CASE STUDY

Improving the Resident Instruction Evaluation System

Joe G. Harper
and Ronald L. Shane

Abstract

Effective evaluation of resident instruction is an on-going and unresolved issue in most departments and colleges. In this case study, a college of agriculture committee of faculty and students was established to address faculty concerns. This committee produced a report consisting of recommendations for changing the college-wide system, a proposal for increased department autonomy, and suggested procedures for implementing the changes. As a result of this experience, the authors offer six recommendations to other colleges considering modifying their teaching evaluation systems.

Introduction

Identifying an effective means by which to evaluate resident instruction faculty is an on-going topic of concern and attention at most academic institutions involved in undergraduate education. The evaluation of instruction is a complex issue with participants often holding passionate positions. The emotional feelings by those being evaluated are aptly stated by Braskamp, Brandenburg, and Ory (1984): "The human side of evaluation is crucial. Evaluation of persons is a deeply personal and sensitive undertaking. We have yet to work with someone who has not been anxious, interested, or concerned about an assessment of his or her work.."

Some college of agriculture faculty perceptions regarding teaching and teaching evaluations are presented in a *National Assessment of Faculty Development Needs in Colleges of Agriculture* (Chudzinski; 1988):

1. One-third of the faculty felt they were not prepared to teach by the time of their first appointment as an assistant professor;
2. Nearly 50 percent of the faculty consider teaching less important than research;
3. Teaching, unlike research, is not adequately evaluated and rewarded;
4. Over-reliance of departments on one teaching evaluation method: student evaluation (60 percent of faculty do not think students are the best judges of how well their professors teach). Other methods, such as formal and informal peer evaluations and postgraduate evaluation are used in only 30 to 40 percent of the departments;
5. Financial and prestigious awards are regarded as very helpful by 6 percent and helpful by 36 percent of the faculty.

Attempts to improve instruction or resolve evaluation problems often leave faculty, administrators, and students frustrated and disillusioned.

Administrators often perceive that the primary purpose of instructor evaluation is to provide information to be used in personnel decisions regarding retention, merit, tenure, and promotion. This form of evaluation has been described as summative evaluation. Faculty often feel that the primary objective of instructional evaluation should be for teaching development to help them improve their teaching, and hence help them keep their jobs, receive merit pay, and achieve tenure and/or promotion. This form of evaluation is traditionally labelled as formative.

The logic for advocating emphasis on summative evaluation is succinctly presented by Scriven (1981): "But summative evaluation is primary because (1) human careers are at stake, not 'mere' improvement; (2) if it is not possible to tell when teaching is bad (or good) it is not possible to tell when it has improved; (3) if it is possible to tell when it is bad or good, personnel decisions can be made even though it is not known how to make improvements. In short, diagnosis is sometimes easier than healing, and an essential preliminary to it." Brock (1981) on the other hand, makes a case for stressing formative evaluation by arguing that evaluation should be viewed as a cycle: "The cycle begins with planning, moves to implementation, and ends with evaluation. Evaluation is the process by which we gauge the effects of the educational plans we implemented; evaluation then serves to refine planning for subsequent educational implementation."
as well as many more, tends to lead to the development of confused, ineffective evaluation systems that do not provide accurate or meaningful information for administrators, faculty, or students. The purpose of this article is to present the process and product of one college of agriculture's attempt to improve its evaluation of resident instruction.

Background

A brief historical review will help the reader understand the change process which occured in the case study college of agriculture relative to the evaluation of instructional effectiveness. Prior to the early 1980's the evaluation of instruction was conducted by each individual department using its own system, instrumentation, and formulae. There was little or no collaboration or common evaluation structure between departments. This departmental based system

Harper is an assistant professor in the Department of Agricultural Education and Shane is an associate professor in the Department of Agricultural Economics at the College of Agriculture, University of Nevada-Reno, Reno, Nevada 89557-0106.

was perceived as having the following disadvantages by both college administrators and some faculty.

