

End of Course Evaluation Analysis

**Daniel Novak
For Marcie Bober
EdTec 798, Spring 2007**

The Purpose

This study focused on surveys that students in the College of Professional Studies and Fine Arts (PSFA) complete at the end of each semester. Specifically, the researcher

- determined whether or not delivery method (print, online) affected survey response rates and
- assessed the impact of delivery method (print, online) on mean instructor ratings.

This second bullet point is of great importance to PSFA faculty since mean ratings on end-of-course surveys factor into the RTP process.

Literature Review

Current research on end of course evaluations focuses on a number of demographic and administrative issues. However, relatively few studies examine the relationship between evaluations and their medium of transmission.

A study performed in 2005 by Roesmary Avery, Bryant, et al., provides the most complete examination of the effects of online and print evaluations on the mean scores of instructors. Using a sample size of 29 courses (including 973 students) in the Department of Policy Analysis and Management at Cornell University, *Electronic Course Evaluations: Does an Online Delivery System Influence Student Evaluations* reports that response rates for web delivery were lower in each of the courses where data was available for both media. The researchers note that demographic factors such as gender and race/ethnicity, as well as factors such as grade expectation and class size, create a statistically significant relationship when compared to the response rates of web-based evaluations.

In the aggregate, however, the researchers assert that the implementation of web-based end of course evaluations leads to lower response rates in the first semester of implementation. The researchers did note, though, that each semester thereafter led to higher response rates, until the web-based response rates approximated those produced by paper evaluations.

It is important to note in examining the study's findings that the sample included far fewer web-based evaluations than paper-based evaluations. Further, the study's sample dated from 1998-2001. Great strides were made in the availability of Internet connectivity and computer availability during the 2001-2004 period. These issues may account for some of the study's findings regarding the low response rates associated with online delivery.

On the surface, the administrative circumstances under which this study was conducted reflects those of the PSFA. Researchers dealt with multiple forms that used similar Likert scales, and had to arrange their data accordingly. Unlike the PSFA, the researchers did not appear to deploy their survey over a single, uniform online delivery system such as SDSU's WebPortal. The confusion attendant with this fact may account for the low response rates encountered by the researchers.

The report also notes that mean course evaluation scores were not significantly different across media, even when disaggregated by the questions contained within the forms. 80.3% of their t-tests returned a lack of statistical significance between online and print delivery (Rosemary, et al., 31). Of the 104 t-values included in their report, 51% revealed greater means from paper surveys, while 49% indicated greater means from online delivery (31). This led the researchers to conclude that "...that there were not systematic differences between the Web and the paper systems." (14)

Another study conducted by researchers at Georgia State University also concluded that mean evaluation scores were not affected by delivery method. *Electronic Versus Traditional Ratings of Instruction* (1999) analyzed 2453 undergraduate and graduate evaluations across 74 courses (Layne, DeCristoforo, and McGinty, 224) and found no significant differences between the means generated by the delivery methods. In fact, researchers determined that the area of study had a greater effect on mean evaluation scores than did delivery method.

The study also identified a number of other findings relating to online evaluation. For example, the researchers found that students felt comfortable with the level of anonymity provided by the online evaluations. Equally, researchers were unable to trace an individual evaluation to an individual student. This finding helps to dispel the potential problem of student distrust of online delivery.

The researchers also found that students felt a greater degree of satisfaction towards online evaluations than towards paper-based evaluations. This indicates that students may be more likely to complete online evaluations, especially when offered an incentive. SDSU's WebPortal currently offers such an incentive in allowing students to view their grades earlier after they have completed the end of course evaluations.

Overall, both studies found that mean evaluation scores were not affected by delivery method. However, they also agreed that online response rates were lower if 1) students do not have access to computers, 2) students do not know how to access and use the online evaluation system, and 3) students do not have an incentive to fill out the evaluations.

The Sample

The PSFA uses five different end-of-course survey forms; each is associated with specific classes. Common to the forms is the response set for scaled items (5=Strongly Agree; 1=Strongly Disagree) and the constructs generally measured. Specific item wording does vary—and some forms feature questions that are somewhat demographic in nature.

For this study, only scaled items were analyzed. As important, the researchers did not focus on either face or content reliability; they assumed the forms adequately measure what they purport to measure.

