Student-Faculty Evaluation

Leon W. Zelby

Recent changes in attitudes toward higher education, particularly after the publication of several reports in the early part of 1971 [for example, by the Carnegie Commission, Newman *et al.* (*1*), and the Assembly on University Goals and Governance (*2*)] have led to a close reexamination of many different facets of education and of educational institutions, with great emphasis on innovations and on quality of teaching. Some of the aspects of education that received particular attention were teaching versus research, teaching innovations, use of technological teaching aids, and cost effectiveness. In addition, much attention has been directed recently toward the student-faculty evaluation (SFE). All of these issues have been very controversial because of the difficulty associated with establishing definite and unambiguous conclusions that could be supported by experimentation.

Controversies on some subjects, for instance on the subject of teaching versus research, have abated a little following the publication of an article by Hayes (*3*). He showed that there was a lack of correlation between the quality of teaching and the quality of research. He also pointed out that the degree of correlation may vary among institutions. Other controversies, however, including those over teaching innovations (*4*), and particularly that over SFE (*5, 6*), are continuing, with the latter increasing in intensity.

One major cause for the growth of the SFE controversy is the trend toward formal, quantitative use of the results of the evaluations in determinations of faculty promotions and salaries. There are at least three reasons for this trend: (i) SFE provides documented, precise, numerical evaluation of instructors; (ii) it tends, thereby, to relieve academic administrators from the responsibility of exercising judgment about teaching performance and ability; and (iii) it tends to constitute proof that something, indeed, is being done to improve teaching.

The overall impact of SFE has not yet been appreciated because not all institutions have been using it and because it has been directed almost exclusively toward the determination of "teaching effectiveness." The SFE's may, however, adversely affect the quality of education if they are inappropriately designed and used in determinations of salary and promotion. It is, therefore, necessary to recognize this potential effect and to consider the broader effects of SFE's on quality of education and quality of educational institutions, as well as their effect on quality of instruction, or of instructors. There may be long-range deleterious results unless more careful consideration—or, better yet, assessment—is made of SFE's before they are used extensively.

SFE and Education

A number of institutions, particularly engineering schools, have had SFE for several decades (*7*). Until quite recently, the primary purpose of SFE was to provide information that could be helpful to students and teachers alike. The current tendency to use these evaluations in a quantitative and formal sense changes the complexion of SFE from that of a helpful collection of information to a device that could become detrimental to education. Its potential for good or bad lies in the formulation of the questionnaire and the interpretation and utilization of the results. The format of SFE depends very much on the character and aspirations of the student body and of the institution. This implies that what may constitute a desirable format in one institution will not necessarily constitute a desirable format in another.

One reason for this is that students' attitudes and expectations are related to their demonstrated abilities. Consequently, students in the more demanding and prestigious schools will tend to favor more challenging forms of instruction than will students in institutions with more relaxed requirements. Since the attitudes of faculties in different schools also vary, the SFE may tend to reflect these attitudes, with possibly adverse results for a large number of schools. This is because the good schools will tend to become better, whereas the others may tend to get worse. For instance, an SFE format that tends to favor challenging, interesting, and innovative forms of instruction will have a more beneficial influence on the quality of education in the long run than a format favoring stereotyped and unimaginative instruction, even though the latter may seem more comfortable to some students and teachers. Most of the reported analyses of SFE's seem to indicate, however, that the emphasis is on the stereotype—on information storage and retrieval—which constitutes only a small part of the educational process. Thus, an uninformed or careless use of a particular SFE will tend to widen the gap between first-rate and second-rate institutions and will tend to spread mediocrity and uniformity among the latter, a tendency already noted in the Carnegie Commission and Newman reports (*1*), irrespective of SFE.

Many articles relating teaching effectiveness to SFE have been written; those of the Rodins (*5*) and Gessner (*6*) are particularly characteristic of the controversy in this area. Their diametrically opposed conclusions, in spite of careful experimentation and meticulous statistical analysis, emphasize the difficulties connected with the resolution of only one issue related to SFE—that of correlating teaching effectiveness and SFE.

