Online Student Evaluations and Response Rates Reconsidered

by Joan Anderson, Gary Brown, and Stephen Spaeth

The increased adoption of online student evaluations from 2% in 2000 to almost 33% in 2005 by a variety of institutions reporting to Brigham Young University's Online Student Evaluation of Teaching in Higher Education (OnSET) Web resource has not been without controversy. The major concern voiced about online course evaluations is reduced response rates from the students. However, our experience with online student course evaluations suggests that such concerns mask a more pressing problem. A diminished response rate to course evaluations has less to do with the method of distribution than it does with faculty and student engagement, particularly since engagement reflects the efficacy of evaluation.

Because student evaluations are so pervasive in use and so predominant in educational research, the lack of engagement with the evaluation process is ironic. For example, Stronge (1997) identifies student ratings as the most frequently studied aspect of education, and yet at the same time, he also notes the root of the challenge—the ineffective use of student ratings. Stronge argues that central to this problem is the lack of faculty development related to the interpetation and requisite responses to student ratings. Theall and Franklin (2000) likewise confirm that, in spite of mountains of research on student evaluations, the use of such evaluations remains inadequate. In this context the role of online delivery in student evaluations deserves further consideration as both a potential means of enhancing the process and making it more relevant to the concerns of instructors as well as students.

In what follows we provide an overview of the reasons why the evaluation process has failed to elicit sufficient involvement from students and instructors as well as the ways in which an online format can help address these problems. We then offer the results of an ad hoc study in which student response rates to an online evaluation tool were measured and assessed at a particular institution in order to determine the factors that influence participation, engagement, and perceived relevance.

The Challenges

Moving student evaluations online inherits challenges unrelated to technology and the Internet; instead, as is often the case, it is the migration online that puts an old issue under a new light. There remains a persistent lack of evidence that student evaluation instruments evince strategies for improvement, ostensibly their principle purpose. In fact, evidence suggests that most student course evaluation instruments may even impede improvement (Birnbaum 2000). The lack of a shared understanding of the purpose of evaluation among students, faculty, and administration underlies this concern. In spite of the fact that "literature from the past 10-15 years emphasizes staff development as key to effective evaluation practices that promote teacher growth and improvement" (Annunziata 1997, 289), there are only a few examples where systematic faculty and student development directly tie into the evaluation process. As a result, instructors and students seldom feel ownership over the evaluation process; moreover, instructors lament that the evaluation instruments do not pertain to (and even punish) pedagogical innovations, do not accomodate student demographics or the specific context in which teaching occurs, and yield inappropriate results that only encourage false comparisons and have little to do with helping improve teaching practice (Stronge 1997; Better Teaching through Assessment n.d).

Such shortcomings in the evaluation process may be compounded by other issues as well. For example, in the minds of instructors and students alike, the evaluation of instructional performance tends to be associated with student satisfaction—often pursued at the expense of learning (Stronge 1997; Better Teaching through Assessment n.d.). Discovering that student satisfaction has declined does not necessarily give the instructor

much guidance for trying to improve the course. Furthermore, the primary use of evaluations typically occurs at the end of the course when it is too late to yield meaningful change; consequently, students and instructors alike often find little use for the evaluation process. For students in particular, the absence of meaningful opportunities to shape their learning experience further contributes to their increased disengagement, identified most starkly by Kuh (2003) who, after reviewing the results of the National Survey of Student Engagement (NSSE) for several years, concludes: "Most students come to college expecting to be more engaged than they are" (28).

The Opportunity

As Elbow (1992) has argued, attending to students' experience is an essential avenue for improving practice: "We must find ways to dignify student evaluation of teachers and to make the process thoughtful and reflective rather than mechanical" (¶ 24). The Teaching, Learning, and Technology (TLT) Group's Better Teaching Through Assessment (BeTA) project, supported by the Fund for the Improvement of Postsecondary Education (FIPSE), illustrates the commitment of several organizations and institutions to Elbow's perspective. The BeTA project embraces the potential of student evaluations to do more to improve student learning experiences, and the project particularly targets students' increasing facility with technologies. Since most students spend substantial portions of time online and have come to regard information technology as an extension of their everyday lives (Kvavik and Caruso 2005), migrating evaluation instruments online and more effectively mediating their use offers the most compelling way to engage students more fully in the evaluation process. Moreover, such a strategy can help address many of the concerns voiced by researchers and faculty members. In addition to the potentially significant cost savings (Bothell and Henderson 2003), Sorenson and Reiner (2003) note improved turnaround time, the provision of rapid feedback to faculty members (essential for promoting changes in the middle of the term), and greater opportunity for elaboration afforded by the ability to type responses online (i.e., students typically type more on open-ended questions than they write by hand). In addition, Sorenson and Reiner argue that online evaluations provide greater convenience for students to respond without using valuable class time, at least for synchronous classes. Others have noted similar advantages in moving student evaluations online (Hmieleski and Champagne 2000). Through such benefits, the online medium provides institutions with the valuable opportunity to reform and revitalize the process of student evaluation.

