Using Teaching Evaluations as a Measurement of Consumer Satisfaction

Should students be consumers and therefore in control of measures of "consumer satisfaction" when evaluating a teacher's performance? This essay promotes the view that students should not be the sole evaluators of teacher performance. The use of student teaching evaluations is discussed, examining the use of statistics derived from the evaluation forms. A qualitative model of teaching evaluation is recommended, using portfolios for conducting summative evaluations of teachers.

Deborah C. Haynes, Montana State University[1]
Holly Hunts, Montana State University[2]

Introduction

At the annual meeting of the American Council on Consumer Interests (ACCI) professionals from a myriad of backgrounds including economics, sociology, demography, business, and family and consumer economics meet to discuss issues facing consumers in America and other countries. Many of the professionals belonging to ACCI come from institutions of learning where they face a relatively well-educated, well informed and well-to-do clientele of consumers--college students. In addition, other members of ACCI employed in extension, government and industry are involved in providing educational services to some type of student-consumer group. These educators face the same pertinent question when it comes time to evaluate their productivity: Is it possible to think of students (broadly defined) as consumers of education and, if so, are teaching evaluations an effective means of measuring consumer satisfaction? The purpose of this essay is to develop some ideas about the appropriate use of student teaching evaluations as a measure of consumer satisfaction.

Background

The evaluation of teaching is a difficult process. Many teachers claim that the evaluation of teaching is an impossible task, since the process of teaching is so complex and multidimensional. Whether the task can be done in a totally representative way or not, the evaluation of teaching happens every day in thousands of pre-school to university-level classrooms. Life changing decisions such as retention, promotion, tenure and merit decisions are made based on the evaluation process; therefore the evaluation process is important to the teacher, the student, the educational institution and society itself.

An evaluation of any process must first begin by setting out the parameters, definitions and assumptions of the object of the evaluation. In the case of using teaching evaluations as a measure of student (consumer) satisfaction, clarity is required as to what outcome is to be evaluated and the meaning of summative (versus formative) evaluations.

One outcome of teaching for which teachers, parents, school administrators and policy makers agree is that learning should take place. A student should know more about a subject matter after having taken the class than before taking the class. The teacher should add value to the learning process over what the student would have learned in an independent learning venture. Some would argue that the teacher only adds value if the student is willing to be taught. The vast majority of students at the university level or adult learners in more informal groups choose to attend class either for the college experience and/or for the credential to further a career advancement later in life. In some relatively small instances the learner is not motivated at all to learn or may have learning problems, and those cases should be discussed in the evaluation process. However, in most cases, the assumption will be made that the students are there to learn and therefore the teacher does have the opportunity to facilitate the learning process.

Secondly, as a refinement to the outcome that "learning takes place," an additional concern is that the learning is not a memorized and regurgitated set of soon-to-be-obsolete facts, but that the learning allows student to

critically think about the subject matter; that is, the that the student uses higher levels of thinking, performing and affective skills according to Bloom's Taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956).

A third element of the outcome measure would include imparting enthusiasm for, at least an appreciation of, the subject matter. A good teacher can communicated enthusiasm for their field of knowledge that can motivate the learner to care about the subject matter. Without this appreciation of the subject matter, the student will leave the classroom without the desire to further their own contact with an understanding of the subject in the future.

To summarize, the minimum outcomes desired from the teaching process are to have learning take place, to develop student ability to critically think about a subject and to have students develop an appreciation for subject. A teaching evaluation process should focus on these outcomes as the overall yardstick for evaluation.

The next step is to define summative evaluations. Summative evaluation is one done for the purposes of personnel decisions. At the university level, personnel decisions are those regarding retention, promotion, tenure, and merit. A commonly held notion among faculty, students, and the general public is that, at a land grant university, teaching does not have much weight in the total personnel decision process by administrators so does not warrant much effort. However, administrators and public pressure indicate that teaching does indeed matter. The problem is in evaluating teaching in a manner that can be fairly compared between faculty members and that can be defended to various publics, not the least of which could be the court system in the case of grievances about the outcome of personnel decisions. So the summative evaluation is one done for the personnel decisions and that can be defended as a just basis for those decisions.

