An Online Evaluation Pilot Study in the College of Arts, Humanities & Social Sciences

Richard W. Bruce

Humboldt State University

February 2013

Table of Contents

Introduction

Institutions of higher education have been using end-of-term course student evaluations for most of this century for a variety of purposes (Avery, Bryant, Mathios, Kang & Bell, 2006). Evaluation use at Humboldt State University (HSU) includes personnel decisions (retention and tenure) and to improve teaching. The traditional in-class, paper-and-pencil format has been the mode of evaluation for classes for decades, so this method is familiar to the campus community for getting student feedback on faculty teaching.

Traditionally, evaluations have been conducted at the end of the semester by paper-and-pencil systems where students take part of class time to respond to surveys about an instructor's teaching. Faculty schedule a time for students to evaluate their classes near the end of the semester, and staff compile the blank evaluation forms, distribute or proctor the in-class evaluations, then scan the evaluation forms and create summary reports to return to faculty (after grades are posted to alleviate concerns of retribution of bad grades by a bad evaluation).

With changes in technology, this paper- and time-intensive process is no longer the only way to get feedback on faculty teaching. A random survey of 500 U.S. colleges and universities in 2003 showed 17% of respondents using the internet with another 10% planning to implement internet evaluations, and another 18% reviewing internet options (Hoffman, 2003); Nowell, Gale & Handley (2010) quote a survey in 2005 that reports 33% of U.S. universities conducted evaluations online. Online delivery of teaching evaluations is increasing. "With the development of the Internet, online course evaluations of teaching are gaining momentum in higher education" (Anderson, Cain & Bird, 2005, p. 34; see also Stowell, Addison & Smith, 2012). Because everyone else is doing it is not a valid reason to change; however, because everyone else is doing it, there must be merit in the change. With the new faculty collective bargaining agreement

requiring all classes to have evaluations (Mullery, 2012), HSU has the opportunity, and the need, to join other universities in evaluating courses online, which could result in a cheaper process in terms of staff time and printing costs, and faster turnaround time to meet lecturer evaluations for reappointment.

Online evaluations, however, do have some problems, as evidenced in the literature. Of primary concern is lower response rates, which seem to be a ubiquitous notation (Avery et al., 2006, p. 29; Donovan, Mader & Shinsky, 2006, p. 287; Nowell et al., 2010, p. 466; Stowell et al., 2012, p. 469; Norris & Conn, 2005, p. 16). The result of this permeation, though, is good studies on improving response rates.

Literature Review

Evaluation Costs

Few studies examined the costs of traditional teaching evaluation administration and compared them to online evaluation administration. However, those that did consider costs found that online evaluations saved considerable amounts in materials (paper, printing and, in some cases, mailing costs) and even more in personnel time. Bothell and Henderson (2003) reported cost savings of up to 97% when moving to web-based evaluations, though the data does vary widely (p. 70). One example they used was BYU, with an enrollment of 30,000 students. BYU mailed evaluations to students, a process that cost BYU an estimated $415,503 annually, over four times their estimated online cost of $101,218 (Bothell & Henderson, 2003, p. 76). Kronholm, Wisher, Curnow & Poker (qtd inWang, 2010) found that paper evaluations were 30 times more costly than online evaluations.

For both administration modes (paper and online), the main expenditures are in development and in ongoing costs. Bothell & Henderson (2003) reported that initial set up time

was twice for online evaluations than for paper evaluations, primarily due to technology costs and higher wages of employees who implement the system (p. 76). Ongoing costs for online evaluations include materials for printing reports and personnel costs (staff time) to set up evaluations for electronic distribution to students and printing the reports. For paper evaluations, material costs include printing and copying the evaluation forms and envelopes to keep a class's evaluations together. Staff time for paper evaluations includes time taken for printing the evaluations, creating evaluation packets, scanning the evaluation forms once the students fill them out, and filing completed evaluation packets. In addition, some departments require a staff member proctor evaluations in each class evaluated. Finally, another associated ongoing cost is storage: paper evaluations require secure, confidential storage space while online evaluations, though with the same security issues, are stored electronically and do not require the same real estate that paper forms require.

Other Advantages to Online Evaluations

In addition to cheaper costs, online evaluations offer additional advantages. There is a faster turnaround time for getting results because no scanning time is required. Depending on how and when online evaluations are administered, they can require less-to-no class time and, with week-long periods to fill out evaluations, allow students as much time as they wish whenever they wish to complete (Anderson et al., 2005, p. 34). Students more likely to leave comments (Anderson et al., 2005, p. 35), and are more likely to leave elaborate comments (Venette, Sellnow, & McIntyre, 2010, p. 111).

Response Rates for Online Evaluation Delivery

The big issue is response rates, and for good reason. "Response rates concern all those who undertake survey research because good response rates help ensure the reliability of the

data" (Ballantyne, 2003, p. 106). As such, response rates to online evaluations are reviewed, evaluated, and scrutinized to differing results. Some studies show that online is lower than on paper evaluations (Avery et al., 2006, p. 29; Donovan et al., 2006, p. 287; Nowell et al., 2010, p. 466; Stowell et al., 2012, p. 469; Norris & Conn, 2005, p. 16), particularly in large classes (Avery et al., 2006, p. 34; Wang, 2010, p. 10), while others show it to be similar (Leung & Kember, 2005, p. 578). However, most do not report if the lower result is "statistically significant."