Response Rates and Online Evaluations

In spite of these potential advantages, the concern that moving student evaluations online reduces response rates persists. Practitioners cite a variety of strategies for improving response rates in this medium: Some educators advocate using extra credit and other incentives such as drawings for prizes, others use systematic reminder e-mails, and still others withhold grades or other information. Yet these strategies, however effective they may or may not be, each reflect a current culture in which student evaluations are external to the day-to-day business of teaching and learning and in which reflection upon practice for both students and teachers is ancillary. More constructive are those strategies that stress the need for faculty members to indicate the importance of results to students (Sorenson and Reiner 2003). Nonetheless, even the importance of the results may be compromised if their asserted significance pertains only to the summative assessment of faculty performance. Not surprisingly, students remain disengaged when the evaluation process consists of using a single number to rank faculty on a truncated scale rather than providing a productive, formative exercise that promotes improvement while honoring students' rich and complex experiences as learners. In this narrow, often high stakes context, it also should not be surprising that instructors (and sometimes students themselves) often voice the concern that students have an inadequate grasp of the subject and subsequently are not qualified to evaluate teaching—which essentially becomes a self-fulfilling prophecy by diminishing educator as well as student engagement with the potential of evaluation. Instead of seizing the opportunity to use student feedback and perspectives about the learning experience as an occasion to engage in a discussion about instructional strategies and student responsibilities, the evaluation is, instead, relegated to a largely dreaded administrative ritual.

The Case at WSU

These issues as they are playing out at Washington State University (WSU) suggest other ways to view the issue of migrating evaluations online. The College of Agricultural, Human, and Natural Resources (CAHNRS) at WSU has pioneered a comprehensive college-wide online evaluation. As part of the TLT-BeTA project, the college has collaborated with WSU's Center for Teaching, Learning, and Technology (CTLT) to develop an instrument based on principles of good practice that strives to be sensitive to multiple teaching strategies. The evaluation project has been designed to address common shortcomings of student evaluations by complementing (and, in time, crossvalidating) the results of the evaluations with peer review and eventually—as assessment processes are developed—with outcomes (Figure 1). In this regard the online instrument was designed as part of a broader institutional strategy of closing the loop between assessment and improved teaching practice at the university.

To the same end, the design of the online evaluation instrument also addresses the need for targeted, detailed, and context-specific information to support the ongoing refinement of instructional practice rather than merely providing summary data at the end of the course. The previous paper-and-pencil student course evaluation instrument focused more broadly on faculty performance and student satisfaction; for instance, all but a few questions invited students to assess the instructor's presentation, essentially presuming that teaching entails a lecture format. In contrast, the new online course evaluation instrument focuses on four discrete constructs: critical engagement or interaction, classroom environment, skills development, and student demographics/study strategies (Exhibit 1). The new instrument is designed to provide feedback responsive to various populations and pedagogies and is based on the assumption that different teachers will have different teaching goals and strategies. In particular, the four constructs were developed to assess students' perceived responsibility for learning as well as to allow clearer distinctions between instructional strategies that focus on presenting content, strategies that focus on skill development, and strategies that promote active and collaborative learning. Finally, the committee was cognizant of the problematic tendency to boil the results down to a single number and subsequently elected to report results according to the four key constructs.

Pilot tests of the new instrument were run in the spring and fall semesters of 2003 and 2004; a college-wide roll-out of the instrument subsequently occured in spring of 2005 during which students completed over 4,500 online surveys. The total number of students participating was smaller than the number of completed surveys because some students were enrolled in more than one surveyed class. More than 20 degree-granting units within the college generated responses, and several units outside CAHNRS also participated by virtue of course crosslisting. CTLSilhouette, the Web-based survey software developed by CTLT at WSU, was used to distribute the survey and collect the data, which was downloaded into SPSS for analysis. In our study, we sought to identify the key factors that influenced student response rates to the new online evaluation survey.

Response Rate Results

Changes from term to term in the instrument, the mode of delivery, and the participants unfortunately made formal comparisons between online and paper-based response rates problematic. The respective designs of the paper-based instrument and the online instrument did not lend themselves to direct comparisons of data with regard to specific items or categories of information. Moreover, in spite of the general perception that the paper-and-pencil survey garnered better response rates, administrators of the survey reported that the results associated with those response rates were not high enough to make summative comparisons with the new instrument tenable. Even in the synchronous, paper-based evaluation process, many students had submitted incomplete surveys, and response rates had not been tracked beyond a few years. As a broad measure of comparison, paper-based responses to the student evaluation instrument had run at about 50% over the past three or four years whereas the response rate to the online instrument was 41%. Given the limitations of the comparisons, the differences in response rates were negligible.

