A REFLECTIVE COMPARISON STUDY OF STATISTICS COURSE EVALUATIONS BY MODALITY

Victoria Ingalls, Tiffin University

ABSTRACT

Statistics for teacher evaluations in seated and online classes show a glaring difference in both quantitative and qualitative scores. The generalized analysis of the quantitative and qualitative elements of the same course, instructor, assignment as grouped by learning environment provide an opportunity to inform future practice. In this four-year study at a small, private Midwestern university (n = 316 students over 20 course sections), the seated students were much more engaged and satisfied as demonstrated by their response rates, Likert-scale ratings of course and professor, and associated comments. Reflection and best practices were employed in the online course for continuous improvement, but the numeric values for online evaluations remained consistently and significantly lower in mean value. This study concludes with a series of takeaways for enhancing connections with students and increasing their associated evaluation scores.

Keywords: course evaluations, IDEA, online students, student perceptions, asynchronous, synchronous learning

INTRODUCTION

Regardless of the size of the university and its location, most professors are evaluated by their students. Some take the student comments lightly and make no changes. Others opt to take the Likert scale rankings and the constructive comments that students provide and make changes to the course and pedagogy the next time they teach the course. In the latter context, what should the professor do when the ratings between different classes for the same coursework are very different? This paper describes one twenty-year veteran professor’s journey through the cognitive dissonance caused by unequal course evaluation results and it provides statistical analysis and conclusions for the reasons for massive discrepancies in student evaluations.

LITERATURE REVIEW

“The goal of every university is to improve faculty to improve student learning,” DeCosta, Bergquist, Holbeck, and Greenberger state (2016, p. 78). Student evaluations can provide information regarding teacher effectiveness as well as teacher preparedness (Wellein, Ragucci, & Lapointe, 2009), though some researchers question the use of traditional classroom evaluation scales for the online modality (Callister & Love, 2016). When directly compared to the traditional seated classroom, Lyke and Frank (2012) examined the learning outcomes in online and face-to-face courses but found no significant differences in scores. However, their research indicated that online learners were significantly less satisfied overall with their course and instructor. More specifically, Baker (2010) found that students in an asynchronous online course rated their instructors lower in terms of immediacy than in synchronous courses. Since teacher-to-student and student-to-student interactions are prerequisites to course satisfaction (Allen, Bourhis, Burrell, & Mabry, 2002), it may be that students have a less favorable perception of online courses because the current evaluation tools might not address all the essential components of online learning (Loveland, 2007). Swan (2002)
asserted “students who rated their level of activity as higher also reported significantly higher levels of course satisfaction and significantly higher levels of perceived learning” (p. 30). Other mitigating factors may include the complexity of class content, type of exam given, grading policies, requirement or elective status, schedule time, class structure, or student interest in the material (Constand & Pace, 2014). The question of whether online students learn as much as face-to-face students has yet to be definitively answered; prior studies have produced differing results with respect to higher achievement of learning outcomes based upon learning modality (Callister & Love, 2016).

Within the notion of virtual learning, Offir, Lev, and Bezalel (2008) offered that there are two methods within the online environment for delivering distance learning: synchronous and asynchronous. The first implies learning in real time and the latter requires no common, real-time meeting. When asked about the ability of web-based instructor tools to assist in achieving learning outcomes, students preferred live synchronous class time sessions, live synchronous chat messaging, but asynchronous email communication (Parenti, 2013). The synchronous chat that attempts to emulate face-to-face discussion is often viewed as inferior to asynchronous online communication exchange (McDonald, 2002). In contrast, Offir et al. (2008) found that students preferred learning via synchronous systems rather than by asynchronous systems. Additionally, they stated that “synchronous learning is more effective among students with higher cognitive ability than those with a low cognitive ability” (p. 1181).

Moving from delivery method to course content, students generally preferred the structure of quantitative courses, but they tended to be more critical of quantitative instructors (Royal and Stockdale, 2015). Many researchers confirm those findings in that professors in the humanities consistently earn higher ratings than do professors in Mathematics and Sciences (Bleske-Rechek & Michels, 2010; Constand & Pace, 2014; Felton, Koper, Mitchell, & Stinson, 2008). Furthermore, other studies asserted that lower teaching ratings might be more indicative of student perceptions concerning course difficulty rather than a reflection of a professor’s teaching ability (Coladarci & Kornfield, 2007). Wright (2014) concluded that some subject matter may be more conducive to online methodologies.

