

Online And Paper Course Evaluations

Faruk Guder, Loyola University Chicago, USA
Mary Malliaris, Loyola University Chicago, USA

ABSTRACT

The purpose of this study is to compare the results of paper and online evaluations. The following analysis examines data from six departments of the School of Business Administration during a programmed switch from paper to online evaluations. The courses that participated in this study were divided and compared in the following manner: advanced and core classes, large and small sections, and courses taught by full-time and part-time faculty. The data was collected over a one-year period and contrasts the Spring 2008 and 2009 semesters, during which a total of 4,424 evaluations were reviewed. In addition, data on the years from 2005 to 2008 are provided as a comparison benchmark of typical responses collected when paper evaluations were used. The conclusions of this study show that while a drop in response rate did occur when the switch was made, no significant change in instructor and course ratings was observed. Furthermore, the students who did complete online evaluations provided lengthier and more numerous comments.

Keywords: course evaluations, online evaluation, instructor evaluation, student comments

INTRODUCTION

As incoming students become an increasingly technologically advanced group, Universities have had to revise and update the way they deliver business education. At the School of Business Administration (SBA) at Loyola University Chicago, recent developments in classroom technology are a testament to this belief: classrooms are equipped with computers, projectors, internet access, Wi-Fi, and clicker systems. All courses have a Blackboard presence. During the previous academic year, several courses conducted a disaster preparedness scenario to see whether they could, in the event of a medical crisis, for example, switch from in-class to online delivery and back again into in-class instruction seamlessly. These advancements have paved the way for a more technologically conscious and up-to-date brand of education. The adoption of online course evaluations is considered to be a logical step in this direction.

Teacher Course Evaluations (TCE) are used in most universities to assess teaching effectiveness. The summarized TCE results not only help faculty members improve their teaching, but they also serve as a basis for promotion, tenure, salary, and merit considerations. The recent introduction of online evaluations has made this process more efficient. However, while an increasing number of academic institutions have replaced traditional paper evaluations with the new online forms, many instructors are concerned that adopting online evaluations will increase both the potential for low response rates and the likelihood of non-response bias.

Research on online evaluations has yielded a wide range of results; response rates have been very good in some cases and abysmal in others. Reports from studies have shown a wide range of results, with rates anywhere from a high of 89% to a low of 31% (Anderson et al (2005), Dommeyer et al (2002, 2003), Layne et al (1999), Liegle and McDonald (2004), Norris and Conn (2005), Schawitch (2005), Thorpe (2002)). Northwestern University moved to online evaluations in 1999 and reported that by 2003 (Hardy, 2003), despite the volume of responses being slightly lower, the number of comments had increased both in number and usefulness. Donovan et al. (2006) found that students who completed the online evaluations wrote more comments.

Whether or not online evaluations yield a different breakdown in positive and negative results cannot be ascertained as of yet. Some studies have shown an increase in positive results (Carini et al (2003), Thorpe (2002)), while others found no change (Layne (1999), Liegle and McDonald (2004)). Still other studies demonstrated an increase in negative evaluations (Donovan et al (2006)). Finally, other studies found the responses to be equivalent

to the opinions expressed on the popular website RateMyProfessor.com (Brown, Baillie, and Fraser (2009)). Rhea et al (2007) compared evaluations collected in face-to-face courses to those from on-line courses and found that students filling out the evaluations in the former offered more constructive comments, while the evaluations for online courses offered more praise and destructive comments.

The relationship between instructor ratings and other variables has also been studied. Hoffmann and Oreopoulos (2009), in a study that included 32,666 students and 1,844 instructors in first-year classes, discovered almost no relationship between the students' evaluation of instructors and the instructor's rank, faculty status, or research focus. The findings of Pan et.al. (2009) demonstrated that high ratings are not dependent on small class size or the expectation of high grades. In contrast, Carrell and West (2008) found that student evaluations of professors are positively related to course achievement in introductory courses. Boysen (2008), on the other hand, stated that only eight percent of students said they gave an instructor a lower evaluation because they were getting a low grade. Hills et al (2009) suggest that the significance of different items on a student evaluation varies by gender, major, and class. Clayson (2009) concluded that there is a small relationship between learning and evaluations that is situational.

