TEACHING MANAGEMENT FACE-TO-FACE VS. ONLINE: LESSONS FROM STUDENTS’ EVALUATIONS OF FACULTY FOR IMPROVED TEACHING AND STUDENT ENGAGEMENT

Amit Shah1* and Michael Monahan1

*Corresponding Author: Amit Shah ashah@frostburg.edu

INTRODUCTION

Student Evaluation of Faculty (SEF) is a common practice in institutions of higher education which allows students to provide anonymous feedback to their instructors on teaching and their experiences. These evaluations are an important tool used to measure teacher effectiveness and used in deciding whether an advancement and tenure are the appropriate courses of action (Otto et al., 2005). These two main purposes measure how the teacher is doing in the classroom and provides feedback which identify higher operative teachers and others to develop and improve (Marzano, 2013).

While student evaluations of faculty are supposed to show teacher effectiveness and provide areas for improvement, studies have shown these evaluations are being used as groundwork for deciding whether professors should get promoted or receive tenure. In fact, some faculty may manipulate the way they teach or grade, in order gain a better evaluation outcome (Simpson and Siguaw, 2000). Further, a study conducted by Washington State University suggests that student evaluations may no longer be a good measuring tool for how well a teacher performs as some teachers who showed more enthusiasm during class received a higher evaluation as well as those who were easier graders (Wilson, 1998).

Nationwide, different institutions are trying to improve their evaluation process. Currently, there are two main ways for students to evaluate their teachers. The first is through a formal document...
given during class. These evaluations, while technically voluntary, are nearly universally completed which yield a much larger and complete sample size (Otto et al., 2005). The second evaluation method consists of students completing their ratings individually in an online environment which is gaining popularity as it doesn’t detract from class time and provides material and administrative time savings. While still voluntary, the response rate is generally lower as only those diligent students or those who had a very positive or negative experience fill out the assessment. However, there is little research to support their effectiveness.

An additional factor is the class delivery method itself. Traditional face to face classes operate at a set time and place with the faculty member present. The advent and growth of online classes is changing the interaction between the faculty and the students. As online learning is asynchronous, anytime and anywhere, there are no set times for the class and the place can be wherever one has Internet access. In a face to face class, on the given evaluation day, the faculty member leaves the room while a colleague administers the evaluation. Online classes and online assessment allow the student at a time of their choosing to complete the evaluation.

The purpose of this study is to ascertain if there are differences in the student evaluations of faculty between traditional face-to-face classes and online classes. The following research hypothesis guided this study.

H1: There is no difference in the Student Evaluation of Faculty scores based on the gender of the faculty member.

H2: There is no difference in the Student Evaluation of Faculty scores based on the level of the class.

H3: There is no difference in the Student Evaluation of Faculty scores based on the delivery method; face to face vs online.

METHODS
Student evaluation of Faculty data was obtained from a small Mid-Atlantic Masters I Comprehensive Institution. Management Department data was selected as seven of the twelve classes are required for all business majors. In addition, the institution offers a Masters of Business Administration (MBA) degree and seven of the twelve classes are in the Management Department. The data file included the student evaluation score based on a Likert like rating scale ranging from one to five, the number of students who completed the evaluation, the gender of the faculty member, and the delivery method; face to face vs. online for a three year period.

LITERATURE REVIEW
There is a variety of research done to find the value of student evaluations of faculty through forms that filled out anonymously by students in the classroom (Read et al., 2001; Martin, 1998). Some key elements that face-to-face evaluations and online evaluations have in common are the quality of the helpfulness of the evaluations and the clarity on them.

The majority of the studies relating to student evaluations of faculty conclude that evaluations are reliable and stable, valid when compared with student learning, and are multidimensional in terms of what they assess, they are useful in
improving teaching, and their minimal bias’s (d’Apollonia and Abrami, 1997; Cashin, 1988; Centra, 1993; Marsh, 1987; McKeachie, 1997). In addition, the results conclude online evaluations were more likely to produce lower response rates.

