# Measuring teacher effectiveness

Dorothy R. Bleyer

## Recommended Citation

# Evaluation of instruction is a complicated activity.

# Measuring teacher effectiveness

## By Dorothy R. Bleyer

The Undergraduate Teaching and Curriculum Committee at Southern Illinois University appointed by the vice president for academic affairs was charged to develop guidelines to be used by the various schools and colleges within the university in formulating procedures for evaluating instruction within the academic units. As a member of the committee, the writer prepared this article which reviews and summarizes recent literature on the evaluation of teaching effectiveness at the tertiary level.

The purpose of this endeavor is not to investigate whether such evaluation should occur. For indeed, it is axiomatic to state that evaluation of instruction always takes place. Teachers constantly are evaluated by students, administrators, colleagues, and the public. It is rather the purpose of this article to provide information which will assist administrators and ad hoc committees for review and evaluation in answering the following questions.

1) Shall the evaluation of instruction be systematized by the development of evaluative methodology?
2) Who shall be the evaluators?
3) What criteria shall be used for measuring teaching effectiveness?
4) How shall the information be collected and processed?
5) How shall the results of evaluation be used?

Teacher evaluation has been with us for as long as teaching has occurred; however, the methods of evaluation and the emphases placed upon it have changed with social and economic factors throughout different periods of our history. Some writers place the search for a valid index of teaching skill among Mankind's Perennial Quests, third in order after the search for the Holy Grail and the Fountain of Youth.

We are still in the **Decade of Accountability** (Austin, 1971). Watergate inquiries, new guidelines for the use of human subjects in research, "Nader's Raiders," cost accounting in the schools, environmentalist groups—these and more reflect the growing concern over the degree to which individuals and institutions should be held responsible for the consequences of their decisions and actions. That teachers need to be accountable is no longer in question. The present debate is over what approaches to accountability are appropriate for the assessment of teaching effectiveness.

Philosophical views of proper methods of teacher evaluation vary from the very informal, subjective, qualitative assessment of a professional (Biddle and Ellena, 1964) to the rigidly structured statistical approach which closely resembles the management-by-objectives technique used by industry (Bolton, 1973). Both of these views have substantive studies and writings to support them.

There are, however, several factors existing at the present time which seem to call for the pragmatic response of some type of formal evaluation of instruction at all levels:

### 1) Governmental controls

The public discontent regarding educational quality has manifested itself in some states as legislatively enacted educational assessment programs. In California, the legislature enacted a mandatory teacher-evaluation system (The Stull Act) for public schools there.

Other governmental agencies at the state and national levels, such as the Illinois Board of Higher Education, Division of Adult, Vocational and Technical Education, and HEW, which control or influence allocations of funds to educational institutions, increasingly are requiring evidence of quality performance which, in many cases, involves teaching competency.

### 2) Institutional policies

Internal pressures also are mandating evaluation of teaching. **The Guidelines for 1976 Promotion and Tenure Recommendations** prepared by the vice president for academic affairs at Southern Illinois University at Carbondale state, "The first step in promotion and tenure decision making is an evaluation of teaching effectiveness . . . It is vital that information concerning teaching effectiveness be included as part of the evaluation."

In an article in a recent issue of the student newspaper, **The Daily Egyptian,** SIU-C President Warren Brandt lists mandatory student evaluation of instructors as one of the important campus issues. Other colleges and universities report similar efforts to require evaluation of instruction.

### 3) Sophistication of research design

The effectiveness of an instructional treatment may be measured by student performance. Since the

outcome (student performance) of any instructional event in which a teacher is involved is influenced by the teacher himself, the individual teacher must be considered an instructional treatment and evaluated as such. Much of the teacher effectiveness research carried on during this century has been directed toward the isolation of some kind of measure of instruction that could be used as a dependent variable. It was hoped that such a dependent variable could

then be used to discern the relative influence of selected independent variables.

**4) Professionalization of teachers**

Because teacher evaluations arrived at in a very vague and perfunctory manner were becoming the basis for salary increases, and in line with their developing professionalism, the NEA, in its resolution in 1961, recognized that "it is a major responsibility of the teaching profession, as of other profes-