1. There was no uniformity across departments.
2. The system did not adequately discriminate among different teaching qualities, that is, a large portion of faculty were being rated as "excellent" teachers.
3. If teaching is to be rewarded by the college, then teaching evaluations must be acceptable to the college administrators.
4. The absence of discriminating teaching evaluation instruments contributes to the problem of teaching not receiving appropriate weight relative to research by promotion and tenure committees at the college and university levels.

In an attempt to address these identified problems, the associate dean of resident instruction for the college of agriculture conducted a search for a research-based teaching evaluation instrument which could be used by all college teaching faculty. The search produced two viable candidates; one from Kansas and another from Arizona, the CIEQ.

A sample of college classes were selected in spring semester 1984. Each college sample class was administered both instruments. A questionnaire was used to obtain faculty and student feedback regarding the relative suitability of each instrument. Based on faculty and student questionnaires, the associate dean mandated the Arizona CIEQ instrument for all teaching faculty for all classes. Hilwig (1982) described the CIEQ as a component of a system of evaluation of teaching in the College of Agriculture at the University of Arizona. The associate dean then established a sixteen point individual teacher rating system. CIEQ scores were worth eight points, teaching load was worth two points, and other factors undergraduate advising, teaching improvement efforts, support of student organizations and activities, teaching diversification, graduate student advising, improvement in teaching were worth one point each. Department chairs could establish a modified system centered on the CIEQ to evaluate faculty on a four category rating system: excellent, commendable, satisfortory, unsatisfactory. As part of the evaluation process, departmental peer review committees review chair evaluations. Discrepancies between final departmental teaching evaluation ratings and associate dean ratings for individual faculty were resolved between the two respective administrators.

Negative feelings of faculty to this mandatory evaluation system were immediate and intense. Some teaching faculty perceptions and conclusions were as follows:

1. CIEQ scores essentially determined teaching evaluation ratings at the department and hence college level.
2. The CIEQ was not accurate for smaller classes since a single student could dramatically alter the calculated CIEQ scores.
3. Challenging students with rigorous material could result in lower teaching scores by upsetting a few students.
4. CIEQ printout results could not be utilized for purposes of improving instruction.
5. The college formula for weighing CIEQ score was not known.
6. CIEQ scores were not valid measures of teaching quality.
7. The CIEQ instrument emphasizes student satisfaction, not course content.
8. The college-wide system was imposed without faculty input.

In addition to these faculty reactions, there were other outcomes associated with the new evaluation system. Teaching faculty received lower teaching evaluation rankings, on average, then previously. Variability of student evaluation scores increased not only between teachers, but, of greater concern to faculty, variablility of scores increased for individual teachers from one course to another and across different classes of the same course.

On the positive side, there were reports that some faculty were improving their classroom behaviors. Students and faculty both observed that some instructors were teaching better classes, primarily in response to low CIEQ scores.

But the price for these gains was quite high. Division between extension and teaching/research faculty increased in the instructional area. Non-classroom extension faculty did not have similar client instruments available which could discriminate between excellent and unsatisfactory extension presentations. Faculty generally felt classroom teaching and extension teaching were not being evaluated with similar rigor.

A major concern was the impact of the new evaluation system on junior faculty. No allowances were made between inexperienced and experienced teachers. Several junior faculty did receive "excellent" ratings; however, a larger proportion of junior faculty, relative to senior faculty, were at the bottom of the scale, and therefore evaluated as "unsatisfactory." As a result, the CIEQ became a threat to junior faculty without tenure. To achieve tenure, a minimum "satisfactory" teaching evaluation is required, by the University System Code.

Without faculty input in the implementation process and without information on what the CIEQ was attempting to measure, an unnecessary amount of faculty time and energy was spent arguing about and denying what the CIEQ scores meant. Senior faculty attacked the instrument out of frustration from being unable to interpret results and thus help themselves or assist junior faculty. In addition, decile rankings of the CIEQ appeared to be sensitive to extraneous variables, e.g., time during the semester relative to exams the instrument was administered.

