

Student Evaluations Of Faculty: Concerns And Possible Solutions

Robert M. Kozub, University of Wisconsin, Milwaukee, USA

ABSTRACT

Student evaluations of university instruction have long been used to evaluate the teaching performance of instructors of all ranks. In spite of the widespread use of the data acquired from student evaluations for the purpose of determining faculty teaching effectiveness, a review of the literature in the area indicates that questions concerning the validity and usefulness of such evaluations remain unanswered. With the pervasiveness use of student evaluations, it is doubtful that substantial changes in the system are going to be implemented across the country soon. This research reviews the major problems that have been identified in student evaluations of faculty, and suggests some possible methods to increase the validity of teaching evaluations without major changes to the current systems of evaluation of faculty. Any changes to teaching evaluations, which would increase the usefulness of student evaluations of faculty, would benefit both college students and faculty. Administrators of colleges and universities for some time have stressed the importance of a marketing orientation (Bush, Ferrell & Thomas, Jr., 1998; Moore 2004). Students are one of the consumer groups interested in the product of a college education, and students' opinions are considered a vital source of information concerning the quality of instruction at colleges and universities.

USE OF STUDENTS' TEACHER EVALUATIONS

Virtually all colleges and universities use student evaluations of instructors as a measure of instructor performance (Magner, p. A18, 1997), therefore, such student evaluations have a significant impact on tenure, promotion and merit pay decisions concerning faculty (Centra, 1979; Ehie & Karathanos, 1994; Kozub, 2006). It is hoped that the response from students may help instructors to improve their teaching performance (Marsh, 1991; Smith 2008). Unfortunately, the use of such ratings for evaluations relating to reward systems of a college or university may be problematic to the extent that student consumers are responding to factors that should be unrelated to teaching quality, such evaluations may be misleading (Cashin, Downey & Sixbury, 1994; Marsh, 1994, 1995; Kozub, 2006), and may have negative consequences on the overall quality of the educational experience for students.

RESEARCH ON THE VALIDITY AND CONSISTENCY OF STUDENT EVALUATIONS

Extensive research has been conducted on student evaluations to determine their validity. The stability and internal consistency (Cashin, Downey & Sixbury, 1994; Costin, Greenough & Menges, 1971; Marsh, 1994, 1995; Gallagher 2000; and Burdsal 2008) as well as variability between instructors (Marsh & Bailey, 1993) have all been demonstrated. Yet, numerous studies have cast doubt on the validity of such instruments. Rodin & Rodin (1972) found a negative relationship between student performance and student ratings. O'Connell & Dickinson (1993) found the amount learned by students was unrelated to the overall ratings of the instructor. Yunker and Yunker (2003) found that students from a class where the instructor was rated higher did worse than students from class where an instructor was rated lower, in a subsequent follow-up class. In another study, Koon and Murray (1995) found that final examination scores had only a .30 correlation with student ratings of instructors. It has been noted that workload and students' grades might also affect student ratings of instructors (Marsh 1987; Centra 2003). Oliver-Hoyo (2008) has found significant variations in teaching ratings for the same instructor teaching multiple sections of

the same course. Conventional teacher supervision and evaluation methods as being inefficient and ineffective have been observed in another study (Marshall 2005).

Students may not have the level of knowledge necessary to properly evaluate their instruction according to Olshavsky & Spreng (1995) which may lead to the use of some other proxy in determination of instructor performance. The "entertainment" level of the classroom experience has been shown to affect over-all instructor ratings (Costin, Greenough, & Menges, 1971). In the famous "Dr. Fox" study (Naftulin, Ware, & Donnelly, 1973) found that an enthusiastic actor was highly rated on teaching quality despite a lecture intentionally devoid of content. Narrative comments noted on student evaluations of faculty in one study demonstrated that many students wanted classes to be more fun and entertaining (Trout, 1997). Williams and Ceci (1997) found that emphasizing communication skills produced significant improvements in ratings of an instructor by students in all areas of teaching for a class, despite the fact that the material and lecture format remained identical, as well as student performance. Wright (2000) found that perceived fairness in grading and instructor appearance were strongly related to student evaluations of professors, despite the fact that these factors may be unrelated to the degree of student learning.

These results suggest that student consumers may rate faculty using a different set of criteria than other constituencies. Student consumers, may, in fact prefer a teaching style that is detrimental to their learning experience. For example, Strom, Hocesvar, and Zimmer (1990) showed that students who preferred easy courses had greater achievement with instructors who had a relatively low orientation toward students' interests, achievement, and satisfaction, but preferred teachers who had a high level of student orientation. In addition, external ratings may differ from student-derived ratings. This was dramatically demonstrated when a Business Week ranking of graduate business schools (Byrne & Leonhardt, 1996, p.112, 113) showed substantial differences for the ratings of teaching performance in terms of an output measure (the perceived level of skills of MBA graduates as rated by corporate recruiters) and a process measure (as rated by the instructional quality of faculty in the same program by MBAs) (Byrne & Leonhardt, 1996, p. 122). Student evaluations of faculty have become increasingly important, thus the process of instructor evaluation needs to be reexamined. Given that the process of using students to evaluate instructors is ingrained in academia, it is unlikely that efforts to modify or replace it will be successful but it may be possible to significantly improve the quality of information gained from student evaluations of faculty with relatively small changes. These changes involve the anonymity of the evaluations, and the massive quantity of data generated.

PROBLEMS WITH THE CURRENT TEACHING EVALUATION SYSTEM

There are four major problems with the current teaching evaluation system. These problems are: (1) the voluntary nature of current teaching evaluation system; (2) anonymous evaluations; (3) survey data; and (4) massive data. Each of these issues is discussed in this section of the research.

Voluntary Nature Of The Current System

At many colleges and universities students do not have to complete the teaching evaluation instrument, but the instructors are held responsible if a significantly large percentage of the students do not complete the instrument. Currently, there is no research on the question: does lack of participation by a student indicate poor teaching, lack of interest in the course, or lack of student confidence in the evaluation system? It is impossible to study this issue due to the anonymous nature of the evaluation system. Thus, administrators are free to interpret the lack of participation by the entire enrolled class any way they chose. This is dangerous as there is no insurance that any administrator will be consistent in their judgment on this issue (Kozub, 2006).

Anonymous Evaluations

Typically, student evaluations of faculty members are anonymous. The idea behind anonymous evaluations is apparently to protect the students from potential reprisal from faculty. This idea certainly has a positive intent, however, anonymity may facilitate reprisal from students to faculty who gave the students lower grades than the students felt that they deserved. Faculty and administrators have no way of determining which students gave a high

or low rating to an instructor. Thus administrators apparently believe that faculty members are less trustworthy than students. Under a system of anonymous evaluations, students need take no responsibility for their opinions. With no possibility for follow-up, students need not consider all the facts in order to come to a valid and justifiable conclusion supported by facts. A student's evaluation could be based solely on latent anger resulting from a recent grade received on an exam, or from a single negative in-class experience with an instructor over the course of an entire semester. In addition, the practice of anonymity provides no way to insure that the evaluations were properly filled out. In many instances the faculty member is required to leave a classroom while the evaluations are being filled out. This leaves the potential for significant abuse. For example, students could enter the room and fill out evaluations that were not even in the class. Another potential problem is lack of ability to follow up on the results. No details can be gathered as to why an evaluation was very positive or very negative. There are several questions that need to be researched. First, do all students with low grades give low evaluations? While a question may be asked concerning the students expected grades, there is no way to insure that students' reports are accurate. The second question is, do students with poor attendance records give low evaluations to an instructor they may have only rarely had contact with in class? There is absolutely no way to determine the relationship between class attendance and student evaluations of instructors. Closely related to the previous point is the lack of ability to track "customers". Businesses use customer-tracking systems to determine if certain individuals are blatantly abusing the complaint system. Such businesses can then determine whether the problem is with the business, or with the customer. Under a system of anonymous evaluations there is no way to determine whether a student is a chronic complainer.

The Problem Of Survey Data

Student evaluations are a result of the students completing a survey instrument. Survey results provides an overview of student feelings concerning a faculty member, they do not provide an in-depth picture of what happened in the classroom. They also do not allow for probing, to determine the factors leading to the evaluation. If it were possible to track evaluations to individual students, in-depth interviews could determine reasons for dissatisfaction. This would allow administrators to determine whether the problems were due to weaknesses in the instructor, or in the student, or in both. It would be possible to determine, for example, if there was a correlation between academic preparation of students and evaluation of instructors. This would permit administrators to answer the question: do instructors get lower evaluations from students who are less prepared?

The Problem Of Massive Data

Regardless of the size of a college or university student evaluations generate an immense quantity of data. If there is evidence of possible problems with instructor performance based on student evaluations, it is imperative that we look closely to determine the extent and nature of such problems. However, such a massive amount of data precludes that possibility. One factor leading to the problem of massive data is the problem of both too much and too little oversight of faculty. It is frequently the case that faculty of all ranks are evaluated in the same fashion, whether they have been teaching for one year, 15 or 30 years. With such a large number of people to evaluate it is very difficult to perform an in-depth evaluation of everyone. Therefore, decisions on such important processes as faculty tenure or merit exercises may be based solely or largely on anonymous student evaluations of faculty.

It would seem evident that much closer scrutiny should be given to new faculty. Given the long-term nature of any tenure decision, and the importance of teaching as a measure of a faculty member's performance, it is important to examine closely all aspects of the teaching performance of untenured faculty members. Without such oversight untenured faculty members may be tempted to "dumb down" their courses, give high grades, and take other extraordinary measures in an attempt to increase student evaluations, and thus maximize a faculty member's chances of being awarded tenure. Conversely, tenured faculty members are, by definition, proven performers in the classroom. They may resent close scrutiny of their teaching performance, and such close scrutiny is unlikely to result in a positive outcome. However, it is not only untenured faculty who attempt to manipulate teaching evaluation scores by a variety of means including giving cash, saving bonds, and other "prizes" to top scoring students on exams or for a term, as well as passing out candy during class (Kozub, 2006).

Another problem with survey data is how and by whom the statistics are generated. At most institutions the statistics are generated by the department, college, or school once the survey response instrument is read by the central campus computer. There have been reported errors in processing the survey response instruments, which produces doubt about the validity of the statistical results (Kozub, 2006). Even more shocking is the use of software written by a faculty member in the unit. It is possible that unintended procedures could be written into the program in order to improperly increase or decrease the teaching scores of individual faculty members. This type of situation cast doubt about the validity of the statistical results (Kozub, 2006).

SUGGESTIONS

Evaluating the teaching performance of faculty members is a critical part of ensuring a high quality education for students. However, given the research showing potential problems with the current method of evaluation, significant changes need to be made. The first suggestion is to change the nature of the student evaluations from being anonymous in nature to confidential in nature. The instructor would not be given access to this database, however, such a database would allow implementation of a customer tracking system by someone in the university. It would also allow for follow-up to investigate particularly high or low evaluations. In addition, this allows investigation into how actual grades are related to evaluations. The second suggestion is to overcome the problems of massive amounts of data, and too little oversight of untenured faculty. Besides the use of traditional student evaluations of faculty, all classes taught by untenured faculty should be subject to sampling and in-depth interviews of selected students. Students could be randomly sampled from among those that had high, medium, and low ratings of the untenured faculty member. These selected students could then be interviewed in depth concerning what really went on in the class, and reasons for the ratings of the professor. In the case of small classes, everyone in the class could be interviewed. This process would serve to determine to what extent rigor and high expectations, or lack of rigor and low expectations affected evaluations in the classroom. Such a system would also serve to protect faculty members who were very demanding in the classroom, but skilled teachers. Again for untenured faculty, it should be an automatic procedure for selected members of the tenured faculty to sit in on a class each semester to evaluate the teaching methods used and to review the syllabus and evaluation instruments used by the instructor to determine rigor, and expectations of that instructor. Since the number of untenured faculty would be expected to be small in relation to the number of total faculty, this plan would be workable.

For tenured faculty, a similar system would be available if requested by the faculty member. For tenured faculty with consistently high ratings, a normal grade distribution, and no student complaints, the standard student evaluations system could be used. For those who seemed to be having problems, closer scrutiny could be paid using the above-described system. However, a strong argument for post-tenure review was made by Wood (2006). It appears that currently universities are not serious about post-tenure reviews.

CONCLUSION

Student evaluations of faculty are a widely used method of evaluating faculty performance in the classroom. However, significant improvements can be made in the system with only minor changes, which may significantly improve the educational experience for students, as well as faculty members.

REFERENCES

1. Bryne, J. A. & Leonhardt, D. (with Bongiorno, L. & Jespersen, F.) (1996, October 21). *The Best B Schools*. *Business Week* 110-116.
2. Burdsal, Charles A. "Further Evidence Supporting The Validity Of Both A Multidimensional Profile And An Overall Evaluation Of Teaching Effectiveness" *Assessment & Evaluation in Higher Education* October 2008, Vol. 33 Issue 5, p567-576
3. Bush, V., Ferrell, O.C., & Thomas, Jr., J. L. (1998). Marketing The Business School: An Exploratory Investigation. *Journal of Marketing Education* 20, 16-23.
4. Cashin, W. E., & Downey, R.G. (1992). Using Global Student Rating Items For Summative Evaluation. *Journal of Educational Psychology*, 84, (4), 563-572.

