

STUDENT EVALUATIONS OF FACULTY REVISITED: ONLINE VERSUS TRADITIONAL METHODS

Marilyn Young

Professor of Management
College of Business and Technology
The University of Texas at Tyler
Tyler, Texas

Sara McCaslin

Assistant Professor of Mechanical Engineering
College of Engineering and Computer Science
The University of Texas at Tyler
Tyler, Texas

ABSTRACT

Past research has focused on many aspects of faculty evaluation from types of question to appropriate application. An increasing number of academic institutions are changing to web-based systems to take advantage of certain efficiencies in the collection process. This research compares results from traditional in-class versus online methods. No significant differences in mean scores were found in the majority of cases. However, when a major paper was required in the course, researchers discovered that student evaluations were significantly more negative. This study proposes a model to show important factors in considering a change from traditional pen and paper evaluations to online evaluations and whether the online process will have a significant effect on means scores and response rates.

INTRODUCTION

Educational institutions have continuously made efforts to determine methods of effective evaluation of faculty, courses, and degree programs. Although progress has been made, the need for continuous improvement is needed. Past research has focused on many aspects of faculty evaluation, such as examination of the appropriate methods, questions, and sample size. Universities often administer student evaluations of faculty to provide feedback to faculty for improvement of teaching effectiveness. Another stream of research assesses the validity and reliability of the instrument, while others focus on the appropriate application (i.e. professional development vs. evaluation of faculty).

Recently, administrators have considered online student evaluations as opposed to the more traditional in-class, paper and pen type evaluations. An increasing number of academic institutions are considering a change to web-based systems to take advantage of the efficiencies in the collection of end-of-semester course evaluations. In considering such a change, it is important that researchers determine whether the method will affect mean evaluation scores and response rates. Even though the literature has criticized reliance on student evaluations of faculty teaching, universities continue to use them to measure faculty

effectiveness. However, since the results may also be used in determining faculty merit, promotion, and tenure, these instruments and methods are of vital importance to faculty. This research attempts to determine if differences exist in results between traditional methods and online evaluations and to assess the resulting implications. Also, this study proposes a model which shows variables, such as class environment which may affect the validity and reliability of online faculty evaluations.

LITERATURE REVIEW

The issue of student evaluation of teaching remains controversial with mixed findings. The advantages of online over traditional in-class, paper methods include reduced costs, ease of reaching representative samples, the ability to validate data during collection, and rapid dissemination of results (Couper, 2000). Also, accounting administrators found a statistically inverse relationship between the weight given the evaluations and the emphasis placed on teaching (Read, Rama, & Raghunandan, 2001). However, some administrators are concerned that the misuse of evaluations may result in faculty members engaging in various activities designed specifically to affect student ratings on teaching evaluations, rather than to improve instruction (Simpson & Sigauw, 2000). In addition, faculty may be concerned that the evaluations do not accurately

reflect their teaching skills. On the other hand, Hobson and Talbot (2001) concluded that well-developed student evaluations with adequate data may provide some of the best measures of teaching effectiveness. Nevertheless, the question remains as to the best method of obtaining student evaluations to obtain valid and reliable data.

The classic method of obtaining student evaluations is the traditional paper and pen survey, as compared to online surveys. In considering a change to online course evaluations, researchers assessed how such a change would affect the quality of course data. The conclusion showed that online evaluation methods lead to lower response rates, but that lower response rates do not affect mean scores. Consequently, a change from traditional methods to online evaluations is unlikely to adversely affect faculty evaluation scores (Avery, Bryant, Mathios, Kang, & Bell, 2006). Furthermore, Donmeyer, Baum, Hanna, and Chapman (2004) discovered that online evaluations produce essentially the same scores as traditional evaluations (Dom-meyer et al., 2004; Layne, DeCristoforo, & McGinty, 1999). Although they found that online response rates are lower, the mean scores of the traditional evaluation are essentially the same as the online course evaluations.

In addition, smaller classes tend to have higher response rates with online surveys while response rates increase over time. Johnson (2002) noted that response rates increased yearly from 40%, 51%, 62%, and finally 71% during the last year of evaluations. Another research project found that students who completed the online surveys responded more favorably toward faculty than students completing the paper format. Consequently, class evaluations results became more positive for faculty (Carini, Hayek, Kuh, Kennedy, & Ouimet, 2003). Interestingly, when students used online surveys, they typed seven times more in comments than with in-class paper evaluations. Likewise, research in a graduate management class found that students typed an average of four times as many words online as they did using a traditional method (Hmieleski & Champagne, 2000).