The result of the summative evaluation has to be a decision that is dichotomous-retain or not; promote or not; tenure or not; merit or not. In addition, summative evaluations may also be required to have an assignment of the level of the success. For example, at one land grant university, the decision to tenure rests upon a combination of ratings in teaching, research, and service. The scale for the end result of the summative evaluation in each of the responsibility areas must be "not effective," "promise of excellence," and "excellence." At least one of the ratings must be "promise of excellence" and the other two ratings must be at least "meets expectations." Therefore, the result of the summative evaluation must not only distinguish dichotomously but also decide the level of success.

With these outcomes and definitions in mind, the discussion now turns to the method that is very commonly used to evaluate the teacher's performance, which is the teaching evaluation completed by the students or adult learners in a given educational experience.

Using Student Teaching Evaluations to Measure "Consumer Satisfaction" in Evaluating an Instructor

For years, dozens of educational institutions, particularly universities, have used student evaluations as a primary measure of teaching effectiveness. A study recently conducted at Oklahoma State University (Weber, 1999) found that of 85 departments in the areas of Human Sciences and Agricultural Sciences, 88.5% of respondents indicated that teaching evaluations by students are submitted as evidence of teaching effectiveness. While it is not known what combination of evidence sources are used in these departments, the research found that only 53% of the departments used teaching portfolios. Some are undoubtedly using only teaching evaluations as evidence of teaching effectiveness. The literature shows, however, that using only student teaching evaluations can lead to grade inflation and progressively easier courses (Washburn & Thornton, 1996). In and of themselves, students do not necessarily rate a course as better if they expect a higher grade or if the course is easier. On the contrary, research has shown that students choose courses in which they will learn more and rate courses according to the result of their learning gains (Shapiro, 1990). However, since student ratings are often the only feedback mechanism teacher have, one response teachers make to poor teaching evaluations is to make the course easier and raise the grades rather than make substantive changes in teaching methods. Research has shown that student ratings used for summative purposes may not be the best mechanisms for formative evaluation, that is, evaluation whose goal is to help the teacher improve the course (Schmelkin, Spencer, & Gellman, 1997).

Further, students can evaluate some areas of teaching well but cannot evaluate the total teaching process. For example, the student is not in the best position to determine if the teacher taught the material that is most useful or important in the domain of the subject matter. Also, the student is not in a position to determine if the teacher is knowledgeable about the subject nor can students comment about whether or not a class will benefit them as a professional. Hundreds of research articles have been written about student ratings. The primary outcomes of these articles is that students definitely should play a part in the teaching evaluation process. However, students should evaluate what they are qualified to evaluate.

The evaluation form to be chosen is another issue. Many of the existing forms do not represent the three areas of teaching outcomes mentioned above. Or they do provide the components but the weighting of the components is not equal. Another issue of choosing a form is that the teaching situation must match the questions used. For example, it is difficult to evaluate a freshman lecture with 150 students with the same teaching evaluation form that is used in a graduate seminar with 10 students. It is unfair to evaluate the teacher on items such as, "willingness to learn my name," in these very different class formats. The evaluation form chosen should have questions appropriate to the characteristics of the course taught.

In summary, the "student-as-consumer" model is appropriate for providing input to the evaluation of a teacher's performance, but the student teaching evaluation should not be the only measure used because it may not be the most valid measure of particular teacher characteristics. Further, the teacher may respond to the student teaching evaluation by catering to the student opinion, even when the educational outcome is compromised in the process. Once the student teaching evaluations have been administered to groups of student, an extremely important area of discussion that has not been thoroughly addressed is how administrators and their institutions should use the numeric ratings produced by the coding of the student teaching evaluation forms. A number of empirical issues arise when using the forms as quantitative tools for statistical analysis.

Analyzing Student Teaching Evaluations Statistically

Many institutions have special data analysis departments in the organization to produce numerical summaries of the student teaching evaluations for faculty members. These numeric reports are often used in summative evaluations of the faculty member without thinking about the statistical assumptions being used in the interpretation of the numbers. A frequently used statistic on the individual questions with the student teaching evaluation instrument is the mean. The reporting of the mean of the components or the mean of the global rating is invalid because it does not represent the variation around the mean. A good teacher wants to be a good teacher to all students in the course, not just to the good students or the poor students. The most desirable situation would be to have high ratings with very little variation around the mean. This would indicate that the teacher was able to reach all students well. Various methods can be used to report the variation, such as the range or the standard deviation.

