

Group

1

The Use of Student Evaluations In The Improvement of Instruction

Lawrence M. Aleamoni

One of the basic objectives of the University could be defined as attempting to insure that the following generation be better, wiser, and more knowledgeable than the present one. This assumes that the University's educational program is capable of changing student behavior in a specified direction as a result of definite courses of instruction. If such a change in student behavior does occur, then one could claim that the courses have been effective.

If the courses have been effective, then there are probably many course elements which contributed to their effectiveness, such as the instructor, textbook, homework, course content, method of instruction, student interest, student attention, and general student attitude toward the course. Assuming that all of these elements can affect, directly or indirectly, student behavior in a course, and assuming that the students are the only ones who are constantly exposed to those elements, then the students appear to be the most logical evaluators of the quality and effectiveness of those course elements. In addition, student opinions should indicate areas of rapport, degrees of communication, or the existence of problems and thereby help instructors as well as instructional developers describe and define the learning environment more concretely and objectively than they could through other types of measurement (Aleamoni & Spencer, 1973). It should be emphasized, however, that student opinions represent only one component of a multi-component system necessary for the evaluation of instructional effectiveness.

Although there are many ways of sampling student opinion, measurement is more useful when comparative results are available. More adequate interpretation may occur when (a) the data have been collected in a standardized fashion with appropriate attention given to sampling, reliability, and validity, and (b) many instructors and instructional programs have been measured with the same instrument so that comparisons can be made.

Over the past 56 years many student opinion questionnaires have been developed and used throughout the country and world. These questionnaires have ranged from the "overnight" student or faculty generated form to the carefully researched and developed form by questionnaire experts. Regardless of how such forms were developed they all seem to have generated common concerns among faculty who use them. In a recent paper

Workshop session presented by Dr. Lawrence M. Aleamoni, Director of the Office of Instructional Research and Development, The University of Arizona, during the 26th annual NACTA Conference at New Mexico State University, June 15-18, 1980.

which deals with eight typical faculty concerns about the appropriateness of using student ratings of instructor and instruction, Aleamoni (1976) points out that very few are supported by the research evidence dating back to 1924. In fact the evidence indicates that college students are very perceptive and reliable judges of the instructional setting. A similar conclusion also was reached by Cooper and Petrosky (1976) in their recent study of secondary school students' perception of mathematics teachers and courses.

The eight typical concerns frequently encountered are:

1. Students cannot make consistent judgments concerning the instructor and instruction because of their immaturity, lack of experience, and capriciousness.
2. Only colleagues with excellent publication records and experience (who are usually also considered to be excellent instructors) are qualified to evaluate their peer's instruction.
3. Most student rating schemes are nothing more than popularity contests with the warm, friendly, humorous, easy-grading instructor emerging as the winner.
4. Students are not able to make accurate judgments until they have been away from the course and possibly away from the university for several years.
5. The student rating forms are both unreliable and invalid.
6. There are extraneous variables or conditions that can affect student ratings. Some of the more common ones are (a) the size of the class, (b) the sex of the student and sex of the instructor, (c) the time of day the course was offered, (d) whether the student was taking the course as a requirement or on an elective basis, (e) whether the student was a major or a nonmajor, (f) the term (or semester) the course was offered, (g) the level of the course (freshman, sophomore, junior, senior, graduate), and (h) the rank of the instructor, (instructor, assistant professor, associate professor, full professor).
7. The grades or marks students receive or expect to receive in the course are highly correlated with their ratings of the course and the instructor.
8. Finally, a question that is frequently raised is "How can student evaluations possibly be used to improve instruction?"

Of the eight concerns mentioned above only part of one (No. 6) seems to be substantiated by research findings. The research indicates that (a) the higher the proportion of students taking the course as a requirement, the lower their rating, (b) as students move from lower level courses (i.e., freshmen) to higher level courses

(i.e., graduate) their ratings become higher. As for the rest of the typical faculty concerns, the research literature is highly supportive of the significance of student evaluations as accurate reflections of student attitudes (Aleamoni & Hexner, 1980; McKeachie, 1979).