Anderson, Cain, and Bird (2005) found that online evaluations are less subject to faculty influence, since students are allowed to have as much time as they desire with flexible time periods to complete the instrument. Another line of research examined student attitudes toward the methods of evaluation and discovered that the online method has a lower response rate than does the in-class method. Further, online respondents (1) expressed concerns about anonymity of their responses; (2) complained that the process was time consuming; and (3) disliked the complicated login procedure (Dommeyer, Baum, & Hanna, 2002). However, Layne et al. (1999) reported that students prefer completing electronic evaluations com-

pared to traditional ones. Ravelli's (2002) research project involving focus groups indicated that online evaluations are easy to use, allow for more confidentiality, and provide more time for comments.

Online evaluations, such as RateMyProfessors.com, may provide some insight into factors affecting evaluation processes. Lawson and Stephenson (2005) discovered several factors that affect student perceptions of professors using the RateMyProfessors.com information. However, they concluded that students are subjective in their evaluations (Lawson & Stephenson, 2005). On the other hand, Kindred and Mohammed (2005) concluded that the student evaluations of professors matched their actual concerns regarding the quality of instruction in terms of competence, knowledge, clarity, and helpfulness. Interestingly, they stated that personality and appearance were not as important.

Other factors influencing student evaluations include class size, difficulty of the class, percent of students responding, and length of the class. Furthermore, expectations of higher grades resulted in more favorable student evaluations; however, the relationship is significantly different depending upon faculty rank (Read et al., 2001).

METHODOLOGY

The basic research question is as follows: Are university online evaluations an effective method to measure class and faculty effectiveness? To answer this question, classes were chosen in a college of business administration and students were asked to evaluate the class in the traditional method as well as online. Eight classes in a college of business administration were selected to participate in the study. The classes were as follows:

MANA 3311, MANA 5320	Organizational Behavior
FINA 3311	Principles of Finance
MARK 3311	Principles of Marketing
MANA 3325	Entrepreneurship
MANA 4310	International Management
FINA 3330	Security Analysis & Portfolio Mgmt.

Students were asked to use the traditional evaluation instrument which consisted of eight questions and was identical to the online evaluation. The traditional pen and paper evaluation was performed at the end of the semester as the instructor left the room and followed standard procedures. The administration requested students to go online to evaluate all of their classes, including those out-

side the college. Students were made aware that the online evaluations would be placed on the university website for public viewing.

To encourage student participation and a high response rate, the university administration used an incentive where students would be able to see their final grades early if they completed the online class evaluations. Surprisingly, the university received a 65% response rate in its first attempt to use the online system.

Students were asked to rank eight questions on a Likert scale using strongly agree, agree, disagree, and strongly disagree (see Table 1). In assessing differences, mean scores for both the traditional method (pen and pencil) and online were calculated. T-tests were conducted to determine if significant differences existed between two sample means for each question in the eight classes.

The following hypothesis was formulated:

H1: No significant differences exist in traditional (pen and paper) method and online class evaluations.

FINDINGS

Means and standard deviations for both the traditional and online version were obtained for each question on the survey (see Table 1). Next, the probability points based on a two-tailed Student's *t*-distribution were obtained using built-in Excel statistical functions.

The results showed that no significant differences in mean scores existed in six of the classes. However, Table 2 shows that two of the classes had significantly lower online course evaluations (*t*-test values ranged from 2.23 to 7.23). These two classes had an end-of-semester project that had not yet been graded when they used the traditional evaluations but had been completed and graded when they completed the online evaluations. Figure 1 further illustrates this phenomena, showing that factors exist that can seriously impact the results of student evaluation results.