Traditional classroom evaluations are generally completed at one time with little distractions in the classroom. Even though they are technically voluntary, there is subtle pressure to complete the evaluation as most of the students’ peers are doing so. Consequently, response rates are higher. With online assessment, you are on your own to complete it during your own time. Even with the best intentions students often may forget or simply not get around to completing it. Also, the computer that the evaluations are being done on can play a part to lower response rates. Slow computers and spam can cause a person to stop completing the evaluation midway through. One more contributor to low response rates could be that students are given a short time at the end of the semester to complete the evaluations but they might have other concerns so that puts the evaluations low on the totem pole.

**ONLINE EVALUATIONS**

Becker and Watts (1999) found many institutions are moving toward online assessment due to their lower cost and greater efficiency. But, since student evaluations of faculty are a critical component in the faculty member’s tenure and promotion, several factors need to be considered. For example, will evaluating online affect the mean course evaluation score, change response rates, which students will actually respond in an electronic medium.

There are many studies that examine teacher effectiveness in online teaching including the validity of student ratings (Feldman, 1989), the effects of instructor personality on student ratings (Murray et al., 1990), the effects of student characteristics on student ratings (Arbuckle and Williams, 2003; Centra and Gaubatz, 2000), the relation between student ratings and student achievement (Cohen, 1981; Greenwald and Gillmore, 1997), and the effects of course characteristics on student ratings (Feldman, 1984; Marsh and Bailey, 1993).

According to researchers, Greenwald and Gillmore (1997) and Marsh (2001), student evaluations are valid and reliable. Marsh has also concluded that student evaluations, when carefully laid out, are not only valid and reliable but also multidimensional, uninfluenced by possible biases, and useful in terms of improving teaching practices. According to Meyer (2002), students that take online courses learn at least as much as students in traditional classrooms (Neuhauser, 2002; Rovai, 2002; Schulman and Sims, 1999; Young et al., 2001). Some researchers believe that online courses can lead to isolation, frustration, boredom, overload, and low course completion rates (Berge, 1999; Hara and Kling, 2000; Northrup, 2002). Northrup found that students in online classes felt that it was very important for the teachers to give opportunities for collaboration and conversation. When active participation is involved, it led to greater learning and also enhanced the motivation of the students (Berge, 1999; Northrup, 2002). In conclusion, even if students in online classes learn just as much as traditional students, it really depends on how the students value different ways of teaching.
In a study conducted by Young et al. (2001), 104 online instructors reported that students were required to take on different responsibilities as students and the instructors were to be more of a facilitator for the students. Northrup (2002) found in his study that online students most valued regulating their own learning and receiving timely responses from instructors. Hara and Kling (2000) found that students were highly concerned when they had issues with communication, including breakdowns and having to constantly check and read through lengthy emails between student and teacher. Students were also frustrated with vague conversations from the teacher and also with the fact that teachers could possibly be busy and can’t email the student back right away, making it difficult to get an answer.

Students are mostly attracted to online courses because of their flexibility (Tricker et al., 2001). Students also care that course materials must be of high quality, assignments must be meaningful, and that feedback and communication between the teacher and students must be of high-quality. A study conducted by Spangle et al. (2002) found that more than 1,200 student evaluations of online classes along with surveys of the teachers rated that written communication skills, appropriate design of activities that allow discussion between one another, and timely feedback were vital components in an effective online course.

