Student Evaluation of Teaching: The Inequity of Faculty Scores in Online versus Face-to-Face Courses

Sam L. Sullivan
Sam Houston State University

Barbara Polnick
Sam Houston State University

Lautrice Nickson
Sam Houston State University

Robert Maninger
Sam Houston State University

J. Yasmine Butler
Sam Houston State University

Follow this and additional works at: https://scholarworks.sfasu.edu/slr

Part of the Educational Leadership Commons, Higher Education Administration Commons, and the Online and Distance Education Commons

Tell us how this article helped you.

Recommended Citation
Sullivan, Sam L.; Polnick, Barbara; Nickson, Lautrice; Maninger, Robert; and Butler, J. Yasmine (2013) "Student Evaluation of Teaching: The Inequity of Faculty Scores in Online versus Face-to-Face Courses," School Leadership Review. Vol. 8 : Iss. 2 , Article 8.
Available at: https://scholarworks.sfasu.edu/slr/vol8/iss2/8

This Article is brought to you for free and open access by the Journals at SFA ScholarWorks. It has been accepted for inclusion in School Leadership Review by an authorized editor of SFA ScholarWorks. For more information, please contact cdsscholarworks@sfasu.edu.
Researchers have conducted a multitude of studies over the last century on Student Evaluations of Teaching (SET); however, very few have been conducted in the new digital age (Loveland, 2007). More work is in progress as researchers try to define the differences in student responses and thoughts about the online teaching environment. The unfortunate side of this can be the administrative decision to use a one-size-fits-all mentality when many authors including Dziuban and Moskal (2011) have outlined several research-based alternatives for evaluation of online instructional effectiveness. SETs are important to faculty because they often are the determining factor in merit pay and tenure/promotion. Faculty use SETs to guide decision-making about their curriculum and instructional strategies used to deliver the course (Sheehan & DuPrey, 1999). The power of the SET is recognized as a driving force in academia. What are often not discussed are the differences between SET for face-to-face courses versus SET for online courses.

The large difference in the number of online courses versus face-to-face courses offered can complicate this issue. The growth in the number of online classes is on the increase and poses challenges for administrators. Deans struggle with hiring faculty with online teaching experience, and they must incorporate new training and faculty incentives. In addition, problems arise associated with the comparison of traditional and online teaching in terms of workload, compensation, and evaluation (Loveland, 2007).

The differences in student evaluations of teaching effectiveness between face-to-face classes and online classes are apparent. Most institutions of higher education use student evaluations to measure faculty effectiveness; sometimes SET is the only measure of teaching effectiveness employed (D'Apollonia & Abrami, 1997). If all other resources of measure are equal, then administrators rate faculty for merit, tenure, and promotion exclusively on student evaluations.

\[\text{Dr. Robert Maninger} \text{ may be reached at rmm023@shsu.edu.}\]
However, the question is whether face-to-face class responses and online class responses can be adequately measured by SET.

**Problem Statement**

The original purpose for implementing student evaluations in higher education was to improve instruction, and these evaluations were considered private between professors and students (Algozzine et al., 2004). Many institutions of higher learning established evaluation instruments to help professors focus on providing quality instruction to their students. This instrument was ideal because, at the time, there were only traditional classes. Currently, student evaluations are not being used to solely improve instruction, and they are certainly not kept private between professors and students.

The Individual Development and Educational Assessment (IDEA) is an instrument used by administrators at institutions of higher learning to assess professors who teach in both traditional and online class settings. In theory, this is an evaluative instrument that should be used to evaluate professors. However, the reality is that the student evaluations used to assess traditional classes do not align specifically to issues addressed in online teaching. The following are questions from the IDEA instrument that students complete to evaluate their professors. These questions are used to appraise professors regardless of the setting in which instruction takes place (Benton, Webster, Gross, & Pallett, 2010): a) displayed a personal interest in students and their learning; b) explained the reasons for criticisms of student academic performance; c) explained course material clearly and concisely; d) introduced stimulating ideas about the subject; e) involved students in “hands on” projects such as research, case studies, or “real life” activities; f) asked students to help each other understand ideas or concepts; and g) encouraged student-faculty interaction outside of class (office visits, phone calls, e-mail, etc.).

Online professors scored lower in the above mentioned areas as compared to professors who taught traditional classes. There is speculation that scores in these areas were lower for online professors because students expect many of the aforementioned elements to occur only in a traditional class. Perhaps scores would have been higher for online professors if questions were reworded to align with online teaching standards. One of the first questions on the IDEA form asks how the course is taught; however, the options offer “distance learning” not “online.” There is an arguable difference between these two terms; certainly distance learning is not the same thing as online.