Major universities need visible ways of convincing their faculty that they (a) want good instructors, (b) sincerely respect good instruction, (c) have specific ways of recognizing and rewarding good instruction, and (d) are interested in improving instruction. Every institution is already committing a major portion of its budget to maintain its instructional program: faculty salaries, classrooms, libraries, laboratories, instructional services, sabbatical leaves, and the like. Interestingly enough, when a new Ph.D. graduate prepares to embark on a professional career at any one of the major universities in the United States or abroad, he or she is told by the department head or dean that the institution embraces the three general objectives of excellence in research, teaching, and service and that rewards are based upon satisfactory to excellent performance in any one or a combination of those objectives. The sad fact is that after a short period of time at the institution, the new faculty member realizes that although the three objectives of research, teaching, and service are appropriate for any institution of higher education, most of the institutions reward faculty primarily for their performance in the research function (Astin & Lee, 1966; Thorne, Scott, & Beard, 1976).

This is a disturbing state of affairs because it indicates, basically, that the institution is only interested in supporting and encouraging excellence in research. Its consequences are obvious in that faculty, regardless of their interest, may neglect their teaching and service activities in order to attain the professional recognition required to remain and succeed at their institution. Students are perhaps the most unfortunate pawns in such a game since they are forced to take courses from faculty who are not able or willing to take the time to prepare and organize their courses or to do such things as spend time outside of class discussing problems and concerns that would help most students learn material better.

Some take the position that there is no inconsistency in this type of skewed reward system since it is suggested that excellent researchers are, in fact, the best teachers (Deming, 1972). The research evidence (Guthrie, 1954; Stallings & Spencer, 1967; Swanson & Sisson, 1971; Aleamoni, 1972; Aleamoni & Yimer, 1973), however, does not support this point and shows that, in general, there is no correlation between scholarly productivity and effective teaching.

It is further suggested that the evaluation of scholarly productivity is much easier, more valid, and more reliable than is the evaluation of instructional effectiveness. This attitude has generally resulted in a policy (dictated by practice) indicating that, regardless of the quality of the evaluative teaching evidence, published research will still take precedence in the reward considerations. Under this policy the relationship between teaching and scholarly productivity is believed to be high and positive even though, as was stated above, the research evidence to date does not support this belief.

Not let us look more closely at the criteria used to judge scholarly achievement. They rest basically on the belief that judgments by colleagues (synonymous with "the academic community at large") provide the final evidence. These judgments typically include (a) publication in journals where expert evaluation is required for acceptance, (b) favorable reviews of books, (c) appointments or awards that require evaluation of professional competence, (d) election to office in learned societies, and (e) receipt of fellowships. If one looks closely at the above judgments, it is obvious that each of these is very subjective; in fact it would be very difficult to find a set of objective criteria on which many colleagues could agree when judging journal publications, books, etc. Those who have read several advisory editors' reviews of their articles or experienced the acceptance of a once rejected article by another equally reputable journal could attest to this subjectivity.

Where then is the rationale for maintaining that the evaluation of teaching based on (a) the testimony of department heads and deans, (b) comments of colleagues who are well acquainted with the teaching performance

26th NACTA CONFERENCE NMSU

Presenting a strong workshop session at the 1980 NACTA Conference, were from the left, E. Grant Moody, Chairman from Arizona State University; Lawrence Aleamoni, University of Arizona; George B. McLeroy, International Animal Resources Consultant from Arizona; Peter Fog, Coordinator of PPP, University of Minnesota Technical College at Waseca; and Donald C. Roush, Acting President of the New Mexico State University.



of the instructor (c) achievement of students, (d) quality of teaching materials prepared by the instructor, and (e) the judgments of students exposed to the instructor, is more subjective, less reliable, and less accurate than the judgments of scholarly achievements? Apparently, the answer lies in the fact that the methods and techniques used in evaluating scholarly achievement have remained unchanged for so long that they have become accepted as standard without question or evaluation. In contrast, the evaluation of teaching has been subject to a wide variety of approaches with no commonly accepted methods or techniques except for student judgments gathered via rating forms, some of which have not, admittedly, been professionally designed.

A recent study by Ladd and Lipset (1975) further emphasizes the need for institutions formally recognizing and rewarding instructional effects. They conducted a survey of U.S. faculty members and found that as a group they are more interested in teaching than research and scholarship. In fact, for every one professor strongly devoted to research, they found that there are nine equally devoted to teaching.

If institutions of higher education (and their departments) are serious about encouraging excellence in teaching, then they must be prepared to reward such excellence in a manner similar to that for excellence in scholarly achievement. One approach toward achieving this goal would be to have departments define what constitutes effective instruction and then establish a system of evaluation which would have as its two major outcomes (a) guidance for improvement, and (b) reward for success.