As online classes are becoming more popular, new ways of teaching and learning are evolving. Similarly, the evaluation method must change as well. In an online class, the instructor is often only experienced by remotely sent keystrokes whereby the interaction is through discussion boards, email, and PowerPoint slides. The evaluation invitation is sent via email where the participants can choose if and when they wish to respond. Dommeye et al. (2004) found the response rate to the online evaluations were usually lower than that to in-class evaluations but when students were pleased with their grades, the online evaluation was comparable to in-class evaluations. However, Bergstrand and Savage (2014) found student’s feel like they have learned less in an online course than if they would have taken that same course face to face which results in a lower student faculty evaluation. Similarly, Russell (1999) asserted there was no real difference between taking a class online and taking classes face to face with the teacher, but students were significantly less satisfied with the online class as demonstrated by lower faculty evaluation scores. Interestingly, Aleamoni (1999) found student faculty evaluations were consistent each semester or year as exemplified by the high correlation of .87-.89 between evaluations. In addition, the correlation between evaluations of the same instructor and course was between .70 and .87.

Perhaps a factor is the demographics of the online learner who are generally nontraditional in that they are older, have jobs, family obligations, and cannot make the time or cannot travel to the face to face class. This can create a fundamental disconnect as research shows that most professors fit the curriculum to the technology, not the other way around (Bennett and Green, 2001; Summers et al., 2005). In addition, these students could be satisfied with the teacher but not satisfied with the class due to its organization and idiosyncrasies of online education. Carnevale (2000) found that students that took online classes seek out many similar things that in face courses have, namely a teacher who knows what they are talking about, interaction, and a feeling of community within the class.
But what constitutes good teaching? Chickering and Gamson (1987), assert seven principles of good, effective teaching that helps evaluate a teacher in a face to face classroom setting. However, Graham et al. (2001) assert that these principles are also a solid framework for evaluating teachers in an online class setting and if followed will result in higher evaluation scores. These principles encompass 1) instructors should provide clear guidelines for interaction with students; 2) well-designed discussion assignments facilitate meaningful cooperation among students; 3) students should present course projects; 4) instructors need to provide two types of feedback: information feedback and acknowledgment feedback; 5) online courses need deadlines; 6) challenging tasks, sample cases, and praise for quality work communicate high expectations; and 7) allowing students to choose project topics incorporates diverse views into online courses. Examples for adapting these guidelines would be telling your students who to contact for technical difficulties and let the students the expected time frame for answering their questions, allowing students to be involved and have a voice, giving proper feedback, setting clear due dates and timelines, providing examples of excellent work, and permitting choice in selection of topics which can match their interests.

While there is administrative pressure to have more evaluations conducted online because of convenience and efficiency, the faculty are resistant to using because of lower return rates and higher percentage of negative responses since only those motivated, positively or negatively respond. Since a lower response rate coupled with some poor evaluations could adversely affect promotion and tenure decisions which would explain why faculty’s reticence. However, to combat this problem Kendall (2001) is a proponent of utilizing management systems such as Blackboard and WebCT since the students had greater overall satisfaction with the software and organization of the content. This finding was supported by Wernet et al. (2000) who found students using WebCT found the site helpful and facilitated greater use of course management tools. Further, Kasiar et al. (2001) found longer and more detailed open ended comments from students. Donavan, Mader, and Shinsky found no difference face to face and online evaluations in the numerical ratings. However, the online open ended responses were more specific on the items the instructor excelled and what was done poorly.

Face-to-face vs. online evaluation has varied little on the quantitative scores (Dommeyer et al., 2004; Layne et al., 1999; Sorenson and Johnson, 2003). However, even though mean scores turned out to be the same the online response rates were lower than traditional response (Avery et al., 2006) but over time, the online response rates increased. Interestingly, Carini et al. (2004) found students who evaluated online were more favorable to the faculty member. In addition, online responses tend to be significantly longer as Kasiar et al. (2001) found the total number of words typed per student more than seven times that of the student using the traditional system. This makes logical sense since in a written face to face evaluation there is limited time and students are writing with a pencil as compared to no time limits and the ease of typing online responses.