The *t*-test results for six of the classes indicated that no statistically significant difference existed between the mean scores for each question in the traditional evaluation and the online evaluation, with probabilities ranging from 0.07 to 1. It is also noteworthy that for the six classes without an end-of-semester project, the instructor and or-

TABLE 1
Student Evaluations Mean Scores according to Traditional and Online Classes
Where No Significant Differences Exist

Questions	MANA 3311.04							MANA 3311.01							MANA 5300									
	In Class			Online				In Class			Online				In Class			Online						
	n=25			n=18				n=32			n=24				n=9			n=5						
	\bar{x}	s		\bar{x}	s	SE	t	(p)	\bar{x}	s		\bar{x}	s	SE	t	(p)	\bar{x}	s		\bar{x}	s	SE	t	(p)
The course was well organized.	4.20	0.71		4.39	0.61	0.20	0.94	0.35	4.50	0.51		4.63	0.49	0.13	0.97	0.34	4.88	0.35		5.00	0.00	0.12	1.06	0.31
The instructor communicated effectively.	4.40	0.64		4.33	0.84	0.24	0.30	0.77	4.50	0.57		4.54	0.59	0.16	0.25	0.80	4.88	0.35		5.00	0.00	0.12	1.06	0.31
The instructor showed interest in progress of students.	4.30	0.75		4.33	0.77	0.24	0.13	0.90	4.40	0.81		4.63	0.58	0.19	1.24	0.22	4.88	0.35		5.00	0.00	0.12	1.06	0.31
Tests/assignments were graded and returned promptly.	4.40	0.58		4.44	0.51	0.17	0.24	0.81	4.80	0.40		4.58	0.58	0.14	1.60	0.12	4.88	0.35		5.00	0.00	0.12	1.06	0.31
Free to ask questions, disagree & express ideas.	4.20	0.82		4.11	0.90	0.27	0.34	0.74	4.50	0.77		4.42	0.93	0.23	0.34	0.73	4.88	0.35		5.00	0.00	0.12	1.06	0.31
Course has been of value.	4.30	0.80		4.06	0.94	0.27	0.88	0.39	4.40	0.55		4.25	0.90	0.21	0.72	0.47	4.88	0.35		5.00	0.00	0.12	1.06	0.31
Overall, this instructor was	4.30	0.79		4.06	1.00	0.28	0.85	0.40	4.30	0.68		4.42	0.65	0.18	0.67	0.51	4.88	0.35		5.00	0.00	0.12	1.06	0.31
Overall, this course was	4.00	0.79		3.89	0.96	0.28	0.40	0.69	4.10	0.67		4.13	0.85	0.21	0.14	0.89	4.88	0.35		4.80	0.45	0.23	0.32	0.75

Questions	FINA 3330							FINA 4330							MARK 3311.01									
	In Class			Online				In Class			Online				In Class			Online						
	n=22			n=22				n=21			n=24				n=36			n=33						
	\bar{x}	s		\bar{x}	s	SE	t	(p)	\bar{x}	s		\bar{x}	s	SE	t	(p)	\bar{x}	s		\bar{x}	s	SE	t	(p)
The course was well organized.	3.68	0.89		3.55	0.96	0.28	0.47	0.64	3.67	0.97		3.33	1.17	0.32	1.06	0.30	4.64	0.49		4.54	0.51	0.12	0.82	0.41
The instructor communicated effectively.	3.27	1.16		3.27	1.35	0.38	0.01	0.99	3.52	1.03		3.38	1.24	0.34	0.42	0.67	4.69	0.47		4.54	0.51	0.12	1.31	0.20
The instructor showed interest in progress of students.	3.59	1.05		3.59	1.05	0.32	0.00	1.00	3.90	0.77		3.67	1.05	0.27	0.86	0.39	4.67	0.63		4.33	0.87	0.18	1.82	0.07
Tests/assignments were graded and returned promptly.	4.36	0.73		4.27	0.55	0.19	0.48	0.63	4.10	0.89		4.04	0.86	0.26	0.21	0.83	4.72	0.51		4.71	0.46	0.12	0.10	0.92
Free to ask questions, disagree & express ideas.	3.95	1.05		3.82	0.91	0.30	0.46	0.65	4.24	0.70		3.96	1.00	0.25	1.09	0.28	4.64	0.54		4.58	0.58	0.14	0.43	0.67
Course has been of value.	3.41	0.91		3.64	1.14	0.31	0.74	0.46	3.33	1.11		3.42	1.14	0.34	0.26	0.80	4.47	0.61		4.46	0.59	0.14	0.08	0.93
Overall, this instructor was	3.27	1.12		3.18	1.30	0.37	0.25	0.80	3.81	1.12		3.42	1.25	0.35	1.10	0.28	4.58	0.69		4.33	0.76	0.18	1.44	0.15
Overall, this course was	3.45	0.91		3.23	1.19	0.32	0.70	0.49	3.38	1.12		3.33	1.09	0.33	0.15	0.88	4.42	0.69		4.21	0.66	0.16	1.27	0.21

