Accounting administrators' perceptions of student evaluation of teaching (SET) information

D. Larry Crumbley and Eugene Fliedner

The authors

D. Larry Crumbley is the KMPG Endowed Professor in the Department of Accounting, Louisiana State University, Baton Rouge, Louisiana, USA.

Eugene Fliedner is a Professor in the Decision and Information Sciences Department, School of Business Administration, Oakland University, Rochester, Michigan, USA.

Keywords

Evaluation, Individual behaviour, Impression management, Students, Instructors

Abstract

Most business schools use student evaluation of teaching (SET) survey data for promotion, tenure, and merit decision-making purposes. Since most SET questionnaires focus on students' perceptions of an instructor rather than learning, there may be an incentive for instructors to resort to dysfunctional behavior in order to manipulate SET scores. The purpose of this article is to report the results of a survey designed to determine if such behavior occurs from an administrative viewpoint. A total of 773 administrative accounting professors were surveyed, with a response rate of 45.3 per cent. Although most administrators believe that a single numerical measure cannot capture all relevant evaluative data, they do believe that SET has caused grade inflation and they are dissatisfied with their current SET system. However, the majority of administrators would not replace the current evaluation system with an alternative evaluation system.

Electronic access

The research register for this journal is available at
http://www.emeraldinsight.com/researchregisters

The current issue and full text archive of this journal is available at
http://www.emeraldinsight.com/0968-4883.htm

Quality Assurance in Education
Volume 10 · Number 4 · 2002 · pp. 213–222

DOI 10.1108/09684880210446884

Administrator reliance upon student evaluation of teaching (SET) surveys to evaluate teaching effectiveness is an important and sensitive issue facing today's college-level faculty and administrators. There is growing controversy in the literature regarding the use of these instruments as they play a vital role in the promotion, tenure, and merit process.

Seldin (1993) found that the number of institutions using SET to evaluate teachers has climbed from 29 per cent in 1973, to 68 per cent in 1983, and to 86 per cent in 1993. Currently, SET instruments are used by more than 94 per cent of accounting departments within schools of business, eight percentage points higher than the national usage (Calderon *et al.*, 1994).

One important reason for the concern regarding the use of SETs is noted by Calderon *et al.* (1994). These authors note that although 82 per cent of accounting administrators use multiple information sources in assessing faculty teaching performance, 18 per cent of accounting administrators utilize only SET information. Most business schools now use SET data for decision making, and 95 per cent of the deans at 220 accredited undergraduate schools "always use them as a source of information" (Accounting Education Change Committee, 1993)[1].

With the growing use of SET information, research has increasingly questioned the validity of these surveys as an indicator of instructor effectiveness (Bauer, 1996; Boex, 2000; Crumbley, 1995; Drowling, 2000; Ellis, 1985). There is growing evidence in the literature that overemphasis on the numerical results of these survey instruments may be contributing to an erosion of quality teaching and scholarship, to a lower level of respect for teachers, and to a weakening of faculty positions (Greenwald, 1977; Haskell, 1997; Sacks, 1996). Registrars at even top schools believe grades have accelerated faster than student talent levels. At Bucknell University, 80 per cent of all grades given are As and Bs, compared to 50 per cent in the 1960s (Bulkeley, 1997). Emery *et al.* (2000) state that "just like bankers do not let customers set interest rates on their loans, the business schools should not allow students to dictate what topics the curriculum should include or what grades they should receive."

The purpose of this research was to elicit from administrators' viewpoint opinions

Quality Assurance in Education
Volume 10 · Number 4 · 2002 · 213–222

regarding the effectiveness of the SET instrument as a means of evaluating college-level instructors. At issue is whether the use of SET data leads to dysfunctional behavior by instructors and to determine administrators' perceptions of the effectiveness of SET for measuring instructor effectiveness. Specifically, this survey was conducted to determine whether administrators believe that teachers dysfunctionally alter their behavior to improve their SET scores.

Results indicate that administrators do believe it is possible for instructors to manipulate student responses within SET data to achieve higher scores on student-reported measures of teaching effectiveness. Additionally, our results indicate that many administrators are aware that this may happen; however, administrators are seemingly satisfied with the present evaluation process.

The remainder of this article is organized as follows. Dysfunctional behavior and relevant research are discussed in the next section, followed by a description of the methods used to survey the opinions of administrators regarding SET effectiveness. The results are then reported, and some conclusions close the article.

Dysfunctional behavior and relevant research

In a performance measurement system judged by student evaluations, classroom behavior and motives of some teachers may be partly explained through an impression management theory (also known as self-representation theory). As noted by Rosenfeld *et al.* (1994), "impression management refers to the many ways by which individuals attempt to control the impressions others have of them: their behavior, motivations, morality, and a host of personal attributes." Schlenker (1980), Schneider (1969), and Swann (1987) note that most individuals desire to be viewed in a favorable manner by others, and they construct a favorable image of themselves in order to maximize rewards, maintain their self-esteem, and create a desired self-identity.