When faculty members were asked what factors bias student evaluations, 72% cited
course difficulty, 68% grade leniency, and 60% identified course workload (Marsh, 1987). However, Centra (2003) found higher grades and a lower level of difficulty were factors in higher student faculty evaluations. Consequently, many faculty believe this is a wise and prudent strategy to ensure higher ratings. However, there are Centra (1998) found smaller class sizes, fifteen and under, the higher the student faculty evaluations will be. Conversely, the larger the class, the lower the evaluations will be.

In a comprehensive study, Centra (1998) investigated whether final grades have an unwarranted influence on student faculty evaluations. However, since evaluations are generally given before the student receives their final grade and there may be bias based on the students’ perceptions on what they think their grade will be. Past studies have shown that response rates percentages in online evaluations are not typically high. In a study of 2,453 undergraduate and graduate course evaluations from students enrolled in 66 different courses Layne et al. (1999) found a response rate of 60.6% for the traditional evaluations and 47.8% for the online evaluations. Cartwright (1999) found on a 20% response rate for online evaluations which increased to 43% with an email request. Even in a rigid structured institution like United States Air Force Academy, the Ewell (2000) found response rates dropping from 82% to 73% online. However, over half the students said they preferred online and over 90% commented the online version was simpler. Students believed that their responses were more anonymous online but completing online took more time. Another factor could be the level of the student both in terms of academic ability and grades. Layne et al. (1999) found that students with higher GPA’s were more likely to fill out the evaluation forms. Students who were seniors were least likely to fill out the forms because they didn’t have anything to gain or lose being so close to graduation.

Johnson (2002) experimented with the evaluation process by giving two evaluations, one before the grade was received and the other after the grade was given. Not surprisingly, students who received a grade higher than what they anticipated gave a more favorable second evaluation. Conversely, if the student’s grade was lower than expected, their subsequent evaluation was lower.

Avery et al. (2006) concluded that online evaluations can be successful if there are sufficient computers available, the students are computer literate, and if assistance if needed is provided. A study conducted by Tallent-Runnels, Lan, Fryer, Thomas, Cooper, and Wang, sought out to find whether or not problems graduate students experience with technology in online courses relates to the type of evaluation given of their instructors. Results found a positive relationship between the two. The more likely a student was to experience technology problems or the worse the problem obstructed learning for the student, the higher they evaluated the instructor and the course. This means that if a teacher got a lower evaluation score, it could not be due to technological problems.

It seems that even there are lower response rates with online evaluations; it still does not affect the average evaluation scores. This means that it will not affect teachers that are midway through the promotion and tenure process. The only big impact that the lower response rates could have on student faculty evaluations is when comparing professors to each other.
As online evaluations become more relevant, response rates will increase. Results from the study conducted by Avery et al. (2006) found that out of 1,317 students, 897 provided evaluations. This gave a response rate of 68.1%. Every course recorded had a response rate greater than 50% and the average was 72%. This percentage is very close to the response rate that the traditional evaluations yielded.

Another bias for consideration is the factor of gender, not only that of the instructor but of the student as well. Some studies have found a very slight difference between the evaluations of female and male teachers on the basis of the gender of the student (Basow and Distenfeld, 1985; Basow and Howe, 1987; Bennett, 1982; Elmore and LaPointe, 1974; Harris, 1975; Kaschak, 1978). Some found gender bias with male students rating their female teacher lower than they would have a male teacher (Basow and Silberg, 1987; Etaugh and Riley, 1983; Kaschak, 1978; Lombardo and Tocci, 1979; Paludi and Bauer, 1983). Conversely female students tend to rate female teachers higher but the differences are not statistically significant (Basow and Silberg, 1987).

Centra and Gaubatz found female students perceived female instructors as more organized, better at communicating, more interactive, and they provided higher quality exams, out of class work, and feedback to the students. However, for the course outcomes scales, there was no significant difference by faculty gender. The female instructors could have got better ratings not because of a gender bias but because they taught differently as they lecture less and promote more discussion. In addition, female teachers and female students were more open to teaching methods that involved connecting the classroom rather than separating it, understanding and accepting rather than testing, and teamwork over argument (Belenky et al., 1986). This is contrary to the older model of “banking” where a teacher’s role is to fill the students by making deposits of information (Freire, 1971).