One question to ponder is what is an adequate response rate? Nulty (2008) argues that it depends on the use of the data: a single constructive comment has the potential to improve an instructor's teaching. However, a great concern is to find a "summative appraisal of the performance of the teacher" (Nulty, 2008, p. 306) for faculty retention, tenure and promotion. He argues that there is no good theoretical rule for a good response rate (p. 307), but does offer guidelines for adequate response rates based on population (class enrollment); see Nulty, 2008, for explanation and methodology, such as the assumption that 70% of evaluations rate 4-5. Table 1 is an abbreviated summary (limited to current in-seat class capacities at HSU) of his estimations for an "adequate" response rate.

Table 1

Required Response Rates by Class Size

Class Enrollment	Liberal Conditions [a]		Stringent Conditions [b]	
	Required Responses	Response Rate	Required Responses	Response Rate
10	7	75%	10	100%
20	12	58%	19	97%
30	14	48%	29	96%
40	16	40%	38	95%
50	17	35%	47	93%
60	18	31%	55	92%
70	19	28%	64	91%
80	20	25%	72	90%
90	21	23%	80	88%
100	21	21%	87	87%
150	23	15%	123	82%
200	23	12%	155	77%
250	24	10%	183	73%

[a]10% sampling error; 80% confidence level; 70:30 split
        responses 4-5 vs 1-3 on Likert Scale
[b]3% sampling error; 95% confidence level; 70:30 split
        responses 4-5 vs 1-3 on Likert Scale
Taken from Nulty, 2008

Reasons for Non-Response on Online Evaluations

With so many articles on response rates, few studies have attempted to evaluate why students don't respond to online evaluations. The difficulty in this type of research is apparent: if a student didn't fill out an anonymous teaching evaluation, how does one identify that student and then get that student to fill out a questionnaire on why? Some of this research began with a profile of the student who does fill out evaluations, and these results should not be surprising. Students with high G.P.A, lower class level (sophomores highest, seniors lowest), and discipline (sciences) were the greatest predictors of respondents (Layne, DeCristoforo & McGinty, 1999, p. 228). Avery et al. (2006) found that the higher grade the student expects in the course, if the student is female, and if it's a smaller class, the student was more inclined to reply; there were

also some considerations on race and ethnicity (p. 32-34). While not studying reasons for response, Johnson (2003) discovered in that pilot program that "students who had more than one course included in the pilot study were more likely to rate all their courses" (p. 54), which shows that students enrolled in multiple classes that require evaluation are more likely to respond, and they are more likely to respond to all the evaluations.

Dommeyer, Baum, & Hanna (2002) did succeed in surveying students who did not respond to online evaluations during the spring 2000 semester at CSU Northridge. Of students who didn't fill out online evaluations, 67% forgot/missed the deadline; the second highest reason was computer problems (13%) followed by "inconvenience" (7%) (p. 13).   Johnson performed a telephone survey of nonrespondents to his pilot program and, though the survey size was small (n=6), "forgot" was the most common reason for non-response (p. 54). Another potential reason for nonresponse was shown in Anderson et al. (2005): paradoxically, students' favorite instructors receive the lowest number of responses (p. 34).

Student Scores for Faculty Teaching Evaluations

Another important question after response rates is about the reliability of responses in online evaluations. Are responses similar for online and paper evaluations? Typically, two types of responses are studied: the mean scores on Likert scales and responses to open ended questions. For the former, most studies compare the average scores for online with those on paper. For the latter, there are more opportunities for analysis. First, studies examine the response rate of open ended questions; secondly, the length of comments can be analyzed to see which delivery method garners in longer comments (with the supposition that longer equals more elaborate and constructive). Finally, some studies examine the tenor of those comments, which attempts to interpret if the comments are positive, negative or neutral.

Likert Scores. Studies on mean scores on Likert scales varied just like those for response rates: some studies found that online evaluation means were higher (Avery et al., 2006, p. 30; Carini, Hayek, Kuh, Kennedy & Ouimet, 2003; Donovan, Mader & Shinsky, 2006, p. 287), others that they were the same as in-class (Dommeyer et al., 2004, p. 620; Donovan et al., 2006, p. 290; Stowell et al., 2012, p. 469; Venette et al., 2010, p. 106). Those studies which resulted in similar scores did have some interesting interpretations. Some stated that there were no differences (Nowell et al., 2010, p. 467) or no significant difference overall (Wang, 2010, p. 10).

Open Ended Questions. Overwhelmingly, students are more likely to respond to open ended questions online. Even though comparisons of overall response rates are lower for online evaluations, the students who do respond to an online evaluation are more likely to leave comments than students who respond to paper evaluations. Only Stowell et al. (2012) reported similar percentages responding with comments (p. 469), while most report greater percentages of students leaving responses online to open ended questions (Hardy, 2003, p. 35; Johnson, 2003, p. 54; Layne et al., 1999, p. 229). Donovan et al. (2006) reported a 2:1 ratio and Kasiar, Schroeder & Holstaad (qtd in Donovan et al., 2006, p. 285) reported a 7:1 ratio of students responding to open ended questions online compared to paper evaluations.

Not only do higher percentages of respondents leave comments to open endeded questions, but they also leave more text. Donovan et al. (2006) found responses to be 54% longer online (p. 288) with more words per student (Stowell et al., 2012, p. 469). Some report more elaboration in the student comments (Venette et al., 2010, p. 109). Hardy's (2003) study found that online evaluations, though incurring a lower overall response rate, generated five times more commentary as paper evaluations (p. 35).

Lastly, only a few studies evaluated content of responses. Of those who did, most of the evaluations returned similar results between online and on paper in content, demeanor and topics (Venette et al., 2010, p. 111; Hardy, 2003, p. 36). Stowell et al. had reported similar percentages of paper and online evaluation respondents making comments, but they found that positive-neutral-negative comments skewed slightly favorably to instructors' benefit (p. 469). It appears that from the studies, the content of the evaluations has little change between online and paper evaluations in regard to negative, neutral and positive comments.

Methods

We examine two main components of the online evaluation: costs of using paper evaluations followed by responses to evaluations. In spring 2012, the College of Arts, Humanities & Social Sciences performed a pilot study of online evaluations. At the end of the semester, data was gathered from the sixteen academic departments (Anthropology; Art; Communication; Critical Race, Gender & Sexuality Studies; English; Geography; History; Journalism & Mass Communications; Music; Native American Studies; Philosophy; Politics; Religious Studies; Sociology; Theatre, Film & Dance; World Languages & Cultures), asking how much storage space they need for evaluations, their estimated annual printing costs and the number of hours to perform the staff component of evaluations (set up, print, proctor, sort, scan, verify, and distribute reports to instructors).

The second part of examining evaluations—response rates, rating averages, and responses to open-ended questions—required faculty volunteers. Department chairs were asked to consult with tenured faculty about performing their end of semester teaching evaluations online. Of particular interest were large classes (over 75), general education classes, major classes, and classes where the instructor was teaching multiple sections of the same class. Once

volunteers were identified, staff created online evaluations in ClassClimate, the university teaching evaluation software designed by Scantron. Students registered in the classes received an email with a link to the evaluation site and a unique password that, once input, expired and was no longer usable; students had one chance to log in and evaluate their instructors during a week period. Staff set up varying reminder emails (typically two days apart) that went to students who had not yet submitted an evaluation; these emails had yet again the link to the survey and the same password as before.

After the end of the evaluation period, the data was extracted from the ClassClimate software into spreadsheets for comparison and entry into SPSS. All the Likert ratings were in ClassClimate, but hand written comments had to be hand counted. Online courses included the text from open-ended comments that downloaded directly into a spreadsheet and allowed for easy word count.

Courses that faculty had volunteered to have evaluated were paired with the same course taught by the same instructor from a prior semester (most were from spring 2011). One online evaluation was excluded because the instructor had not had been evaluated for that class previously. All ratings and response rates were evaluated using a paired-samples t-test.

<div align="center">**Results**</div>

Summary of Costs

Costs, as stated previously, were divided into printing/duplicating/materials costs and to staff time. Most staff only reported paper costs or how much the Graphics Department charged to duplicate evaluations. Some of the estimates are lower than actual costs because some departments used leftover blank evaluations from the previous semester, so the baseline for a year's estimates for printing should be higher. One department (Geography/Religious Studies)

had recently changed to allow students to proctor teaching evaluations, reducing time

commitment by half of the staff; some other departments already had this as a practice, while

others continue to require a staff member or faculty member to proctor the evaluations in class.

The reported costs from spring 2012 are shown in Table 2 with adjustments for staff who only

listed one semester's commitment. These amounts do not consider the practice in Distance and

Extended Education.

Table 2
Estimated Annual Costs for Paper Evaluations

Department [a]	Storage (in number of file drawers)	Printing Costs (in dollars)	Staff Hours	Staff Costs (salary + benefits)	Total Costs (not including storage)
Anth, Soc, Phil[b]	7	330	51		
Comm[b]	5	300	36		
CRGS, NAS, WLC[b]	9	200	50		
English	6	248	50		
Geog,RS[b]	15	100	40		
JMC[b]	4	240	36		
Music	5	300	54		
TFD[b]	13	176	20		
Totals	64	$1,894	337	$9,066	$10,960

[a] Some departments that share staff reported a combined amount for those departments.
[b] Anthropology, Philosophy, Sociology (Anth, Phil, Soc); Communications (Comm); Critical Race, Gender & Sexuality Studies (CRGS); Native American Studies (NAS); World Languages & Cultures (WLC); Religious Studies (RS); Journalism and Mass Communications (JMC); Theatre, Film & Dance (TFD)

The College of Arts, Humanities (CAHSS) has sixteen academic departments of varying size;

this report represents thirteen departments, and is a healthy mix of department sizes (Art, Music

and English are some of the largest on campus in terms of instructors and sections while

Philosophy, Religious Studies, and Journalism & Mass Communication are relatively small). The

university hosts thirty-six departments and schools, almost three times those reported in CAHSS,

of comparable sizes. For university-wide considerations, a multiplier of 3 will be used. Based on

that multiplier, the university spent $32,881 in production costs alone in 2011-2012. This does not include "behind the scenes" setup performed by IT staff or, as in the fall 2012 semester, troubleshooting time from a problem loading class rosters into ClassClimate (the latter the result of a problem that would have been avoided using online evaluations).

Staff in Anthropology/Philosophy/Sociology also reported the breakdown of hours for each step of setting up evaluations. Their estimate was that 70% of staff time could be eliminated by moving to online evaluations. With the savings 70% estimate, the annual cost of online evaluations university-wide, assuming $5,682 material cost and salary costs of $27,199 for a total process cost of $32,881, would be $8,160, or a savings of 75% ($24,721).

Response Rates

As expected, response rates were lower for online evaluations. Only one instructor taught two sections of the same course in the same semester; for that course, online evaluations garnered a response rate of 78.95% while the paper evaluation resulted in an 85.19% response rate, a difference of 6.24%. All data below compares the spring 2012 online evaluation for that course to the previous year's course offering (spring 2011) so it aligns with the other courses. The other courses evaluated online (all from spring 2012) were paired with the same course taught by the same instructor the most recent time it was evaluated (most were in spring 2011). Some courses (N=9) had little variation (0.22% less for online) while many online evaluations had a response rate lowered by 4%-10%. Figure 1 shows the paired courses in terms of response rate.

Figure 1. Paired response rates for online and paper evaluations
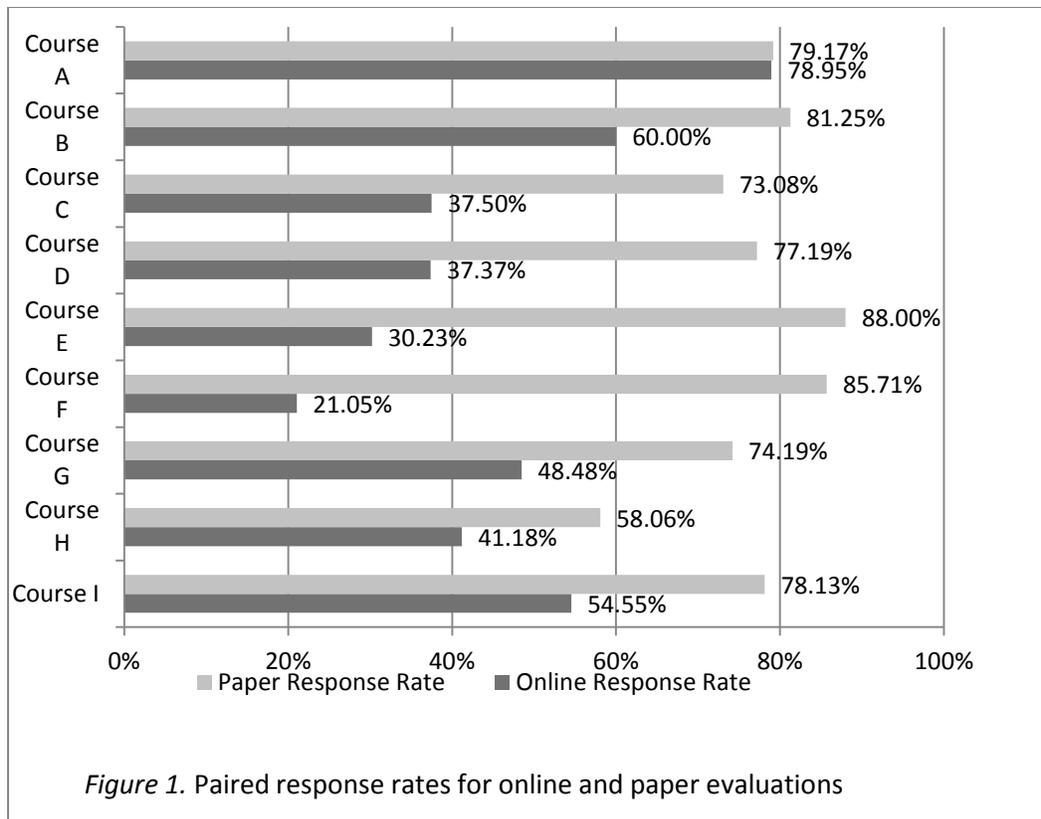
The mean online response rate is 45.47%; the mean paper response rate is 77.86%. An interesting note is the standard deviation, at 17.31 and 9.07 respectively, varies greatly, meaning that there was a wide range of response rates for online evaluation while paper evaluations are mostly consistent.

Running a paired-samples t-test of these courses, there was a significant difference between response rates for online evaluations (M=45.47, s=17.31) and paper evaluations (M=77.87, s=9.07); $t(8)$=-5.084, p=0.001. This states that response rates were not the same for online and paper delivery.

Following Nulty's (2008) recommendations for "adequate" response rates, none of the paper evaluations met the requirements for the stringent conditions; however, all did meet the liberal conditions. Online evaluation response rates were not as fortunate. None attained a

stringent condition response rate, and only four online evaluations garnered the response to meet

his conditions for a liberal response rate based on population (enrollment) size.

Table 3

"Adequate" Response Rate Comparisons for Online Evaluations

Course	Enrollment	Responses	Rate	Liberal/ Conservative/ Neither
Course A	19	15	78.95%	Liberal
Course B	15	9	60.00%	Neither
Course C	24	9	37.50%	Neither
Course D	99	37	37.37%	Liberal
Course E	43	13	30.23%	Neither
Course F	19	4	21.05%	Neither
Course G	33	16	48.48%	Liberal
Course H	34	14	41.18%	Neither
Course I	33	18	54.55%	Liberal

Likert Scores

The university's core evaluation consists of ten questions on a scale of 1-5, with 1 being

low and 5 being high. Departments are able to include additional questions and also open ended

questions. Some departments surveyed had questions that elicit a Likert scale response, but none

of those is included in this data in order to fairly compare courses to between departments.

Overall Online-to-Paper Mean. The first step was to compare the mean of the responses

to all online answers to questions 1-10 for every course and the mean of the responses to all the

paper answers to questions 1-10 for every course. A paired-samples t-test was conducted to

compare the overall mean score for the non-demographic core questions data. A paired-samples

t-test did not show a significant difference ($p=0.312$) between the combined online means

(M=4.19, $s=0.526$) and paper evaluation means (M=4.30, $s=0.341$); $t(8)=-1.078$.

Question-by-Question Online-to-Paper. In Appendix B, we compare all the scores of

all nine courses to the first question online (OnlineQ1) and the first question on paper (PaperQ1)

through the tenth question both online (OnlineQ10) and on paper (PaperQ10). In each case, the mean for all nine classes was lower for the same question on the online evaluation than for the paper evaluation. Online evaluations also had a wider range (lower lows, higher highs) for eight of the ten questions, most up to a full point variation, showing greater fluctuation in high and low responses.

A similar paired-samples t-test of individual question means resulted in similar findings. Comparing all the means for each of the nine courses in the online evaluation and the paper evaluation resulted in only one question being close to, but still failing in, showing a statistically significant difference between the online (M=4.06, s=0.569) and paper (M=4.31, s=0.397) conditions; t(8)=-2.111, p=0.068. There is no significant difference in the mean online scores per question when considering the classes surveyed together.

Course-to-Course Online-to-Paper. At this point, the class-by-class comparison is important in illuminating the difference in individual instructors. The computation of the difference between means of the overall score (sum of all individual question ratings in the class divided by the total number of responses) shows that as many online courses have higher online evaluation ratings, as evidenced by a positive difference, as courses evaluated in the traditional paper method. Table 4 is a summary of the difference in means of paired courses. If the online evaluation was higher than the paper evaluation, the difference is positive; if the paper evaluation overall was higher, the difference is negative. The total of those differences, however, is heavily weighted toward paper evaluation scores (1.273 points on a five point scale, or 25.46% of the rating scale).

Table 4

Difference Between Mean Scores of Teaching Ratings

Course	Difference	Percentage Difference
Course A	-1.105	-22.10%
Course B	0.067	1.34%
Course C	-0.228	-4.56%
Course D	0.146	2.92%
Course E	0.164	3.28%
Course F	-0.159	-3.18%
Course G	-0.115	-2.30%
Course H	0.117	2.34%
Course I	-0.16	-3.20%
Sum of Differences	-1.273	-25.46%
Mean	-0.141	-2.83%
Sum of Differences Without Course A	-0.168	-3.36%
Mean without Course A	-0.021	-0.42%

Course A shows a significant difference in the mean score for the online evaluation (over one point, or a 20% skew toward the paper evaluation), which could simply indicate a "bad" semester and is therefore sampling error. Assuming that such a significant difference is an outlier, the mean difference is -0.021 points, or a 0.42% skew in favor of paper evaluations. Omitting the first course (Course A) from the list, the sum of the differences is -0.168; on a 5 point rating scale, a 3.36% difference.

Open-Ended Questions

The literature review showed that more students comment on online evaluations than on paper evaluations. Because of how this data is collected on paper forms, a complication in collecting this data is in knowing how many students left comments. For paper evaluation reports, each question has a summary of comments that are captured as an image; the report places these images containing text in the students' handwriting in a list with the other responses

to that question. There is no way to tell if a student responded to multiple questions or if multiple students responded to only one question. It is impossible to know which students left comments to the first question and the second question, and the third, etc. Therefore, the number used for students commenting on open-ended questions is the highest count of the total number of students responding to a single question.

Online evaluation reports, however, separate this data. If a student replies to all open-ended questions, there is a response in each question associated to that student (though the responses cannot be tied to the student); if a student replies to only one question, there is data in that one question and not the others. Some departments have a single open-ended question, one department has five. For the department with five questions, the last two questions are simplistic enough that they are often single-word responses ("How well does this course follow the course description?" and "How did the workload required for this course compare to that of similar courses?"); they are omitted from this data so as not to bias results when comparing across departments.

The next table (Table 5) compares response rates to open-ended questions. This data is a calculation of the minimum number of students responding to questions over the number of students responding to the evaluation. For paper evaluations, the number of students could be smaller than actuals because comments aren't associated with single response sheet. A negative difference means a higher percentage of respondents to the paper evaluation; a positive difference means a higher response rate to online evaluations.

Table 5
Open-ended question response rates

Course	Online Comment Rate	Paper Comment Rate	Difference
Course A	73.33%	89.47%	-16.14%
Course B	77.78%	84.62%	-6.84%
Course C	100.00%	78.95%	21.05%
Course D	72.97%	64.77%	8.20%
Course E	84.62%	79.55%	5.07%
Course F	75.00%	83.33%	-8.33%
Course G	37.50%	86.96%	-49.46%
Course H	85.71%	88.89%	-3.17%
Course I	44.44%	92.00%	-47.56%

Overall, fewer students left comments on online evaluations, in some cases by what appears to be a significantly lower amount. However, when running a paired-samples t-test, there is not sufficient evidence that there is a difference between online comment rates overall (M=69.90, s=17.97) and paper comment rates (M=83.17, s=8.18); $t(8)$=-1.785, p=0.112.

The next section of open-ended questions studied was the word count. As response rates to open-ended questions were lower, there is no surprise that there were fewer words to open-ended questions overall in online evaluations. In every case, there were fewer students responding to open ended questions on online evaluations than on paper evaluations (though in some cases still resulting in a higher response rate, per Table 5); however, these students wrote more words per student than students responding to paper evaluations. Running a paired-samples t-test with these means showed a significant difference in the average words per student (p=0.003). The mean words per student for online evaluations (M=52.63, s=20.35) is greater than for paper evaluation (M=38.05, s=15.95); $t(8)$=4.318.

Analysis

Most of the results from this pilot program in CAHSS were not unexpected. Paper evaluations are costly in terms of resources, both materials and staff time, with an expected significant savings if converting to online evaluations. Response rates were lower, as predicted by the literature review, though the significant differences could be attributed to the disciplinary variations found by Layne et al. (1999, p. 228). The Likert scales showed no significant difference even though in looking at the scores when dismissing the outlier, they are generally slightly lower for online evaluations. Comment rates also showed no significant difference, though they were expected to increase per the literature. The average words per student did reflect an expected increase, though perhaps not as large as reported by other studies.

