Clinical Sociology Review

Volume 2 | Issue 1 Article 18

1-1-1984

Evaluating Teaching Effectiveness

Thomas J. Rice

Evaluating Teaching Effectiveness

Thomas J. Rice

The literature on effective teaching is, like most other scholarship on teaching, becoming both sophisticated and voluminous. However, empirical findings on teaching effectiveness are notably inconsistent (Wagenaar 1979) and are not likely to support reliable generalizations for some time into the future. Given that the assessment of teaching is becoming central to administrative decision making in this era of academic "hard times," the following questions must be addressed: (1) What assumptions lie behind the evaluation of teaching? (2) Why evaluate in the first place? (3) What do we know about evaluation at the present time? (4) Who should do the evaluation of teaching? (5) When should it be done and how often? (6) How can it be done to ensure both validity and reliability?

Assumptions

Before any evaluation can begin, it seems critical that all participants articulate and share a set of assumptions from which the evaluation process may go forward. A number of scholars (for example, Eble 1976; McGee 1974; Wagenaar 1979) have advanced inventories of assumptions. Several seem to be rather commonly accepted; for example, it seems to be agreed upon that (1) effective teaching is a socially learned competence, not an inborn attribute; (2) effective teaching can be analyzed into its component parts and reproduced in a range of settings; (3) major components can be measured through research methodologies; (4) while "great" teachers may have some qualitatively ascribed characteristics, "good" teachers can achieve this assignation through conventional learning formats (that is, analysis of methodologies and feedback toward a change in behavior); (5) no single method of evaluation is adequate to the task of assessing the effectiveness of teaching, but a triangulation, or multimethod approach, should be employed; and (6) teaching effectiveness should not focus on the professor alone, but on the interaction of forces and factors in the social space that encompasses the learning context.

Reasons for Evaluations

The question "why evaluate?" is answered most thoroughly by Grasha (1977). This question is more complicated than any other question posed, since it involves both a practical pedagogical aspect and a political consequence. Grasha assumes that feedback improves performance; to the degree that evaluation provides feedback, teaching should improve. The application of this relatively

simple principle is, however, subject to a great deal of faculty resistance because of the political and professional sensitivity of such processes.

We evaluate teaching because it: (1) provides for an atmosphere of participation and cooperation between faculty and students in the quest for excellence in teaching; (2) sets standards for excellence that are shared in the community of intellectuals and that aid in the sifting-sorting processes of the "academic marketplace" in ways that do not perpetrate arbitrary decision making; (3) motivates teachers to make teaching the top priority of their institutional responsibilities; and (4) furnishes reliable feedback to teachers as to their impact on students and colleagues so that they may modify their teaching repertoire toward greater effectiveness in the classroom and outside it.

The efficacy of teaching evaluation programs in reaching such a desirable set of goals remains an empirical question.

The Nature of Teaching Effectiveness

The third question actually asks, What does our research tell us about teaching effectiveness? In the most meticulous review of the research to date, Wagenaar (1979) concluded that the literature supports ten major generalizations:

1. No significant relationship exists between Student Evaluation of Teaching (SET) and age, sex of instructor, sex of student, experience, or severity of grading.
2. Student characteristics *are* crucial, such as ability, predicted achievement, etc.
3. There is a negligible relationship of SET with research productivity; i.e., productive professors are no less, nor are they more, effective than their less productive colleagues, as measured by publications.
4. A high level of agreement exists between faculty and students on the basic characteristics that make for excellence in teaching.
5. There is a modest negative relationship between size of class and teacher ratings, though this is reduced somewhat when a control is introduced for required courses and course level (both are negatively related to SET).
6. There is no consistent relationship between SET and grades, final exam scores, and other tests, although such dimensions as student interest and faculty/student interaction seem to be moderately positively correlated with achievement.
7. The reliability for student evaluations of teaching is high (87% but quite low for colleague ratings (.57).
8. Students are rather generous in their rating of teaching: the mean is 3.8 - 4.0 on the typical 5.0 scale.

9. Several underlying dimensions repeatedly emerge as absolutely relating to good teacher performance, usually based on factor analysis:

a. structure/organization/preparation
b. interpersonal (tolerance, sensitivity)
c. professional (knowledge, examples, interrelationships)
d. skill/communication/clarity
e. motivation (stimulation, encouragement, enthusiasm)
f. assignment/evaluation (perceived fairness)

10. The greater the salience of an item of evaluation for the student, the higher the correlation of that item with the overall rating of teaching.

Participation in the Evaluation

The question, Who should do the evaluation? has a fairly consistent response in the literature. The answer is: *everyone affected by the outcome.* This clearly includes students, colleagues, professional evaluators, administrators, and the teacher in question. Such "triangulation" is the most valid and most acceptable approach, though it is also the most expensive.