A question that naturally arises in developing systems of instructional evaluation is "Can student ratings of instruction and instructor be useful in improving college teaching once they are made available to the instructor?" Although there has been a great deal of anecdotal evidence from instructors and researchers to suggest that student evaluations do have a positive effect, very few studies are available that deal with that effect on college-level instruction. McKeachie (1979) felt that the most impressive results were those reported by Overall and Marsh (1979). They found that instructors who received feedback from student ratings not only received more favorable ratings at the end of the year, but their students also scored higher on an achievement test and on a measure of motivation for further learning and application of the material learned, when compared with a control group of instructors.

In a study at the University of Arizona, Aleamoni (1978b) investigated the effect on faculty performance of a combination of feedback of ratings and personal consultations. This approach proved to be superior to simply providing a printed report of the results. McKeachie and Lin (1975) obtained similar results.

In his review of studies on the validity of student ratings in achieving cognitive, attitudinal, and motivational goals, McKeachie (1979) concluded that, taken as a whole, the results indicate that the highly rated instruc-

tors and courses result in higher student achievement of cognitive, attitudinal, and motivational goals. More research, however, needs to be conducted in these areas.

Instructors would be willing to devote the time and resources necessary to achieve excellence in teaching if they felt such efforts would be appropriately rewarded. However, the limited research available (Aleamoni, 1978a; Centra, 1977; McKeachie, 1979; Thorne et al., 1976) indicates that very little if any weight is applied to evidence of teaching effectiveness in promotion and tenure decisions. Where such evidence is used it tends to be based on (a) anecdotal or testimonial information, or (b) student ratings. However, there are very few "systems of instructional evaluation" that have been developed to provide systematic, objective, and comprehensive instructional evaluation evidence. Where such "systems" have been developed, the satisfaction on the part of the participant faculty and administrators appears to be very high (Harrell, 1979; O'Connell & Smartt, 1979). It, therefore, seems logical to encourage the development of comprehensive systems of instructional evaluation that have guidance for improvement and reward for success as their two major outcomes. When student ratings are introduced as a component of such systems then faculty acceptance of their use will be greatly enhanced.

REFERENCES

- Aleamoni, L. M. 1972. Response to Professor W. Edward Deming's "memorandum on teaching." *The American Statistician*. 26(4), 54.
- Aleamoni, L. M. 1976. Typical faculty concerns about student evaluation of instruction. *NACTA Journal*. 20(1), 16-21.
- Aleamoni, L. M. 1978. *Report on the investigation of the criteria and guidelines used by various departments and colleges at the University of Arizona for promotion and tenure decisions (RM 4)*. Tucson, AZ. University of Arizona, Office of Instructional Research and Development. (a).
- Aleamoni, L. M. 1978. The usefulness of student evaluations in improving college teaching. *Instructional Science*. 7, 95-105. (b).
- Aleamoni, L. M., & Hexner, P. Z. 1980. A review of the research on student evaluation and a report on the effect of different sets of instructions on student course and instructor evaluation. *Instructional Science*. 67-84.
- Aleamoni, L. M., & Spencer, R. E. 1973. The Illinois course evaluation questionnaire: A description of its development and a report of some of its results. *Educational and Psychological Measurement*. 33(3), 669-684.
- Aleamoni, L. M., & Yimer, M. 1973. An investigation of the relationship between colleague rating, student rating, research productivity, and academic rank in rating instructional effectiveness. *Journal of Educational Psychology*. 64(3), 274-277.
- Astin, A. W., & Lee, C. B. T. 1966. Current practices in the evaluation and training of college teachers. *Education Record*. 47, 361-375.
- Centra, J. A. 1977. *How universities evaluate faculty performance: A survey of department heads* (GREB No. 75-5bR). Princeton, NJ: Educational Testing Service.
- Cooper, E. R., & Petrosky, A. 1976. Secondary school students' perception of math teachers and math classes. *Mathematics Teacher*. 69(3), 226-233.
- Deming, W. E. 1972. Memorandum on teaching. *The American Statistician*. 26(1), 47.
- Guthrie, E. R. 1954. *The evaluation of teaching: A progress report*. Seattle, Washington. University of Washington.
- Harrell, R. A. April, 1979. *The development of a comprehensive faculty evaluation system*. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco.