

MAKING EFFECTIVE USE OF STUDENT EVALUATIONS TO IMPROVE TEACHING PERFORMANCE

Brian T. Engelland, Mississippi State University

ABSTRACT

Most universities have adopted a system to obtain student evaluation of teaching (SET) data at the end of each semester. Scores are utilized to provide “objective” data in support of annual faculty performance evaluations, but few enjoy being evaluated, and many faculty have an aversion to this process. Problematic are concerns over the objectivity and validity of SET scores and the fact that the data rarely provide clear guidance to assist faculty in improving their teaching. This paper offers information and encouragement for faculty to manage the SET process more effectively, so that the information obtained will benefit on-going efforts to improve instruction and teaching effectiveness.

INTRODUCTION

The idea that student performance should be graded is as old as the teaching profession. The idea that faculty performance should also be graded is almost as old. For instance, recall that the quality of Socrates’ teaching was judged to be deficient, and he lost his life as a result. Sanctions for ineffective teaching are somewhat less severe in the 21st century, but many consider the denial of tenure just as life-threatening. Most colleges and universities have adopted some form of teaching evaluation system (Simpson and Siguaw 2000), and these systems have begun to play an increasingly important role as input to promotion decisions, salary increases and tenure determinations. Consequently, faculty have a vested interest in scoring well on their student evaluation of teaching (SET) performance.

The purpose of this paper is to provide information and encouragement for marketing faculty to manage the SET process more effectively, so that the information obtained will benefit on-going efforts to improve teaching effectiveness. The paper begins with a discussion of what SET instruments really measure. Then, the paper identifies important considerations involved in selecting good SET questions as part of a total program of teaching assessment.

WHAT DO SET INSTRUMENTS REALLY MEASURE?

There has been considerable debate in the literature about what constitutes good measurement of effective teaching (Bosshardt and Watts 2001; Young and Shaw 1999). No disagreement exists regarding the idea that the art of teaching is about transferring knowledge to stu-

dents, and that the central measure of teaching effectiveness is whether students mastered the material they were supposed to learn. However, scholars have labored to develop measures of teaching effectiveness that are more global in application, that is, measures that may be used to compare teaching performance from course to course and instructor to instructor across the campus. This stream of research has focused not on whether specific course objectives are accomplished, but on the identification of global surrogate factors – characteristics, behaviors and traits – that correlate highly with knowledge transfer measures. While much of the literature has modeled teaching effectiveness as a complex, multidimensional construct, the literature has been inconclusive with respect to what these global dimensions are (Marsh 1987; Marsh and Roche 1997).

For instance, Marsh (1984, 1987) identified nine dimensions, including learning/value, enthusiasm, organization, group interaction, individual rapport, breadth of coverage, examination/grading, assignments, and workload/difficulty. Seiler and Seiler (2002) collapsed those dimensions to four, to include course delivery style, professor characteristics, course characteristics, and the workload required to pass the course. Others have modeled many of the dimensions identified in the literature as either antecedent or derivative constructs, and therefore not a part of the teaching effectiveness construct domain (Abrami and d’Apollonia 1997). Consequently, Abrami and d’Apollonia (1997) suggested that dimensions of good teaching vary across courses and instructors, and argued against the use of multidimensional global measures. On the other hand, Ryan and Harrison (1995) recommended a weighting approach in which the importance of each dimension is adjusted according to course and instructor characteristics. All of this work has resulted

in a very confused understanding of what constitutes teaching effectiveness and how it should be measured. Because the literature has failed to reach any consistent agreement as to the dimensions of effective teaching, faculty committees at colleges and universities have done the best they could with the tools available and have adopted a wide variety of instruments.