The wide range of results from past research, as such, did not provide any not conclusive prediction about the outcome of switching from paper to online course evaluations. With respect to the SBA, the first concern anticipated was the drop in response rates when converting to online evaluations. Although some studies had reported no change in response rates, the majority found a decrease. A second area of focus was course and instructor ratings; some faculty expressed concern that their ratings would decrease since students typically used online formats to complain, not praise. In other words, students who were happy with the course delivery would be the ones not to respond. The last concern was the comments made by the students. Were students going to be more likely to provide comments on the open-ended questions? Would students spend more or less time on this important (to instructors) part of the questionnaire?

In order to evaluate these questions, this study compares the response rate, instructor and course rating, and student comments provided by these two methods of course evaluations. A baseline for comparison between the two methods was established, which includes data collected from both the transition year (during which the programmed switch to online evaluations occurred) and the four years preceding the switch. Monitoring the change in responses and ratings over multiple years not only allows for a more concrete analysis of yearly fluctuations, but it also helps put the transition year in perspective. Furthermore, to increase the scope and comprehensiveness of the study, all faculty and courses in the undergraduate program were included. Furthermore, the collected data was sorted by dividing the courses participating in the study as such: advanced and core classes, large and small sections, and courses taught by full-time and part-time faculty.

DATA SET

Evaluations are conducted for each course/section at the end of each semester. The questionnaire used by the SBA contains 20 multiple choice questions rated on the Likert-scale from 1 (strongly disagree, or very poor, depending on the format of the statement) to 5 (strongly agree, or excellent). The last two questions, of most interest to the faculty, are those which ask the student to give an overall rating of the instructor and the course. These multiple choice questions are followed by three open-ended response questions, prompting students to identify the best aspects of the course, possible areas of improvement, and any other comments. As such, the data set for this study is derived from these portions of the evaluations.

The switch to online forms occurred for the Spring 2009 term. Core classes are taught both in Fall and Spring terms, but many of the advanced classes are held during only one of the semesters. Thus, we used only Spring paper evaluations as a point of comparison. The included data covered the years from 2005 through 2009 and was taken from the set of classes taught in the Undergraduate division. In order to make the data set as comparable as possible, only classes offered in consecutive Spring terms and faculty who taught in the same consecutive Spring terms were used in the comparison. In other words, when comparing 2005 and 2006, for example, only faculty who taught both years and courses offered both years were used in calculating any changes from 2005 to 2006.

Furthermore, classes are divided by size in the following manner: below 50 (small) and 50-and-above (large). Large classes were not used prior to 2007, so this class-size comparison occurs only in the last two pairs of years. The number of sections analyzed in each pair of years along with the number of student evaluations used in the analysis are shown in Table 1.

Table 1: Count of Sections and Responses used in Analysis

	2005 to 2006	2006 to 2007	2007 to 2008	2008 to 2009
Number of Sections	152	138	185	182
Number of Responses	3701	3728	5308	4424

The data set also includes course descriptive information such as department, number of enrolled students, whether the instructor is full-time or part-time, and the number of valid responses received. The final data set contained evaluations from years 2005 through 2009, representing 67 separate instructors and 66 courses taught in the SBA (14 of these are core, the remainder advanced). The percent of representation across the subcategories is shown in Table 2.

Table 2: Percent of Sections in Each Sub-category

	2005 to 2006	2006 to 2007	2007 to 2008	2008 to 2009
Advanced	36	46	46	47
Core	64	54	54	53
Full-time	96	83	72	84
Part-Time	4	17	28	16
Large	N/A	N/A	34	39
Small	N/A	N/A	66	61

RESPONSE RATES

We anticipated a drop in response rate as this had been reported in a number of other studies, but we wanted to take this analysis one step further and analyze the change by class type. The changes in response rates between the pairs of years are shown below in Table 3.