5. _____, Downey, R.G. & Sixbury, G.R. (1994). Global And Specific Ratings Of Teaching Effectiveness And Their Relation To Course Objectives: Reply To Marsh (1994). *Journal of Educational Psychology*, 86, (4), 649 – 657.
6. Centra, J. A. “Will Teachers Receive Higher Student Evaluations by Giving Higher Grades and Less Course Work?” *Research in Higher Education* 2003
7. _____, (1979). *Determining Faculty Effectiveness: Assessing Teaching, Research, And Service For Personnel Decisions And Improvement*. San Francisco: Jossey-Bass Inc.
8. Costin, E, Greenough, W.T., & Menges, R.J., (1971). Student Ratings Of College Teaching: Reliability, Validity, And Usefulness. *Review of Educational Research*, 41, 511-535.
9. Ehie, I. C. & Karathanos, D. (1994). Business Faculty Performance Evaluation Based On The New Aacsb Accreditation Standards. *Journal of Education for Business*. 69, (No. 5), 257-262
10. Gallagher, Timothy J. “Embracing Student Evaluations of Teaching: A Case Study” *Teaching Sociology* 2000 pp. 140-147
11. Hativa, N. & Raviv, A. (1993). Using A Single Score For Summative Teacher Evaluation By Students. *Research in Higher Education*. 34 (5), 625-646.
12. Koon, J. & Murray, H.G., (1983). Low-Inference Classroom Teaching Behaviors And Student Ratings Of College Teaching Effectiveness. *Journal of Educational Psychology*, 75, 138-149.
13. Kozub, R.M., (2008) “Relationship Of Course, Instructor, And Student Characteristics To Dimensions Of Student Ratings Of Teaching Effectiveness. ” Working Paper
14. _____, (2006). Relative Importance Of Research, Service, And Teaching In Tenure, Promotion, And Merit Decisions. Working Paper.
15. Magner, D. K., (1997) Report Says Standards Used To Evaluated Research Should Also Be Used For Teaching And Service. *The Chronicle of Higher Education*, 44, (2), p. A18-A19.
16. Marsh, H. (1987). Students' Evaluations Of University Teaching: Research Findings, Methodological Issues, And Directions For Future Research. *International Journal of Educational Research* 11, #3, 253-388.
17. _____ (1991). Multidimensional Students' Evaluations Of Teaching Effectiveness: A Test Of Alternative Higher-Order Structures. *Journal of Educational Psychology*, 83 (2), 285-296.
18. _____ & Bailey, M., (1993). Multidimensional Students' Evaluations Of Teaching Effectiveness. *Journal of Higher Education*, 64, (1), 1-18.
19. _____ (1994). Weighting The Right Criteria In The Instructional Development And Effectiveness Assessment (Idea) System: Global And Specific Ratings Of Teaching Effectiveness And Their Relation To Course Objectives. *Journal of Educational Psychology*, 86, (4), 631-648.
20. _____ (1995). Still Weighting For The Right Criteria To Validate Student Evaluations Of Teaching In The IDEA System. *Journal of Educational Psychology*, 87, (4), 666-679.
21. Marshall, Kim “It's Time to Rethink Teacher Supervision and Evaluation” *Phi Delta Kappan* v. 86 no. 10 (June 2005) p. 727-35
22. Moore, Robert M. “The Rising Tide: "Branding" and the Academic Marketplace: *Change* v. 36 no. 3 (May/June 2004) p. 56-61
23. Naftulin, D.H., M.D., Ware, J.E., Jr., & Donnelly, EA., (1983). The Doctor Fox Lecture: A Paradigm Of Educational Seduction. *Journal of Medical Education*, 75, 138-149.
24. O'Connell, D. Q. & Dickinson, D. J., (1993). Student Ratings Of Instruction As A Function Of Testing Conditions And Perceptions Of Amount Learned. *Journal of Research and Development in Education*, 27, (1), 18-23.
25. Oliver-Hoyo, Maria “Student Evaluations That Generate More Questions Than Answers” *Journal of College Science Teaching* v. 38 no. 1 (September/October 2008) p. 37-9
26. Olshavsky, R. W. & Spreng, R.A. (1995). Consumer Satisfaction And Students: Some Pitfalls Of Being Customer Driven. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior* 8, 69-77
27. Parasuraman, A., Zeithaml, V., & Berry, L. (1985). A Conceptual Model Of Service Quality And Its Implications For Future Research. *Journal of Marketing*, 49 (Fall), 41 50.
28. Rodin, M. & Rodin, B., (1972). Student Evaluations Of Teachers. *Science*, 177, 1164-1166.

29. Smith, Calvin “Building Effectiveness In Teaching Through Targeted Evaluation And Response: Connecting Evaluation To Teaching Improvement In Higher Education” *Assessment & Evaluation in Higher Education* October 2008, Vol. 33 Issue 5, p517-533
30. Strom, B., Hocevar, D. & Zimmer, J., (1990). Satisfaction And Achievement: Antagonists In ATI Research On Student-Oriented Instruction. *Educational Research Quarterly*, 14, (4), 15-21.
31. Trout, P. A. (1997). What The Numbers Mean: Providing A Context For Numerical Student Evaluations Of Courses. *Change*, 29, 25-30.
32. Williams, W.M. & Ceci, S.J. (1997). "How'm I Doing?" Problems With Student Ratings Of Instructors And Courses. *Change*, 29, 13-23.
33. Wood, M. “When Post-Tenure Review Policy And Practice Diverge Making The Case For Congruence” *The Journal of Higher Education* Volume 77, Number 4, July/August 2006 pp. 562-588
34. Wright, R.E. (2000). Student Evaluations And Consumer Orientation Of Universities. *Journal of Nonprofit and Public Sector Marketing*, Vol 8, p. 33-40.
35. Yunker, P. & Yunker, J. (2003). Are Student Evaluations Of Teaching Valid? Evidence From An Analytical Business Core Course. *Journal of Education for Business*, Vol. 78, p. 313-317.