POD Quarterly: The Journal of the Professional and Organizational Development Network in Higher Education

Professional and Organizational Development Network in Higher Education

Guidelines for Successful Evaluation Programs

Peter Seldin

Guidelines for Successful Evaluation Programs

PETER SELDIN

Abstract

In recent years, it has become as commonplace in American colleges and universities for students to grade professors as for professors to grade students. Despite the impressive growth of student evaluations, controversy over their use continues without letup. Unfortunately, it has been exacerbated by the not infrequent misuse of student feedback. To facilitate more appropriate and equitable use of student evaluation, this paper will present specific strategy suggestions and general guidelines to help implement and maintain successful programs.

Introduction

In theory, colleges and universities have a number of sources of student opinion at their disposal. An exit interview, face-to-face discussion, student testimonial, a suggestion box and questionnaire to alumni are all useful approaches. In practice, however, a written questionnaire or rating scale generally serves as the only source of student feedback on teaching performance.

Within the past decade, use of student ratings has dramatically increased. Today, for perhaps the first time, they are an accepted component in personnel decisions at a majority of institutions.

Some institutions have highly successful experiences using student ratings. Others live through experiences that are little short of complete failure. For many institutions, the difference between success and failure is determined by whether or not they recognize and effectively deal with several important tasks and options.

Choosing an Instrument for Student Ratings

Typically, when a college or university begins a program of student evaluation, a committee is named to weigh the possibilities and then devise an assessment instrument. The process is often lengthy and frequently divisive. And in the end, the rating form that is developed often looks like a carbon copy of forms used at other institutions.

Instead of devising an original student rating instrument, and in effect rediscovering the wheel, institutions have at their disposal the experience and rating instruments developed at other colleges and universities. For example, the Purdue Rating Scale for Instruction and the Illinois Course Evaluation Questionnaire have been tested over many years and a rich inventory of normative data has been developed. The universities of Washington, Princeton, Texas Christian and California at Berkeley are among many institutions that have amassed considerable experience in the use of student rating forms.

The rating scale, of course, must be compatible with the student evaluation goals on a particular campus. And the selection of the questionnaire items will depend on local conditions. As a general rule, therefore, it is better to adapt—not adopt—an already existing student rating scale and reshape it to create a better fit with local needs and goals. Since most rating scales contain a surplus of questionnaire items this is not a difficult task.

Probably the best place to begin the reshaping process is with a clear definition of teaching effectiveness as related to the goals of the institution. Most large and medium-sized colleges and universities have well-qualified faculty members who can assist in the designing and testing of the adapted student evaluation questionnaire. Their expertise should be tapped.

Open discussions on the subjects of teaching effectiveness and student ratings are essential if trust and confidence in the system is to be achieved. Administrators, faculty and even students should be encouraged to participate and share their views. Disagreements should be discussed in an atmosphere devoid of rancor. At all stages of program development, decisions must be arrived at openly and the reasoning behind them must be clearly known.

Form of the Questionnaire

Is there a single questionnaire suitable to every course, department or institution? Probably not, since different instruments are needed to evaluate different courses and produce different information.

It is virtually impossible to design a single student evaluation questionnaire that is equally effective for a large lecture, a seminar and a laboratory course. On the other hand, meaningful comparative data is generated when a common instrument is used to assess a range of teaching styles and subject areas.

At the SUNY College at Brockport (New York), for example, the faculty members select one of six versions of a questionnaire as most suitable for their course. Each version, reports Humm (1978), contains general questions that are common to all forms. One section of the questionnaire is designed to generate diagnostic feedback and contains six different sets of questions applicable to six different learning environments: A) Lecture/discussion, emphasis on content; B) Lecture format with a minimum of class participation; C) Seminar/discussion report; D) Lecture/discussion format, emphasis on process; E) Apprentice format, skill development; and F) Material-centered format for student self-study or mediated courses.

What is the appropriate length of the questionnaire? If it is too long, it may cause student resentment and present sizable logistical problems in tabulating results. If the questionnaire is too short, it may not produce enough information. Some standard instruments (for example, those developed by Wilson and Dienst at the University of California, Berkeley) are available in long, medium and short forms.

