A LITERATURE REVIEW:

ELECTRONIC VS. PAPER SYSTEMS OF COURSE EVALUATIONS

Dick Collier, Institutional Research, Planning, and Effectiveness, October 2008

Moving the Student Instructional Rating System (SIRF) to the Web, Memorandum, Wendell Lorang, Director of Institutional Research, UAlbany, 6/15/2005.

The benefits cited in this in-house memo of three years ago included possibility to add questions tailor-made for the course or of general concern to a department or school (particularly those that use their own instrument rather than SIRF); written comments from students need not be transcribed (digitalized) after the fact; much faster turnaround of results to faculty; and savings efficiencies for the University (forms, mailings, secretarial work in departments, mailroom, IR). The following concerns were also identified. A review of the literature strongly suggests these concerns, while warranted, do not appear to diminish the utility of a web-based SIRF administration.

Potential Concerns
1) faculty concerns about a fundamentally different course evaluation system;
2) low response rate from students;
3) response bias in respondents- negative or positive;
4) implementation and/or technology does not work well;
5) students demand access to results;
6) faculty concern about security of website/results;
7) faculty concern about results being released to students.

CONCERNS ABOUT A FUNDAMENTALLY DIFFERENT SYSTEM

Numerous articles in the 1990’s studied the use of web-based or other electronic questionnaires and, as the years progressed, found less and less reluctance, resistance and confusion on the part of the faculty. Perhaps more significant, studies also not surprising showed markedly waning student suspicion concerning online surveys and markedly increased student facility in handling them. There appears little concern about the mode of the response system on technical or “user-friendliness” grounds in the 21st century.

As for the questions themselves, they are the same as those on the paper version of SIRF, though the web-based SIRF allows additional questions to be added. The open-ended responses from students remain, but will no longer have to be typed within departments to protect students’ confidentiality. There is some evidence, though not recent, that students may be more voluble in their responses when they can type them into a computer and are under no time constraints to do so. Unfortunately, there is no current study suggesting this speed and volubility has continued to increase, though that is a reasonable expectation.
While response rates differ from study to study, no recent study (i.e., no study from a period when students and faculty are used to online surveys from campuses and otherwise) suggest diminished response rate has seriously impacted the results or their usefulness of student evaluations of courses and instructors. One of the most recent studies is that by Avery et al.:

- **Electronic Course Evaluations: Does an Online Delivery System Influence Student Evaluations?**
  The study’s “results found that Web-based evaluation methods led to lower response rates, but that lower response rates did not appear to affect mean evaluation scores. They suggested that faculty evaluation scores will not be adversely affected by switching from paper to Web-based evaluations.”

  A second invitation to participate in a web evaluation is often necessary (as is planned for web-based SIRF), which may not be as necessary in the 9 years since the Cartwright study was done: “Cartwright (1999) used electronic evaluations in his distance learning course and found a 20 percent response rate the first time, when the call for evaluations was e-mailed, and a 43 percent response rate at the second request for evaluations.”

  Response rate, as noted by Avery et al. and other studies, not surprisingly increases with student familiarity: “Rosenberg et al. (2001) reported response rates to evaluations in their medical residency program between 81 and 92 percent within the first 12 months of instituting the new electronic system. Ewell (2000) used data from the United States Air Force Academy (USAFA) during a period in which it was considering converting its mid- and end-of-course evaluations from paper to a computer-administered method. The study found an 82 percent response rate for paper delivery and 73 percent for online delivery.”

  Perhaps most significant, students tend to prefer this mode of responding, even those who may remain a bit cynical about the confidentiality promise: “When asked if they preferred online or paper delivery, over half the students in the Ewell (2000) study said they preferred online delivery, and over 90 percent of the online students said the mechanism was easy to use. Slightly over half of the students believed that their responses were anonymous, nearly 30 percent did not, and 14 percent were undecided. “

  And this was true even though “Completing an online evaluation also took longer than completing a paper one.” The study points out that the time for completion of the web based survey is “free time” (at least for the instructor not wishing to use class time for the survey, and for the proctor who had to deliver and take back the survey in the classroom administration of a paper evaluation.

- **“Gathering faculty teaching evaluations by in-class and online surveys: their effects on response rates and evaluations”,** Curt J. Dommeyer*, Paul Baum, Robert W. Hanna & Kenneth S. Chapman. *California State University, Northridge. Assessment & Evaluation in Higher Education Vol. 29, No. 5, October 2004*

  Similar to the preceding [NOTE: no one is suggesting a “grade incentive” at Albany. RLC]: “It was found that the response rate to the online survey was generally lower than that to the in-class survey. When a grade incentive was used to encourage response to the online survey, a response rate was achieved that was comparable with that to the in-class survey. Additionally, the study found that online evaluations do not produce significantly different mean evaluation scores than traditional in-class evaluations, even when different incentives are offered to students who are asked to complete online evaluations.”
RESPONSE BIAS