Unfortunately, performance review committees within universities might interpret the lower ratings as inferior classroom instruction despite the more difficult content area (Constand & Pace, 2014). This mistake could affect merit, tenure, and promotion evaluations. Since student evaluations have been found to be one of the most commonly utilized methods of assessing faculty performance (Wellein et al., 2009) and play such a large role in the tenure and promotion decisions (Loveland, 2007; Rothman, Romeo, Brennan & Mitchell, 2011), relying on student evaluations of teaching when making judgments of teaching effectiveness poses great concern for faculty who teach online courses (Rothman et al., 2011).

THEORETICAL FRAMEWORK

The theoretical framework for this reflective study employed Simonson, Schlosser, and Hanson’s (1999) Equivalency Theory. The essence of the theory is captured in the following metaphorical quote:

*A triangle and a square that have the same area are considered equivalent even though they are different geometrical shapes. Similarly, the experiences of the local learner and the distance learner should have equal value even though specific experiences might be quite different* (Simonson, Schlosser, & Hanson, 1999, p. 7).

Based upon educational theories emphasizing learner independence and autonomy, the industrialization of teaching, and interaction and communication in the classroom, the research trio developed their theory on the assertion that learning experiences should be the same for all learners, regardless of modality. “The more equivalent the learning experiences of distant learners are to those of local learners, the more equivalent will be the outcomes of the educational experiences for all learners” (Simonson et al., 1999, p. 6).

More specifically, their research determined that learning outcomes should be equivalent for both local and distance learners, and that the delivery of instructional ideas fit the expectations and facilities available to each learner and the learning situation. Since seated and online students have fundamentally different environments in which
they learn (Simonson et al., 1999), Equivalency Theory advocates that the design for distance and local learners be a collection of different, but ultimately equivalent, learning experiences. Thus, the major components of the theory are equivalency, learning experiences, appropriate application, students, and measurable cognitive outcomes. (Simonson et al., 1999).

The purpose of this reflection is to analyze and explore the reasons for the strong differences in opinion of teaching and course quality, which seem dependent solely upon modality. Similar to Hurst’s (2015) analysis concerning online teaching, one purpose of this article is to add to the knowledge base “concerning the difference between or pros and cons of face-to-face and online formats” (p. 36). The emphasis and direction of this paper provides a reflective and personal glimpse into the empirical evidence concerning some major discrepancies in student evaluations despite the basis of Equivalency Theory in education. Taking the call for action from other studies, “The same course in both online and on-campus environments makes for an extended experimental comparison of learning outcomes, while controlling for two important variables: the instructor and the course content” (Koory, 2003, p. 1). The generalized analysis of the quantitative and qualitative elements of the same course, instructor, assignment, etc., as grouped by learning environment provide an opportunity to inform practice far beyond one small, private Midwestern university.

**METHODOLOGY**

The standardized IDEA (www.ideaedu.org/) teaching evaluations for 316 seated and online students across 20 sections of a traditional undergraduate statistics course were gathered and analyzed across four years of fall, spring, and summer semesters for one professor. The IDEA course evaluations were identical, regardless of the learning modality. Data gathered included the Likert scale ratings for excellent teacher rating, excellent course rating, progress on relevant objectives, and course response rate. The IDEA system uses the classic five-point Likert scale where 5 = almost always/definitely true, 4 = frequently/more true than false, 3 = neutral/in between, 2 = occasionally/more false than true, and 1 = hardly ever/mostly false. When applicable, the numerical ratings were examined by independent t-tests, assuming unequal variance, for the comparison of teaching evaluations differing only by modality (i.e., same course, content, assignments, professor, etc.). It is noted that the IDEA student ratings inventory produces reliability coefficients above .70, though most approach or exceed .80 (Benton, Li, Brown, Guo, & Sullivan, 2015). Finally, a phenomenological approach for establishing thematic elements in the qualitative comment portions of the student evaluations was used. Representative comments were chosen as evidence of each general theme in each modality.

**RESULTS**

In contrast to Dziuban and Moskal’s (2011) findings that course modality does not impact student evaluations of learning experiences, the two modalities in this study displayed a strong difference concerning student perceptions of quality. For instance, the seated coursework students rated an average of 4.7 for excellent teacher, with a very small range of .4. The mean for excellent course rating was 4.3, while the response rates were approximately 78%. Common qualitative statements included statements such as “You are perfect!” “She was an awesome teacher that I enjoyed having as a professor.” “She actually cares about her students and you don’t see that a lot. She wants her students to achieve progress and knowledge and that really shows,” and finally, … is by far one of the best professors. She is extremely active and engaging with students and course material. It is obvious that she enjoys teaching and has a great passion for learning. She makes students WANT to do well.