TABLE 2
STUDENT EVALUATIONS MEAN SCORES ACCORDING TO
TRADITIONAL AND ONLINE CLASSES
WHERE SIGNIFICANT DIFFERENCES EXIST

Questions	MANA 4310						MANA 3325							
	In Class		Online		SE	t	(p)	In Class		Online		SE	t	(p)
	n=25	n=14	n=30	n=24										
\bar{x}	s	\bar{x}	s	\bar{x}	s	\bar{x}	s							
The course was well organized.	4.64	0.49	3.21	1.05	0.30	4.81	0.0000	4.63	0.56	2.60	1.47	0.32	6.42	0.0000
The instructor communicated effectively.	4.69	0.47	3.14	1.35	0.37	4.17	0.0002	4.57	0.73	2.32	1.55	0.34	6.55	0.0000
The instructor showed interest in progress of students.	4.67	0.63	3.43	1.28	0.36	3.39	0.0017	4.67	0.55	2.88	1.51	0.32	5.51	0.0000
Tests/assignments were graded and returned promptly.	4.72	0.51	3.79	0.70	0.21	4.37	0.0001	4.53	0.63	3.68	1.28	0.29	2.99	0.0043
Free to ask questions, disagree & express ideas.	4.64	0.54	3.86	0.86	0.25	3.06	0.0041	4.63	0.61	3.04	1.37	0.30	5.29	0.0000
Course has been of value.	4.47	0.61	3.79	1.05	0.31	2.23	0.0319	4.53	0.68	2.40	1.38	0.31	6.93	0.0000
Overall, this instructor was	4.58	0.69	3.00	1.24	0.36	4.41	0.0001	4.63	0.61	2.40	1.41	0.31	7.23	0.0000
Overall, this course was	4.42	0.69	2.93	1.21	0.35	4.23	0.0001	4.57	0.63	2.48	1.36	0.30	6.95	0.0000

ganization of the course had the highest probability of statistically significant differences, both with probabilities of less than 0.4.

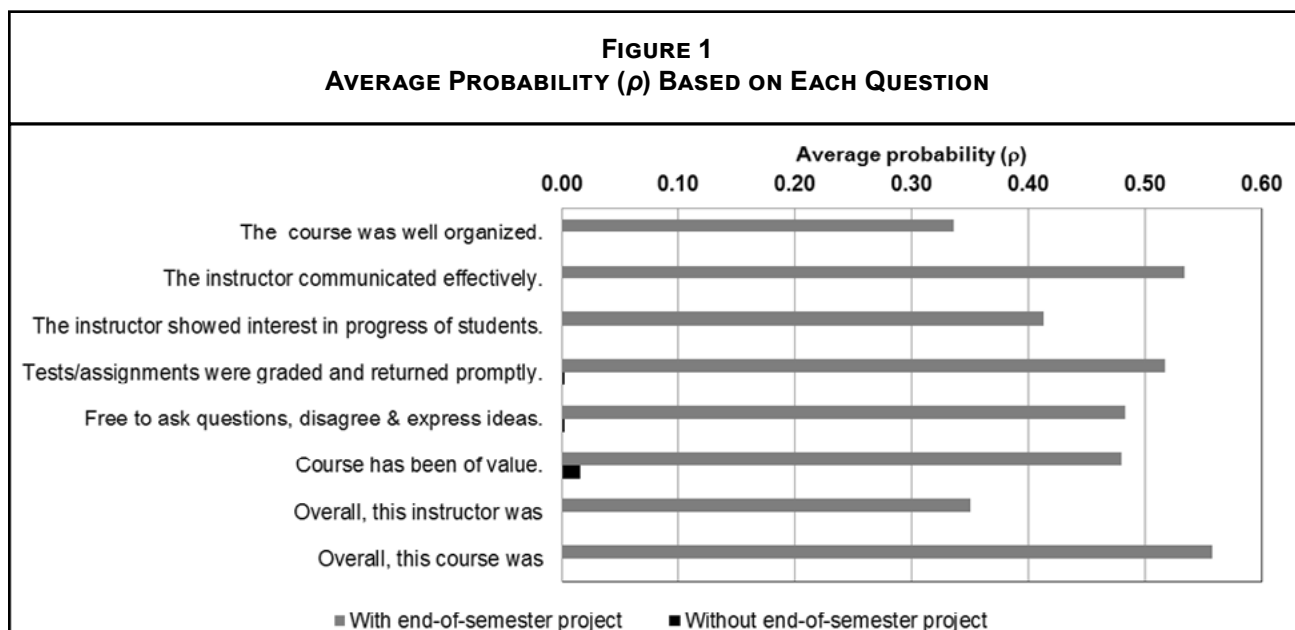
The remaining two classes involved in the study, however, showed very different results: the maximum probability was calculated to be 0.0043, indicating a statistically significant difference between mean values for the traditional and online evaluation. This phenomenon is further illustrated in Figure 1, which shows average probability (p) based on question type for classes with end-of-semester project and without end-of-semester project. In short, the