For personnel decisions, a short form containing from five to ten items is generally sufficient. For improving classroom performance, a medium (15–20 items) or long form (30–35 items) on no more than two pages is probably best. Between ten and twenty minutes is about the time needed by students to complete the questionnaire.

Some global questions such as "How do you rate this teacher in comparison to others you have had in the department?" should be included in every form. They tend to be more closely related to student learning than questions on specific instructional behaviors and, importantly, are not limited to any instructional style.

If student ratings are to be used to improve classroom perform-

ance, it is advisable to include several open-end questions so that students can respond in their own words. Examples: "What has your instructor done especially well in his teaching of this course?" "What should your instructor do to improve her teaching of this course?" "In what ways did this course meet or fail to meet your expectations?"

Planning Programs of Instructor Evaluation—Key Decisions

To develop a successful student evaluation program requires a series of key decisions. Any program which neglects to spell out these decisions will almost certainly falter in its execution:

1. Should a single school-wide instrument be used or should each department develop its own?
2. Should participation be mandatory for all faculty?
3. How frequently should student evaluations be conducted? Once per term? Once per year?
4. How should student views be obtained in very small classes which are unsuitable for typical student ratings?
5. Are results of student evaluations to be published?
6. Are results to be used for personnel decisions? If so, how much weight will they be given in decisions for promotion and tenure?
7. Who will have overall responsibility for administering and maintaining the student evaluation system?
8. Who will pay for it? Student government? Administration?
9. What is the minimal acceptable percentage of participating students in any one class? 60%? 70%? 80%?
10. What procedures will be used for conducting the student evaluation? Should faculty be involved?
11. Should an institution develop its own form or adapt one already in use elsewhere?
12. Will the results of student evaluation feed directly into a follow-up program of improvement? How?

Administering the Rating Forms

Many student rating programs are plagued by a lack of systematic administrative procedures. Among the more common defects are a sporadic rating schedule, a biasing effect of improper instructions and a lack of standards which voids the possibility of meaningful interpretation (Scott, 1975). Data generated from such unsystematic procedures is clearly of limited value for either purposes of personnel decision or teaching improvement.

The usefulness of student ratings can also be sharply diminished if the forms are administered immediately before, after or during examination periods, if less than 70% of the students in a class fail to participate or if student ratings are conducted too frequently. (Evaluation fatigue is likely to set in if every class is evaluated every semester. Students and faculty can suffer if there is too much evaluation.)

How and when should the questionnaire be administered? One procedure in use at many institutions has proved simple and effective. Students fill out the rating forms two weeks before the term ends, and before final examinations. The forms are distributed and completed within a single week. A statement of instructions is read to the students by a class assistant. During the reading of the instructions and the completion of the forms by the students, the instructor remains outside the classroom. At the end of a fixed period, ordinarily no more than 20 minutes, the assistant collects the forms, places them in an envelope which is labeled with course title and number, and in view of the students seals the envelope. The envelope is then secured in the dean's office or other controlled place. Although the forms are processed within two weeks, the results are not publicized until after the issuance of final grades.

Interpreting the Rating Forms

A chronic and frustrating problem accompanying student ratings lies in their interpretation. Frequently, instructors have no idea whether their ratings are good, average or poor, or how their ratings compare with other instructors' ratings.

Some institutions have solved this problem by issuing to each instructor an interpretation manuel, which contains the norms (average scores, percentiles, etc.) and makes possible performance comparisons. To be most meaningful, however, the norms should be broken down into different disciplines—or courses within a discipline—and different course formats as well.

If student ratings are going to be used by administrators for personnel decisions, then global or overview ratings should be emphasized. Decisions should be based on several courses over several semesters, as a rating of a single course in one semester may be influenced, either positively or negatively, by special circumstances.

An increasing number of institutions today provide the oppor-

tunity for faculty members to review and comment on their evaluations. The comments are attached to the ratings so that faculty committees and administrators using the ratings for personnel decisions also have the benefit of the comments.