Recommendations for Implementing Online Evaluations

The discipline of the students has an impact on response rate for online evaluations (Layne et al., 1999, p. 228), so it would follow that response rates would vary among departments as disciplines as well as among colleges. One suggestion is to review the response rates, mean scores, comment rates, and average words per student in the other two colleges' online evaluations to see if these conditions hold, as well as to get an overall view of the university. This data likely already exists.

Another recommendation is to have more controlled conditions for analyzing response rates: a class that receives email reminders every two days for online evaluations, a class that receives emails and reminders from the instructor, and scheduling one class period in a computer lab to perform evaluations and do other class work. Other important considerations would be to follow the suggestions of increasing grades by a small percentage or some other sort of prize/reward.

If evaluation rates and responses are truly of concern, then the university should attempt as many methods to improve online response rates as much as possible; that is where we stand now, looking at the potential of evaluating all 5,500 class sections per year with the new CBA instead of our present estimated 2,500, a significant workload increase for staff of more than double.

The overall evaluation procedure should be the same across the university, however, to provide consistency for students. As a liberal arts institution that requires students to take classes from many disciplines, students may get inconsistent procedures with the present decentralized system. Sending a pre-notification email from the president to all students and integrating evaluations onto the student portal would increase visibility, create a culture change, and provide consistency for students. The faculty body demonstrating their support for the evaluation process has shown to be a necessary component of response rates, and this, too, would have to be consistent across the university from all faculty. Additionally, there are additive effects for using multiple methods to encourage a greater response rate, including consistency, when implemented over a period of time. (Appendix A has a list of strategies on improving response rates.)

Norris & Conn (2005) summarize their observations from the literature about online evaluations quite well: "These benefits [of online course evaluations] include time and cost savings, quick return of responses, longer and more in-depth responses to open-ended questions, and simple convenience for participants who are becoming increasingly technologically literate" (p. 14). For the CAHSS study, the response rates decreased, faculty ratings remained not significantly different in spite of the lower response rate, and the average words per student in response to open-ended questions increased. With persistent effort, response rates should increase to familiar levels. As with many processes and change in culture, it would have been

best to change to online evaluations ten years ago. This would have improved response rates by

having trained students to consistently respond to online evaluations, created a culture of faculty

demonstrating support for students to respond, and given us the opportunity to find out which

methods work best. As a campus, we would also have become accustom to the response rates,

evaluation means, and comments on open-ended questions. Instead, we are now forced to

strenuously consider this change as a response to a significantly increased, and unfunded,

workload resulting from the new faculty CBA.

References

Anderson, H.M., Cain, J., & Bird, E. (2005, January 31). Online student course evaluations:

Review of literature and a pilot study. *American Journal of Pharmaceutical Education,*

69(1), 34-43. Retrieved from http://www.ajpe.org/doi/pdf/10.5688/aj690105

Avery, R.J, Bryant, W.K., Mathios, A., Kang, H., & Bell, D. (2006 Winter). Electronic course

evaluations: Does an online delivery system influence student evaluations? *Journal of*

Economic Education, 37(1), 21-37.

Ballantyne, C. (2003, December 15). Online evaluations of teaching: An examination of current

practice and considerations of the future. *New Directions for Teaching and Learning,*

2003(96), 103-112. doi: 10.1002/tl.127

Bothell, T., & Henderson, T. (2003). Do online ratings of instruction make $ense?. *New*

Directions for Teaching and Learning, *2003*(96), 69-79.

Carini, R.M., Hayek, J.C., Kuh, G.D., Kennedy, J.M., & Ouimet, J.A. (2003, February). College

student responses to web and paper surveys: Does mode matter? *Research in Higher*

Education, 44(1), 1-19.

Dommeyer, C.J., Baum, P., & Hanna, R.W. (2002). College students' attitudes toward methods

of collecting teaching evaluations: In-class versus on-line. *Journal of Education for*

Business, *78*(1), 11-15.

Dommeyer, C.J., Baum, P., Hanna, R.W., & Chapman, K. (2004, October). Gathering faculty

teaching evaluations by in-class and online surveys: their effects on response rates and

evaluations. *Assessment & Evaluation in Higher Education, 29*(5), 611-623. doi:

10.1080/02602930410001689171

Donovan, J., Mader, C.E., & Shinsky, J. (2006, Winter). Constructive student feedback: Online

    vs. traditional course evaluations. *Journal of Interactive Online Learning, 5*(3), 283-296.

    Retrieved from http://www.ncolr.org/jiol/issues/pdf/5.3.5.pdf

Hardy, N. (2003). Online ratings: Fact and fiction. *New Directions for Teaching and Learning*,

    2003(96), 31-38.

Hoffman, K. (2003, Winter). Online course evaluation and reporting in higher education. *New*

    Directions for Teaching and Learning, 2003(96), 25-29.

Johnson, T.D. (2003). Online student ratings: Will students respond?. *New Directions for*

    Teaching and Learning, *2003*(96), 49-59.

Joint Council of Deans and Faculty Leaders Ad Hoc Committee (2002, September 14).

    Recommendations to the Joint Council on the future of optical scanning of course

    evaluations and grade reports.

Layne, B.H., DeCristoforo, J.R., & McGinty, D. (1999). Electronic versus traditional student

    ratings of instruction. *Research in Higher Education*, 40, 221-232.

Leung, D.Y.P., & Kember, D. (2005, August). Comparability of data gathered from evaluation

    questionnaires on paper and through the internet. *Research in Higher Education, 46*(5),

    571-591. doi: 10.1007/s11162-005-3365-3

Mullery, C. (2012). Summary of selected changes to 2012-2014 faculty collective bargaining

    agreement Especially relevant to the work of department chairs. Retrieved from

    http://humboldt.edu/aps/docs/Policies/CBA_Changes_2012.pdf

Norris, J. & Conn, C. (2005). Investigating strategies for increasing student response rates to

    online-delivered course evaluations. *The Quarterly Review of Distance Education, 6*(1),

    13-29.