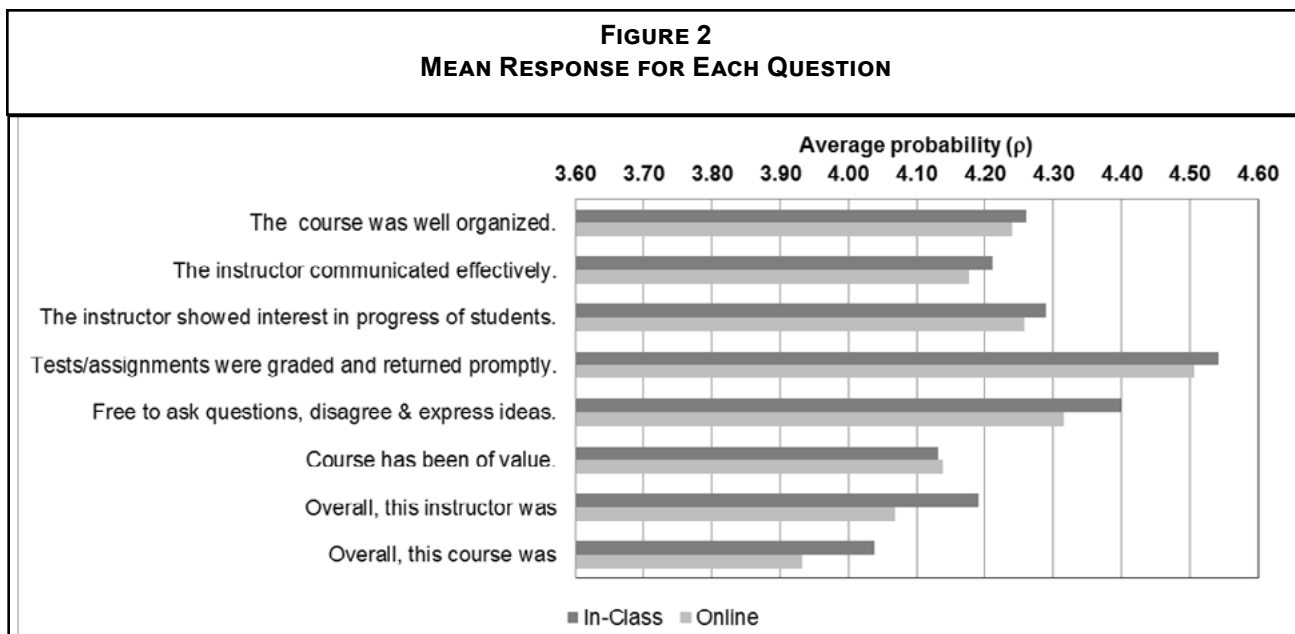
results of this study indicate that, barring additional factors, the results of online evaluations do not significantly differ from those obtained using traditional methods. Figure 2 shows the mean responses according to average probability (p) values.

PROPOSED MODEL AND IMPLICATIONS

The following is a discussion of several variables, which may affect the success of online evaluations. Next, a proposed model is offered.