If the purpose of the student ratings is to improve teaching performance, then they should be interpreted for specific teaching behaviors. Just as students need specific feedback to correct errors, faculties need specific data to point the direction to self-improvement. But whether such improvement actually takes place depends on the teacher genuinely caring about the evaluative process and realistically being able to make the necessary changes.

Simply handing the professor a computer print-out containing the results of his/her student evaluation is not likely to motivate self-improvement. A follow-up injection of short periods of faculty counseling is often needed.

To provide that counseling, many colleges and universities have set up faculty development centers to bring advice and guidance to faculty members interested in converting evaluative feedback into teaching improvement. At Kansas State University, for instance, a trained consultant helps instructors interpret their ratings, helps clarify and resolve problems, and assures that corrective action is planned. A study by Aleamoni (1974) found that student ratings improved teaching performance when the ratings were analyzed and discussed with a trained evaluation person.

General Guidelines and Strategy Suggestions

Following are some general guidelines and strategy suggestions to help implement a successful student evaluation program:

1. Obtain administrative support and faculty cooperation by proceeding cautiously with frequent, open discussions of all issues.
2. Anticipate faculty resistance and deal with it sincerely, positively, calmly and realistically. A sound rationale and a solid research base are important in coping with opposition.
3. Arrive at working definitions of the purposes and goals of student ratings that are acceptable to all parties.
4. Make the crucial decisions regarding the what, who, where, how, and when of student evaluation.
5. Hold open faculty forums during the developmental stage of the rating instrument and encourage student groups to attend.
6. Conduct dry runs to improve the questionnaire, tighten the pro-

cedures of administration and help reduce the anxiety level of faculty.
7. Keep all faculty and students adequately informed on a continuous basis.
8. Provide sufficient time for the overall process of implementation—a year or even two is reasonable and typical.
9. Establish an active research program to permit assessment of the evaluation system itself to provide a basis for future improvement.

Conclusion

If student evaluation is to be properly used as one component in the assessment of overall faculty performance, careful consideration must be given to certain key decisions. These include:

1. Choice of an assessment instrument.
2. Format of the questionnaire.
3. How to administer the rating form.
4. How to interpret the data.
5. Who should receive feedback.

Failure to give adequate attention to each of these matters will almost certainly prove fatal to the student evaluation program.

Student assessment of teaching falls far short of a complete assessment of an instructor's teaching contribution. Colleague and administrator appraisal, as well as self-appraisal, are other obvious sources of information which should be taken into account.

But, as Costin, Greenough and Menges (1971, p. 531) suggest, "If teaching performance is to be evaluated, either for purposes of pay and promotion, or for individual improvement, a systematic measure of student attitudes, opinions and observations can hardly be ignored."

In summary, student ratings that are carefully planned, properly administered and judiciously interpreted can be a particularly useful component in the evaluation of faculty performance.

BIBLIOGRAPHY

Aleamoni, L. M. *Typical Faculty Concerns about Student Evaluation of Instruction.* Unpublished paper delivered at the Technion-Israel, Institute of Technology, Haifa, Israel, March 24, 1974.
Costin, F., Greenough, W. T., and Menges, R. J. "Student Ratings of

College Teaching: Reliability, Validity, and Usefulness." *Review of Educational Research,* 41, 511–535, 1971.

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

Humm, H. L. "Instructions for Use of the Instructional Assessment System." SUNY College at Brockport, August 1978.

Miller, R. I. *Evaluating Faculty Performance.* San Francisco: Jossey-Bass, Inc., Publishers, 1972.

Scott, C. "Student Input Relative to Professorial Performance." In C. S. Scott and G. L. Thorne (eds.) *Professorial Assessment in Higher Education.* Monmouth, Oregon: Oregon State System of Higher Education, 1975.

Seldin, P., *Successful Faculty Evaluation Programs.* New York: Coventry Press, 1980.

Seldin, P., *Teaching Professors to Teach.* New York: Blythe-Pennington, Ltd., 1977.