Table 3: Change in Response Rates for Each Pair of Spring Terms

	Response Rate Percent Change From			
	2005 to 2006 Paper to paper	2006 to 2007 Paper to paper	2007 to 2008 Paper to paper	2008 to 2009 Paper to online
Total	1.51	-0.49	0.72	-25.99
Advanced	2.39	-3.16	0.23	-21.63
Core	0.98	1.25	1.17	-29.55
Full-time	1.71	-0.43	-0.39	-24.57
Part-time	-3.48	-0.72	3.62	-33.56
Large	---	---	3.13	-25.09
Small	1.51	3.40	-0.52	-26.45

The overall response rates during the paper-to-paper years (from 2005 to 2008) show a normal up and down fluctuation of up to 3 ½ percent across all categories. The change from paper to online shows a large drop across all categories. Although anticipated, this is an area we want to improve significantly. The subcategory breakdown shown in the above table indicates there was a greater drop in the response rates in core classes, part-time faculty, and in smaller classes.

The correlations between the response rates, ratings, and class size for the years in which large classes were taught are shown in Table 4. The correlations between response rate and class size across all three years are small and negative.

Table 4: Correlation between Response Rate and Several Variables

Variables	2007	2008	2009
Class Size	-0.35	-0.11	-0.18
Instructor Rating	0.34	0.25	0.37
Course Rating	0.32	0.27	0.37

Instructor and course ratings have positive, though not large, correlations with response rates, and they are slightly lower during 2008 than in 2007 or 2009. Based on this, response rates do not have strong relationships with the instructor rating, course rating, or class size.

INSTRUCTOR AND COURSE RATINGS

In addition to response rates, we wanted to see whether students gave lower instructor and course evaluations in online evaluations than they did in paper evaluations. Also, we wanted to compare these changes to those seen in the paper-to-paper sets of years. Average scores for both instructor and course can vary from 1 (poor) to 5 (excellent). We calculated the difference in average from year to year. The results for these differences are shown in Table 5. Note that the value of 0.02 for the Total from 2005 to 2006 means that the average instructor rating for all courses in spring 2006 was 0.02 higher than that of all the courses in spring 2005. As seen in Table 5, all changes, whether positive or negative, are below 0.15. In the paper to electronic year, there was also a drop in overall instructor ratings, but this accounted for less than two percent. Although a slight drop was seen for both full-time and part-time faculty, in both core and advanced classes, the ratings for the instructors of larger classes actually increased by one percent. It is important to note that the changes in instructor ratings observed from paper-to-online evaluations are similar to the variations observed within the paper-to-paper evaluation years.

Table 5: Change in Ratings for Instructor

	Instructor Average Change From			
	2005 to 2006	2006 to 2007	2007 to 2008	2008 to 2009
Total	0.02	0.03	-0.08	-0.07
Advanced	0.12	0.03	0.01	-0.04
Core	-0.03	-0.01	-0.13	-0.07
Full-time	0.03	0.05	-0.07	-0.08
Part-time	-0.13	-0.11	-0.10	-0.11
Large	---	---	0.00	0.05
Small	0.00	0.09	-0.12	-0.15

As shown in Table 6, course ratings during the paper years indicate a mix of small up and down moves, all less than 0.15. In comparing the course ratings from 2008 to 2009, there was an insignificant drop of three-quarters of a percent overall. This drop was seen in classes taught by both full-time and part-time faculty in both advanced and core areas, but was more specifically concentrated in smaller classes. Classes of 50 or more students showed an increase in course rating from 2008 to 2009. Again, the variations in the observed course ratings from paper-to-online evaluation year (2008 to 2009) are similar to the year to year variations observed prior to online evaluation in 2009.

Table 6: Change in Ratings for Course

	Course Average Change From			
	2005 to 2006	2006 to 2007	2007 to 2008	2008 to 2009
Total	0.00	0.01	-0.04	-0.03
Advanced	0.08	-0.04	0.10	-0.02
Core	-0.04	0.03	-0.15	-0.01
Full-time	0.00	0.01	-0.01	-0.03
Part-time	-0.01	0.02	-0.13	-0.07
Large			-0.01	0.01
Small	0.00	0.05	-0.06	-0.06

We wanted also to look closer at all courses taught in the Spring of 2009, the term where all evaluations were done online. For this more specific analysis, we looked at all classes, regardless of whether they had been taught before or whether the faculty was new or not. These results for the response rate, instructor rating, and course rating are shown in Table 7. This table gives the average score across all sections of classes. Then, for each subcategory, the table displays the difference in that subcategory from the average score. For example, the average response rate over all classes was 44.15, but advanced classes averaged 3.38 points higher.