In order to gather an appropriate sample, Dr. Kathryn LaMaster (Associate Dean of the College) drew from courses where data had been continuously collected over a two-year

period: Fall 2004 and Spring 2005 (print), and Fall 2005 and Spring 2006 (online). Each four-evaluation cycle set was associated with a single professor/instructor. To the extent possible, Dr. LaMaster compiled a stratified sample that proportionally represented each form's actual use (by percentage and by class type). The end result was a sample of 44 courses with 176 evaluation cycles, and a total of 5972 individual responses.

Methodology

Using the reports provided by the PSFA, researchers ran a series of one-way Anova tests that examined the data from a progressively precise perspective. The first section examines the basic response rate information culled from the evaluation summaries. The second section examines the means produced by these evaluations in the aggregate, by semester, by form and semester, by delivery method, and by form and delivery. This systematic approach to the data analysis helped to account for potential unforeseen variables.

Response Rates

Table 1 - Response Rates

Form Number	N (student responses)	N (courses)	N (sets of responses)	Response Rate (aggregate)	Response Rate (Print)	Response Rate (Online)
Form 1	174	2	8	70%	77%	62%
Form 2	1425	13	52	78%	76%	80%
Form 3	875	7	28	78%	72%	85%
Form 4	1054	10	40	81%	77%	85%
Form 5	2444	12	48	76%	67%	84%
Total	5972	44	176	76.6%	73%	82%

Table 2 contains the answer to the first research question. The response rates listed above appear as both aggregates (the mean response rates for both delivery methods) and disaggregated by delivery method (print and online) and Form. The response rates for online evaluations are higher than their print counterparts (73% versus 82%), except in the case of the smallest sample (Form 1, at 62%). It is important to note that above number of student responses may include duplicates from students taking multiple courses in their departments over the length of the sample period. This, however, should not affect the mean results of the evaluations that we will discuss later.

Overall, the online evaluations drew a 9% higher response rate than print forms. This indicates that students feel comfortable using the online evaluation forms, and that the medium of evaluation does not negatively affect the response rate.

Evaluations Means and Delivery Method

This section examines the effects of the delivery method on evaluation forms from multiple perspectives. The results of the broadest of the One-Way Anova tests, the test of the aggregated means, appears below.

Table 1 - Aggregated Means by Form

Form Number	Mean	Min	Max	Std. Deviation	F	Sig
Form 1	4.260	3.97	4.63	.19900		
Form 2	4.475	3.58	4.91	.31033		
Form 3	4.217	3.40	4.86	.40487		
Form 4	4.040	3.43	4.78	.38651		
Form 5	4.102	3.38	4.89	.37285		
Total	4.224	3.38	4.91	.39529	10.45	.000

As Table 1 and the Post-Hoc analysis (Appendix A) show, the mean scores produced by four of the forms differ in a statistically significant way when researchers applied a One Way ANOVA test with a Tukey Post-Hoc analysis. The Post-Hoc analysis did reveal that the means in Form 2 demonstrate a $>.05$ level of significance when compared to Forms 3, 4, and 5. In general, the PSFA accounts for these differences in means by comparing instructors and faculty against other instructors who were evaluated using the same form, then comparing these results to the rest of the college.

Comparison by Semester

To evaluate the possibility of erroneous results caused by the statistical differences between semesters, the researchers used a One-Way Anova test to determine if there differences between the evaluations submitted via print (Fall 2004 and Spring 2005) and online (Fall 2005 and Spring 2006). The test generated the results contained in Table 3.

Table 2 - Results by Semester

Semester	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Fall 04	44	4.2436	.40451	3.40	4.88		
Spring 05	44	4.3451	.41883	3.28	4.92		
Fall 05	44	4.2241	.36804	3.43	4.91		
Spring 06	44	4.2513	.32733	3.40	4.79		
Total	176	4.2660	.38093	3.28	4.92	.879	.453

As the table shows, the Anova test did not find a significant difference amongst the course means when they are dissected by semester. This indicates that any differences in means cannot be accounted for by the differences amongst semesters.

Comparison by Form, by Semester

Researchers determined that a further test by form and semester was necessary to determine if any significant relationships existed amongst the mean evaluation scores for each semester at the form level. Researchers used a series of One-Way Anova tests to generate the data contained in the following tables.