Process

In June 1986, the associate dean established an ad hoc teaching evaluation committee consisting of seven teaching faculty, one graduate student, and one undergraduate student. Each faculty member represented one of the seven college departments and all were generally recognized as "good" undergraduate teachers. The committee established three goals:

1. Improve the quality of instruction in the college of agriculture.
2. Clearly define the current and future role of the excellent teacher/satisfactory researcher in the college of agriculture.
3. Improve department and college of agriculture procedures for evaluating teaching, achieving an acceptable balance between flexibility and equity among departments.

In an attempt to achieve these goals, two activities were undertaken. First the committee chair and a committee member visited with faculty in five of the seven departments in the college of agriculture. The second activity was a two part colloquium consisting of comments from administrators, faculty, and students.

Product

The final report of the committee consisted of a series of recommendations, a proposed model for departmental evaluations, and a process for implementing department models. The recommendations made by the committee were grouped into four categories: general recommendations, teaching recommendations, student related recommendations, and teaching evaluation recommendations.

The teaching evaluation recommendations are as follows:

1. Teaching evaluation should be department based.
2. Departments should establish general weighting schemes for various teaching model components for a "representative" department faculty situation.
3. If a diversity of teaching assignments exist in a department, a single weighting scheme formula need not be applied to all faculty; rather, a weighting scheme appropriate to special individual faculty situations should be applied.
4. Incorporate into the teaching evaluation criteria those non-classroom teaching/research activities by faculty which contributed to student learning but which have no direct research output.
5. Use faculty development and improvement as a criterion in teaching evaluation.
6. Relative to the current Arizona CIEQ student opinion survey:
   a. for the present, the Arizona form be used on a college-wide basis for evaluation, with modifications in use and application which may vary across departments;
   b. departments may select different sets of questions for evaluating classroom presentations;
   c. departments may use deciles or raw scores for evaluation;
   d. for the purpose of determining current teaching performance levels, comparison across faculty or against a known standard, is appropriate;
   e. for the purpose of determining improvement, comparison of previous performance levels for the faculty member being evaluated is appropriate;
   f. if faculty want to use alternative questions or instruments to improve teaching, they should be encouraged to do so; and
   g. if some departments remain dissatisfied with the Arizona CIEQ as a source of evaluation information even after modifications are implemented, those dissatisfied departments should be encouraged to experiment with alternative student survey instruments.

The second component of the report, the proposed evaluation model, outlined various information departments could use to evaluate instruction. The model is composed of four types of input:

1. Faculty Self-Evaluation - consisting of individual faculty perceptions of their teaching. Criteria suggested were the faculty perceptions of their course content, course objectives, teaching strategies, areas for improvement, student behaviors, assignments.
2. Student Evaluations - student perceptions of classroom activities, teaching methods, course content, evaluations procedures for student achievement, teacher enthusiasm, clarity, and other teacher behaviors.
3. Peer Evaluation - the perceptions of faculty peers concerning such areas as course structure, student evaluation procedures (tests, etc.), course content, classroom/-laboratory organization, student assignments, and student perceptions.
4. Department Chair Evaluation - the perceptions of the department chair regarding the faculty member's goals, self-evaluation plan, faculty member efforts to improve, student and peer evaluations, course loads, availability for students, advisement activities, and other contributions to student learning.

The intent of the proposed model was to provide more information than the present system for each

faculty member which the departments could use to establish a more balanced approach to evaluating instruction.