**Fig. I. Numbers in parentheses indicate references.**

| SOURCE | STRENGTHS | WEAKNESSES |
|---|---|---|
| Administrators | a) Meet with less resistance because they are traditional and expected. (10, 11, 14)<br>b) Convenient to secure. (11)<br>c) In a position to act upon results. | a) If not engaged in actual teaching for some time, he may not be capable of judging teaching competency.<br>b) May base his evaluation on indirect information, due to lack of time for observation.<br>c) May be influenced by "halo effect." (3) |
| Students (Present)<br><br>Increasing in use.<br><br><br><br><br><br><br><br><br><br><br>(Former) | a) Results tend to show consistency. (10, 14)<br>b) In direct contact with teaching process. (14)<br>c) Studies show that college teachers are responsive to students' ratings (subsequent evaluations show improvement). (15)<br>d) Two-way evaluation at college level develops a mutual feeling of trust. (1, 15)<br>e) Positive addition to communication process.<br>f) Correlate highly with administrative ratings. (9)<br>g) In keeping with "consumer satisfaction" concept. (5)<br>h) Increased validity when fear of reciprocity is removed. | a) Considerable "halo effect" found. (6)<br>b) Tangential factors (grades, age of instructor, etc.) may affect ratings. (4, 12)<br>c) Some studies show that student ratings at higher education level correlate negatively with student learning gain. (6)<br><br>d) Costly and difficult to secure. |
| Peers | a) Most aware of teaching conditions and expectations.<br>b) Exchange of ideas may contribute to improvement of instruction.<br>c) Ranks by peers give valid results. (10) | a) "Halo effect" influences peer ratings. (10)<br>b) Peers dislike evaluating colleagues for salary, promotion, and tenure decisions.<br>c) No time for observation and conferences. |
| Self | a) Self-identification of weaknesses should lead directly to improvement.<br>b) Agrees with professionalism resolution. | a) Tendency for instructors to overrate themselves. (10)<br>b) Shows negligible correlation with administrative and student ratings. |
| Outside experts | a) Trained observers. (2)<br>b) Objective evaluation. (2)<br>c) Each teacher evaluated by same standard. (2) | a) Costly. |

sionals, to evaluate the quality of services." (**NEA Proceedings,** 1961, 189-93)

Potentially, administrators, students, peers, self, an outside group, or any combination of these can engage in formal evaluation. All of these groups informally evaluate teachers now. Each of the potential evaluators brings a different perspective to the evaluation—a perspective which may limit or enhance the validity of his assessment.

A review of the literature shows there are strengths and weaknesses of each source of evaluation. These strengths and weaknesses are summarized and presented in tabular form. (See Fig. I)

Establishing the criteria for assessing teacher effectiveness may be the most complex element of the entire evaluation process. The writers are in general agreement that there is diversity in criteria according to level of instruction, type of subject matter, situational constraints, in addition to other factors. McNeil says, "Increasingly those in college are recognizing that good teaching is not a phenomenon, but a class of diverse phenomena, with various criteria and sometimes incompatible traits." (McNeil, 1971, p. 27)

Most sources consulted included the following as possible criteria for teacher evaluation: professional qualifications, techniques of instruction, teaching results (measured by student performance), classroom management, social relations (attitudes toward students, colleagues, administrators), and personal characteristics. It is a general recommendation that the criteria for evaluation be developed jointly by those (or their representatives) who are to be involved in the evaluation process, using a systematic and comprehensive approach. Ryans (1957) found that when criteria were developed from empirically supported and rational considerations, they were likely to be relevant and usable.

In selecting measures for evaluations, a major rule of thumb is "select the instrument that best fits your purpose," i.e., identify the measurement techniques and strategies that provide the data desired. Practical considerations in the choice of instruments are the (1) cost factor, (2) time factor, and (3) source factor. Other considerations in the choice of instruments are relevance, reliability, validity, and ease of administration.

Instruments which are being used with varying degrees of success include rating scales, structured and non-structured comments, systematic observation, pupil-test performance, follow-up studies of students, and video tape or audio tape recordings of classroom presentations. There is overwhelming evidence that the first two are used most often and possibly least reliable. Their advantage is the low cost and the ease of administration. Reliability of rating scales may be increased by including low-inference items and by training the evaluators.

Systematic observations minimize the influence of observer bias. The observer records whether a specific behavior occurred but makes no value judgment as to whether the behavior is "good" or "bad." Use has shown this instrument to be reliable by a high degree of interobserver agreement. There are weaknesses of this instrument. Negative factors not accounted for may be so potent that they cancel out the teacher's positive action. Another weakness of all observation instruments is that tendency-type research studies are being used to make particular judgments about an individual teacher. Most writers feel pupil-test performance should not be used for purposes of teacher evaluation as studies indicate that pupil-test performance tends to be a function of intelligence rather than teacher effectiveness.

A follow-up study of former students in the form of a questionnaire might be one of the most valuable measures of teacher competence. However, the relatively high cost and difficulty of implementation has limited its use.

A rather recent innovation in teacher evaluation is video and audio tape recordings of mini-presentations in the classroom. This measure has real potential for use in self-evaluation for purposes of instructional improvement. This evaluation tool was used by the writer in a mathematics class during the previous semester along with feedback from a student evaluation team. The team of students volunteered to meet regularly with a resource person from the University's Learning Resources Center to discuss the instructor's strengths and weaknesses. The learning specialist relayed the students' remarks to the instructor with suggestions for improvement as appropriate. The exercise was found to be constructive and non-threatening.