These articles are also characteristic in that they follow the trend of attempting to establish some definite, general, and unique relation between SFE and teaching effectiveness. Teaching effectiveness is, however, only one factor in the quality of education, which is not necessarily identical to the quality of instruction. A more significant aspect of SFE than simply the quality of instruction is its potential impact on the overall quality of education.

The conflict between the Rodins' (*5*)

The author is professor of electrical engineering, School of Electrical Engineering, University of Oklahoma, Norman 73069.

and Gessner's (6) conclusions can be attributed to several factors. The obvious factors include differences in sample size (119 and 293 students, respectively) and make-up: for example, the Rodins' sample consisted of undergraduate students in a calculus course, whereas Gessner's sample consisted of second-year medical students—a highly selected and motivated group—in a basic science course. The institutional and national examinations used to determine the correlation between teaching effectiveness and SFE may have been designed for completely different purposes, just as the courses could have been. Finally, the formats of the respective SFE's, which were not reported, may have been quite different.

The subtle factors include fundamental differences between the two institutions—for example, admission requirements and, consequently, the caliber of students, their maturity and motivation; popularity of certain courses, whether required or elective; the carry-over of an instructor's reputation, deserved or not, because of a particularly good or bad single performance; the "halo" (8) and Disneyland effects and periodic fluctuations in student interests and performance; geography and class scheduling; and many others that may be extremely important, although not readily quantifiable.

Admittedly, it would be preposterous to suggest—much less advocate—that SFE be postponed until all factors that may affect its results can be included in such evaluations. It is equally preposterous, however, and possibly more dangerous, to use the results of an SFE that has an inappropriate format.

In this article I demonstrate the importance of the format by showing that, given a specific format, it is possible to adapt one's teaching technique to obtain a good or bad evaluation and that a good evaluation may be associated with a teaching technique of lesser educational value than a poor evaluation.

Table 1. SFE data for E (electromagnetic fields) and T (technology in society) courses (1 is best and 5 is worst, except for questions 8 through 11).

Questions*	Evaluation			
	Teaching methods in E courses		Teaching methods in T courses	
	a	b	c	d
1. The instructor seemed to have made adequate preparation for class	1.30	1.67	1.36	1.86
2. The instructor expressed himself clearly and concisely	1.91	2.61	1.86	2.21
3. The instructor clarified material in reading assignments†	3.00	4.65	2.00	4.27
4. The instructor was available and helpful outside of class	1.48	2.06	1.64	2.38
5. The instructor gave well thought-out assignments	2.33	3.48	2.57	3.86
6. The exams and quizzes offered a reasonable coverage of material	2.04	3.11	2.18	2.86
7. The results of the exams and quizzes were an adequate measure of your knowledge in this course	2.35	3.39	1.80	3.00
8. The knowledge assumed about the prerequisite subject was	3.72	3.96	3.00	3.00
9. The number of exams and quizzes was	1.70	2.82	2.20	2.93
10. The time required for homework assignments was	2.57	3.56	3.00	3.21
11. The credit given for homework assignments was	2.13	3.06	2.57	2.71
12. Considering the instructor's effectiveness, how would you rate him compared to other engineering faculty members?	1.96	2.44	1.71	2.67
13. Considering the course content only, how would you rate this course?	2.48	2.61	2.36	2.64
14. Considering the amount of knowledge you gained in this course, how would you rate it compared to other engineering courses?	2.52	2.56	2.00	3.42
15. How would you rate yourself as a student in this course?	2.74	2.94	2.43	2.36

* Evaluate the statements by using for questions 1 through 7: (1) definitely yes, (2) yes, (3) unsure, (4) no, and (5) definitely no; for questions 8 through 11: (1) far too few (little), (2) too few (little), (3) about right, (4) too many (much), and (5) far too many (much); for questions 12 through 15: (1) excellent, (2) good, (3) average, (4) mediocre, and (5) poor. † The original version of this question was, "The instructor merely clarified material in reading assignments," but the word "merely" was removed.