Table 3 - Results by Semester, Form 1

Semester	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Fall 04	2	4.4169	.30140	4.20	4.63		
Spring 05	2	4.3728	.04640	4.34	4.41		
Fall 05	2	4.1478	.00751	4.14	4.15		
Spring 06	2	4.1100	.19799	3.97	4.25		
Total	8	4.2619	.19895	3.97	4.63	1.460	.352

Table 4 - Results by Semester, Form 2

Semester	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Fall 04	13	4.4001	.30678	3.83	4.85		
Spring 05	13	4.6212	.20266	4.26	4.91		
Fall 05	13	4.5278	.32873	3.86	4.91		
Spring 06	13	4.3191	.33425	3.58	4.73		
Total	52	4.4671	.31185	3.58	4.91	2.629	.061

Table 5 - Results by Semester, Form 3

Semester	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Fall 04	7	4.2905	.49407	3.40	4.86		
Spring 05	7	3.9720	.50534	3.28	4.82		
Fall 05	7	4.0484	.29686	3.80	4.49		
Spring 06	7	4.2523	.27600	3.85	4.61		
Total	28	4.1408	.40760	3.28	4.86	1.010	.405

Table 6 - Results by Semester, Form 4

Semester	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Fall 04	10	4.2104	.38261	3.51	4.88		
Spring 05	10	4.5556	.31766	4.00	4.92		
Fall 05	10	4.1948	.41215	3.43	4.65		
Spring 06	10	4.3788	.27274	3.97	4.79		
Total	40	4.3099	.35501	3.43	4.92	1.332	.279

Table 7 - Results by Semester, Form 5

Semester	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Fall 04	12	4.0454	.44285	3.50	4.67		
Spring 05	12	4.1667	.44072	3.51	4.89		
Fall 05	12	4.0349	.23461	3.65	4.41		
Spring 06	12	4.0946	.37465	3.40	4.77		
Total	48	4.0854	.37398	3.40	4.89	.296	.828

As shown in the tables above, none of the data return a statistically significant result at the $\alpha > .05$ level when disaggregated by evaluation Form and semester.

Comparison by Aggregated Delivery Method

To answer the second research question, researchers examined the relationship between delivery methods to determine if the delivery medium affects students end of course evaluations. Table 4 contains the results of a One-Way Anova test of the means of the evaluation scores based on the delivery method.

Table 8 - Aggregated Delivery Method

Delivery Method	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Print	88	4.3152	.38845	3.40	4.91		
Online	88	4.2414	.34872	3.40	4.91		
Total	176	4.2783	.36992	3.40	4.91	1.761	.186

The significance level of .186 generated by the Anova test indicates that the delivery method (in the aggregate) does not affect the mean scores collected by the end of course evaluation.

Comparison by Delivery Method, Disaggregated by Form

Though the previous test confirmed that the aggregated means of the forms did not constitute a significant degree of difference when compared by delivery method, researchers determined that a second test was necessary. This concern stemmed from the different means produced by each form, and the fact that the PSFA never compares the results of one evaluation form to another. As such, researchers performed a One-Way Anova test on the means of each Form (1-5). The results are contained in Tables 5 through 9.

Table 9 - Form 1, by Delivery Method

Delivery Method	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Print	4	4.3948	.17789	4.20	4.63		
Online	4	4.1289	.11646	3.97	4.25		
Total	8	4.2619	.19895	3.97	4.63	6.257	.046

Though Form 1's level of significance registers above the .05 cutoff, this could be attributed to the small size of the sample. Considering this fact, and the .004 difference, the researchers hazard to say that there is no significant difference between the mean scores of the online and print evaluations.

Table 10 - Form 2, by Delivery Method

Delivery Method	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Print	26	4.5107	.27857	3.83	4.91		
Online	26	4.4234	.34180	3.58	4.91		
Total	52	4.4671	.31185	3.58	4.91	1.019	.318

Unlike Form 1, Form 2 covers a greater number of courses. It does not, however, indicate a significant difference between the two delivery methods, with a p value of .318.