Another concern with the use of the mean is the degree to which the mean actually represents a true quantitative measure of teaching performance. When the course has small numbers and/or a wide variation, the most statistically correct measure of the mean would be to include the confidence interval in which the true means lies. If one chooses a 95% confidence level, the confidence interval for the true mean can be calculated by the mean plus or minus the standard deviation divided by the square root of the number of students completing the evaluation. For example, if the mean of a seven point scale was 5 and the standard deviation was 1.2, the confidence interval in a response set of 40 students would be 5 plus or minus 1.2 divided by the square root of 40. The confidence interval in which the true mean lies 95% of the time would be 4.81 to 5.19. The confidence interval makes the overall reporting of the mean of the student ratings more difficult. For example, if the confidence interval spans the categories of "effective" to "excellent" what does the teacher report? No literature is available to guide this decision. One suggested method would be the following: The logic of choice should be that the teacher chooses the category to report that is spanned most completely by the confidence interval. In the case of the confidence interval spanning two categories completely, the teacher can choose the higher category, that is the one that provides a better evaluation of the teacher. However, if the confidence interval completely spans more than one category, this would suggest that the teacher is not meeting all student needs evenly and the other evidence in the portfolio could be used to evaluate the teacher at a lower level or a higher level of effectiveness.

It is clear that a simple reading of the descriptive statistics of teaching evaluation analyses is inadequate because of the above mentioned issues involving the mean and spread around the mean. However, more fundamental issues of quantitative analysis arise when considering the idea that many institutions base their entire summative evaluation of a given faculty member on the student's responses to teaching evaluation forms. This paper has argued that students should be one input to the final summative evaluation of a faculty member's productivity, but that they should not be the only source of information for the evaluation. The next section will examine the model more appropriately used for summative evaluation from the quantitative paradigm.

Quantitatively Modeling the Summative Faculty Evaluation

One way to think about student evaluation forms is to envision them as a measure of student satisfaction. If a statistical model were to be specified correctly, the functional form might look something like:

Student satisfaction with Course X = f (t, s, p, ph, ea, tq, error term)

where:

t = teacher performance (fair examinations, organized and clear instruction, variety of teaching methods, homeworks that are relevant, accessibility to students, enthusiasm)

s = supplemental materials (text, outside readings, recitation, labs)

p = preparedness to study topic being taught (adequate math background, vocabulary, theoretical understanding, student maturity, etc.)

ph = physical characteristics of the classroom (size of classroom, size of the class, location on campus, temperature, lighting, able to see the visuals, etc.)

ea = emotional aspects of the class (individual interest in topic being taught, "personality" of the class, time of day)

tq = true quality (material presented is up-to-date, accurate, unbiased)

This functional form would be a superior form of determining the role the instructor plays in student satisfaction than a simple score because it would hold constant independent variables which are intuitively important and look specifically at the effect of teacher performance. This model, in practice however, is fraught with problems. Kennedy (1985) enumerates five basic assumptions in all classical linear regression models and all of these assumption are violated in this model, as discussed below.

Assumption 1: The first assumption is that the dependent variable can be calculated as a linear function of a specific set of independent variables, plus a disturbance term. The omission of relevant independent variables or the inclusion of irrelevant independent variables is a common violation of this assumption and is certainly true in the teacher evaluation scenario. As often, in practice, the function is specified simply as student satisfaction= f(teacher performance), omitting many of the other relevant variables. Secondly, often irrelevant independent variables are added to the vector of teacher performance. For example, asking students to assess the knowledge of the instructor is an invalid question that adding it to the equation is essentially adding an irrelevant variable.

Assumption 2: The expected value of the disturbance term is zero. This assumes that there is no bias in the data. If one accepts the idea that teaching evaluations are not given at random times but at times when faculty gauge they will receive the best evaluations, the data should be considered biased. For example, one professor stated that they give their evaluations out two class periods after the last mid-term exam a time when a minimum number of students are attending the class. In this way, this professor is self-selecting the most serious students and biasing upward the teaching evaluation scores.