SET data gathered from these instruments have been studied broadly with a view toward understanding which student characteristics and course factors best correlate with SET scores. This research shows that age, year in school, major, GPA, expected grade in course, and measured aptitude are the student characteristics that significantly correlate with SET scores; that years of experience, academic rank, sex, age, grading leniency, allocation of effort to the course, and measured personality characteristics are correlated instructor characteristics, and that whether the course is required, class size, and the method of administering the questionnaire, are correlated situational characteristics (Greenwald and Gillmore 1997; Fandt and Stevens 1991; Wallace and Schwab 1973; Seiler, Seiler, and Chang 2001). This research demonstrates that extraneous factors have a great deal to do with observed differences in SET scores, and suggests that a kindly old "Yoda-like" male professor teaching a small elective class to a group of high performing graduate students would have an advantage in achieving high SET numbers, regardless whether students learn anything. Consequently, there appears to be some reliability and validity issues associated with the typical SET instrument used on campuses today.

ARE SET ITEMS WELL CONCEIVED?

In order to understand the composition and structure of current SET instruments a little better, a convenience sample of nine instruments currently in use were gathered from colleagues at universities across the country. Items contained on these instruments were then given to a panel of eight marketing doctoral students enrolled in a measurement seminar, and the panel was asked to evaluate the quality of the items used. Results indicated that 50.6 percent of the items included on these instruments were deemed unacceptable because of spurious content, double-barrel phrasing or confusing terminology. The presence of these items tend to reduce reliability and validity of the overall measure of which they are a part, and undermine the saliency of the whole SET process. A selection of these problem items is listed in Table 1.

Seven of the instruments contained items that were deemed by the panel to be spurious because the item tapped something that on the face had little to do with teaching effectiveness. This accounted for 29 percent of the poor items identified by the panel. An example of this type of item is, "I have had to work hard in this course." While hard work is a noble undertaking, and we all hope

that our students work hard, the hard work could be because the instructor presented the materials ineffectively, and the student had to learn it on his own! Or, perhaps the hard work was necessary to overcome poor prior instruction in a prerequisite course. Consequently, the item is not directly related to teaching effectiveness. Another spurious example is, "The instructor asked students to help each other to understand the concepts presented." Asking students to help one another may be an appropriate teaching approach in courses where the integration of concepts with real-world examples is necessary. But is it appropriate in every course? Was the reason for involving students in explaining the material because the instructor had difficulty explaining the material himself? The panel suggested that involving students as instructors for other students appeared as an abdication of good student-teacher roles, and did not address teaching effectiveness. A third example of a potentially spurious item is, "The instructor's presentations are well-organized." While research has shown a correlation between organized presentations and teacher effectiveness, research has also shown that effective teaching does not always require organization (Young and Shaw 1999). There are many effective teachers who deliver spontaneous classes to great effect. Certainly good organization, as desirable as it may be, should not be considered a central dimension of teaching effectiveness.

Double-barreled items were found in eight of the nine instruments studied, accounting for 37 percent of the offending items. Typical was an item which stated, "The instructor is willing to listen to student questions and opinions." This is really two questions rolled into one, as student questions and student opinions could have far different connotations. Certainly an effective instructor is attentive to questions that students have about the course material, but that same effective instructor may engage in behaviors that stifle those disruptive students who interrupt discussion with ill-formed opinions. Double-barreled items do not provide clear guidance as to which of the barrels the student is addressing, and should be recast into two separate items.

Confusing items were also present in eight of the nine instruments, accounting for 36 percent of the offenders. Typical was the item, "I would like to have this instructor again." On the face of it, the item sounds like an appropriate way of assessing effectiveness. If students indicate that they want to have the instructor again, then the instructor must certainly be effective, right? Perhaps, but on the other hand, student enthusiasm may have to do more with grading leniency, classroom humor, or other extraneous factors, than with teaching effectiveness. Furthermore, if the instructor only teaches this particular class, a student respondent may be moved to disagree with the item, since it would mean that he or she would be taking the same class all over again. Items such as these should be recast to eliminate ambiguity.