Human nature suggests that if you are in the position to evaluate an individual's work and do not provide a superior evaluation (e.g. an instructor giving a student low grades), such an individual may not evaluate you highly on an anonymous questionnaire. Centra and Creech (1976) report a moderately strong, statistically significant relationship between student grade expectations and the students' rating of instructor effectiveness. Haladyna and Hess (1994) found that 38 percent of evaluations were a result of bias. Students expecting an A grade evaluated instructor effectiveness with a mean of 3.95 while those students expecting a D grade gave a mean rating of 3.02. An instructor's grading policy and course rigor may be significant factors in determining student responses on instructor evaluations. Certainly many instructors believe that Newton's (1988) leniency hypothesis is valid and take corrective actions to improve their evaluations. Ryan *et al.* (1990) note that at least one-third of their survey respondents indicated they have substantially decreased their grading standards and level of course difficulty. Bures *et al.* (1988) found that only 20.4 per cent of 559 accounting professors agreed with the statement that SETs are indicative of an instructor's teaching and should be used directly in calculating annual salary increases.

If an instructor can choose teaching styles, grading difficulty, and course content, he or she may prefer the choices that are expected to result in higher SET scores. According to Medley (1979), "if teachers know the criteria on which decisions affecting their careers are based, they will meet the criteria if it is humanly possible to do so." Worthington *et al.* (1979) argue that "as an instructor inflates grades, he or she will be much more likely to receive positive evaluations." Many SET enhancement choices have the potential to be dysfunctional or anti-learning, resulting in grade inflation, course work deflation, and "pander pollution" behavior. Pander pollution may be defined as purposeful intervention by an instructor inside and outside the classroom with the intention of increasing SET scores, which is counterproductive to the learning process. Increasing use of the SET has the potential for professors to engage in pander pollution in an attempt to enhance their SET scores[2].

The research findings of Yunker and Sterner (1988) and Bures *et al.* (1990) provide support for the objective of this research. These two nationwide studies indicate that accounting department

Quality Assurance in Education
Volume 10 · Number 4 · 2002 · 213–222

chairpersons rely heavily upon SET performance for evaluating faculty job performance. Reliance upon this factor was second only to research publications in professional journals.

Administrator reliance upon student ratings of teaching effectiveness for job performance evaluation is an important issue. For example, Yunker and Stern observe that 37 per cent of their respondents were dissatisfied with the present evaluation system. In a survey of 561 accounting professors, Bures *et al.* (1990) report that 15.3 per cent indicated "strong disagreement", 31.7 per cent "disagree", and 25.5 per cent indicated a "neutral" response to the statement that the present system for student evaluation of faculty is well-designed and properly implemented.

Research methods

Although many college-level administrators and instructors support the use of SETs in the evaluation process, there is clearly considerable disagreement and dissatisfaction among many academicians that student evaluations are effective tools for assessing instructor effectiveness and teaching quality. To elicit the opinion from administrators' viewpoint as to whether dysfunctional behavior exists on behalf of instructors in an effort to improve SET scores and to determine administrators' perceived effectiveness of the SET for measuring instructor effectiveness, a mail survey was sent to all of the 773 administrative accounting professors employed at four-year universities and colleges in the USA. These individuals were identified using the *Accounting Faculty Directory* (Hasselback, 1998). A total of 350 usable surveys were returned for an overall response rate of 45.3 percent. Nonresponses on a few individual items resulted in a slight variation in the number of responses for each question. Descriptive statistics (rankings, counts, means and standard deviations) are reported and discussed below.

General survey information

Descriptive university/college information was solicited to indicate the generalizability of the survey results. For example, 57.6 per cent of the respondents were at public institutions, while 42.4 per cent were at private schools. Although 45.9 per cent of the respondents noted AACSB accreditation, 54.1 per cent noted the lack of such accreditation.

General teaching survey information was also solicited, including administration information and usage information for both summative (numerical responses to simple questions typically employing a Likert-type scale ranging from 1, strongly disagree, to 5, strongly agree) and formative (opened ended questions which solicit students' suggestions for instructor improvement) teaching evaluation surveys. An overwhelming 95.1 per cent of the respondents indicated summative surveys are administered to evaluate teaching effectiveness. A smaller 41.7 per cent indicated that either a separate formative survey instrument or a combined summative/formative instrument is also used.

In addition to providing faculty with feedback for improvement, the information from SET surveys also serves students and administrators. For example, some advocates report that students find this information useful in course selection decisions. However, only 11.4 per cent of the respondents indicated that SET scores are made available to students.

An overwhelming majority (95.1 per cent) of the respondents said that the SET was a useful tool for evaluating teacher effectiveness. A total of 90.6 per cent of the respondents said SET scores are used in promotion decisions, with 69.3 per cent indicating use of SET scores in merit pay decisions. Respondents also were asked to indicate other methods for evaluating teaching effectiveness. The most frequently cited tools along with the frequency that they were mentioned included evidence provided by the instructor (24.0 per cent), student comments (16.6 per cent), classroom visits (16.0 per cent), peer reviews (15.1 per cent), classroom visits by department head (9.1 per cent), former student comments (4.3 per cent), and student performance (4.0 per cent).

Dysfunctional behavior acknowledged

To determine whether administrators believe teachers alter their behavior in order to improve their SET scores, responses to

several questions were solicited. A minority of 39 per cent of the respondents were aware of instructors that have altered their classroom behavior (reduced their grading standards and course work content) in order to improve SET scores. This corresponds closely with the finding reported by Ryan *et al.* (1980), that one-third of their survey respondents decreased their grading standards and level of course difficulty.

In an effort to determine factors that are most important, from an administrator's perspective, in shaping student responses and the resultant SET scores, the administrators were asked to rate on a five-point, Likert-type scale (1 being strongly disagree and 5 being strongly agree) 17 factors that may affect SET scores. Results of the administrators' responses are tabulated in Table I. As reported, an instructor's preparation and organization was ranked first, followed closely by enthusiasm, presentation, and teaching style. Factors such as class size, time of day the class meets, and giving students free time were felt to have negligible effects on SET scores.

Another question asked respondents to rank 15 factors on a five-point, Likert-type scale regarding their perceived importance for improving SET scores. This information is reported in Table II. The results indicate administrators clearly do believe an instructor should avoid embarrassing students in the classroom. Easy exams, providing students with sample examinations, keeping students' grade expectations high, grading on a curve, inflating grades, and course work deflation were perceived to be not nearly as influential for improving SET scores. Teaching in a regulated class (common exams, books, and course syllabi) was perceived to be of little importance.

In another question, administrators were asked to indicate the actions they pursue given knowledge or suspicion that an instructor has reduced the course work load or used more lenient grading standards in an attempt to improve SET scores. The actions cited (and the response frequency) were as follows: counsel the instructor (26.0 per cent); unaware that this dysfunctional behavior happens (21.4 per cent); nothing, as it cannot be proven (8.3 per cent); use it in annual instructor evaluations (3.1 per cent); and screen course syllabi/grades in an effort to prevent it in the future (1.1 per cent). Numerous additional responses were also reported. These comments included responses such as "reward instructor as having improved their teaching effectiveness", "praise the instructor", and "fire the instructor if possible". The summed frequency of these additional responses was 5.4 per cent.

The administrators were asked to respond to 15 additional questions using a five-point, Likert-type response scale. These results are reported in Table III. With a mean of 4.60,

Table I Administrators' ranking of factors affecting SET scores (five-point scale)

	Rank	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	σ	Number of responses
1. Preparation and organization	1	0.0	0.6	11.3	37.3	50.8	4.38	0.71	327
2. Enthusiasm	2	0.3	0.9	9.2	47.6	42.1	4.30	0.70	328
3. Presentation	3	0.3	1.2	9.5	51.2	37.7	4.24	0.70	326
4. Teaching style	4	1.9	4.0	15.4	46.0	32.7	4.03	0.90	324
5. Instructor availability	5	1.5	9.5	30.8	44.5	13.7	3.59	0.89	328
6. Course difficulty	6	3.3	12.0	35.3	31.0	18.1	3.47	1.04	326
7. Instructor niceness	7	7.2	13.1	28.0	34.3	17.4	3.41	1.13	321
8. Student learning	8	5.6	12.8	30.5	36.8	14.3	3.41	1.06	321
9. Grading policy	9	3.7	13.8	36.8	32.2	13.5	3.38	1.00	326
10. Tough grading	10	5.2	16.6	33.8	26.5	17.8	3.35	1.11	325
11. Heavy course workload	11	11.4	17.6	34.6	24.7	11.7	3.07	1.16	324
12. Higher the level of course	12	10.3	20.1	29.8	32.0	7.8	3.06	1.12	319
13. Required or elective	13	15.1	17.3	33.6	25.3	8.6	2.95	1.17	324
14. Major/non-major	14	16.3	18.8	35.1	23.5	6.3	2.84	1.14	319
15. Class size	15	12.9	31.1	35.4	14.5	6.2	2.69	1.06	325
16. Class time of day	16	24.4	28.7	33.0	11.4	2.5	2.38	1.05	324
17. Giving students free time	17	43.7	24.6	19.7	10.4	1.6	2.01	1.09	309

Table II Administrators' ranking of factors for improving SET scores (five-point scale)

	Rank	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	σ	Number of responses
1. Avoid embarrassing students	1	5.4	8.5	27.5	39.2	19.3	3.58	1.06	316
2. Easy examinations	2	16.2	20.6	25.4	28.3	9.5	2.94	1.23	315
3. Provide sample examinations	3	12.8	26.2	30.0	24.6	6.4	2.85	1.12	313
4. High student grade expectations	4	19.0	19.7	27.7	24.2	9.4	2.85	1.25	310
5. Grade on a curve	5	14.3	23.9	30.9	25.2	5.7	2.84	1.13	314
6. Inflating grades	6	21.2	20.3	26.6	21.2	10.8	2.80	1.29	316
7. Cover less material (course work deflation)	7	18.5	23.2	28.0	25.5	4.8	2.74	1.16	314
8. Avoid cumulative final exam	8	29.4	24.2	22.3	24.8	6.8	2.48	1.26	310
9. Teach during banker's hours (9.00-3.00)	9	32.9	23.5	23.5	14.7	5.5	2.36	1.23	307
10. Avoid trying to teach students to think	10	42.8	20.6	17.0	13.1	6.5	2.19	1.29	306
11. Allow students to determine grade, coverage, and difficulty	11	42.6	22.0	15.1	13.7	6.5	2.19	1.30	291
12. Give same examination each semester	12	44.0	24.8	14.9	10.3	6.0	2.09	1.24	302
13. Giving parties (e.g. food, donuts, drinks)	13	45.1	26.3	16.1	8.2	4.3	2.00	1.15	304
14. Teach regulated classes (common exams)	14	45.2	26.7	25.4	6.2	3.1	1.95	1.08	292
15. More free time	15	47.9	26.2	13.8	9.2	3.0	1.93	1.12	305

the administrators were confident that a single numerical measure cannot capture all relevant aspects of an instructor's teaching ability. Yet they were slightly below neutral (mean of 2.89) with respect to replacing SET with an administrator and/or peer review system. However, the majority were clearly dissatisfied with their current teaching evaluation system (2.58 mean). But the majority also prefer to continue to use summative SET for evaluation and administrative decision-making purposes (mean of 2.55).

Although 39.0 per cent of the respondents indicated knowledge of instructors altering behavior to improve SET scores and 38.6 per cent indicated agreement or strong agreement that administrators understand such behavior occurs, 29.3 per cent disagreed or strongly disagreed that administrators understand such behavior occurs. The administrators did not seem to believe or acknowledge that dysfunctional techniques are used by instructors to improve SET scores (mean of 3.11). Yet with a mean of 3.29, the respondents felt that summative SETs have caused grade inflation. The administrators also generally believe that students reward an instructor for good mid-term grades (mean of 3.32), instructors do not reduce their grading standards to improve their SET scores (mean of 2.96), instructors do not reduce their course content or student workload to improve SET scores (mean of 2.87), and instructors try to keep their students' grade expectations high to improve SET scores (mean of 3.01).

Conclusions

Clearly, one objective for the use of SETs is to measure instructor effectiveness or the quality of instruction. Towards this end, student evaluations must provide meaningful and useful feedback for faculty. In order to make improvement in class content and presentation, for example, student comments and feedback using both summative and formative evaluation instruments are clearly warranted and usually sought by most, if not all, instructors. The research literature suggests SETs are capable of providing instructors useful feedback for improving class performance.

However, instructors can manipulate student responses on SET questionnaires, then these questionnaires should not be used for administrative decision-making purposes. Since most SET questionnaires focus on students' perceptions of an instructor rather than student learning, there may be an incentive for instructors to resort to dysfunctional behavior in order to manipulate SET scores. As reported, results of our survey indicate 39.0 per cent of the administrator respondents know of instructors that have altered their behavior to improve SET scores.

Table III Administrators' selective opinions

	Rank	Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Mean	σ	Number of responses
1. A single numerical measure cannot completely capture all relevant aspects of an instructor's teaching ability	1	0.3	0.3	4.2	28.6	66.6	4.60	0.61	332
2. An important goal of my institution is to improve student retention rates	2	2.4	8.8	15.1	48.0	25.7	3.85	0.98	331
3. Students are not qualified to judge areas of teaching skills (e.g. curriculum content, comprehensiveness, etc.)	3	6.0	12.3	15.7	41.6	30.1	3.77	1.18	332
4. Instructors should be allowed to evaluate department heads on a SET-type form at least annually	4	4.0	4.9	23.7	45.2	22.2	3.76	0.98	325
5. Students may reward an instructor on a SET for good mid-term	5	4.0	18.3	29.6	38.1	10.1	3.32	1.01	328
6. Use of summative SET by administrators has caused grade inflation	6	7.0	19.5	25.2	33.4	14.9	3.29	1.15	329
7. Administrators understand that dysfunctional techniques are used by instructors to improve their SET scores	7	7.7	21.6	32.1	28.4	10.2	3.11	1.10	324
8. Instructors try to keep their student grade expectations high to improve their SET scores	8	8.2	23.5	32.6	29.6	6.1	3.01	1.05	328
9. Instructors reduce their grading standards to improve their SET scores	9	10.0	26.3	25.4	28.7	7.3	2.96	1.11	331
10. SET should be replaced by an evaluation system based upon administrator and/or peer evaluation of instructors	10	10.2	28.9	31.4	20.0	9.5	2.89	1.13	325
11. Instructors reduce their course content or student workload in order to improve their SET scores	11	11.1	30.1	25.0	27.1	6.6	2.87	1.13	332
12. More than 50 per cent of my judgement of the effectiveness of an instructor is based upon his/her SET scores	12	15.7	35.3	18.7	26.0	4.2	2.67	1.14	331
13. I am satisfied with our current teaching evaluation system	13	19.3	35.3	16.6	24.5	4.3	2.58	1.17	331
14. Summative SET should not be used by administrators for control purposes (e.g. merit pay, tenure decisions and promotion)	14	22.3	35.4	26.8	14.9	10.4	2.55	1.27	328
15. The major goal of higher education is no longer learning	15	34.1	31.4	28.7	14.0	6.1	2.26	1.24	328

Although the majority of administrators in our survey responded that a single numerical measure cannot capture all relevant evaluative data, results indicate 18 per cent of accounting administrators utilize only SET information in order to evaluate instructor effectiveness. Additionally, 71.7 per cent of the administrative respondents "agreed" or "strongly agreed" that students are not qualified to judge many areas of teaching skills. The possibility that instructors may be able to manipulate responses on SET questionnaires is an important issue. This manipulation is especially alarming given the results of our survey which indicate that administrators believe there are many factors that influence student responses on SETs to varying degrees.

Use of the SET instrument is also intended to serve additional objectives, including assisting both administrators in personnel decisions and students in course selection decisions. To meet these objectives, these instruments should be:

- multidimensional and valid for a variety of indicators of teaching effectiveness;

Quality Assurance in Education
Volume 10 · Number 4 · 2002 · 213–222

- statistically reliable; and
- able to control for a variety of hypothesized sources of bias (e.g. results must not be dependent upon the course being taught).

There is growing controversy in the literature regarding the ability of SETs in meeting these objectives. As noted by Wright *et al.* (1984), numerical results of the student ratings often are used with little consideration of their validity and reliability for these purposes. These authors note that overall an instructor's expressiveness has a substantial impact on student ratings, but a small impact on student achievement. As evidenced by their association with learning criteria, these authors found the validity of student evaluations to be weak.

When SET rating forms are not carefully constructed, they may be statistically unreliable and invalid. Research literature indicates numerous confounding environmental variables or factors which are not controllable in student evaluations of teaching effectiveness. No clear guidelines have been proposed to account for statistically confounding variables including, for example, class size, class time, nature of course (required class, elective class, graduate-level, undergraduate-level, lower-level or upper-level), course in student's major field, student GPA, academic rank of professor, and so on.

Everly and Aleamoni (1972) observe that some SET forms are faculty-generated and their reliability may be so low as to "negate completely the evaluation effects and its results." One respondent of their study said that "most instruments have been developed to provide feedback to instructors to improve delivery, content, etc. and not for promotion/tenure/merit decisions. Most instruments have not been properly validated and surprisingly there have been few lawsuits on this matter." Another respondent asserted that unfortunately our central administration believes the student's responses to one question is the 'magic' measure of teaching effectiveness." Even carefully constructed commercial questionnaires may be suspect. One department head stated that his small school uses the IDEA evaluation system "which I consider costly, biased, and unfair. It is not even used extensively by the Kansas school that developed it."

Aleamoni and Hexner (1980) mention at least 28 studies which have reported significant positive relationships between grades, both expected and relative, and the ratings of the course and the instructor. Even staunch defenders of SET admit to a positive correlation from 0.10 to 0.52 between student ratings and expected grades (Cashin, 1988; Gillmore and Greenwald, 1994)[3]. Gillmore and Greenwald (1994) found that the student ratings general factor is influenced by three factors: "students perceptions of the ratio of valuable hours to total hours, the challenge of the course, and their grades in the course." These researchers worry about a "cycle of grade inflation – giving higher grades leads to higher ratings and the averages of both slowly creep upward."

There may be many potential sources of bias in student ratings. The following potential SET biases have been cited in the literature:

- Older faculty receive lower ratings (Feldman, 1983).
- A smaller class will tend to receive higher ratings (Cashin and Slawson, 1977; Feldman, 1984; Aleamoni and Hexner, 1980).
- Freshman students tend to rate faculty more harshly than sophomores and so on, so that graduate students rate faculty more generously (Arreola, 1994).
- An instructor's style of presentation is more important than the substance or the content (Naftulin *et al.*, 1973).
- Non-anonymous ratings tend to be higher (Braskemp *et al.*, 1984; Feldman, 1979; Marsh, 1984).
- An instructor's presence while students are completing the forms tends to result in higher scores (Feldman, 1979; Marsh, 1984).
- If students are told the results will be used for making rank, pay, or tenure decisions, ratings will be more positively biased and have less variability than those collected for feedback alone (Aleamoni and Hexner, 1980; Braskemp *et al.*, 1984; Feldman, 1979; Marsh, 1984).
- Higher ratings occur where students have a prior interest in the subject matter (Marsh, 1984) or are taking the course as an elective (Aleamoni, 1981; Feldman, 1978).
- A preference for female teachers over male instructors when statistically

Quality Assurance in Education
Volume 10 · Number 4 · 2002 · 213–222

significant differences are found in studies (Feldman, 1993).

- A single evaluation instrument which is not uniformly applicable to all discipline areas (Barnes and Barnes, 1993).
- Instructors of certain identified courses are more likely to receive unfavorable ratings (DeBerg and Wilson, 1990).

Although administrators do believe that SET has caused grade inflation and they are dissatisfied with their current SET system, the majority of our survey respondents would not replace the current evaluation system with an alternative evaluation system. Our survey results indicate future research efforts should be directed towards the design of an evaluation system which is less subject to potential dysfunctional behavior influences and relies upon additional measures of effectiveness.

In addition to evaluations by current students, the evaluation system used to monitor the effectiveness of college-level instructors ought to rely upon additional measurement tools such as peer evaluations, in-class evaluations by an independent and qualified third-party, documentation through a teaching portfolio, review of teaching materials and or content and rigor, and former student interviews.

Although some may view this as an intrusion and the additional data collection effort as a nuisance, with the importance placed upon quality instruction today, the dividends from enhanced student learning and more satisfied customers would be well worth the effort.

Notes

1 One Texas dean in 1993 said that "students are the best judge of teaching competence", and a Massachusetts dean said that "we rely on student ratings more than any other source of data on teaching" (Accounting Education Change Committee, 1993).

2 Laws highly regulate financial statements to reduce income manipulation and opportunistic behavior; yet there is no regulation of SET. Most administrators blindly accept them as truth. Instructors have a high incentive to manage SET, even more so than managers have the incentive to enhance earnings. See for example, Dechow *et al.* (1995).

3 Cashin (1988) suggests that in the social sciences validity correlations above 0.70 are unusual, especially when studying complex phenomena (e.g. learning). Thus, correlations between 0.20 and 0.49 are practically useful.

References

Accounting Education Change Commission (1993), "Evaluating and rewarding effective teaching: issues statement no. 5", *Issues in Accounting Education*, Vol. 8 No. 2, Fall, pp. 436-9.

Aleamoni, L.M. (1981), "Student ratings of instruction", in Millman, J. (Ed.), *Handbook of Teaching Evaluation*, Sage Publishing, Beverly Hills, CA.

Aleamoni, L.M. and 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*, Vol. 9, pp. 67-84.

Arreola, R.A. (1994), *Developing a Comprehensive Faculty Evaluation System*, CEDA, Memphis, TN, 288 pp.

Barnes, L. and Barnes, M. (1993), "Academic discipline and generalizability of student evaluations of instruction", *Research in Higher Education*, Vol. 34 No. 2, pp. 135-49.

Bauer, H.H. (1996), "The new generations: students who don't study", *The Technological Society at Risk Symposium*, Orlando, FL, 10 September, pp. 1-37.

Boex, L.F. (2000), "Attributes of effective economics instructors: an analysis of student evaluations", *Journal of Economic Education*, Vol. 31 No. 3, pp. 211-28.

Braskemp, L.A., Brandenberg, D.C. and Ory, J.C. (1984), *Evaluating Teaching Effectiveness: A Practical Guide*, Sage Publications, Beverly Hills, CA.

Bulkeley, W.M. (1997), "Would tax plan further inflate college grades?", *Wall Street Journal*, 22 April, pp. B1, B7.

Bures, A.L., DeRidder, J.J. and Tong, H.-M. (1990), "An empirical study of accounting faculty evaluation systems", *The Accounting Educators' Journal*, Summer, pp. 68-76.