RESULTS

The data covered a three year period academic years 2010-2013 and was comprised of nearly six thousand responses from four hundred and ten classes (see Table 1). The four nominal variables were nearly all at a ratio of two-thirds to one-third.

The Management Faculty scores were extremely high averaging 4.58 out of a possible 5.0. A comparison of the Student Evaluation of Faculty Scores (SEF) on the basis of gender revealed no difference in the faculty evaluation scores by the students (see Table 2).

The average undergraduate SEF scores were .37 points higher than graduate scores which resulted in a moderate Pearson’s $r$ correlation which was statistically different at the .000 level (see Table 3). Interestingly, the standard deviation of the graduate responses was nearly two and half times greater than the undergraduate ratings indicating a much greater range of scores at the graduate level.

When comparing the SEF scores between face to face and online a modest statistically significant difference was found as online scores ratings was .385 or 8.1% lower than face to face (Table 4).
### Table 1: Demographics

<table>
<thead>
<tr>
<th>Delivery (# Classes)</th>
<th>N</th>
<th>%</th>
<th>Level</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face to Face</td>
<td>253</td>
<td>62%</td>
<td>Undergraduate</td>
<td>284</td>
<td>69%</td>
</tr>
<tr>
<td>Online</td>
<td>157</td>
<td>38%</td>
<td>Graduate</td>
<td>126</td>
<td>31%</td>
</tr>
<tr>
<td>Total</td>
<td>410</td>
<td></td>
<td>Total</td>
<td>410</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of Student Responses</th>
<th>N</th>
<th>%</th>
<th>Faculty Gender</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face to Face</td>
<td>4031</td>
<td>68%</td>
<td>Male</td>
<td>274</td>
<td>67%</td>
</tr>
<tr>
<td>Online</td>
<td>1871</td>
<td>32%</td>
<td>Female</td>
<td>136</td>
<td>33%</td>
</tr>
<tr>
<td>Total</td>
<td>5902</td>
<td></td>
<td>Total</td>
<td>410</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Faculty Gender and SEF Scores

<table>
<thead>
<tr>
<th>Faculty Gender</th>
<th>Number of Classes</th>
<th>SEF Mean</th>
<th>Standard Deviation</th>
<th>r</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>136</td>
<td>4.55</td>
<td>0.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>275</td>
<td>4.6</td>
<td>0.39</td>
<td>0.052</td>
<td>0.298</td>
</tr>
</tbody>
</table>

### Table 3: Level of Class and SEF Scores

<table>
<thead>
<tr>
<th>Level</th>
<th>Number of Classes</th>
<th>SEF Mean</th>
<th>Standard Deviation</th>
<th>r</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate</td>
<td>284</td>
<td>4.7</td>
<td>0.25</td>
<td>0.398</td>
<td>0</td>
</tr>
<tr>
<td>Graduate</td>
<td>126</td>
<td>4.33</td>
<td>0.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4: Mode of Delivery and SEF Scores

<table>
<thead>
<tr>
<th>Mode of Delivery</th>
<th>Number of Classes</th>
<th># of Students</th>
<th>SEF Mean</th>
<th>Standard Deviation</th>
<th>r</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face to Face</td>
<td>253</td>
<td>4031</td>
<td>4.73</td>
<td>0.2</td>
<td>0.436</td>
<td>0</td>
</tr>
<tr>
<td>Online</td>
<td>157</td>
<td>1871</td>
<td>4.35</td>
<td>0.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>410</td>
<td>5902</td>
<td>4.58</td>
<td>0.43</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 5: Comparison of Level and Delivery