Assumption 3: Disturbance terms all have the same variance and are not correlated with one another. Across observations (and in this case we are referring to the unit of analysis being different courses) one would expect the disturbance terms to have much greater variance in larger courses than in smaller courses, simply due to the broader range of personalities and perceptions possible.

Assumption 4: Observations on the independent variables can be considered fixed in repeated samples. Again, thinking about the unit of analysis being different courses one would not expect the independent variables to be fixed in repeated samples. For example, think of a university where Professor A teaches Personal Finance using the Garman and Forgue text to a class of 80 at 8:00 a.m. Mondays, Wednesdays and Fridays and Professor B teaches Chemistry using the standard chemistry text to a class of 500 at 2:00 p.m. Tuesdays and Thursdays and so on and so forth until you have listed all of the different courses, days, places, etc. that people teach. If the idea is to hold all independent variables constant so that only those related to teacher performance can be assessed, one can see that across observations it is impossible to repeat these values. The problem leads to the conclusion that if teacher performance can not be isolated, then comparing the scores of teaching performance across classes is faulty.

Assumption 5: The number of observations must be greater than the number of independent variables and there must be no linear relationships between the independent variables. This assumption is clearly violated in cases where the class size is very small (say under 20).

Because of the problems identified in evaluating teaching based on student teacher evaluations alone (a very common practice) and because of the statistical problems identified with using student teaching evaluations in a quantitative paradigm, one conclusion that can be drawn is that the summative evaluation of teacher effectiveness cannot be easily done in the quantitative model because of the violation of assumptions and the cost and time required to actually conduct the quantitative assessment in a valid and reliable manner. In practice, evaluating teaching effectiveness is essentially a qualitative process that should involve multiple sources of information. The appropriate model agreed upon in current literature is a qualitative use of a portfolio model of summative evaluation. This model is described in the next section.

The Teaching Portfolio as a Paradigm of Appropriate Evaluation of Teaching Performance

Returning to the goal of teaching evaluation once again, the goal of the evaluation is to determine how well a teacher aids the learner in knowing more about a subject, thinking critically about a subject, and developing an appreciation for the subject. The summative teaching evaluation must result in a single summative measure such as that the teacher "not effective," "effective," "shows promise of excellence," or "is excellent."

Because teaching is complex and multifaceted, no single measure of teaching effectiveness can fairly give a single summative rating. According to Seldin (1991) and other researchers (Diamond, 1987; Centra, 1994) an appropriate tool to represent the complex teaching task is the portfolio. In the portfolio, the teacher is requested to gather a variety of viewpoints as to his or her teaching effectiveness. The evidence of good teaching should include self, others and the products of good teaching (Selden, 1991).

According to Selden (1991), the material from the teacher, or the self-evaluation, could include a statement of teaching responsibilities, a statement of teaching philosophy, a statement of teaching short and/or long-term teaching goals, copies of course syllabi, a history of efforts designed to improve teaching, a description of curricular revisions undertaken, research on teaching undertaken, evidence of advising honors students, graduate students and other research learning activities, and a self-evaluation by the professor. Materials from others could include evaluations by peers (both inside and outside the institution), student course evaluations, statements from the administration such as the department head or other administrators, evaluations from student advisees, honors received as a teacher, evidence of participation in curriculum creation or teaching improvement efforts within the discipline, and videotapes of a typical class. Products of good teaching could include the scores students receive on tests, particularly standardized tests, student essays or reports, records of student achievement after leaving the course and/or institution, records of student achievement in more advanced courses, and statements from alumni on the quality of the instruction.

Arreola (1995) recommends that the faculty have the decision authority in deciding what should go into the portfolio. He offers a thoughtful system for the development of a faculty evaluation system that allows faculty to decide on the trade-offs between comprehensiveness and the time requirements. Because teaching is complex and multifaceted, the teacher could spend enormous time and energy evaluating the teaching. Given the full schedules of faculty members, the evaluation prepared should not be unduly time-consuming for the faculty members or others. The evaluation prepared should be able to be prepared within five to seven pages (Seldin, 1991) and with a few hours of effort to report preparation process.