TABLE 1
TYPICAL PROBLEM ITEMS USED IN SET INSTRUMENTS

Potentially Spurious Questions

1. I have had to work hard in this course.
2. The instructor asked students to help each other to understand the concepts presented.
3. The instructor's presentations are well-organized.
4. The overall quality of the textbook was appropriate for the course.
5. The course increased my interest in the subject matter.
6. The course content is consistent with my prior expectations.
7. The instructor had command of spoken English.
8. This instructor encourages divergent thinking.
9. The instructor encouraged student-faculty interaction outside of class.

Double-Barreled Questions

1. The instructor is willing to listen to student questions and opinions.
2. The instructor used approaches to teaching that were interesting and creative.
3. The instructor inspired students to set and achieve goals which really challenged them.
4. I always felt challenged and motivated to learn.
5. The instructor was available during office hours or by appointment.
6. The instructor was warm and friendly.
7. Assignments and projects are clearly defined.
8. The assignments were appropriate in amount and level.
9. The instructor's way of summarizing or emphasizing important points is effective.

Confusing Questions

1. I would like to have this instructor again.
2. I was challenged by this course.
3. I was challenged by this instructor.
4. The instructor found ways to help students answer their own questions.
5. This instructor willingly considers questions from students.
6. The instructor knows if the course content is being understood.
7. This person teaches at too high a level for the class.
8. Tests, quizzes, and projects challenge the student to express his/her understanding and knowledge of the subject matter.
9. The professor is a skilled lecturer or skilled discussion leader depending on which type is appropriate for the subject matter.

Other problems the panel observed included the confounding of student, course, and instructor characteristics as part of effectiveness scoring; questionnaires that were too long and tiresome to complete in a reasonable time, no reverse-coded items to control for acquiescence bias; and inadequate open-ended questions or reply space to encourage student response. In conclusion then, the SET instruments the panel evaluated seem to be poorly designed and ill-prepared to do the job for which they were intended. Unfortunately, these instruments are generally inadequate in reflecting the quality of the learning

experience, or indicating whether the students believe they accomplished the established learning objectives.

OTHER INDICATIONS OF A "PROBLEM"

There are other indications that SET scores do not serve the function for which they were intended. Research shows that individual instructor's self-rating of teaching effectiveness does not correlate highly with SET scores (Bosshardt and Watts 2001), suggesting that a wide gap exists between what professors believe they are delivering

and what SET scores suggest students receive. In addition, faculty have a real concern whether students have the ability to evaluate faculty accurately and effectively (Simpson and Sigauw 2000). Naturally, faculty members may be uncomfortable with the results.

Results of another group of studies are most troubling. Active teaching techniques, such as group assignments, discussion problems, and group case studies, have been shown to improve student cognitive processing of course materials, increase retention, and enhance the rate of learning. However, extant research has shown no significant differences in SET scores for courses taught with active learning techniques versus courses taught with more traditional lecture and test methods (Leeds, Stull, and Westbrook 1998). If students are really learning better, shouldn't this improvement show up in SET scores? This can only happen if SET instruments really assess teaching effectiveness, but it appears many do not. The standard teaching evaluation instruments and procedures fail to do an adequate job of assessing teaching effectiveness. Rather, they measure, in aggregate, how well students like certain characteristics of the instructor and course. As a consequence, SET administrations may offer little actionable advice on how the course or instruction can be improved, focusing the attention on cosmetics rather than knowledge transmission. No wonder faculty often fail to take the process seriously.

HOW DO FACULTY RESPOND TO SET MEASUREMENT?

Often faculty sense that university sponsored SET instruments are not completely serving their purposes, so they are faced with a choice between two options – either fight a battle to get something more useful in place (not easy for junior faculty), or play the “teach-to-the-instrument” game, in which SET scores are manipulated by instructor shenanigans (Wilson 1998). A number of creative strategies have been developed to support the latter choice. These include inducements (serving pizza the day of evaluations to put students in a good mood), pre-evaluation actions (telling the class how successfully they have performed, then passing out the evaluations), grading leniency (cancel announced quizzes or hand back high-scored assignments immediately prior to SET administration), watching (remaining in the room during SET administration), and manipulation of expectations or achievements (repeatedly tell students that they're outperforming other classes) (Simpson and Sigauw 2000; Neath 1996). But all of these approaches ultimately backfire for the faculty member who is genuinely interested in improving teaching performance. By artificially raising scores through these actions, a true understanding of the instructor's teaching performance is masked and real opportunities for improving the craft of teaching are lost.