In contrast, the online ratings provided a very different vision. For excellent teacher, the average was 4.1 and the excellent course rating average was 3.5, with respective variances of 1.73 and 2.15. The average response rate for the online courses was 37%. Representative negative comments in the virtual environment included statements such as “…She has very little flexibility…,” and “… made this subject so difficult that I had to take time from other courses just to balance her workload…” Across the four summer classes, there were only a total of 18 comments: 12 were attempting...
constructive criticism with regard to lowering the amount of work, four remarks reflected the notion of “great teacher” similar to the seated environment comments, and two were scourging personal attacks.

When compared for inference, the independent t-test assuming unequal variances was chosen since the standard deviation for the online evaluations was more than double that of the seated scores. As expected, the results were significant. For each of the respective elements that were measured, the independent t-values assuming unequal variance were $t(38) = 2.63$, $p < .01$ for excellent teacher and $t(38) = 3.22$, $p < .001$ for excellent course. Noting the small sample sizes, which were based on the number of classes rather than the number of students, the effect sizes were calculated as well. The respective Cohen’s $d$ values were .55 and .65, both considered medium effect sizes by the Cohen definition and huge by the Sawilowsky expansion of the defining values.

**DISCUSSION**

Despite employing the Equivalency Theory in the online course development and adhering to the points of best practice, the descriptive statistics formed a definite pattern dependent upon modality. The analysis proved the trends, as each test demonstrated significant differences in student ratings with medium effect sizes. The seated students uniformly rated the course, professor, and progress at much higher means, although these results contrast other studies (Hurst, 2015; Olson, 2015). The seated students also had much higher response rates.

My interpretation was that when courses are taught asynchronously, as in this study, there was less of an opportunity to build rapport and relationship with the students despite the multiple strategic attempts to do so. While deeper relationships would help to increase levels of satisfaction with the course and professor, they would also better support student learning (Callister & Love, 2016). According to Offir et al. (2008), when trust is built, the students want to succeed and feel that the support is there for them to do so.

The online students in this study apparently did not feel a strong interpersonal connection. One possible reason is that the lack of synchronicity was a large contributor in hindering both the students’ motivation and their association with me in the online course. Falloon (2011), Patillo (2007), and Hrastinski (2008) found similar results concerning real-time interactions and sense of community. Due to the asynchronous format marketed by the university, the online students did not get a real glimpse of my caring, compassionate, and quirky personality. Such a perspective is echoed by Alsharif and Qi (2014). The rating deficit might have been overcome by my multiple, concerted efforts to connect with students, but sadly, my efforts were ignored in many circumstances. This research concurs with Yanetske & Oates (2012), who emphasized the necessity of richer mediums of communication where use of pitch, tone, emphasis, and nonverbal reflection can increase personalization and feelings of trust and empathy. The results may have been quite different if the course were offered solely in a synchronous format.

Perhaps the differentiated scores and response percentages were directly related to the relationships and community-building interactions inherent in face-to-face coursework, which would have built the trust necessary to improve outcomes (Callister & Love, 2016). Meyer (2003) described comments from face-to-face coursework interactions as energetic and enthusiastic, while the asynchronous postings were slow and emotionless. While it is easy to hypothesize that the reason for the statistically significant differences in response rates was the lack of enforcement to fill out course evaluations in the virtual space, the reason may be tacitly implied. In short, if faculty social presence is critical for students to feel engaged and motivated (Bangert, 2004; Cole

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Means</th>
<th>Variance</th>
<th>Independent $t$-value</th>
<th>Cohen's Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent Teacher</td>
<td>Online 4.13</td>
<td>Seated 4.71</td>
<td>Seated .44</td>
<td>Online 1.73</td>
</tr>
<tr>
<td>Excellent course</td>
<td>Online 3.51</td>
<td>Seated 4.31</td>
<td>Seated .83</td>
<td>Online 2.15</td>
</tr>
</tbody>
</table>

*p<.01, **p<.001
& Kritzer, 2009; Perry, Dalton, & Edwards, 2009; Rothman et al., 2011), then the converse may be true as well. Thus, the implied emotional distance may make students feel that they are not known or cared for to the same degree that a seated student might feel. Moreover, because of the lack of connection, there is no social obligation to express their opinion on the Likert scale.

In conclusion, the comparative results for the statistics course show a glaring difference with respect to the learning platform. The seated students felt more connected, as demonstrated in their comments. The online students were missing the relationship network nestled in trust and rapport with the professor. They did not take the time to complete the evaluations and when they did their numeric values were consistently lower in mean value and were higher in degree of variation.