FIGURE 1
AVERAGE PROBABILITY (p) BASED ON EACH QUESTION



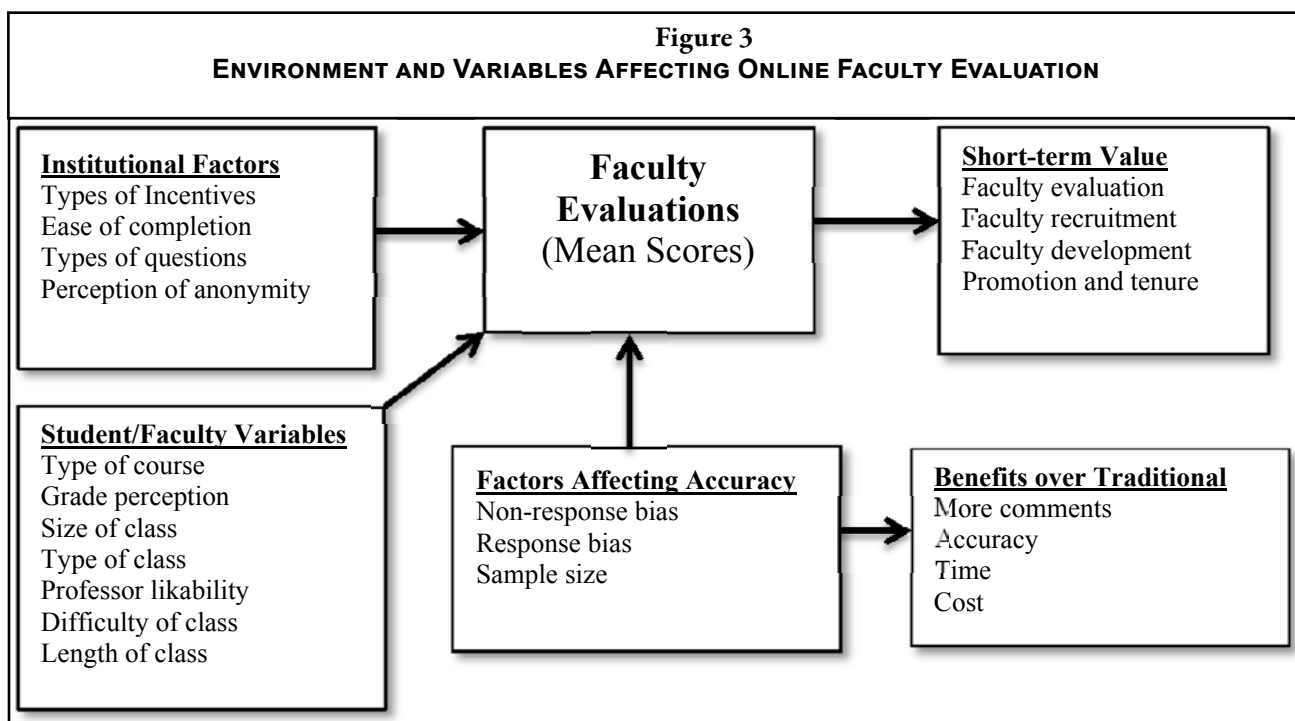


Institutional Variables

In order to increase the response rate, incentives are used, such as a drawing for an iPad or priority viewing of exams. If the questions are short and the process is easy to answer, students should be more prone to complete the online evaluations. The survey data show no significant differences between the two methods with the exception of those classes where a large end-of-semester project was involved. However, other variables which could affect the online effectiveness include incentives, ease of comple-

tion, types of question, and grade perception as shown in the proposed model (Figure 3).

A major issue in assessing the reliability of the method is in terms of obtaining a large participation percent. If the sample size is not large enough, non-response bias could exist. In other words, care should be taken to increase the sample size and include a cross section of students within the class. However, it is possible that students who complete the online evaluations are biased in that they have



either an extremely positive or negative perception of the course, and the results are, therefore, skewed.

Student/Faculty Factors

A major finding revealed that no significant differences existed in the two methods with the exception of courses with a major project. Students may have an idea of what their project grade was before the evaluation. After the evaluation, the grade may indeed be lower and thus students gave much lower evaluations. It seems that a professor must give students short-term feedback with potential grades during the semester in order that a student may have adequate assessment of his/her performance.

Certainly, whether the course is graduate/undergraduate, online, or distance learning would have a bearing on the student's evaluation. Also, students' perception of anonymity and likeability may affect the results. Furthermore, the size, length and difficulty of class may be variables in the final evaluation.

Benefits

Faculty evaluations are important for a number of reasons including merit raises, faculty recruitment, and promotion and tenure decisions. For this reason, administrators should be cognizant of their strengths and limitations. Many faculty and administrators believe the major purpose of such evaluations is for faculty development only.