However, one significant pattern in the student responses to the online instrument was noticeable at the

outset. In contrast to the 41% response rate during the first roll-out of the instrument, response rates during the pilot phases conducted by faculty and chairs who had served on the project committee (and were subsequently engaged in the process) had been better than 80%. The drastic drop in response rates as we moved from selected and volunteer faculty members to the college-wide faculty population prompted further examination and was the first indication that the response rate phenomenon was one that a focus on technology alone might not explain.

In our analysis we therefore decided to focus more closely on differences in response rates to the online instrument across different kinds of classes within a single term. We also examined faculty characteristics in order to understand the extent that differences in response rates reflected not just student engagement but faculty engagement as well. We looked at individual student class level (academic maturity), faculty member rank, class size, class location, and academic discipline as additional factors and, when the information was available, whether extra credit predicted significant improvements in response rates. It is important to note that traditional and statistical analysis was not applicable in this case since the instrument was distributed not to a sample but to the entire population of teaching faculty in the college and all enrolled students. The variability of response rates inherently violates statistical assumptions of normality underlying sampling and statistical procedures; in this case, probability is subsumed by actuality. Nevertheless, as the results demonstrate, the distribution of responses provides a revealing picture of the response rate phenomenon at one institution.

Response Rate Fluctuations

The graph in Figure 2 illustrates a selection of 12 classes of four types. Note the widely disparate variation in response rates. Although the overall pattern might suggest that large introductory classes have lower response rates, such a generalization clearly remains untenable in light of the poor response rates in two of the three graduate seminar courses and two of the three small science courses; contrary to what one might expect, one graduate course garnered only a 12% response rate while a large social science course obtained an 84% response rate. Perhaps the most important finding to emerge from an overall response rate online of 41% is the conclusion that using a mean response rate is appropriate since the range or deviation of response rates varies dramatically in no obvious or predictable pattern.

What this level of fluctuation further suggests is that neither class size, discipline, nor the online distribution of the evaluation instrument predicts response patterns; instead, the individual attributes of each instructor-student cohort appear most significant. For example, we noted a 30% drop in response rate when the survey was introduced to faculty members who were not involved with its development—a fluctuation that hardly seemed coincidental (Figure 3). Equally notable, faculty who were in programs where the chair of the department was involved in the project had generally higher response rates than programs in which leadership was not involved (Figure 4). In addition, faculty members who were active on other teaching and learning committees, who frequently sought assistance from teaching and learning resource centers on campus, and who otherwise had a history of participation in faculty development activities had higher than average response rates.

Response rates also mirror student engagement. Although response rates fluctuated from class to class, upper division courses had higher response rates than lower division courses (51% for the former versus 39% for the latter), and at WSU, National Survey of Student Engagement results clearly indicate less student engagement among lower division students compared with upper division students (an average difference of 21% across the five NSSE benchmarks). Although we were unable to derive a statistical correlation between the two, it is unlikely that faculty engagement and student engagement are unrelated.

What Students Say

To further explore the factors related to student response rates, we invited several colleagues to take a few minutes of class time and ask their students if they responded to an online midterm evaluation and, if not,

why. Student responses were categorized as follows:

- Students who reported they did not respond to the midterm evaluation because "It was not required," because "they forgot," and because they "did not have time" were identified as *disengaged* (67%).
- Students who reported that they do not "do e-mail" or who reported link problems, however dubious, were identified as those *who had issues related to technology* (34%).
- Students who reported that they felt their response would not be attended to were identified as *no benefit* (10%).
- The remaining responses were attributed to *other* (14%).

Exhibit 2 shows a sample of our responses, and Figure 5 presents the consolidated distribution mentioned above. The variety of responses suggest that technology—the online aspect of the evaluation—is not the salient reason behind low student response rates; rather there is a clear sense of disengagement with the evaluation process itself, regardless of the medium. Further, the responses suggest that revising the evaluation instrument alone does not in itself enhance student engagement, however much it may solicit more nuanced, reflective, and relevant information regarding the teaching and learning process.

Commentary

With the essential caveats appropriate to a limited ad hoc study at a single institution, our experience suggests that the reporting of student response rates to online evaluations in the aggregate may be misleading. The sizeable fluctuation between classes is more interesting and has more urgent implications, none more so than the confirmation that more can be done to promote and nurture a culture that engages both instructors and students more fully in the evaluation process.