The evaluation of teachers may serve many purposes: to improve teaching, to reward superior performance, to supply information for modifying assignments, to protect both the individual and the institution in legal matters, and to generate plans for individual growth and development. There seems to be general agreement among educators that improvement of instruction is the most important purpose. Teachers' reception to formal evaluation efforts tends to be far more positive if a formative evaluation program is developed which includes opportunities and facilities to correct weaknesses and deficiencies. It is considered virtually unethical to subject teachers to the intense scrutiny of current evaluation procedures without offering developmental programs for their use.

Since there is increasing pressure from boards of education and taxpayers to reward superior performance, evaluation may serve to identify those deserving salary increases based on merit. However, writers claim this use of evaluation is in direct conflict with the viewpoint of the majority of teachers. They suggest the teachers' major objection to evaluation for this purpose stems from the subjective nature of most evaluation systems. The results of a formalized evaluation process surely are more objective and to be preferred over other measures in use at the present time. In a recent study reported with tongue-in-cheek, Clifford Hooker (1978) found physical proximity to the merit rater (distances between offices) to be a better predictor of salary increase than teaching load, quantity of publications, or number of graduate students supervised.

Information gathered in the evaluation process may be used to modify teachers' assignments, either by promotion, changes in teaching load, or release. While these are necessary activities in educational institutions, when evaluation emphasizes the summative aspect, it tends to be viewed negatively and to undermine staff morale. Some writers contend, however, that better staff morale and a better instructional program result from a well-defined system of evaluation and orderly dismissal procedures for incompetent teachers.

Emphasis on the legal aspects of teacher evaluation can be viewed negatively by teachers unless they realize that their own protection against unjust charges as well as that of the institution can be assured by documentation of performance.

## SUMMARY

Researchers agree evaluation of instruction is a complicated activity, difficult to conceptualize fully in all its ramifications, and even more difficult to implement with sound substance and fair process. The writings reviewed by the author agree upon the following general recommendations:

1) that evaluators using standard techniques recognize their weaknesses and interpret the results accordingly;

2) that researchers continue to study and refine the more promising techniques;

3) that all persons who are to be involved in the evaluation system also participate in the development of it;

4) that the evaluation process include multiple, rather than single, indicators of a teacher's skill, and

5) that the emphasis be on helping an individual to improve his contribution to the learning experience.

## SOURCES

Austin, Gilbert. "Educational Accountability—Hallmark of the 1970s." **The Science Teacher,** April 1971, p. 26. (1)

Biddle, Bruce J., and Ellena, William. **Contemporary Research on Teacher Effectiveness.** Chicago, Ill.: Holt, Rinehart and Winston, 1964. (2)

Bolton, Dale L. **Selection and Evaluation of Teachers.** Berkeley, California: McCutchan Publishing Corp., 1973, p. 24. (3)

Broadwell, Martin. "Broadwell on Instructor Evaluation." **Training in Business and Industry.** October 1973, pp. 25-28. (4)

Diener, Thomas J. "Can the Student Voice Help Improve Teaching." **Improving College and University Teaching.** Winter 1971, pp. 35-37. (5)

Good, Thomas; Biddle, Bruce; and Brophy, Jere. **Teachers Make a Difference.** Chicago, Ill.: Holt, Rinehart, and Winston, 1975. (6)

Hooker, Clifford P. "A Behavior Modification Model for Merit U.," **Phi Delta Kappan,** 59:7, March 1978, p. 481. (7)

Jacobson, Joan. "Should Students Evaluate Teachers?" **Today's Education.** May 1973, p. 49. (8)

Krull, George and Crooch, Michael. "Measuring Teaching Effectiveness." **Collegiate News and Views.** Fall 1973, pp. 9-13. (9)

McNeil, John D. **Toward Accountable Teachers.** Chicago, Ill.: Holt, Rinehart, and Winston, 1971. (10)

Popham, James W. **Educational Evaluation.** Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975. (11)

Rodin, Miriam, and Rodin, Burton. "Student Evaluation of Teachers." **Science.** November 1972, pp. 1164-1166. (12)

Welch, David; Richards, Fred; and Richards, Anne. **Educational Accountability: A Humanistic Approach.** Fort Collins, Colorado: Shields Publishing Co., 1974. (13)

Wicks, Larry. "Teacher Evaluation." **Today's Education.** March 1973, p. 42. (14)

"Values of Evaluation by Students." **Intellect.** April 1973, pp. 416-417. (15)

**Numbers in parentheses correspond with numbers in Figure I.**