The third component of the committee report was to propose a process for implementing department models. The associate dean of resident instruction was encouraged to work with department chairs to develop "appropriate" teaching evaluation programs based upon the proposed models which best represent the individual faculty and department teaching situations. This process would include:

1. Each department develop teaching evaluation models which are consistent with department and college teaching and research objectives.
2. The final department version be acceptable to the associate dean of resident instruction.
3. The associate dean of resident instruction should then evaluate the fairness of department evaluation models between individuals and across departments.
4. The associate dean of resident instruction should then use the department evaluation model criteria to determine the college level evaluations of the teaching component of faculty evaluations.

Recommendations

Based on this experience, the authors have the following recommendations for other colleges wanting to change their existing teaching evaluation systems.

1. Before making significant changes in existing evaluation procedures, there should be broad faculty input and support.

Faculty at your institution are currently being evaluated under an existing system and have adjusted to it. Your faculty accept, with varying enthusiasm, the system and have certain expectations that it will continue as is.

2. Weigh trade-offs of proposed changes carefully.

A popular phrase in economics is, "There is no such thing as a free lunch." This is also true for changes in existing procedures for evaluation. There are always advantages and disadvantages of the new system, alternatively these can be labelled as benefits and costs. A change which may have noble and glorious advantages (benefits), may have large disadvantages (costs).

3. Emphasize final goals, not means to goals.

One generally accepted goal for teaching is to maintain or improve the quality of instruction. One means (not a goal in itself) to achieve this may be a single college-wide student survey instrument. In this case study, implementing a single student instrument across all departments was a specific goal.

4. Related to three above, emphasis is on teaching development, not evaluation.

There are three interrelated components to the challenge of improving the quality of instruction: reward, evaluation, and development. The evaluation process can emphasis categorizing faculty in appropriate monetary reward categories or it can emphasize providing information useful for development. The latter approach is preferred by faculty and is more consistent, in the authors' opinion, with a long term program of better instruction. Even under such an evaluation, it is possible to simultaneously obtain the information necessary to categorize faculty into unacceptable and acceptable categories, including degrees of acceptability. In addition, however, it is also possible to undertake two types of teaching developing programs: a) for teachers identified as unacceptable, remedial development to improve unacceptable behaviors, and b) for satisfactory faculty, improvement programs aimed at "weaker" teaching behaviors.

5. Do not reduce the teaching evaluation system to the point where primary weight is given to a single measurable value.

College teaching is a multidimensional activity. If faculty must be categorized beyond unsatisfactory and satisfactory, then the evaluation process should include perceptions from a wide range of alternative sources. Granted, weights must be assigned to information obtained from these alternative sources. But this teaching evaluation process will have much greater faculty acceptability and contribute to a more productive scholarly environment than a rigid one which accounts for only limited aspects of teaching.

6. Treat inexperienced teaching faculty differently from experienced faculty in the evaluation process.

New, inexperienced university faculty typically do not have any training in how to teach or much previous involvement in classroom instruction. An appropriate environment of support can exist if there is emphasis on teaching development and a recognition by administration that it generally takes time for inexperienced faculty to identify and develop their most effective teaching styles.

References

Braskamp, L., Brandenburg, D., and Ory, J. (1984). *Evaluating Teaching Effectiveness: A Practival Guide.* Beverly Hills, CA: Sage Publications.

Brock, S. (1981) "Evaluation-based Teacher Development." In J. Millman (Ed), *Handbook of Teacher Evaluation* (pp. 229-243). Beverly Hills, CA: Sage Publications.

Chudzinski, L., Simerly, C., & George, W. (1988). *National Assessment of Faculty Development Needs in Colleges of Agriculture.* Champaign: University of Illinois at Urbana-Champaign, College of Agriculture.

Hilwig, R. & Jacobs C. (1982) "A College of Agriculture Develops and Evaluates a System for Instructional Improvement." *NACTA Journal,* 26(2), 11-14.

Scriven, M. (1981) "Summative Teacher Evaluation." In J. Millman (Ed), *Handbook of Teacher Evaluation* (pp. 244-271). Beverly Hills, CA: Sage Publications.