In addition to the evaluated instrument not corresponding with online class presentation, the other concern is the low response rates which cause online scores to be invalid. According to Benton et al. (2010), on average, the proportion of students responding to the paper version of IDEA is higher than the online version. The overall mean student response rate for online survey delivery declined from a high of 56% in 2002 to 51% in 2008. The general decline has been somewhat more dramatic for scores related to online courses (Benton et al., 2010).

Scores from student evaluations are used to decide how much merit pay professors receive as well as if professors will obtain tenure and promotion. Because student evaluations are often used for high-stakes personnel decisions, it is vital that they accurately assess teaching
effectiveness (Kelly, Ponton, & Rovai, 2007). In order to protect the validity of faculty evaluations, the instrument used to evaluate professors should correspond with either traditional or online classes. Part of the disparity in online scores can be traced to lower response rates. Higher response rates can increase the total pool of student scores and decrease the risk of bias from students who are highly motivated to respond with overly positive or negative views (Faculty Senate University Affairs Committee, 2012).

The problem addressed in this study is the possibility that the evaluation instrument used for merit, tenure, and promotion for traditional and online professors is not equitable. Therefore, professors in a College of Education at a regional institution in Texas hypothesize that there is disparity within many departments. Professors who teach in departments with a large number of online classes are receiving lower student evaluation scores. Consequently, the purpose of this study is to examine the degree to which a student evaluation system (IDEA) is an equitable instrument for measuring teaching effectiveness in online versus face-to-face classes.

**Literature Review**

In his analysis of the research conducted on student evaluations of teaching, Aleamoni (1999) reviewed over 150 studies spanning a 75-year period, and found 16 myths that have remained myths over time. These myths, contrary to the research available, are often accepted as true among higher education faculty. Aleamoni (1999) makes two points that are specifically relevant to this study: (a) student ratings tend to be stable and result in substantial correlations both over time and across the same instructor, and (b) student ratings can be useful to the instructors for the purpose of enriching and improving their courses as well as to document instructional effectiveness for administrative purposes. Since this research was published, recent studies have been conducted regarding these myths and the results were similar. Specifically, Balam and Shannon (2010) found that although student ratings on single general items are accurate measures of teaching effectiveness, faculty still believed that student ratings were invalid and unreliable. Not all researchers agree regarding the accuracy of student ratings. Marsh (2007), for example, found that ratings could be biased and subject to external factors over which instructors may have little control.

Even if student evaluations are effective measures of teaching effectiveness, issues still exist regarding their use in higher education. There are opposing views of the usefulness of student course evaluations to assess teaching effectiveness. Aleamoni (1999) asserted that there was a downside to using student evaluations to improve teaching effectiveness including misuse and misinterpretation. Specifically, when administrators use the ratings for punitive purposes, faculty often find ways to undermine their use, causing many to doubt the credibility of the process (Aleamoni, 1999).

When analyzing studies published within the last ten years, the researchers found that data on student evaluations of teaching addressed a wide variety of areas. These include factors related to a) effective teaching (Balam & Shannon, 2010; March, 2007); b) personal characteristics of instructors (e.g. gender, position, age, and rank) (Isely & Singh, 2007; Kogan, Schoenfield-Tacher, & Hellyer, 2010; Kozub, 2010; Kyriakides, 2005; Slocome, Miller, & Hite, 2011); c) student characteristics (e.g. gender, age) (Heckert, Latier, Ringwald, & Silvey, 2006; March
2007); d) impact upon tenure, promotion, and merit decisions (Irons, Carlson, Kirk, & Monk, 2011); e) grading and student evaluations (Addison, Best, & Warrington, 2006; Bembenutty, 2009; Centra, 2003; Germain & Scandura, 2005; Heckert, Latier, Ringwald-Burton & Drazen, 2006; Isely & Singh, 2005; Liegle & McDonald, 2004); f) course difficulty (Heckert, Latier, Ringwald-Burton & Drazen, 2006); g) use of student evaluations to improve instruction (Finelli et al., 2008; Hallinger, 2010; Read, Rama, & Raghunandan, 2001); h) online technology (Hossain, 2010; Keefe, 2003; Lan et al., 2003; Tallent-Runnes et al., 2005); and i) evaluations across disciplines (Kember & Leung, 2011).

Measures of Effective Teaching

If student evaluations are to serve the purpose of providing feedback to instructors for improvement, then these evaluations should be tied to measures of effective teaching. Research on how students evaluate teaching is important to analyze when trying to address a problem such as the one in this study—the equitable use of student evaluations in online versus face-to-face courses as measures of teacher effectiveness. When administrators use teacher effectiveness as one of the components for determining merit, tenure, and promotion, the need to understand these SETs across all delivery methods (online and face-to-face) is critical. While student evaluations of instructors have been found to be either highly reliable or at least moderately valid in measuring student perceptions of teachers (Aleamoni, 1999; Centra, 1993; Hobson & Talbot, 2001), certain areas can be quite challenging in online courses, such as organization, rapport, and technology challenges. These topics have been found to have a moderate to high impact on how students evaluate instructors (Jirovec, Chathapuram, Ramanathan, & Rosegrant-Alvarez, 1998; Tang & Chamberlain, 2003).