Table 7: All Spring 2009 Courses, with Sub-Categories Compared to Average

	Response Rate	Instructor Rating	Course Rating
Average Score - All	44.15	4.19	4.08
Advanced	3.38	0.16	0.16
Core	-3.38	-0.16	-0.16
Full-time	1.01	0.09	0.07
Part-time	-1.92	-0.18	-0.12
Class >=50	-2.51	0.05	-0.02
Class < 50	1.49	-0.03	0.01

The observed differences within core and advanced classes, regardless of the subcategory, are noticeably consistent. The advanced class numbers are all slightly higher than average, and the core class numbers are slightly lower. A similar result occurs in the full-time (slightly higher) and part-time (slightly lower) results. However, for class size, there is a mixed result. Larger classes had a lower than average response rate, while smaller classes had a slightly higher response rate. The instructor rating for large classes was slightly higher with a lower than average course rating. The reverse occurred in small classes.

COMMENTS BY STUDENTS

In addition to 20 multiple choice questions, students were asked the following essay type questions:

- Question 1. What were the best aspects of this course?
- Question 2. How would you improve this course?
- Question 3. What other comments do you have?

Among those who completed the online evaluations in Spring 2008 and 2009, the percentage of students providing comments in the essay type questions is given in Table 8. When compared to the students who provided comments in paper evaluations, approximately 19% more of the students provided comments in online evaluations.

Table 8: Percentage of Responses to Essay-Type Questions

	Question 1	Question 2	Question 3
Online (Spring 2009)	59.4	50.9	37.6
Paper (Spring 2008)	49.5	40.8	33.7

Table 9 gives the average number of words written in the comments to the essay-type questions. The average number of words written in online comments was about 149% greater than the average number of words written in paper evaluations.

Table 9: Average Number of Words Written per Student

	Question 1	Question 2	Question 3
Online (Spring 2009)	19.64	23.80	27.29
Paper (Spring 2008)	8.10	9.46	10.74

CONCLUSIONS AND FUTURE PLANS

The primary concern in switching to an electronic survey system was that response rates would be lower. This proved to be true. Compared to the previous paper evaluation years, we observed a drop of about 25 percent in responses during the online evaluation switch of 2009.

The second concern was that online evaluations would result in lower instructor and course rankings. Faculty had expressed a belief that average ratings of instructors and courses would be lower due to decreased participation. Furthermore, many believed students would only use the anonymity of the online evaluations to voice complaints. We did not see any significant drop in either area. The observed variation in average instructor ratings was down by 0.15, and course ratings were down by 0.03 during the online evaluation switch. However, similar fluctuations were observed in the previous years when paper evaluations were used.

The third concern dealt with the comments provided by students to the open-ended questions. This study has shown that the students who completed the online evaluations wrote more comments and provided lengthier feedback in their statements.

As other studies have also noted, we need to address the issue of lower response rates. There are a number of measures other schools have tried to rectify this problem, and we have a committee looking into ways in which student interest and involvement in the process can be increased. Moreover, although the average response rate in online evaluations was lower than that of the paper evaluations, there were significant differences in the response rates among different courses, with response rates varying from 22 to 76 percent for individual courses in online evaluations. Further study is necessary to understand the reasons for these large fluctuations. This, in turn, will provide us with a better understanding of how to increase the overall response rate.

AUTHOR INFORMATION

Faruk Guder is Associate Dean and professor of Operations Management in the School of Business Administration at Loyola University Chicago. He teaches courses in the areas of operations management, quality management, and statistics. His research interests are in the areas of quality management, inventory management, and distribution facilities location in a price-responsive environment.

Mary Malliaris is an Associate Professor of Information Systems & Operations Management in the School of Business Administration at Loyola University Chicago. She teaches courses in the area of data analysis including data mining, informatics, financial modeling, and statistics. She has research interests in several areas including applications of neural networks, modeling financial markets, hospital professional practice environments and gender in IT.

