

Other results of this study indicate there were limited differences between group variances and more within group variances in marking individual items for each dependent measure. Tables 2 through 7, however, reveal that there were differences among faculty preferences in rating various items: uses of methods of instruction, uses of instructional resources, attitudes toward use of audio-visual materials and devices, perceived deterrents, experiences and needs in operating equipment and devices, and in preparing instructional materials. Thus, faculty members have different perceptions concerning individual items within categories of instructional technology.

Each academic discipline category has certain potential strengths and weaknesses. For example, members of the Plant and Soil Technology Group and Animal Technology Group utilize most instructional resources, while members of the Horticulture and Forestry and Home Food and Technology utilize most methods of instruction, and members of the Applied Social Science Group have the most favorable attitude toward audiovisual equipment but need the most assistance in operating equipment. These results suggest that each group has certain skills that need to be shared with other groups.

Conclusion

This study was designed to gather information about habits, attitudes, preferences, and needs of faculty members. The findings as identified will be used to set the goals and objectives of the Instructional Development Laboratory that recently moved into its new facility. The facility includes three media classrooms, equipped with the latest presentation and recording technology, including television and telelecture. The classrooms are used for regularly-scheduled classes that require media support. They are also used as training facilities for faculty and graduate students to improve their communication and teaching skills. The Laboratory initiates and promotes workshops and seminars related to teaching and learning resources and methods, e.g., microcomputers as an instructional tool. Along with the in-service function there is a materials development laboratory for the faculty and graduate students to assist them in identifying and producing resource materials in meeting their specific teaching-learning needs. For example, slide sets are developed for use in the classroom and in the individualized learning laboratories. The facility is located in the new Central Library of the St. Paul Campus and through networking relationships supports and provides guidance in the utilization of the library's nonprint distribution system for resident instruction. The limitations so far have been program support funds, for as in most large institutions the education dollar is limited. In retrenchment periods instruction returns to the basics and teaching-learning resources assume the luxury syndrome. We hope to identify through the Instructional Development Laboratory and studies such as this one the priority resource needs of the faculty and resident teaching. We also believe each teacher

should know his tools and techniques of instruction. To know resources is to use them effectively. A do-it-yourself program for resources (with assistance) is one answer to the lack of funds. There is a need for more interaction among faculty members and disciplines relating to the teaching-learning process. There is also the need for coordination and leadership in developing seminars and introducing the faculty to new technology materials and methods. We hope the Instructional Development Laboratory will fill some of these gaps. As in all previous studies at this University, this study reinforces faculty desires to use good instructional alternatives and resources if they are adequately supported with competent personnel and funds.

References

- "Learning Resources for Colleges and Universities," Sept. 1964, an advisory and assistant project, No. OE-3-16-025, of the Educational Media Branch of the Office of Education. Principle Investigator, Fred Harcleroad.
- "The Development of the Audio Visual Education Program at the University of Minnesota," Ph.D. Dissertation 1970, 809 pages, Wesley J. F. Grabow
- "Does Anybody Care?" *Change*, the Magazine of Higher Learning, Report on Teaching No. 4, 4-5, July, 1977.
- "Resources for Teaching and Learning" *NACTA Journal* Vol. XXV No. 2 June 1981.

Toward Establishing A Record Of Teaching Performance

Chauncey T. Ching

Evaluating classroom instruction is continually perplexing and controversial. One can find support in the literature for just about any viewpoint. Seemingly, there is always something to write or talk about whenever teaching evaluation is discussed. See, for example, the articles by Aleamoni, Foth, Moody, Deaver, and Shrode.

As a common starting point, I first suggest that whether teaching can be evaluated is irrelevant. We must evaluate teaching. Second, there are at least two major objectives for evaluating teaching: to use the information to make personnel decisions and to improve teaching quality. Third, at least two items are evaluated during the process: the input resources going into teaching such as instructor characteristics, and the output or student learning that comes about from application of the teaching process.

This paper will describe briefly four main methods of evaluating teaching. Then it makes a case for utilizing several methods of teaching evaluation to establish a record of teaching performance over time. These records can provide sufficient information to distinguish between instructors doing a "good" job and those doing a "poor" job.

Ching is Professor and chairman of the Agricultural Economics Department, University of Hawaii, Bilger Hall, 2545 The Mall, Honolulu, Hawaii 96822.