Table 11 - Form 3, by Delivery Method

Delivery Method	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Print	14	4.1313	.50778	3.28	4.86		
Online	14	4.1504	.29500	3.80	4.61		
Total	28	4.1408	.40760	3.28	4.86	.015	.904

Form 3 also registers a degree of significance above the .05 threshold ($p = .904$). This indicates that there is no significant difference between the means produced by either delivery method using Form 3.

Table 12 - Form 4, by Delivery Method

Delivery Method	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Print	20	4.3330	.36464	3.51	4.92		
Online	20	4.2868	.35300	3.43	4.79		
Total	40	4.3099	.35501	3.43	4.92	.166	.828

As in Forms 2 and 3, the degree of significance between the means generated by the end of course evaluations for Form 4 are not statistically significant when broken down by delivery method ($p = .828$).

Table 13 - Form 5, by Delivery Method

Delivery Method	N	Mean	Std. Deviation	Minimum	Maximum	F	Significance
Print	24	4.1061	.43649	3.50	4.89		
Online	24	4.0648	.30722	3.40	4.77		
Total	48	4.0854	.37398	3.50	4.89	.144	.706

Since four of the five forms return a value greater than .05, one can see that the medium of delivery does not have a significant impact on the mean scores. After having examined the evaluations by semester, form, and delivery method, it appears that there is no significant difference between the means produced by print evaluations and online evaluations.

Conclusion

After analyzing the relevant data, the researchers feel confident in making the following assertions:

1. Online end of course evaluations do not lead to lower student response rates. In fact, the data available from the PSFA consistently shows that students are more likely to respond to end of course evaluations through the Portal than in class.

The switch to online surveys has a number of potential benefits for students, faculty, and administration. Instructors can use the class time associated with paper evaluations for reviewing course material before the final exam. These reviews could in turn help students study more effectively for their finals. University administration will save both money (in terms of the costs associated with scantrons, photocopying, and delivering the evaluations) and human resources (in the form of the time spent scanning, converting, entering, and organizing the data).

2. Online end of course evaluations do not lead to lower mean evaluation scores for faculty. This study did not find any significant differences between the scores produced by print evaluations and those produced by online evaluations.

Faculty need not worry that their mean evaluation scores will decline after the adoption of online evaluations, as this study consistently found otherwise. Faculty may also discover a number of secondary effects from the switch. While this study did not examine the effect of online evaluations on qualitative questions, there is reason to believe that students will offer better feedback if they are able to type their responses. Further, faculty may receive their evaluations more quickly if the testing office does not have to retype or otherwise capture and store written comments.

Further, online surveys may provide a number of benefits regarding the implementation of end of course evaluations. Online surveys offer greater security and anonymity (as noted above). They also reduce the possibility of faculty influence on the final results. In addition, they offer consistency in implementation, which is essential to generating meaningful evaluation results.

In conclusion, online evaluations hold the promise of streamlining the efficiency of the evaluation process without lowering either response rates or mean faculty scores.

Sources

Avery, R.J., Bryant, W.K., Mathios, A., Kang, H., and Bell, D.. (Winter 2006). Electronic Course Evaluations: Does an Online Delivery System Influence Student Evaluations? (*Avery 2006*), Volume 37, no.1, 21.

Layne, B.H., DeCristoforo, J.R., and McGinty, D.. (1999). Electronic Versus Traditional Student Ratings of Instruction. *Research in Higher Education*, Volume 40, no. 2, 221.

Appendix A – Inter-form Tukey Post-Hoc Test

Survey Form (I)	Survey Form (J)	Mean Difference (I-J)	Significance
Form 1	Form 2	-.21577	.509
	Form 3	.04286	.998
	Form 4	.21975	.510
	Form 5	.15771	.778
Form 2	Form 1	.21577	.509
	Form 3	.25863*	.020
	Form 4	.43552*	.000
	Form 5	.37346*	.000
Form 3	Form 1	-.04286	.998
	Form 2	-.25863*	.020
	Form 4	.17689	.269
	Form 5	.11485	.662
Form 4	Form 1	-.21975	.510
	Form 2	-.43552*	.000
	Form 3	-.17689	.269
	Form 5	-.06204	.928
Form 5	Form 1	-.15711	.778
	Form 2	-.37348*	.000
	Form 3	-.11485	.662
	Form 4	.06204	.928
* indicates the mean difference is significant at the .05 level			