HOW SHOULD TEACHING EFFECTIVENESS BE MEASURED?

Faculty should keep in mind that the best global measure is still merely a surrogate for whether the student has accomplished the learning goals of the course. A complete assessment of teaching effectiveness requires more than just SET scores. Recent literature has suggested that a full assessment of teaching effectiveness should not solely rely on SET scores, but instead should require two additional types of assessment, (1) authentic real world assessment, and (2) student self-assessment (Schrock 1997). If colleges and universities do not provide the systems to gather needed metrics on a timely basis, then it is incumbent on the faculty member to take responsibility for gathering and assessing the kinds of information needed to assess their courses. It can't be left up to the institution's official process, especially if that process and its measures are deficient. Each of the three types of assessment are discussed in the following paragraphs and in Table 2.

Authentic assessment requires that students demonstrate they can actually apply the course material to a “real world” task, and do so to the satisfaction of an outside judge. For instance, as a result of the class, can students write a marketing plan for a new product, construct a research questionnaire, make an effective sales pitch, etc., depending on the objectives of the course? Authentic assessment can be implemented with the help of practitioner alumni who volunteer as judges and help rate student's submissions according to some scale. Alumni judges who visit the class when team presentations are made can also assist the instructor in establishing a high level of importance and professionalism to the exercise.

Student self-assessment requires that students take an introspective look at their own performance and learning. Self-assessment can be made after key learning tests (such as examinations and papers), at the end of the course (through supplementary course evaluation questions) and after graduation, through administration of an instrument containing appropriate satisfaction items. Alumni surveys may be particularly effective in identifying meaningful course activities and learning achievements, especially after the student has been in a real-world job for some period of time. Self-assessment measures attempt to discern how the student rates his or her own achievement relative to the objectives established for the course or degree program.

Good self-assessment questions should relate to students' performance relative to the learning objectives established. Several such questions are presented in Table 3. If students do not have confidence that they know something, this may indicate a deficiency in instruction that should be resolved. In addition to questions related to course objectives, it is also appropriate to include items

TABLE 2
SUGGESTIONS FOR EFFECTIVE TEACHING ASSESSMENT

1. Each instructors should take responsibility to gather and assess the kinds of information they need to assess their courses. It can't be left up to the institution's official process, especially if that process and its measures are deficient.
2. Even if you are required to use a university or college adopted SET instrument, you can augment the process with an instrument of your own design, one that both measures teaching effectiveness and also provides the information you need for improvement.
3. Authentic assessment requires that students demonstrate they can actually apply the course material to a "real world" task.
 - Write a Marketing Plan
 - Perform a Case Analysis
 - Create an Integrated Communications Plan
 - Develop a Positioning Statement
4. Alumni and local business leaders can visit classes and otherwise serve as "outside judges" of student performance.
5. Student self-assessment requires that students take an introspective look at their own performance and learning. Self-assessment can be made after key learning tests (such as examinations and papers), at the end of the course (through supplementary course evaluation questions) and after graduation, through administration of an instrument containing appropriate satisfaction items.
6. Questions that ask the respondent to recall certain in-class situations when learning was most/least facilitated provide effective diagnostics. This approach is referred to as the "critical incident technique" (Sautter and Hanna 1995).
7. Open-ended questions that allow the student to state what they liked and disliked about the course and instructor are very helpful.
8. Students exhibit a great diversity in expectations about how courses should be conducted and what behaviors on the part of their instructors constitute excellence (Chonko, Tanner, and Davis 2002), so ongoing assessment is absolutely necessary.

related to other aspects of student performance, including class attendance. Personal experience suggests that students who do not come to class do not perform well on exams and other forms of learning assessment. Students with high absenteeism typically rate their performance low against learning objectives.