<table>
<thead>
<tr>
<th></th>
<th>Undergraduate</th>
<th>Standard Deviation</th>
<th>Graduate</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face to face</td>
<td>4.73</td>
<td>0.18</td>
<td>4.74</td>
<td>0.37</td>
</tr>
<tr>
<td>Online</td>
<td>4.53</td>
<td>0.44</td>
<td>4.27</td>
<td>0.61</td>
</tr>
<tr>
<td>Diff</td>
<td>-0.2</td>
<td>0.26</td>
<td>-0.47</td>
<td>0.24</td>
</tr>
<tr>
<td>Difference %</td>
<td>-4.20%</td>
<td>144.40%</td>
<td>-9.90%</td>
<td>64.90%</td>
</tr>
</tbody>
</table>
When comparing the level of the class and the delivery methods several interesting findings emerged. At the Undergraduate level a -.20 or 4.2% difference between face to face and online with a tremendous variance in the standard deviations of each. Conversely, at the Graduate level a difference of -.47 which was a drop of approximately 10% (see Table 5). Interestingly, when comparing level and delivery a strong Pearson’s $r = .672$ at the .000 significance level was found.

CONCLUSIONS AND RECOMMENDATIONS

Hypothesis H1 was supported as there was no difference in the Student Evaluation of Faculty scores based on the gender of the faculty member. This finding aligns with the findings of Basow and Distenfeld (1985); Basow and Howe (1987); Bennett (1982); Elmore and LaPointe (1974); Harris (1975); Kaschak (1978). Hypothesis H2 was not supported as there was a statistically significant difference in the Student Evaluation of Faculty scores based on the level of the class. Hypothesis H3 was not supported as there was a statistically significant difference in the Student Evaluation of Faculty scores based on the delivery method; face to face vs. online.

When examining the literature and the results three major findings emerged which were consistent with the literature.

1. The faculty evaluations received from undergraduate and graduate classes were nearly identical since they varied by only .01 point.

2. Response rates for faculty evaluations of any class were lower if the evaluation was completed online instead of face to face.

3. The evaluations for classes conducted online were significantly lower than face to face classes.

The authors contend a multi-pronged approach is needed to address the issues of lower evaluation and improve teaching and student engagement:

First and foremost is the lack of personal engagement between the students and the faculty. In an online environment the instructor’s personality, humor and caring is not easily conveyed to the students. The lessons are often impersonal words and it is the instructors’ responsibility to humanize themselves. Solutions can include embedding sound files in the PowerPoint slides, posting photographs of themselves and asking the class to do the same, and through the use of webcam, self-videos and online tools such as Illuminate in Blackboard where there can be live interaction. Further, since online learning is asynchronous when a student has a question they need a prompt response. While it is not practical to have the instructor be available 24/7, the instructor could make it a priority to respond to questions within 24 hours. The response could be much shorter if the request was made at the time the instructor was online, but knowing that a response would be received at a maximum period of one day would be helpful.

While many classes have discussion boards pertinent to the topics being covered, the instructor may wish to post an open discussion board where the students can hang out and discuss items outside of the class content.

In order to increase the response rates, the faculty should make personal appeals to the students to voluntarily complete the evaluation.
forms. By appealing to the students for their suggestions for improvement and feedback on what was done correctly, the psychological contract between teacher and student would be reinforced and hopefully would yield a greater response.

Finally, since this is an issue in and among most institutions, best practices could be shared and benchmarked. These techniques could be developed with the aid of Qualitymatters.org, a web site dedicated to improving online education.

FURTHER RESEARCH

Studies could be expanded to include other Business classes such as Accounting, Finance, and Marketing. In addition, the non-Business areas of Education and Arts and Sciences could also be compared. Finally, determining the gender and gpa of the student could bring insights if these factors were related to the faculty evaluation.

REFERENCES


42. Stone M (2003), “EC Professors fair (somewhat) well on ratemyprofessors.com”, *The Leader*.