Several of the following also address the “low response rate” concern:

  Instructor attractiveness and personality, ease of the course, ease of grading, the discipline (math and science classes tend to be rated lower than social sciences and humanities) are discussed and there is strong evidence that these do influence student ratings of courses and professors. There is no evidence, however, that these biases are exacerbated if the survey is administered online. Cf. also: *Grade Inflation: A Crisis in College Education*, Valen E. Johnson, 2003, Springer-verlag and Mark A Shapiro’s “Irascible Professor” website, [http://irascibleprofessor.com/comments-06-23-03.htm](http://irascibleprofessor.com/comments-06-23-03.htm)

  Shouping includes the suggestion that response biases could be measured by comparing responses to students’ grades in the course and overall GPA. This would violate confidentiality if students had to identify themselves on paper SIRS forms, of course, but such *massaging* could be done through an online survey *before* the data were turned over to the instructor and department (thus protecting confidentiality, if not anonymity).

  Whereas Felton et al. above begin talking of RateMyProfessor.com but are concerned with institutional assessments, this article cites a study by Michael E. Stonntag of 126 students who rated their professors at Lander University, South Carolina, correlated with student rankings of the same courses and professors on RateMyProfessor.com. Rankings on “ease of courses were consistent,” the rankings correlated on “clarity” and “helpfulness”, and the study concluded this offers “preliminary support for the validity of the evaluations on RateMyProfessors.com.”
  “Sonntag said that there are two ways to read the results: One is to say that RateMyProfessors.com is as good as an educationally devised system and the other would be to say that the latter is as poor as the former. But either way, he suggested, it should give pause to critics to know that the students’ Web site ‘does correlate with a respected tool.’”

  Of numerous blogs and articles questioning whether students are *qualified* to judge instructors, the mode of instruction, the objectives of the course, etc., Martinson suggests (four years before the Stonntag data) that sites such as RateMyProfessor.com are sufficient as an outlet for student opinions.

  This study concludes that students with higher GPA’s are more likely to complete electronic surveys. That was true across all class years, though sophomores were more likely to respond and seniors were the least likely to do so.
This study also concludes that those with higher GPA’s are more likely to respond to web-based evaluations, as were those doing better in the class. Minority status was not a factor. However, gender was: “Interestingly, women were more likely than men to complete the course evaluation process. Women completed about 54% of the potential course evaluations whereas the male students completed only 49% of their assigned course evaluations."

For numerous additional takes concerning those who feel student evaluations in general tend to result in popularity contests with resulting grade inflation and or “dumbing down” of courses, The Chronicle of Higher Education had a “colloquium” asking about both effects and the many faculty responses are interesting. http://chronicle.com/colloquy/98/evaluation/re.htm. Unfortunately, no more recent open call for faculty opinions on this has occurred.

[NOTE: It seems very plausible that the ease of students’ adding open-ended comments, and the transmission of these comments without requiring tedious and time-consuming transcription in departmental offices to protect confidentiality, would tend to provide useful feedback to the professor that may also serve to balance a student’s possibly biased response. A student is qualified to express her or his opinions that a course was graded too hard or easy, that the text and materials seemed woefully out of date, that some inclusions among “required” texts seemed odd particularly if little used, or that the professor—again, in the opinion of a student-- repeatedly used the lectern as a pulpit or platform for matters seemingly unrelated to the course. RLC]

In a review of literature, mostly from the 1990’s, the Suffolk report reaches conclusions that may mitigate or put in perspective the evidence that fewer students doing less well in a class or overall will respond to a web-based than a paper survey (this of course assuming all students who show up for a class when the paper evaluation is given are a captive audience and perforce will respond):

The effect of expected grades or grading leniency is perhaps the most controversial and most researched of the potential biases to student ratings (Arreola, 1995). To the degree that higher grades reflect greater learning, a positive relationship between grades and learning, is appropriate and should be expected. Research on the grading leniency effect indicates that the effect is both weak and unsubstantial (Braskamp & Ory, 1994; Feldman, 1976a; Marsh & Dunkin, 1992; Marsh & Roche, 1997). Most of the correlation between grades and ratings can be accounted for by self-reported student learning (Howard & Maxwell, 1980, 1982), which supports the hypothesis that teaching effectiveness influences both grades and ratings, therefore student ratings are valid. However, other possible hypotheses have been posed to explain this association (Cashin, 1995: Greenwald & Gillmore, 1997): 1) student motivation (general or course specific) influences both learning and ratings, and is controlled for statistically; or 2) students give high ratings in appreciation for lenient grading. Rather than statistical control for possible leniency, Cashin (1995) recommends peer review of the course material, exams, graded samples of essays and projects, etc… to determine grade inflation.

Back in 1987, Aleamoni made the perhaps surprising argument that the bias of student grades isn’t worth worrying about, since not consistently supported in the literature and “since grades are notoriously unreliable anyway.”