Even if the faculty member is required to use a university or college-adopted SET instrument, the process can be augmented with an instrument of the faculty member's own design, one that both measures teaching effectiveness and also provides the information needed to identify any weaknesses and guide improvement. The key is to structure the questions so that the answers provide guidance to the faculty member. Questions that ask the respondent to recall certain in-class situations when learning was most/least facilitated are effective in understanding what activities worked and what didn't work. This

approach is referred to as the "critical incident technique" (Sautter and Hanna 1995). Supplementary questionnaires should continue from semester to semester, because students exhibit a great diversity in expectations about how courses should be conducted and what behaviors on the part of their instructors constitute excellence (Chonko, Tanner, and Davis 2002).

Assessment is an on-going process and should not be relegated to an end-of-a-semester activity. Faculty may conduct assessments through-out the semester, especially when a new project, case, or lecture is tried. Assessment may be conducted through the traditional pencil-and-paper approach, class discussion approach, or through an on-line survey. The on-line format offers some advantages in terms of encouraging well-constructed written comments and a depth of discussion that is hard to achieve in the typical pencil-and-paper in-class exercise. However,

TABLE 3
POSSIBLE SELF-ASSESSMENT ITEMS

1. The first objective of the course is to “improve the student’s abilities to identify marketing problems and opportunities confronting an organization.” How well do you believe you achieved this objective?
2. The second objective of the course is “to develop an ability to create a marketing plan that resolves problems and capitalizes on opportunities.” How well do you believe you accomplished this objective?”
3. The third objective of the course is “to apply marketing management concepts to personal career planning efforts.” How well do you believe you accomplished this objective?
4. Other objectives for the course are to improve the student’s comprehension of the role of marketing in an organization, the steps involved in marketing planning, the control process for evaluating marketing performance, and the effects of social, legal, ethical, and technological forces. How well do you believe you accomplished these objectives?
5. Approximately, how many classes did you miss this semester? _____
6. How do you rate your class attendance for the course?
7. How do you rate the quality of your class participation and discussion?
8. Overall, how do you rate your performance in the course?

there are some drawbacks to on-line formats, most especially respondent concern over anonymity (Dommeyer, Baum, and Hanna 2002).

SUMMARY AND CONCLUSIONS

This paper began by reviewing the SET assessment literature, and concluded that the construct domain for effective teaching is open to a fairly broad interpretation. Efforts to develop robust global measures that can be used across the campus have focused on factors and items that only indirectly relate to the central question: whether students under the instructor’s influence have achieved the learning objectives established. Further, an examination of nine instruments currently in use reveals widespread problems in item wording which can adversely affect reliability and validity. It is no surprise that faculty tend to mistrust the SET process, discount the validity of ratings received, and are drawn toward engaging in behaviors designed to manipulate ratings rather than improve authentic student learning.

The marketing discipline has long recognized that businesses need to collect effective performance metrics such as market share, customer satisfaction, brand attitude, response time, market coverage, and stock-out rates so as to provide a running evaluation of a marketing

performance (Best 2000). Marketing faculty commonly teach the value of these performance metrics and how market research can be used to guide organizations toward continuous performance improvement. Unfortunately, when it comes to improving teaching performance, many marketing faculty fail to gather sufficient performance metrics to guide their own teachingimprovement.

As presented in this paper, a well-rounded assessment of teaching performance requires authentic assessment to determine if students learned what was intended, student self-assessment to determine student satisfaction with their own performance, and student evaluation of teaching to determine student satisfaction with the quality of the teaching they received. Marketing faculty should take the initiative to ensure that all three of these assessments are gathered and analyzed so that the faculty continue to improve real teaching performance from class to class, and semester to semester. SET instruments are only effective if they assist professors in improving teaching performance by providing diagnostic information that can result in actionable changes. Professors have the responsibility to improve what they do, and if the SET process is properly constructed, it can provide valuable information to help make this possible.