Students’ perceptions of organization such as understanding exactly what they need to do can be impacted by students’ abilities to interface with the online platform and maneuver in and out of different resources in online courses. Two other areas that are closely linked to perceptions of teaching effectiveness are how much students feel they learned in the course (Bard, 1997) and how much they feel they were stimulated by the class (Remedios & Lieberman, 2008; Tang & Chamberlain, 2003). These factors may impact online courses more than face-to-face courses. Additionally, Centra (2003) found that students were quick to rate instructors lower if the courses seemed too easy or too difficult. In the case of online courses, the technology as well as the content can shape students’ perceptions of course difficulty. Lastly, the degree to which instructors are motivated, answer questions, and treat students courteously are factors linked to measures of teaching effectiveness which can also be challenging to address when courses are not conducted in person (Tang & Chamberlain, 2003). Often communication via e-mail or course feedback on assignments may seem less emotional and defining precisely what is courteous and motivating to an individual student in this environment can vary widely.

Other Factors Related to Student Evaluations of Online Courses

Two areas impacting the validity of evaluations of online courses (outside of teaching effectiveness) relevant to this study include: (a) low returns and non-response bias, and (b) factors on evaluation instruments which do not align with online instruction. While we found several studies addressing the first area, the second is in need of further exploration.
Low return response and non-response bias. Professors of online classes are often concerned that face-to-face classes are less likely than online courses to suffer the effects of non-response bias because most students are assumed to be in attendance when in-class evaluations are conducted. Thorpe (2005) reported that some studies have found several factors that might influence an individual’s decision to complete an online survey, including familiarity with the internet, the ease of completing the survey, and concerns for privacy and confidentiality. In their study of 2,057 student evaluations from 32 instructors over two semesters, Stowell, Addison, and Smith (2012) reported that online evaluations had a significantly lower response rate than classroom evaluations. In contrast, other studies have found that there was no significant difference in using a paper-based method or the web-based evaluation process in terms of non-response bias (Thorpe, 2005).

Factors on evaluation instruments not aligning with online instruction. The IDEA instrument used in this study has several items that instructors in online courses often find more challenging to replicate in online classes versus those conducted face-to-face. For example, students are asked to rate the degree to which their instructor encouraged student-faculty interaction outside of class as well as to rate how well their instructor fostered collaboration by asking students to help each other understand ideas or concepts. There is some research emerging that would alter or completely create alternative ways to evaluate peer learning and evaluation, for example, as well as collaborative learning outcomes as they are structured in online courses (Gazi, 2011). The expectations for organization and relevance of content may be even higher for students in online classes. Jones (2012) found:

Students in online courses want high quality and rigorous courses that are well developed and organized, and that provide them with engaging learning experiences. Students expect their online instructors to develop and deliver challenging and worthwhile courses that offer alternatives to the traditional classroom, but not at the risk of losing high-quality learning experiences (p. 56).

Results

The data reported were from the 2012 spring semester of the College of Education at a regional institution in Texas and were collected from the Institutional Research database. Data from previous semesters were available, but there was no distinction between a course taught and evaluated online versus a course taught face-to-face and evaluated online. Starting with the spring 2012 semester, this delineation could be made and provided a more realistic data set for face-to-face and online courses. We calculated the total number of course offerings by department (Curriculum and Instruction, Educational Leadership and Counseling, Health and Kinesiology, and Literacy, Language and Special Populations) and separated these according to mode of delivery (face-to-face or online). We also calculated the number of courses below the 65% response rate, the average response rate, and the average class size. See Table 1 for the representation.
Table 1

\textit{Response Rate}

<table>
<thead>
<tr>
<th>Department</th>
<th>Total</th>
<th>Face-to-Face</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>&lt;65%</td>
<td>Avg %</td>
</tr>
<tr>
<td>C&amp;I</td>
<td>59</td>
<td>2</td>
<td>87</td>
</tr>
<tr>
<td>ELC</td>
<td>61</td>
<td>1</td>
<td>90</td>
</tr>
<tr>
<td>HK</td>
<td>98</td>
<td>21</td>
<td>74</td>
</tr>
<tr>
<td>LLSP</td>
<td>93</td>
<td>4</td>
<td>86</td>
</tr>
<tr>
<td>\textbf{Means}</td>
<td>\text{77.8}</td>
<td>\text{7}</td>
<td>\text{84.25}</td>
</tr>
</tbody>
</table>

\textit{Note.} More face-to-face courses offered, but with near equivalent numbers of students in each. Response rates are the notable differences between categories.