The Experiment

To test the hypothesis that it is possible to teach toward a specific SFE and to point out, thereby, the need for its careful design, I have been alternating my teaching techniques for several semesters since 1970, when SFE was first adopted by the college of engineering. The hypothesis suggested itself after I had been teaching the second course in each of two different sequences. I noticed that there was, on the average, no discernible difference between the performance of students who had had "good" teachers in the first course of the sequence and students who had not.

The following constraints were imposed on the experiment: the students were not informed about the experiment, in order to avoid the Hawthorne effect; the requirements of each course, as specified in the catalogue, had to be fulfilled, regardless of the teaching technique—that is, the topics required had to be covered and the course level maintained; and the time demands upon the students should not exceed, on the average, the standard 2 hours of homework for each hour of lecture.

The results of the initial experiment, performed in the Spring of 1970, have been reported elsewhere (9). Since then, continuing the experiment to test reproducibility of the results, I have taught two different courses each semester for several successive semesters: "technology in society" (T) and "electromagnetic fields" (E). During one semester, the E course was replaced by "electronic properties of materials," a course essentially equivalent to E in content and in the kind of student it attracts. The courses were junior to junior-senior level, with the T course an elective, without prerequisites, and open to all students. The E course was required in electrical engineering and had prerequisites in mathematics and physics. The population in this course was very homogeneous, rarely including students not majoring in electrical engineering, and averaged

22 students per semester; the population in the T course was very heterogeneous, including majors from all but the colleges of fine arts and education, and averaged 12 students per semester.

The administrative procedures in each course—distribution of syllabuses, lists of required and recommended reading, homework, solutions, examination dates, and so forth—were the same, regardless of the teaching technique. Required and recommended reading materials were on reserve in the appropriate libraries. The first lecture of each semester in each course was devoted to an explanation of the nature of the subject matter, its relation to the curriculum, and its purpose. The teaching method to be used in the course was discussed and justified without indicating the specific reasons for it.

In the E courses the two teaching techniques were as follows: (a) lectures were based primarily on the assigned textbook, with few deviations from the text in respect to approach, derivation, or point of view; assigned problems were usually solved in detail on the blackboard during class time; and examination problems were very similar in nature to homework assignments and (b) lectures were designed primarily to augment the textbook; problems were rarely solved in detail in class, but various possible approaches to solutions were presented (in each case, written solutions to the homework and examination problems were distributed); and examination problems were designed to require a somewhat different application of concepts from that required in the homework assignments. In both cases, reasonable familiarity with prerequisites was assumed (see Table 1).

In the T courses the techniques were as follows: (c) lectures were based on the assigned reading material and were mainly in the nature of a critique, with discussions of a fairly general nature; data were used sparingly, and then only to illustrate, rather than prove, a particular point; assignments consisted of essays evaluating and criticizing material read and discussed and (d) lectures were designed primarily to augment reading assignments; discussions were more specific, requiring careful analyses of diverse points of view; and assignments required logical and consistent developments of points of view or opinions, supported by relevant data; and additional reading material was suggested by directing attention to types of journals rather than specific issues and page numbers (see Table 1). The primary aim of methods a and c was information storage and retrieval; that of methods b and d, the development of an ability to learn more independently and to cope more successfully with novel or different situations. A possible analogy between the two methods would be that of writing a finished (methods a and c) or a self-teaching (methods b and d) computer program for the solution of some inexact problem—for example, in education, or chess (*10*).

Good SFE's followed methods a and c, placing me in the upper quarter of the electrical engineering faculty; poor SFE's followed methods b and d, placing me in the third quarter. A sample of the data is given in Table 1 next to the appropriate questions. Responses show that methods a and c were consistently voted better than methods b and d, even though the latter required much more time and effort on the part of the instructor (see question 1); yet responses to question 10 indicate that time demands on students were not excessive in b and d, although greater than in a and c. Responses to question 4 reflect the differences in teaching methods (interpretation of "helpful"), since availability was the same in all cases.

In the T courses, the students preferred vague discussions, without utilization of data, and writers like McLuhan, Reich, or Vonnegut to a systematic and reasonably rigorous consideration of technological effects, based on documentation and data, and writers like Commoner, Drucker, or Schrag. The preference was for, besides critiques of material, originality alone rather than creativity, in the sense that the latter includes some degree of feasibility in addition to novel or different ideas (*11*).