Despite the limited response rates of the initial roll-out, we believe that a properly designed, distributed, and reported online evaluation instrument can still play a key role in supporting a broader commitment to engage student voices regularly throughout a term. By using online evaluations to guarantee rapid turnaround times, provide custom questions that focus on current or even planned activities, and address the innovative teaching strategies and individual characteristics that distinguish different courses, faculty and students can come to reflect more easily upon those unique aspects of instruction that matter to them. Moreover, the flexibility of online surveys can help administrators make more productive decisions by helping them consider the significant distinctions between a first-year lecture course and an upper-division, writing-intensive course or between undergraduate lab courses and graduate seminars. Online surveys could also be used to provide opportunities for students to develop questions for each other and even, perhaps, for their faculty members—an especially valuable trait. Assessment is, after all, critical engagement, thinking, and learning.

At the same time, the results of our study also suggest that it is no less misleading to conclude that online technology will provide a quick remedy to the perceived disconnect between the evaluation process and the teaching and learning process. Insofar as such a disconnect has become deeply ingrained in academic culture, addressing this problem will require much of the same sustained, long-term institutional and administrative support that would be necessary even without the many advantages afforded by the online medium. Even a well-designed online survey will remain compromised in its value if the survey is not strongly supported at the local level of the department, if it is not regularly open to further input and refinement by faculty committees, and if it is not explicitly tied to institutional practices that consistently foster a culture of ongoing faculty development. Moreover, if online surveys merely replace rather than supplement a process of productive, open dialogue between instructors and students about what happens in the classroom, many students are still likely to regard such surveys as no less arbitrary and mechanical than the paper-based

evaluations forms of the past.

Conclusion

As various coalitions of constituencies increasingly require educational institutions to post evaluations online in the new learning market, there will be more pressures to reduce the complexity of teaching to simplistic and useless comparisons. But the opportunity to understand for ourselves and present for others the implications of the evaluation process in all of its complexity remains. Learning is not a simple phenomenon, and mere numerical ratings do not adequately represent or measure it; it is in this context that well-designed online student evaluations can offer a substantial contribution to how we assess teaching and learning. It is in this context, too, that response rates are themselves a critical indicator of student engagement and, whether or not we are prepared to embrace the full and final implications, a critical indicator of faculty engagement as well.

References

Annunziata, J. 1997. Linking teacher evaluation and professional development. In *Evaluating teaching: A guide to current thinking and best practice,* ed. J. H. Stronge, 288-301. Thousand Oaks, CA: Corwin Press.

Better Teaching through Assessment (BeTA). n.d. Course evaluation and the BeTA project. http://tltgroup.org/Beta/betahome.htm. (accessed July 23, 2006).

Birnbaum, M. H. 2000. A survey of faculty opinions concerning student evaluations of teaching. *The Senate Forum* 15 (2). http://faculty.fullerton.edu/senatenews/page2.html (accessed July 23, 2006).

Bothell, T. W., and T. Henderson. 2003. Do online ratings of instruction make sense? In *Online student ratings of instruction*, ed. D. L. Sorenson and T. D. Johnson, 69-79.

Elbow, P. 1992. Making better use of student evaluations of teachers. *ADE Bulletin* 101:2-8. http://www.ade.org/ade/bulletin/n101/101002.htm (accessed July 23, 2006).

Hmieleski, K., and M. V. Champagne. 2000. Plugging into course evaluation. *The Technology Source,* September-October. http://distance.wsu.edu/facultyresources/savedfromweb/pluggingin.htm (accessed July 23, 2006).

Kuh, G. 2003. The messy NSSE. *Change*, March-April:25-32.

Kvavik, R. B., and J. B. Caruso. 2005. EDUCAUSE Center for Applied Research (ECAR) study of students and information technology, 2005: Convenience, connection, control, and learning, 6. http://www.educause.edu/ir/library/pdf/ERS0506/ekf0506.pdf (accessed July 23, 2006).

Sorenson, D. L. and C. Reiner. 2003. Charting the uncharted seas of online student ratings of instruction. In *Online student ratings of instruction*, ed. D. L. Sorenson and T. D. Johnson, 1-21.

Stronge, J. H. 1997. Improving schools through teacher evaluation. In *Evaluating teaching: A guide to current thinking and best practice,* ed. J. H. Stronge, 1-23. Thousand Oaks, CA: Corwin Press.

Theall, M., and Franklin, J. 2000. Creating responsive student ratings systems to improve evaluation practice. In *Evaluating teaching in higher education: A vision for the future,* ed. K. Ryan, 95-107. San Francisco: Jossey-Bass.