Many faculty are convinced that the grades or marks the students receive, or expect to receive, are highly related to their ratings. Much research has been conducted in this area of concern. If we plotted the correlations from these studies, we would see a nice bell-shaped curve, where the mean, median, and mode would be close to zero correlation with a standard deviation of approximately 0.16. This should not be surprising, since grades are notoriously unreliable anyway, and they do not necessarily reflect what the student has actually learned.

• “The Effect of Grades on Student Ratings of Teaching: Implication for Understanding Student Motivation”. Kevin S. Carlson, Centre for Educational Development, Republic Polytechnic, Singapore

Somewhat in contrast to Aleamoni, the Singapore study concludes: “Student self-reports of learning was indexed by one simple item on the student evaluation – ‘I think I can do well in this module.’” Some may argue that this is hardly an index of real learning; however, most of the previous measures of this aspect emphasize achievement and such an item potentially matches such achievement evaluations.”

Further, “The results of the analyses replicate some of the findings of previous researchers. Namely, there is indeed a moderate correlation between grades and ratings of teachers (r = .29, p < .0001); however, controlling for feelings of doing well substantially reduces the effect to practically zero (pr = .12, p < .0001). However, the strength of feelings doing well was quite remarkable. This one item was highly related to teaching ratings (r = .74, p < .0001), and the strength of this relationship was extremely robust when controlling for grades (pr = .71, p < .0001).”


This piece denies that grades (or gender, class size, etc., etc.) need necessarily impact on faculty ratings as long as the questionnaire is “well designed.” Of course the author, a consultant, presumably makes his living showing campuses just how to “properly construct” their questionnaires.

For example, “One very common impression is that student ratings of instructors correspond strongly with the actual or expected grade students receive in that class. Over 400 studies of this issue indicate that no significant correlation exists between grades and properly constructed ratings.”

Further, “Factors such as the gender of the instructor or the rater, the time of day when class is held, a student's major, and an instructor's rank have no impact on properly designed student ratings. Moreover, no consistent relationship exists between class size and evaluation-students do not consistently rate small classes higher than large classes simply because of the lower teacher-to-student ratio.”

**CONCERN THAT “TECHNOLOGY DOESN’T WORK WELL”**

There were generic concerns early on but this seems not a concern in this century. The particular process and system to be used at Albany, of course, was already tested in a pilot this summer.
STUDENTS DEMAND RESULTS; FACULTY FEAR RELEASE OF CERTAIN PORTIONS OF THE RESULTS AND THE SECURITY OF THOSE RESULTS


“After decades of paper-pencil evaluations, Northwestern University implemented a system for online student ratings of instruction in 1999. Since then, students have evaluated instruction online and have obtained a certain portion of the compiled reports through the Web. More portions of the evaluation report are made available online to faculty and administrators.”

[NOTE: Similar systems are to be employed at Albany, with ample opportunity for faculty feedback. The IRPE office can explain at length the processes, security, and how results can be disaggregated to assuage the fears and doubts of faculty and students alike and how the system will quickly provide the appropriate information, and in what form, to different interested parties.


The following is quoted at length as an example to demonstrate that safeguarding the data, protecting student confidentiality, and assuring faculty that information deemed inappropriate is not made public are not “new territory”, unprecedented challenges, etc.

We hope the new online system will improve our response rate from the current 75% to our target of 90%. During our pilot of the online system this summer, we received a response rate of 92%. Although the system provides anonymous feedback, we are able to monitor which students have and have not completed the evaluation and can build incentives for completion. Automatic email reminders sent to students who have yet to fill out their course's survey assist in raising the response rate.

The system functions as follows. Faculty enter their course learning objectives at the beginning of the semester. Students are also sent an email introducing them to the new system at the beginning of the semester. Students are sent a second email near the end of the term (or semester), notifying them that they can access their course evaluation the following morning. The online system is available to them for a period of 8 days to 2 weeks prior to their final exam depending on the length of the course. During the evaluation period, faculty can obtain a list of students who have not completed the evaluation (but not the actual responses of students) and encourage responses. The evaluation is closed on the day before the course final exam. The evaluation results are compiled automatically on the daily basis and summary reports are made available to the faculty and administrators after the course grades are submitted.

For the protection of our data, our servers are backed up nightly and a solid disaster recovery plan is in place in the event of a technical system failure. The online system adds security features that were not in place in the paper system. There is no longer any concern of evaluations being tampered with before delivery to administrative offices and the logistical challenge of receiving all evaluations is eliminated. The new system allows only enrolled students to complete a course survey, and it prohibits multiple evaluations from a single individual. A time-out feature on the system protects students and their work in the event that they walk away from a lab computer while still logged on to their course evaluation. Additional safeguards of the system help to further protect both faculty and students, such as, faculty are able to view their survey results only after grades are released; and the evaluation for a particular class is closed prior to the start of the final exam.