Timing of Evaluation

As mentioned earlier, ongoing evaluation provides more effective feedback than end-point questionnaires: teaching adjustments can be made while the course is still in process. The most promising model seems to be one of regular, but brief, evaluations of lectures, discussions, and other formats. Coupled with these evaluations, I suggest a comprehensive mid-semester review of the teaching process; similarly, the end of the semester evaluation should be comprehensive. Since students and classes differ in their responses to various teaching methods and learning experiences, discrete assessment and feedback can take place and adjustments can be made.

Evaluation Methods

Questions about evaluation methods are probably the most interesting and useful to raise for most teachers. There is little consistent instrumentation for the introductory course, though a few principles have been advanced to aid in the development of instruments. These may be summarized as follows:

1. A combination of objective and subjective questions provides the richest and most reliable data.

2. Questions should be developed that relate to the learner, the instructor, and the substantive content of the course, as well as any other relevant factors that may have entered the learning context.
3. Short evaluation instruments are more effective for soliciting information than are long ones, provided that these are concentrated on salient parts of the teaching process.
4. With regard to the questions about the instructor, the six dimensions (structure, interpersonal, professional, skill, motivation, and assignment) should be made explicit.

One of the most comprehensive sources of instruments for evaluating teaching can be found in *Passing On Sociology* (Goldsmid and Wilson 1980). Most of the instruments have been fully tested for reliability and validity, thus resolving technical issues. However, the most difficult part of teaching evaluation has probably never been the technical aspect. As long as human institutions are arranged so that the destiny of one group is tied to the evaluation of another, the major problems will be political and social in nature. This is clearly more complicated than simply getting an instrument that works.

Guiding Principles

Grasha approaches the problem of guidance by suggesting eight principles of evaluation that, if activated, would serve to reduce resistance and render the idea of evaluation both practically and politically attractive (elicited from Malec 1978):

1. *There is no perfect evaluation system.* This does not mean that evaluation is impossible. It suggests that "current procedures will improve faster if people are willing to assess the advantages and disadvantages of their practical experience" (12).
2. *Assessment activity should be broadly based.* Teaching is multidimensional as an activity. Too often only one activity and one data source (students) become the sole basis of evaluation. The learning *gestalt* should be included — teachers, students, content, institutional context, faculty advisers, and so on.
3. *Evaluation must be linked to the reward structure of the institution.* In other words, the evaluation must not be a perfunctory exercise carried out to meet unknown bureaucratic imperatives. It must be meaningfully linked to recognition in the form of salary, promotion, or other tangible rewards.
4. *Evaluation is best considered a process rather than a terminal event.* This principle implies that the model of evaluating a course through one student questionnaire given at the end of the semester is unsatisfactory. It seems more appropriate to gather several data sets as the course unfolds so that the inclination and possibility to respond are immediate.

5. *Evaluation systems should allow for the assessment of interindividual differences and for the assessment of personal goals and objectives.* In other words, the evaluation process must recognize that each teacher has unique features that should not be diminished in a monolithic evaluation process.
6. *Evaluation systems should maintain personal and career growth and development focuses.* A system that attempts only to provide a *data base* for promotion, reappointment, or tenure or to "hire and fire" is an empty system and eventually encounters resistance. "Development must be an issue from the beginning to insure the long term survival of evaluation procedures" (15).
7. *People who are affected by a system of evaluation should participate in its development.*
8. *Attempts to make personnel decisions on the basis of an objective quantification of assessment data should be avoided.*

While these principles serve as a blueprint for the ideal teaching situation, and are not likely to prevail in the experience of most teachers, they point us toward some of the best reasons for having a program of evaluation.

REFERENCES

Eble, Kenneth E.
1976 The Craft of Teaching. San Francisco: Jossey-Bass.
Goldsmid, Charles, and E. K. Wilson.
1980 Passing On Sociology. Belmont: Wadsworth.
Grasha, Anthony F.
1977 "Assessing and developing faculty performance: principles and models." Communication and Education Associates, Cincinnati, Ohio.
Malec, Michael.
1978 "Teacher/learner effectiveness." Unpublished paper presented to the Knowledge Available and Knowledge Needed in the Teaching of Sociology Conference, Southfield, Michigan.
McGee, Reece.
1974 "Does teaching make any difference: a perplexed essay," Teaching Sociology, April, 210-23).
Wagenaar, Theodore C.
1979 "Assessing teaching effectiveness." Pp. 251-72 in Knowledge Available and Knowledge Needed in the Teaching of Sociology. Washington, D.C.: American Sociological Association.