REFERENCES

- Abrami, Philip C. and Sylvia d'Apollonia (1997), "Navigating Student Ratings of Instruction," *American Psychologist*, 52 (11), 1198–1207.
- Best, Roger J. (2000), *Market-Based Management: Strategies for Growing Customer Value and Profitability*, 2nd Ed. Upper Saddle River, NJ: Prentice Hall.
- Bosshardt, William and Michael Watts (2001), "Comparing Student and Instructor Evaluations of Teaching," *Journal of Economic Education*, (Winter), 3–17.
- Chonko, Lawrence B., John F. Tanner, and Roger Davis (2002), "What Are They Thinking? Students' Expectations and Self-Assessments," *Journal of Education for Business*, (May/June), 271–81.
- Dommeyer, Curt J., Paul Baum, and Robert W. Hanna (2002), "College Students' Attitudes Toward Methods of Collecting Teaching Evaluations: In-Class Versus On-Line," *Journal of Education for Business*, (September/October), 11–15.
- Fandt, Patricia M. and George E. Stevens (1991), "Evaluation Bias in the Business Classroom: Evidence Relating to the Effects of Previous Experiences," *Journal of Psychology*, 125 (4), 469–77.
- Greenwald, Anthony G. and Gerald M. Gillmore (1997), "Grading Leniency Is a Removable Contaminant of Student Ratings," *American Psychologist*, 52, 1209–17.
- Leeds, Michael, William Stull, and Jilleen Westbrook (1998), "Do Changes in Classroom Techniques Matter? Teaching Strategies and Their Effects on Teaching Evaluations," *Journal of Education for Business*, (November/December), 75–78.
- Marsh, Herbert W. (1984), "Students' Evaluations of University Teaching: Dimensionality, Reliability, Validity, Potential Biases, and Utility," *Journal of Educational Psychology*, 76, 551–63.
- _____ (1987), "Students' Evaluations of University Teaching: Research Findings, Methodological Issues, and Directions for Further Research," *Journal of Educational Research*, 11, 253–388.
- _____ and Lawrence A. Roche (1997), "Making Students' Evaluations of Teaching Effectiveness Effective," *American Psychologist*, 52, 1187–97.
- Neath, Ian (1996), "How to Improve Your Teaching Evaluations without Improving Your Teaching," *Psychological Reports*, 78 (3), 1363–72.
- Ryan, Joseph M. and Paul D. Harrison (1995), "The Relationship Between Individual Instructional Characteristics and the Overall Assessment of Teaching Effectiveness Across Different Instructional Contexts," *Research in Higher Education*, 36 (5), 577–94.
- Sautter, Elise Truly and Joe Hanna (1995), "Instructional Development Using the Critical Incident Technique," *Marketing Education Review*, 5 (1), (Spring), 33–40.
- Seiler, Michael J., Vicky L. Seiler, and Dalen Chiang (2001), "Professor, Student, and Course Attributes that Contribute to Successful Teaching Evaluations," *Financial Practice and Education*, (Fall/Winter), 91–99.
- Seiler, Vicky L. and Michael J. Seiler (2002), "Professors Who Make the Grade," *Review of Business*, (Spring), 39–44.
- Schrock, Kathy (1997), "The Many Layers of Assessment," *Technology Connection*, 4 (7), 5.
- Simpson, Penny M. and Judy A. Siguaw (2000), "Student Evaluations of Teaching: An Exploratory Study of the Faculty Response," *Journal of Marketing Education*, 22 (3), (December), 199–213.
- Wallace, Marc J., Jr. and Donald P. Schwab (1973), "The Validation of a Teaching-Effectiveness Measure in Two Business Schools," *Academy of Management Proceedings*, 229–36.
- Wilson, Robin (1998), "New Research Casts Doubt on Value of Student Evaluations of Professors," *Chronicle of Higher Education*, 44 (19), A12–A15.
- Young, Suzanne and Dale G. Shaw (1999), "Profiles of Effective College and University Teachers," *The Journal of Higher Education*, 70 (6), (November/December), 670–86.