Four Methods of Evaluating Teaching

Criterion Referenced Instruction¹

Criterion referenced teaching is best described by contrasting it to norm referenced instruction. Norm referenced instruction identifies an individual's performance in relation to others in the same class. For example, in norm referenced instruction, test scores are meaningful only relative to other scores. A good example of a norm referenced test is a standardized achievement test such as the Graduate Record Examination.

In contrast, a criterion referenced instruction or test is less common. Here an individual's status is identified relative to established performance standards (criterion). The criterion is a set of learning objectives specified by the instructor. Thus the meaning of an individual's score is not dependent on comparisons with others in the class, but on how well the individual scored in terms of the intended objectives.

An example of a criterion referenced test involves a dog owner who would like to keep his dog in the backyard. He administers a fence jumping test to his dog to see how high the dog can jump. Once he determines this, he builds a fence slightly higher than the height the dog can jump. The dog owner is not interested in how high his dog jumps relative to other dogs, but only relative to the criterion, height of the backyard fence.

Because the norm referenced measures are devised to make comparisons between individuals, they serve as a rationing device to select entrants to the next level of learning. One good example, especially in agriculture is to use norm referenced measures to identify those students "qualified" to attend veterinary schools.

In contrast, criterion referenced testing is used to make decisions about individuals and treatments. In this instance, "treatments" means the teaching process. In making decisions about individuals, criterion referenced testing would indicate whether an individual has learned enough to advance to new material. For example, before a student is qualified to study calculus, he must have attained certain minimum standards of performance in algebra.

For this discussion, criterion referenced testing can be used to make decisions about the teaching process or the instructional program. In other words, we might use criterion referenced testing to determine how well the teaching process has contributed to learning certain instructional objectives. After the instructional program has been completed, the test would indicate its effectiveness — relative to the stated instructional objectives.

Thus when there is a restriction on how many achievers can advance to the next level of learning (e.g. veterinary school), norm referenced tests might be used. In contrast, when there are no constraints on the number who can possess that skill, criterion referenced tests would be used. And, when the treatment or the teaching

process is questioned, only criterion referenced tests apply. Specifically, when tests would indicate how well the instructional program had contributed to learning specific instructional objectives. As such, criterion referenced instruction as an evaluation method is the only technique described here which addresses the output or student learning phase of teaching evaluation.

Student Evaluation

Student evaluation is a process followed in most institutions. Usually, a questionnaire is prepared and administered by the instructor — typically at the end of the semester or term.

The literature shows considerable evidence that students can provide reliable and meaningful evaluation of teaching. For example, Foth reports that there is no correlation between students rating of teaching and their age, sex, years in college, class size, major or non-major or grade point average. Also length of time out of school and the sex of the instructor have no influence. Foth also reports the teachers who were particularly effective in terms of student performance on tests were rated higher by their students than were less effective teachers.

In contrast, there is also support for the argument that students' grade point averages affect their rating of instructors (Shrode). Further, there is some evidence that a student's major also affects instructor ratings (Burger and Seif). Also, evidence suggests that freshmen tend to rate courses higher than juniors (Marshall). Thus, the literature demonstrates that student evaluations must be corrected for a number of factors before they can be reliable indicators of teaching effectiveness.

Perhaps one of the more serious criticisms of student evaluation is that the evaluation is usually administered at the end of a semester. Consequently, students have very little motivation for filling out the questionnaire. The questionnaire is impersonal and creates nominal student involvement. Also students have very little to gain from completing the questionnaire. If they effect improvements, those improvements are for future "generations" of students and of no benefit to current students (Casavant).

Another criticism of student evaluation has been offered by Marg. He reports that it is easy to achieve high ratings by students. But, such ratings are meaningless when student learning has not been increased. Only the instructor's popularity has increased.

Student evaluation is a typical instrument used to evaluate teaching. As such, it usually evaluates the input or resources going into teaching and not the output or learning.

Peer or Administrative Evaluation

Again, the literature contains support for and against teacher evaluation by either peers or administrators. Foth reported, for example, that the personality of the evaluator becomes involved when such evaluations are used. Thus, a teacher is rated effective if the procedures he follows are consistent with the evaluator's biases.

¹This section relies heavily on the work by Popham.

Peers tend to rate an instructor relative to knowledge or the subject matter and his contribution to research and professional activities (Foth). On the other hand, administrators tend to rate the instructor based on the instructor's impact on the school's image or enrollment. Thus, it is usual to find different evaluations by peers versus administrators.