Calderon, T.G., Green, B.P. and Reider, B.P. (1994), "Extent of use of multiple information sources in accessing accounting faculty teaching performance," working paper, May, pp. 1-22.

Cashin, W.E. (1988), "Student ratings of teaching: a summary of the research", *IDEA Paper No. 20*, September, pp. 1-6.

Cashin, W.E. and Slawson, H.M. (1977), *IDEA Technical Report No. 2: Description of Data Base*, Kansas State University, Center for Faculty Evaluation and Development, Manhattan, NY.

Centra, J.A. and Creech, F.R. (1976), "The relationship between student, teacher, and course characteristics and student ratings of teacher effectiveness", *SIR Report No. 4*, Educational Testing Service, Princeton, NJ, pp. 24-7.

Crumbley, D.L. (1995), "The dysfunctional atmosphere of higher education: games professors play", *Accounting Perspectives*, Spring, pp. 67-76.

DeBerg, C.L. and Wilson, J.R. (1990), "An empirical investigation of the potential confounding variables in student evaluation of teaching", *Journal of Accounting Education*, Vol. 8, pp. 37-62.

Dechow, P.M., Sloan, R.G. and Sweeney, A.P. (1995), "Detecting earnings management", *The Accounting Review*, April, pp. 193-225.

Drowling, W.C. (2000), "Why we should abolish teaching evaluations", *The Daily Targum*, 3 December.

Ellis, R. (1985), "Ratings of teachers by their students should be used wisely – or not at all," *The Chronicle of Higher Education*, 20 November, p. 88.

Emery, C., Kramer, T. and Tian, R. (2000), "Customers vs products: adopting an effective approach to business", *Society for a Return to Academic Freedom*, available at: bus.lsu.edu/accounting/faculty/lcrumbley/customersVSproducts.htm

Everly, J.C. and Aleamoni, I.M. (1972), "The rise and fall of the advisor", *Journal of National Association of Colleges and Teachers of Agriculture*, Vol. 16 No. 2, pp. 43-5.

Feldman, K.A. (1978), "Course characteristics and college students' ratings of their teachers: what we know and what we don't", *Research in Higher Education*, Vol. 9, pp. 199-242.

Feldman, K.A. (1979), "The significance of circumstances for college students' ratings of their teachers and courses", *Research in Higher Education*, Vol. 10, pp. 149-72.

Feldman, K.A. (1983), "Seniority and experience of college teachers as related to evaluations they receive from students", *Research in Higher Education*, Vol. 18, pp. 3-124.

Feldman, K.A. (1984), "Class size and college students' evaluations of teachers and courses: a closer look", *Research in Higher Education*, Vol. 21, pp. 45-116.

Feldman, K.A. (1993), "College students views of male and female college teachers: part II – evidence from students' evaluations of their classroom teachers", *Research in Higher Education*, Vol. 34 No. 2, pp. 151-91.

Gillmore, G.M. and Greenwald, A. (1994), "The effects of course demands and grading leniency on student ratings of instruction", working paper, March, pp. 1-17.

Greenwald, A.G. (1997), "Applying social psychology to reveal a major flaw in student evaluations of teaching", pp. 1-15, available at: weber.u.washington.edu/~agg/asptramf.htm

Haladyna, T. and Hess. R.K. (1994), "The detection and correction of bias in student ratings of instructors", *Research in Higher Education*, Vol. 3, pp. 210-40.

Haskell, R.E. (1997), "Academic freedom, tenure, and student evaluation of faculty: galloping polls in the 21st century", *Education Policy Analysis Archives*, Vol. 5 No. 6, pp. 1-32.

Hasselback, J.R. (1998), *Accounting Faculty Directory*, Prentice-Hall, Englewood Cliffs, NJ.

Marsh, H.W. (1984), "Students' evaluations of university teaching: dimensionality, reliability, validity, potential biases, and utility", *Journal of Educational Psychology*, Vol. 46, pp. 707-54.

Medley, D.M. (1979), "The effectiveness of teachers", in Peterson, P.O. and Walberg, H.J. (Eds), *Research on Teaching: Concepts, Findings, and Implications*, McCutchan Publishing, pp. 11-27.

Naftulin, D.H., Ware, J.E. and Donnelly, F.A. (1973), "The Doctor Fox lecture: a paradigm of education seduction", *Journal of Medical Education*, Vol. 48, pp. 630-5.

Newton, J.D. (1988), "Using student evaluation of teaching in administrative control: the validity problem", *Journal of Accounting Education*, Vol. 6, pp. 1-14.

Rosenfeld, P., Giacalone, R.A. and Riordan, C.A. (1994), "Impression management theory and diversity", *American Behavioral Scientist*, March, Vol. 37, pp. 601-20.