REFERENCES

1. Anderson, H., J. Cain and E. Bird, Online Student Course Evaluations: Review of Literature and a Pilot Study. *American Journal of Pharmaceutical Education*, 69(1) Article 5, pp. 34-43, 2005.
2. Boysen, G., Revenge and Student Evaluations of Teaching, *Teaching of Psychology*, 35(3), pp. 218-222, Jul 2008.

3. Brown, M., M. Baillie and S. Fraser, Rating Ratemyprofessors.com: A Comparison of Online and Official Student Evaluations of Teaching, *College Teaching*, 57(2), pp. 89-92, Spring 2009.
4. Clayson, D., Student Evaluations of Teaching: Are They Related to What Students Learn? –A Meta-Analysis and Review of the Literature, *Journal of Marketing Education*, 31(1), pp. 16-30, 2009.
5. Carini, R. M., J. C. Hayek, G. D. Kuh and J. A. Ouimet, College student responses to web and paper surveys: Does mode matter? *Research in Higher Education*, 44 (1), pp. 1-19, 2003.
6. Carrell, S. and J. West, J, Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors, NBER Working Paper No. 14081, National Bureau of Economic Research, 2008.
7. Donmeyer, C. J., P. Baum, and R. W. Hanna, College students' attitudes toward methods of collecting teaching evaluation: In-class versus on-line. *Journal of Education for Business*, 78 (1), pp. 11-15, 2002.
8. Dommeyer C. J., P. Baum, K. S. Chapman and R. W. Hanna, An experimental investigation of student response rates to faculty evaluations: The effect of the online method and online treatments. Paper presented at Decision Sciences Institute; Nov. 22-25, Washington, DC, 2003.
9. Donovan, J., C. E. Mader and J. Shinsky, Constructive student feedback: Online vs. traditional course evaluations, *Journal of Interactive Online Learning*, 5(3), pp. 283-296, 2006.
10. Hills, S., N. Naegle and K. Bartkus, How Important Are Items on a Student Evaluation? A Study of Item Salience, *Journal of Education for Business*, 84(5), pp. 297-303, 2009.
11. Hoffmann, F., and P. Oreopoulos, Professor Qualities and Student Achievement, *Review of Economics and Statistics*, 91(1), pp. 83-92, February 2009.
12. Layne B. H., J. R. DeCristofor and D. McGinty, Electronic versus traditional student ratings of instruction. *Research in Higher Education*; 40(2), pp. 221-32, 1999.
13. Liegle, J. and D. McDonald, Lessons Learned From Online vs. Paper-based Computer Information Students' Evaluation System Paper (refereed) presented at 21st Annual Information Systems Education Conference, 2004.
14. Mangan, K., Professors Compete for Bonuses Based on Student Evaluations, *Chronicle of Higher Education*, 55(21), A10, 2009.
15. Hardy, N., Online Ratings: Fact and Fiction, *New Directions for Teaching and Learning*, No 96, pp. 31-38, 2003.
16. Norris, J., and C. Conn, C., Investigating Strategies for Increasing Student Response rates to Online-delivered Course Evaluations. *Quarterly Review of Distance Education* 6(1) pp. 13-32 (ProQuest document ID 975834871), 2005.
17. Pan, D., G. Tan, K. Ragupathi, K. Booluck, R. Roop and Y. Ip, Profiling Teacher/Teaching Using Descriptors Derived from Qualitative Feedback: Formative and Summative Applications, *Research in Higher Education*, 50(1), pp. 73-100, February 2009.
18. Rhea, N., A. Rovai, M. Ponton, G. Derrick and J. Davis, The Effect of Computer-Mediated Communication on Anonymous End-of-Course Teaching Evaluations, *International Journal on E-Learning*, 6(4), pp. 581-592, October 2007.
19. Schawitch, M., Online Course Evaluations: One Institute's Success in Transitioning from a Paper Process to a Completely Electronic Process! Presentation at the Association for Institutional Research Forum, June 2005.
20. Thorpe, S., Online Student Evaluation of Instruction: An Investigation of Non-Response Bias. Paper presented at the 42nd Annual Forum of the Association of Institutional Research, June 2002.
21. Wilhelm, W., The Relative Influence of Published Teaching Evaluations and Other Instructor Attributes on Course Choice, *Journal of Marketing Education*, 26(1), pp. 17-30, 2004.

NOTES