Since the model is going to produce a summative measure for judgements about personnel, the teacher should have the opportunity to produce their work in the best light possible. Therefore, the model allows the teacher to intensively evaluate a course that is the most developed as a large part of the evaluation package. Additionally, the teacher has the responsibility of communicating their efforts to improve those courses that are less developed.

Portfolios also offer the advantage of presenting material outside the classroom as evidence of the teacher's commitment to and skill in teaching. For example, teaching rewards, research in teaching methodology, serving on national committees to improve teaching in the discipline, responding to requests for assistance in improving a colleague's teaching, and mentoring former students are all activities that excellent teachers do, but are not typically reflected in the student teaching evaluations. Portfolios can show this type of teaching activity in a favorable light.

Portfolios are suggested as the best method of showcasing the multifaceted aspects of teaching. However, portfolios, just as other method of evaluation, can be constructed poorly or well. A rigorous approach to constructing portfolios is to incorporate reflection of the teacher. The teacher must provide evidence of growth and change, success and failure, additional activities in the teaching function, and plans for the future. The teacher must communicate the logic and depth of their teaching performance. Teachers must have concrete evidence and communicate their work in order for the portfolio to be credible (Koon & Murray, 1995).

As pieces of qualitative research, the philosophies, theories and methodologies of qualitative research should be followed to produce the information necessary for assigning a valid and reliable summative result. Much more work should be done to instruct faculty and administrators on applying the qualitative paradigm to this very important performance evaluation--this evaluation that directly affects the livelihoods of faculty and other educators as well as the real and perceived success of the sponsoring institutions for these education workers.

Summary and Conclusions

This essay has highlighted the concept of the student as the ultimate consumer of education, examining the role of the student in summative teacher evaluation. While the student is an important player in the educational process, the student cannot be the all-knowing consumer in the short run. Therefore, the student should be one voice of many in the summative evaluation of a teacher. Further, examination of the quantitative model of analyzing student teaching evaluations statistically shows that violating the assumptions of parametric linear statistical models is nearly always the case in educational institutions. Because of these problems, a qualitative model of summative evaluation should be used, incorporating a multi-faceted portfolio that is well constructed following sound principles of qualitative research. Sound research methods in the conducting of summative teaching evaluations can do much to ensure justice and fairness of the summative evaluation and to increase the probability that sound educational experiences will be designed for students.

References

Arreola, R. A. (1995). Developing a comprehensive faculty evaluation system. Bolton, MA: Anker Publishing Company, Inc.

Bloom, B., Engelhart, M., Furst, E., Hill, W. & Krathwohl, D. (1956). Taxonomy of educational objectives. New York: David McKay Company, Inc.

Centra, J.A. (1994). The use of the teaching portfolio and student evaluations for summative evaluation. Journal of Higher Education, 65(5), 555-570.

Diamond, R. M. (Ed.). (1987). A guide to evaluating teaching for promotion and tenure. Center for Instructional Development, Syracuse University.

Kennedy, P. (1985). A guide to econometrics. Cambridge, MA: The MIT Press.

Koon, J. & Murray, H.G. (1995). Using multiple outcomes to validate student ratings of overall teacher effectiveness. Journal of Higher Education, 66(1), 61-81.

Schmelkin, L. P., Spencer, K. J. & Gellman, E. S. (1997). Faculty perspectives on course and teacher evaluations. Research in Higher Education, 38(5), 575-592.

Seldin, P. & Angelo, T. A. (1997). Evaluating teaching performance: When will we ever learn to use what we know? Videoconference Sponsored by the Virginia Tidewater Consortium for Higher Education and the National University Telecommunications Network, Presented November 14, 1997.

Shapiro, E. G. (1990). Effect of instructor and class characteristics on students' class evaluations. Research in Higher Education, 31(2), 135-148.

Washburn, K. & Thornton, J. F. (Ed.). (1996). Dumbing down: Essays on the strip mining of American culture. New York: W.W. Norton & Company.

Weber, Margaret J. (1999). InVEST study. Stillwater, OK: Oklahoma State University.

Endnotes

[1] Assistant Professor, Department of Health and Human Development

[2] Assistant Professor, Department of Health and Human Development