We then ran a one-way ANOVA between groups and found that only one factor had a significant difference (p < .05). Table 2 details the “average response rate by percentages” comparison.

Table 2

\textit{ANOVA One Way}

<table>
<thead>
<tr>
<th></th>
<th>SS</th>
<th>df</th>
<th>Mean$^2$</th>
<th>F</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Groups</td>
<td>1225.125</td>
<td>1</td>
<td>1225.125</td>
<td>14.197</td>
<td>.009</td>
</tr>
<tr>
<td>Within Groups</td>
<td>517.750</td>
<td>6</td>
<td>86.292</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1742.875</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\textit{Note.} This demonstrates a statistically significant difference (p < .05) between the average response rate of online versus face-to-face courses.

After running the one-way ANOVA, we calculated the four major general categories (Progress on Relevant Objectives, Excellent Teacher, Excellent Course, and Summary) for final scoring on the IDEA data report using the raw data category. The data below represent a reporting of each department in the College of Education, displayed by both face-to-face and online classes. The percentages reflect the number of courses that were at or above the IDEA database average across all the institutions served by IDEA (Benton et al., 2010). There were no statistically significant results, but the differences in means are worth reporting (Table 3).

Table 3

\textit{Percentage of Classes at or Above IDEA Database Average}

<table>
<thead>
<tr>
<th>Department</th>
<th>Progress</th>
<th>Face-to-Face</th>
<th>Online</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Teacher</td>
<td>Course</td>
</tr>
<tr>
<td>C&amp;I</td>
<td>83</td>
<td>76</td>
<td>80</td>
</tr>
<tr>
<td>ELC</td>
<td>85</td>
<td>77</td>
<td>80</td>
</tr>
<tr>
<td>HK</td>
<td>89</td>
<td>83</td>
<td>87</td>
</tr>
<tr>
<td>LLSP</td>
<td>74</td>
<td>63</td>
<td>72</td>
</tr>
<tr>
<td>\textbf{Mean}</td>
<td>82.75</td>
<td>74.75</td>
<td>79.75</td>
</tr>
</tbody>
</table>

\textit{Note.} This details the four major general categories reported by IDEA; Progress on Relevant Objectives, Excellent Teacher, Excellent Course, and Summary.
Although Table 3 may not include any statistical significant results, it is noteworthy that all of the mean scores for online courses are below the mean scores of the face-to-face courses. This should allow points of discussion later. The largest deficit from face-to-face appears in the Excellent Teacher category.

**Discussion**

There are multiple layers of concern to discuss, but relative to this research, the field does narrow. One concern would be the use of the same evaluation form for an online course and a face-to-face course. Because these formats are so different, some consideration should be given to changing the actual tool to better represent the course delivery. Our research highlighted a statistically significant difference in regard to response rate between online and face-to-face courses. New ways to encourage online students to respond should be investigated. The hesitancy to “bribe” a student to complete the evaluation is understood, but other considerations should be investigated. There should be consideration by department chairs and college deans for a formula approach to weight the scores of an online course to better represent consistency between online and face-to-face evaluation scores, especially when merit, tenure and/or promotion are being considered. Faculty should be provided with more training and information about how to teach online so that courses provide a deeper sense of community for students. Faculty should also receive training on how to better communicate the evaluation process to students.

**Conclusion**

The study of online instruction in higher education is in its infancy. The research on student evaluation of teaching (both past and current) is at times conflicting, offering challenges to instructors whose careers depend on these measures. This is especially difficult for instructors who teach online, where issues regarding students’ relationships with their instructors, students’ abilities to understand and maneuver through the organization of the online course, and students’ perceptions of how well their teachers engaged them in both learning and collaborating with others, can be significant factors when evaluating teaching effectiveness.

Overall, the findings of our research suggest that an assessment should be developed to measure teacher performance and effectiveness in online settings exclusively. Given that the IDEA does not correlate with the best practices of online teaching environments, instructors miss out on valuable feedback that could potentially inform their course revision decisions and, subsequently, enhance the quality of digital classrooms. Additionally, assessments such as IDEA should not be used as the sole or prominent indicator of teacher effectiveness – particularly when instruction takes place solely online. Using an invalidated instrument to make decisions regarding high stakes matters such as merit pay, promotion, and tenure seems at best, absurd.


References


Faculty Senate University Affairs Committee. (2012, April). *Issues related to IDEA evaluation in online classes*. Huntsville, TX: Sam Houston State University Faculty Senate.