Nowell, C., Gale, L., & Handley, B. (2010). Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment & Evaluation in Higher Education*, *35*(4), 463-475.

Nulty, D.D. (2008, June). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education, 33*(3), 301-314.

Stowell, J.R., Addison, W.E., & Smith, J.L. (2012). Comparison of online and classroom-based student evaluations of instruction. *Assessment & Evaluation in Higher Education, 37*(4), 465-473. doi: 10.1080/02602938.2010.545869

University Senate. (2004, Mar 9). Resolution on On-Line Course Evaluations (#13-03/04-EP).

Venette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: Assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education*, *35*(1), 97-111.

Wang, J. (2010, Fall). Pilot study for assessing the viability of using online course evaluations at California State University Sacramento: A report from the electronic Course Evaluation Task Force. Retrieved from

http://www.csus.edu/acse/archive/1011/ECETF%20Distributed%20Report.docx

Appendix A

Strategies for Increasing Response Rates

If increasing response rates is still a priority, Ballantyne (2003) reports that "some high response rates have been reported … for individual campuses and individual courses; by using appropriate strategies, it *is* possible to obtain good response rates" (emphasis in the original) (p. 110). There are many particular things to implement, which will be described below, but the most important include having faculty support and consistency. Ballantyne (2003) reported that Goodman and Campbell (1999) found that response rates to online evaluations dropped in the second semester of their study after a dip in faculty support. BYU represents a success where response rates in the fifth year of online evaluation implementation reached 70% (Johnson, 2003). Norris and Conn (2005), though not a particularly encouraging note, decided that with effective strategies, the highest response rates for online evaluations were nearly the same as the lowest response rates for in-seat evaluations (p. 25). Many reported that over time, response rates increased as culture changed (Avery et al., 2006 p. 36; Nully, 2008, p. 304).

Since response rates seem to be such an important aspect of online evaluations, there is an enormous corpus of literature. Suggestions for increasing response rates for online evaluations can be summed in three general ways: procedural and logistical recommendations to elicit more responses, punishment/rewards to alter behavior, and explaining the purpose of evaluations to garner student support.

Finally, incorporating multiple recommendations has an additive advantage (Nulty, 2008, p. 305). According to the study by Norris and Conn (2005) of online courses, two or more postings (in a syllabus, email reminders) resulted in higher response rates (p. 18). Wang (2010) had found that "when students receive e-mail messages with a link to conduct their teaching

evaluations, are reminded on several occasions by their teachers to complete the teaching evaluation, are informed by the faculty member that their teaching evaluation feedback is valuable to the teaching process, and are reminded by e-mail to complete the teaching evaluation, that the response rate nears those of paper and pencil rates." (p. 4). Multiple approaches and reminders seemed to have a significant impact, particularly when coming from instructors.

Procedural and Logistical Recommendations. These recommendations begin when the instructors begin planning their syllabi. Norris and Conn (2005) suggest announcing class evaluation in multiple places. During their study on when the evaluation was announced, timing played an important part. On the syllabus at the beginning of the class only resulted in a 25% response rate; announcements 2-3 weeks before the evaluation begins (also a recommendation from Anderson et al., 2005) increased the response rate to 48%; the week of the evaluation (finals week during his study) lowered the response rate to 44%, presumably because of the intensity and stress of finals week. Posting on a class discussion board had the greatest single response rate of 50%.

Other recommendations that impact planning come from Ballantyne (2003), which is to increase access to internet/computers (p. 107). Faculty can directly impact this by scheduling a computer lab for a class meeting that allows students to fill out the evaluations (Johnson, 2003, p. 55; Joint Council of Deans and Faculty Leaders Ad Hoc Committee, 2002, p. 2; Nulty, 2008, p. 306); however, this negates the class time savings by offering the convenience of online evaluations. Holding the evaluation period in a computer lab does offer the opportunity to give an in-class demonstration of how to fill out evaluations, another recommendation to enormously improve response rates (Dommeyer, 2003, qtd in Anderson et al., 2005, p36; Johnson, 2003, p. 55).

For students who haven't responded to an evaluation, repeat reminder emails increase participation and thereby increase response rates (Dommeyer et al., 2004, p. 621; Norris & Conn, 2005, p. 26; Nulty, 2008, p.303; Stowell et al., 2012, p. 469). This practice appears to be well understood in surveying and common in online evaluation, so it hardly merits mention.

The most important recommendation is to have faculty demonstrate support for online evaluations. Faculty have influence over students, and if their attitudes are dismissive of evaluations, it will be reflected in response rates. It is important to have faculty support and promote the online evaluation process.

Finally, it should be noted that students are functionally an online generation. For their time in academia, the entire process is online. Students apply to the university online. They register online. They access their grades are online. They perform library research online. They email their instructors. Many of their classes have an online component. "And while students may find it somewhat inconvenient to evaluate faculty online, they should have little difficulty understanding the online procedures, since students, by now, are quite comfortable with the Internet and web sites" (Dommeyer et al., 2004, p. 620). Students in general thrive in an online environment, so it is not surprising that they would be comfortable evaluating their teachers online. In fact, many may prefer it.