In the E courses, the students preferred a very highly structured format that adhered extremely closely to the textbook in all aspects of the course. They also preferred presentation of homework problems to other forms of lecture, in spite of the fact that written solutions to homework problems and examinations were regularly distributed. There was a very distinct preference for a "how to" technique rather than for a thorough understanding of the basic concepts and their range of application. In both courses, the students were reluctant to read material in excess of specific assignments.

Although it has not been possible to conduct independent examinations at the end of each semester to determine the differences in learning resulting from differences in technique, it is my opinion that the students learned more and better with techniques b and d, in spite of a relatively poor SFE. This opinion is based on comparisons of methods and solutions to homework and examination problems by the E students, and of analyses of issues by the T students. Groups b and d exhibited a better degree of understanding of the respective material and underlying concepts than did groups a and c. Groups b and d did as well as a and c in stereotyped, but much better in more challenging, assignments. In this respect, the experiments tend to support the Rodins' conclusions (*5*, *7*).

Discussion

In view of the fact that the basic requirements of the two courses were satisfied with each scheme, one may be tempted to suggest that the best solution would be to teach for a good SFE. This would be appropriate, given a questionnaire that was designed to meet the needs and the aspirations of the students and the institution, as well as the community at large. A number of questionnaires, however, similar to that given in Table 1, have a format suggesting that teaching for a good SFE may not be consistent with the best educational practices and may not "challenge them [students] to go where they have never been" (*12*, p. 609), one of the major functions of an educational institution [for other formats, see (*7*)]. Except possibly for question 1 and part of question 2, the questionnaire is conducive to stereotyped teaching, certainly not the most desirable form of teaching at the college level. How concisely concepts or phenomena can be explained is not nearly as important as how well they can be explained. Questions must be less vague and deal more specifically with such factors as stimulation versus boredom and challenge versus stereotype (*7*). Thus, the potential dangers of SFE in decreasing the effectiveness of education lie not in the evaluation proper, but in the format of the evaluation.

Data in my experiment have not been rigorously quantitative, despite the three-digit precision in the table. This merely represents an unavoidable victory of precision over accuracy: computerized compilations. Andreski cautions that "The gravest kind of danger stems from the illusion that, because certain data can be quantified and processed by a computer, therefore they must be more important than those which cannot be measured (*13*).

It is difficult to overemphasize the importance of a careful approach to SFE, from which much good can accrue. Careless SFE, concerned only with the narrow aspect of teaching effectiveness—if this indeed can be unequivocally established—will inhibit educational experimentation and development, particularly if SFE is used formally in the determination of salaries and promotions.

There is little doubt, however, that SFE in almost any form will become widely and rapidly accepted because it will permit academic administrators to shirk the responsibility of exercising judgment in the evaluation of teaching performance, and at the same time to use SFE as tangible proof that something is being done about improving teaching.

Conclusions

Desirable attributes of SFE can be vitiated by its premature utilization in a formal and quantitative sense. This is because it is possible to teach for a specific student evaluation, given a particular questionnaire. Consequently, because of the importance of SFE, the questionnaire must be designed to meet the expectations of the students, as well as the aspirations of the respective institutions. Indiscriminate use of SFE will increase the gap between first-rate and second-rate institutions—first-rate institutions will continue to attract more demanding students, a fact that will be reflected in SFE's, whereas second-rate institutions, in an effort to maintain levels of enrollment, may tend to formulate SFE's that emphasize popularity and mediocrity of education. Careful construction of the format of SFE, on the other hand, could do much toward increasing the quality of teaching, as well as the motivation of students and teachers, in many institutions.

