for the Advancement of Science, Commission on Science Education. Miscellaneous Publication 69-11 (rev. ed., 70-5).

Principles and Practice of Instructional Technology. Palo Alto, California: General Programmed Teaching, 424 University Avenue, P.O. Box 402. 1970.

Prophet, C. W. *Laboratory Instructional Units for Biology of Populations.* Emporia, Kansas: Teachers College Press, 1970.

Ransom, J. D. and E. B. Kurtz *Field and Lab Biology: An Individualized Approach.* Emporia, Kansas: Teachers College Press, 1970.

Rosenshine, Barak. "Evaluation of Classroom Instruction." *Review of Educational Research* 40: 279. Contains an extensive bibliography.

Seelye, H. Ned, "Performance Objectives for Teaching Cultural Concepts." *Foreign Language Annals* 3:566-78.

Smith, B. Othanel and Robert H. Ennis, *Language and Concepts in Edueation.* Chicago: Rand McNally and Co., 1961.

Smith, G. Kerry, ed. *Current Issues in Higher Education, 1955. Proceedings of the 10th Annual Conference on Higher Education.* Washington: National Education

Association, 1955. See Group 26 chapter on "How Can Institutions Evaluate the Effectiveness of Teaching and Other Faculty Services?"

Smock, Richard and Edward Kelly, "The Evaluation of Collegiate Instruction: How to Open a Closed System." Washington, D.C.: American Educational Research Association, June 1970. ED 046 329. 13 pp.

Spalding, W. B. and E. Hummel. "New Answers to an Old Question: Problem of Teacher Salary Schedules and Merit Rating." *American School Board Journal* 132:21-22.

Springer, E. K. "Evaluating the Services of the Faculty Member." *Higher Education* 7:117.

Steiner, Florence. "Performance Objectives in the Teaching of Foreign Languages." *Foreign Language Annals* 3:579-91.

Strong, M. "Merit Rating: Points to Consider for a Pilot Test." *The Clearing House* 29:402-04

Tiedman, D. V. *Teacher Competence and Its Relation to Salary*. Newton, Mass.: New England School Development Council, 1956.

Tyler, Ralph W. "The Evaluation of Teaching." In *Preparing College Teachers*, edited by Albright, A. D. and John E. Barrows. Louisville: University of Kentucky, 1960.

Tyler, Robert, *et al. Perspectives of Curriculum Evaluation*. Chicago: Rand McNally and Co., 1967. AERA Monograph Series, vol. 1.

Walbesser, Henry H. *Science – A Process Approach: An Evaluation Model and Its Application, Second Report*. Washington, D.C.: American Association for the Advancement of Science. Commission on Science Education, Miscellaneous Publication 68-4.

——, Kurtz, Edwin B., *et al. Constructing Instruction Based on Behavioral Objectives*. Stillwater, Okla.: Oklahoma State University, 1971. Available from Dr. Larry Goss, Department of Aerospace and Mechanical Engineering, Oklahoma State University, Stillwater, Oklahoma.

Werdell, Philip, *Course and Teacher Evaluation*. Washington, D.C.: United States National Student Association, 1966. Contains an extensive collection of student rating forms.

Winthrop, H. "Worth of a Colleague." *Improving College and University Teaching* 14-262-64.

Witheiler, Paula and Harold E. Yuker, *Course Evaluation at Hofstra University, 1969*. Hempstead, New York: Center for the Study of Higher Education, March 1970. ED 040 661. 24 pp.

Young, David B. "Developing Practicum Experiences for Prospective Teachers." Mimeographed. Baltimore: The University of Maryland, Baltimore County, 1970.