Punishment or Reward Recommendations. Though there are ethical issues involved with rewarding or punishing students for filling out evaluations on faculty (see Dommeyer et al., 2004), these methods can be very effective. Many studies show that giving incentives (or rewards) as the most effective method to improve response rates (Anderson et al., 2005, p34; Johnson, 2003, p. 55; Nulty, 2008, p. 303).

Incentives and rewards are not necessarily grade related, but they do seem to be some of the most popular (Ballantyne, 2003, 107; Johnson, 2003, 55). Nulty (2008) found that "[w]hen grade incentive [as low as a quarter of 1%] was applied, the response rates for both online and on-paper surveys were high—and almost identical (86.67% and 86.99% respectively)" (p. 303). Dommeyer (2003, qtd in Anderson et al., 2005) suggested a fourth of a percent grade increase (p. 36).

Dommeyer found that either the grade increase resulted in the greatest response rate bump or an early grade reporting did. Other incentives (like early grade reporting) can be effective (Ballantyne, 2003, p. 107), and they can be performed differently. Early grade reporting can done on an individual level (instructors tell students their grades before they're officially posted on university grading reports) or at the course level by the instructor telling students their grades early if the class achieves a 67% response rate (Anderson et al., 2005, p. 36).

Other incentives that increase response rate are drawings or lotteries for prizes (Dommeyer et al., 2004, p. 621; Nulty, 2008, p.303) and earlier registration times (Layne et al., 1999, p. 230). These can be cash prizes, gift certificates, reserved parking spaces, entering student participantes in a drawing for a cash prize led to incremental rises (Ballantyne, 2003, p. 105)

Punishments are also often considered to improve response rates. The most common is to count the evaluation as an assignment (Johnson, qtd in Anderson et al., 2005, p. 36; Joint Council of Deans and Faculty Leaders Ad Hoc Committee, 2002, p. 2), as opposed to offering extra credit for completion. A related suggestion, which came from a student questionnaire, was to withhold early access to grades until students at least log in to the system, allowing them to

opt out of the evaluation (Johnson, 2003, p. 55) or assigning a grade of "I" until completed (Anderson et al., 2005, p. 41).

        Explaining Purposes of Evaluations. The third main set of recommendations is the faculty expressing how important evaluations are for improving teaching. When students see that their opinions are being heard, an extremely important interaction for the digital generation, they are more inclined to respond to evaluations. Students must understand how their voice impacts teaching and retention. Faculty must explain the value of the process and student feedback (Johnson, qtd in Anderson et al., 2005, p. 36; Norris & Conn, 2005, p. 26) and explain how faculty use that feedback to improve teaching and to influence tenure and retention. Johnson (2003) believes that instructors must show a personal interest in students completing the rating forms by mentioning evaluations, letting students know they're paying attention to responses, and sending emails to students as reminders (p. 55). Faculty support of the process and explanation of how this information is used is paramount.

Appendix B

One-Sample Statistics for Question-to-Question Comparisons

	N	Mean	Minimum	Maximum	Std. Deviation	Std. Error Mean
OnlineQ1	9	3.9863	2.47	4.50	.63032	.21011
PaperQ1	9	4.2122	3.55	4.73	.33647	.11216
OnlineQ2	9	4.2369	3.13	5.00	.52894	.17631
PaperQ2	9	4.2744	3.92	4.59	.22095	.07365
OnlineQ3	9	4.0327	2.87	4.54	.51382	.17127
PaperQ3	9	4.0709	3.26	4.63	.38800	.12933
OnlineQ4	9	3.8940	1.93	4.54	.78390	.26130
PaperQ4	9	4.1531	3.44	4.75	.45412	.15137
OnlineQ5	9	4.1602	2.93	5.00	.59057	.19686
PaperQ5	9	4.3909	3.77	4.74	.30980	.10327
OnlineQ6	9	4.2266	2.80	4.75	.59150	.19717
PaperQ6	9	4.5297	4.00	5.00	.35712	.11904
OnlineQ7	9	4.4379	3.53	5.00	.44138	.14713
PaperQ7	9	4.6172	4.29	4.83	.17270	.05757
OnlineQ8	9	4.4410	3.53	4.93	.42752	.14251
PaperQ8	9	4.5768	3.82	4.88	.35426	.11809
OnlineQ9	9	4.0632	2.87	4.78	.56928	.18976
PaperQ9	9	4.3093	1.93	4.44	.39688	.13229
OnlineQ10	9	3.8859	3.74	4.89	.77151	.25717
PaperQ10	9	4.2113	3.47	4.70	.48183	.16061

Appendix C

Open-Ended Question Average Words per Student and Ranges

Course	Evaluation Type	Mean Words per student	Range		
Course A	Online	40.55	14	-	50
	Paper	27.24	6	-	86
	Difference	13.31	8	-	-36
Course B	Online	32.29	12	-	47
	Paper	27.64	4	-	52
	Difference	4.65	8	-	-5
Course C	Online	32.33	1	-	50
	Paper	27.67	1	-	75
	Difference	4.67	0	-	-25
Course D	Online	49.11	2	-	51
	Paper	33.58	1	-	63
	Difference	15.53	1	-	-12
Course E	Online	47.09	1	-	47
	Paper	28.03	2	-	47
	Difference	19.06	-1	-	0
Course F	Online	39.33	33	-	45
	Paper	22.70	6	-	29
	Difference	16.63	27	-	16
Course G	Online	79.50	2	-	82
	Paper	63.25	6	-	51
	Difference	16.25	-4	-	31
Course H	Online	65.83	6	-	49
	Paper	61.44	5	-	56
	Difference	4.40	1	-	-7
Course I	Online	87.63	9	-	52
	Paper	50.91	1	-	58
	Difference	36.71	8	-	-6