Ryan, J.J., Anderson, J.A. and Birchler, A.B. (1980), "Student evaluation: the faculty responds", *Research in Higher Education*, Vol. 12 No. 4, pp. 317-33.

Sacks, P. (1996), *Generation X Goes To College*, Open Court, Chicago, IL.

Schlender, B.R. (1980), *Impression Management: The Self-Concept, Social Identity, and Interpersonal Relations*, Brooke/Cole, Monterey, CA.

Schneider, D.J. (1969), "Tactical self-presentation after success and failure", *Journal of Personality and Social Psychology*, Vol. 13, pp. 262-8.

Seldin, P. (1993), "The use and abuses of student evaluation of professors", *The Chronicle of Higher Education*, 12 June, p. A-40.

Swann, W.B. (1987), "Identity negotiation: where two roads meet", *Journal of Personality and Social Psychology*, Vol. 53, pp. 1038-51.

Worthington, A.G. and Wong, P.T.P. (1979), "Effects of earned and assigned grades on student evaluations of an instructor", *Journal of Educational Psychology*, Vol. 71 No. 6, pp. 764-75.

Wright, P., Whittenburg, R. and Whittenburg, G.E. (1984), "Student ratings of teaching effectiveness: what the research reveals", *Journal of Accounting Education*, Vol. 2 No. 2, Fall, pp. 5-30.

Yunker, B.J. and Sterner, J. (1988), "A survey of faculty performance evaluation in accounting", *Accounting Educators' Journal*, Fall, pp. 63-74.

Further reading

Brown, D.L. (1976), "Faculty ratings and student grades: a university-wide multiple regression analysis", *Journal of Educational Psychology*, Vol. 68, pp. 573-8.

Cornwell, R.C. (1984), "The evaluation of faculty performance", *Collegiate News and Views*, Fall-Winter, pp. 9-13.

Ditts, D.A. (1980), "A statistical interpretation of student evaluation feedback", *Journal of Economic Education*, Spring, pp. 10-15.

Dowell, D.A. and Neal, J.A. (1983), "The validity and accuracy of student ratings of instructions: a reply to Peter A. Cohen", *Journal of Higher Education*, July/August, pp. 459-63, p. 3.

DuCette, J. and Kenney, J. (1982), "Do grading standards affect student evaluations of teaching: some new evidence on an old question", *Journal of Educational Psychology*, Vol. 74 No. 3, pp. 308-14.

Howard, G.S. and Maxwell, S.R. (1982), "Do grades contaminate student evaluations of instruction?", *Research in Higher Education*, Vol. 16 No. 2, pp. 175-88.

Kennedy, W.R. (1975), "Grades expected and grades received – their relationship to students' evaluations of faculty performance", *Journal of Educational Psychology*, Vol. 67 No. 1, pp. 109-15.

Kirsch, R.J., Leathers, P.E. and Snead, K.C. (1993), "Student versus recruiter perceptions of the importance of staff auditor performance variables", *Accounting Horizons*, Vol. 7 No. 4, December, pp. 58-69.

Lawler, E.E. and Rhode, J.G. (1976), *Information and Control in Organizations*, Goodyear Publishing, Pacific Palisades, p. 102.

Marsh, H.W. (1982), "Factors affecting students' evaluations of the same course taught by the same instructor on different occasions", *American Educational Research Journal*, Vol. 19 No. 4, pp. 485-97.

Porcano, T.M. (1984), "An empirical analysis of some factors affecting student performance", *Journal of Accounting Education*, Vol. 2 No. 2, Fall, pp. 111-26.

Powell, R.W. (1977), "Grades, learning, and student evaluation of instruction", *Research in Higher Education*, Vol. 7, pp. 193-205.

Renner, R.R. (1981), "Comparing professors: how student ratings contribute to the decline in quality of higher education", *Phi Delta Kappan*, October, pp. 128-31.

Rideway, V.F. (1956), "Dysfunctional consequences of performance measurements", *Administrative Science Quarterly*, pp. 240-7.

Seldin, P. (1978), "The use and abuse of student ratings of professors", *Journal of Educational Psychology*, Vol. 68 No. 1, pp. 75-82.

Seldin, P. (1984), *Changing Practices in Faculty Evaluation*, Jossey-Bass, San Francisco, CA.

Stumpf, S.A. and Freeman, R.D. (1979), "Expected grade covariation with student ratings of instruction: individual versus class effects", *Journal of Educational Psychology*, Vol. 71 No. 3, pp. 293-302.

Winsor, J.L. (1977), "As Bs but not Cs?: a comment", *Contemporary Education*, Vol. 48 No. 2, Winter, pp. 82-4.