DISCUSSION

While faculty evaluation methods continue to expand, providing up-to-date, accurate information in a timely manner presents an additional challenge. Administrators should enhance their efforts to develop and coordinate the most efficient methods for faculty evaluation to achieve institutional, departmental, and faculty objectives. Therefore, it is vital to have accurate information to measure teaching effectiveness. This research shows that online evaluations have positive implications with certain variables being examined. Although no significant differences existed in the majority of the classes sampled, a 3.6 (in-class) or 3.3 (online) mean value may be different in the minds of the students and other viewers. Also, it is important to note the type of class may be a major influence on evaluations. If the professor gives no feedback until the end of the semester, it could be a potential problem in overall evaluations.

Administrators may need to enhance new methods to increase sample size. Various incentives should be examined, such as immediate access to grades or a drawing for several iPads. Also, faculty and staff could provide an announce-

ment on Blackboard or other sites explaining the procedure. This communication should emphasize and clarify the purpose, outline, and purpose of these evaluations.

LIMITATIONS OF RESEARCH

The major limitation of this study was the use of only eight classes within one college and for one semester. Future research would do well to have a formal study of a large number of classes within the entire university to determine where differences exist.

Information derived from this research may be used to assess the purpose and methodology of faculty evaluation and provide better insight into improving the instrument and overall procedure. Both methods have been examined as to their role in effective teaching, but this study compares the results of two methods of obtaining student evaluations and determines if significant differences exist and their implications. Although other research examines the two methods, this research is different as faculty and students were aware that the results will be placed on the Internet for viewing. Therefore, obtaining accurate results appears to be even more important.

REFERENCES

- Anderson, H. M., Cain, J., & Bird, E. (2005). Online course evaluations: Review of literature and a pilot study. *American Journal of Pharmaceutical Education*, 69(1), 34-43.
- Avery, R. J., Bryant, W. K., Mathios, A, Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *Journal of Economic Education*, 37(1) Winter, 21-37.
- Carini, R. M., Hayek, J. C., Kuh, G. D., Kennedy, J. M., & Ouimet, J. A. (2003). College student responses to web and paper surveys: Does mode matter? *Research in Higher Education*, 44, 1-19.
- Couper, M. (2000). Web surveys: A review of issues and approaches. *Public Opinion Quarterly*, 64, 464-494.
- Donmeyer, C. J., Baum, P., & Hanna, R. W. (2002). College students' attitudes toward methods of collecting teaching evaluations: In-class versus online. *Journal of Education for Business*, 78(1), 11-15.
- Dommeier, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). 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.

- Hmieleski, K., & Champagne, M. V. (2000). Plugging in to course evaluation. *The Technology Sources Archive*. Retrieved from http://technologysource.org/article/plugging_in_to_course_evaluation/
- Hobson, S. M., & Talbot, D. N. (2001). Understanding student evaluations: What all faculty should know. *College Teaching*, 49(1), 26–31.
- Kasiar, J. B., Schroeder, S. L., & Holstad, S. G. (2002). Comparison of traditional and web-based course evaluation processes in a required, team-taught pharmacotherapy course. *American Journal of Pharmaceutical Education*, 66, 268–270.
- Kindred, J., & Mohammed, S. N. (2005). “He will crush you like an academic ninja!”: Exploring teacher ratings on RateMyProfessors.com. *Journal of Computer-Mediated Communication*, 10(3).
- Kuhtman, M. (2004). Review of online student ratings of instruction. *College and University Journal*, 80(1), 64–67.
- Lawson, R. A., & Stephenson, E. F. (2005). Easiness, attractiveness, and faculty evaluations: Evidence from RateMyProfessors.com. *Atlantic Economic Journal*, 33(4), 485–486.
- Layne, B., DeCristoforo, J., & McGinty, D. (1999). Electronic versus traditional student ratings of instruction. *Research in Higher Education*, 40(2), 221–232.
- Read, W. J., Rama, D. V., & Raghunandan, K. (2001). The relationship between student evaluations of teaching and faculty evaluations. *Journal of Education for Business*, 76(4), 189–192.
- Simpson, P. M., & Siguaw, J. A. (2000). An exploratory study of the faculty response. *Journal of Marketing Education*, 22(3), 199–213.