References and Notes

1. Carnegie Commission Report, *The Capitol and the Campus* (McGraw-Hill, New York, 1971); F. Newman *et al.*, *Report on Higher Education* (Department of Health, Education, and Welfare, Government Printing Office, Washington, D.C., 1971).
2. Assembly on University Goals and Governance, *First Report of the Assembly on University Goals and Governance* (American Academy of Arts and Sciences, Cambridge, Mass., 1971).
3. J. R. Hayes, *Science* **172**, 227 (1971).
4. R. Dubin and T. C. Taveggia, *The Teaching-Learning Paradox: Comparative Analysis of College Teaching* (Univ. of Oregon Center for Advanced Study of Educational Administration, Eugene, 1969); S. G. Tickton, Ed., *To Improve Learning* (Bowker, New York, 1970–1971), vols. 1 and 2.
5. M. Rodin and B. Rodin, *Science* **177**, 1164 (1972).
6. P. K. Gessner, *ibid.* **180**, 566 (1973).
7. W. J. McKeachie, *AAUP Bull.* **55**, 439 (1969); O. E. Lancaster, *Eng. Educ.* **62**, 95 (1971); *IEEE (Inst. Electr. Electron. Eng.) Trans. Educ.* **E-16**, 138 (1973).
8. J. P. Guilford, *Psychometric Methods* (McGraw-Hill, New York, 1954), p. 297.
9. L. W. Zelby, *IEEE (Inst. Electr. Electron. Eng.) Trans. Educ.* **E-15**, 30 (1972).
10. M. M. Botvinnik, *Computer, Chess and Long-Range Planning* (Springer-Verlag, Berlin, 1970), pp. 5–7.
11. C. S. Whiting, *Creative Thinking* (Reinhold, New York, 1958), p. 3.
12. D. R. Goddard and L. C. Koons, *Science* **173**, 607 (1971).
13. S. Andreski, *Social Sciences as Sorcery* (Deutsch, London, 1972), p. 120.

NEWS AND COMMENT

Uranium Enrichment: U.S. "One Ups" European Centrifuge Effort

The countries of Western Europe seem determined to end their dependence on the United States for a large part of the enriched uranium to fuel their nuclear reactors, but disagreement over the form that the European effort should take has produced a serious conflict of technical issues and national interests.

France has recently announced plans to build a $1.4 billion new plant for uranium enrichment based on the expensive and slow but proved method—gaseous diffusion. Britain, the Netherlands, and West Germany, on the other hand, plan to build a plant based on a newer and more uncertain technology—the gas centrifuge. The advantage of the gas centrifuge method is that it requires far less electrical power and offers much more flexibility in the size of the plant.

As uncertainties about foreign oil supplies make nuclear power more and more appealing, the competition between diffusion and centrifuge methods is becoming a game with high stakes, not only for Europe but also for the United States. Three diffusion plants now easily supply all the U.S. demands for reactor fuel as well as foreign requirements, but new plants will soon be needed. In the next 2 years, the United States must decide whether to stick with the old technology or gamble with the new one.

At a recent press tour of the diffusion plant operated by the U.S. Atomic Energy Commission at Oak Ridge, Tennessee, AEC chairman Dixy Lee Ray told reporters that the United States has a substantial lead over Europe in the development of the gas centrifuge method. The tour marked the first time that reporters had ever been allowed to see the inside of a uranium enrichment plant, and Ray's remarks provided a clearer picture of the AEC's progress in developing gas centrifuge technology than had been publicly available. Because any country with the centrifuge technology could produce weapons-grade uranium in a small, easily concealed facility, the technology is closely guarded by the AEC as well as by Urenco, Ltd., the production arm of the collaborative British-Dutch-West German effort.

"Statements by Urenco officials would indicate that large European production plants would need hundreds of thousands of centrifuges," Ray said. "On the other hand, U.S. technology would require only tens of thousands of centrifuges for large-scale plants. It is this U.S. technology that is now being demonstrated in AEC facilities." Since both capital costs and operating costs of centrifuge enrichment plants are expected to be heavily dependent on the number of units needed, the statements of the AEC chairman indicate that the U.S. process will be many times cheaper.

Urenco has announced plans to have two pilot plants operational by the end of 1976. They will have a combined capacity of 400 metric tons of separative work, or about 5 percent of the capacity of one of the large U.S. diffusion plants. In an apparently coordinated effort to "one up" Urenco, another AEC spokesman repeated Ray's statement 2 weeks later to the Joint Com-