There is also some concern that administrators and peers may not be qualified to review an instructor. Further, if an administrator or peer reviews just one class of an instructor the instructor may have had a good or a bad day and the sample of one is not especially reliable. Further, the criteria used for evaluating teaching by the peer or administrator are not clear to either party. Finally, peers may be reluctant to be critical for fear of retribution.

Peers and/or administrators generally evaluate the input or resources going into teaching and not the output or learning.

Mid-semester Review

There are several forms of mid-semester teacher reviews. Casavant reported on a system called Colleague Related Evaluation. Ching and Garrett have more recently reported on a technique known as Mid-semester Evaluation of teaching. These procedures are more similar than different.

The author's department uses the following kind of mid-semester review: First, the department chairman and the faculty develop a questionnaire which emphasizes the important criteria of effective teaching. Second, at the middle of the semester or term, the department chairman goes to each instructor's class and asks him to leave. The chairman then discusses the reasons for the evaluation and the procedure to be followed. The chairman asks for candid comments on the specific points contained in the questionnaire. Next, the department chairman reviews the results of the discussion with the instructor and also provides the instructor with a written statement regarding the review.

This system has several advantages. First, the criteria for evaluating teaching are known to both the instructor and the reviewer. Second, students are motivated to respond to the review because they see the potential of benefitting from the review during the second half of the class. Third, the process avoids the "sample of one" criticism of peer or administrative review. The questionnaire does not ask for the instructor's performance during any one class period, but over the first half of the semester. Fourth, the approach is relatively inexpensive in terms of the reviewer's time. The reviewer spends approximately 15 or 20 minutes with each class. The reviewer also spends an hour preparing a written statement and debriefing the instructor. Fifth, the procedure is less impersonal than a student questionnaire. The reviewer is present. He is talking to the students and asking them to get involved. There are also opportunities for students to respond to the reviewer on a one-to-one basis should they be reluctant to discuss items before the entire class.

The approach used is also subject to some of the same criticisms as the student evaluation form. Again, students evaluating the instructor during the mid-semester review are likely to have different ratings depending on their grade point average, their major, and whether the course is required. And, the mid-semester review process again evaluates the input or resources going into the teaching process and not the output learning.

Monitoring Teaching Effectiveness

If nothing else, the foregoing discussion emphasizes that evaluating teaching is complex and that different methods of evaluating teaching address various aspects of teaching. For example, criterion referenced instruction evaluates teaching outputs. Other methods evaluate inputs into teaching; but, they evaluate these inputs in different ways. Since the methods of evaluating teaching already described provide information on different aspects of teaching or provide information in different ways on the teaching process, it is not unreasonable to view them as being complementary rather than mutually exclusive procedures.

Given the complexity of the evaluation process and the complementary nature of the techniques, it seems logical to utilize all (or at least several) techniques to evaluate teaching. Further, to be truly effective, evaluations should be conducted periodically. Then the results should be viewed after sufficient time has passed to provide an overall indication of the instructor's effectiveness.

When the various methods are utilized in a complementary manner and results of the evaluation are viewed after several years of performance ratings, they provide two types of information. First, persistent problems of the instructor or desirable qualities of the instructor or the instruction process are isolated. Second, the information provided after several years of evaluation allows administrators to place an instructor in one of two categories: (a) doing a good job, (b) not doing a good job. In effect, several years of evaluations can be used to build a history of teaching effectiveness. Evaluations monitor instructor performance. In this sense, such a history is not unlike having one's blood pressure taken periodically to provide a general trend of one's health. Minor deviations are not alarming. But consistent deviations or trends could be critical and should receive specific treatment.

The importance of a history of performance can be illustrated with examples from the author's institution where all four techniques described above are used, although some are not used very often. Criterion referenced instruction is used by some instructors. No documentation exists since the technique is administered by the instructor. Peer or administrative review is also used, but sparingly. Typically the associate dean of the college will review an instructor's class by attending the class. But, this is done perhaps once in five years for any one instructor — usually prior to a promotion/tenure decision.

In the author's department, student evaluation of instruction and mid-semester review of instruction by the department chairman is used regularly. At the end of each semester, instructors administer a standardized student evaluation form used by the College of Agriculture. These evaluations are summarized and most instructors file a summary copy with the department chairman. In addition, in the author's department, mid-semester reviews provide some interesting information to both the instructor and departmental and/or college administrators.

For this study all student evaluation summaries for faculty members who have been teaching in the department for more than three years were used. The student evaluation form has two parts. One part concentrates on evaluating the teacher and the second on evaluating the course.

One question from the evaluation of the teacher was used as a general measure of the instructor's effectiveness. This question concerns the "general rating of the instructor." At one end of the spectrum is the category, "outstanding", meaning "truly an exceptionally fine teacher in every way, one of the best I have had." In the middle is "average" which means "about average." At the other end of the spectrum is "terrible" or "one of the worst instructors I have had." The scale ranges from zero for terrible to 100 for outstanding. Average is 50.

As an example of how the rating works, summary student evaluations were available for one instructor who taught lower division, upper division, and graduate courses. For the lower division course the instructor started with a rating on the question of 77, dropped to 66 in his second year, and then rose to 85 or 89 in the subsequent two years. The same instructor on an upper division course increased from 79 in 1977 to 90 in 1978. In a graduate course, the same instructor had a 61 in 1977, 87 in 1978, and 90 in 1979. Thus the evaluation in any one year is not especially significant. But viewed over time for the same course or course level (i.e., graduate vs. undergraduate), evaluations provide some definite indication of instructor performances. In this instance, the instructor is clearly doing a good job.

One of the advantages of the mid-semester evaluation is that a written record for each course evaluated is provided. The reviewer prepares a memo to the instructor whose class is being reviewed, and these are filed in the instructor's personal file. Review of these evaluations for the past five years reveals obvious trends as well as some perplexing problems. First, students gave fewer suggestions for improvement in the courses. Five years ago the memos typically contained eight to ten suggestions for improvement. Reviews for the same instructor in recent years offer fewer suggestions for improvement. Problems that were easy to overcome such as providing course outlines and giving students clear explanations or descriptions of the course objectives were "solved." In other words, suggestions for improving

teaching were followed, a fact that would not be known without comparing several years of mid-semester reviews.

In contrast, review of these same documents showed persistent criticisms appearing for some instructors. In 1976, one instructor was criticized because he mumbled. The same comments appeared in 1977 and 1979. The same instructor was also criticized for not emphasizing important points of his lecture in 1976 and 1977. The same comment appeared in 1979. Another instructor was criticized early in 1976 for administering a test but not grading it for several weeks. This past spring's review of him contained the same comment. The point is that these unresolved problems were not apparent from a one-time teaching review. They are pronounced, however, after a comparison of several years' reviews. Consequently the department chairman can more emphatically point out these persistent problems to the instructor.

Concluding Comment

Providing a history of teaching effectiveness and placing less emphasis on one-time evaluation of teaching are valid evaluation methods. Histories or trends in evaluations accomplish two objectives. They provide input to the faculty to improve their teaching. Points that did not appear significant during annual evaluations become glaring criticisms when viewed over time. Further, these trends provide administrators with a substantial amount of information for making personnel decisions. The trends from the various methods of evaluating teaching provide administrators with enough information to place an instructor in either of two categories: instructors doing a good job or those doing poorly. Based on this classification, appropriate personnel recommendations can be made. Should both of these objectives be attained by monitoring teaching evaluations over time, this would be a major accomplishment. Fulfilling both objectives is the main reason for evaluating teaching.

References

- Aleamoni, Lawrence M. March, 1976. "Typical Faculty Concerns About Student Evaluation of Instruction," *NACTA Journal*.
- Burger, A.W. and R.D. Seif. September, 1975. "Teacher Rating Versus Backgrounds of Enrollees," *NACTA Journal*.
- Casavant, Ken. June, 1972. "Colleague Aided Evaluation (CAE) as an Experiment in Improved Teaching," *NACTA Journal*.
- Ching, C.T.K. and J.R. Garrett. March, 1978. "Mid-semester Review of Teaching," *NACTA Journal*.
- Foth, Henry D. October, 1972. "Effective Teacher Evaluation," *Journal of Agronomic Education*.
- Marg, Elwin. Spring, 1979. "How to Sacrifice Academic Quality," *Improving College and University Teaching*.
- Marshall, Max S. September, 1975. "What Price Evaluation," *NACTA Journal*.
- Moody, E. Grant and R. Merwin Deaver. March, 1973. "Evaluating Competence of College Teachers," *NACTA Journal*.
- Popham, W. James. 1973. *Criterion-Referenced Measurement, an Introduction*, Educational Technology Publications, Englewood Cliffs, New Jersey.
- Shrode, Robert R. March, 1975. "Does a Student's Status as a Scholar Affect His Evaluation of Instruction?" *NACTA Journal*.