The Equivalency Theory was referenced as the theoretical framework for this study, and it still applies as an outline for increasing the scores for online evaluations (or simply to get them to par with seated evaluations). Instructors should research and adapt many of the best practices for teaching online replete throughout the internet with respect to instructional design, assessments, and interactions with students. Best practices in online instruction include providing an authentic learning environment where collaboration enhances both synchronous and asynchronous distance platforms (Lee & Duncan-Howell, 2007). For instance, according to Brown University (n.d.), the five major concepts for best practice are:

- As you design your online course, consult with an Instructional Designer (ID).
- Help your students be successful online learners.
- Engage with your students early and often.
- Review and comment in discussion forums daily.
- Provide effective feedback.
- Another website stated that best practices for teaching online are:
  - Make it a group effort.
  - Focus on ‘active’ learning.
  - ‘Chunk’ the lessons.
  - Keep group sizes small.
  - Be present.

- Parse your time.
- Embrace multimedia assignments (O’Malley, 2017).

The purpose of this paper was to provide empirical evidence of the inequality of evaluation scores between asynchronous online courses and traditional seated courses. With that done, the question is, how the scores can be equalized? As a reflective teacher, the four-year longitudinal nature of the study allowed for both course and pedagogical modifications. However, when attempting continuous improvement toward reflecting best practices as stated above, the advice in different research articles and forums was often stated in generalized, nonspecific statements that lack direct examples for online education. The remainder of this discussion therefore focuses on some specific tweaks to the online version of the course and pedagogy that either have been implemented during the course of the study or will be done to minimize negative qualitative comments and possibly increase the quantitative evaluation scores.

In that spirit, there are a few specific takeaways that others may implement to improve their online courses. First, students need to know that the university supports their learning efforts in the form of tutorial services, access to supplementary materials, and information technology assistance. A direct example would be to place links and phone numbers for access to such services front and center on the course web page when the class begins, then frequently remind the students of their availability as the term progresses. Second, since online students expect a quality course that mirrors the seated version, professors and instructional designers should collaborate to align the content directly to course learning outcomes and then deliver the identical content in ways that best fit the online environment. For instance, have the students find specific research articles related to their own career goal and data for analysis instead of relying on the professor to provide scholarly readings that are not pertinent to students’ majors or interests. As stated by Merrillat and Scheibmeir (2016), “The LMS can be used to supplement content delivery so more time can be spent on the all-important interaction between professors and students” (p. 169). Moreover, if limited to a solely asynchronous environment, Riggs and Linder (2016) offer many
specific points for conceptualizing active learning in the asynchronous learning environment. Third, it is very important that the professor lends personal support. This may take the form of checking in to the online course more frequently than is expected from the university. By checking email and help forums in the morning and the evening can help catch the students early in their descent into cognitive dissonance and develop a greater degree of trust than having them wait 24 hours for a response from the instructor.

I also suggest holding virtual office hours for “just-in-time teaching” that may or may not be of a group nature. My experiences have been that the synchronous environment is the better way to connect with the students to build quality rapport in a quantitative course. For example, students who attended live virtual teaching sessions responded favorably to video conferencing to go over homework and answer pertinent questions, explain assignments, and just generally get to know where the students are. Though only a few of the students attended (an average of 3 of 25 potential students per session), they generally did better on assignments and appeared more than satisfied with the interactions as evidenced by multiple email comments. I suggest recording these sessions and posting them in the learning management system for future student reference, regardless of their attendance in the sessions. Interestingly, of the positive qualitative comments found in the online evaluations, four of them came after implementing a greater degree of phone communication. In fact, many of the students who I spoke to on the phone were the same students who attended the live sessions. Clearly, some of them felt connected and cared for by virtue of one of the lines of synchronous communication.

In contrast to the aforementioned suggestions, my experiences can also provide a specific list of what not to do. For instance, refrain from asking about previous experiences in the content area, especially in math. In short, by asking about their previous experience in mathematics courses and their comfort level with the general content, I inadvertently formed a group with the mob mentality of “I hate math.” This was quite counterproductive: in some of the video conferencing sessions students wanted to tell me how much they hated math and did not understand it, rather than interacting with the content, each other, and me to come up with the best solution process and increased understanding.

Another concern is the amount of electronic communication given to students during the term. Although due dates should be clearly stated on the syllabus and reiterated in the weekly announcements, a large number of announcements and associated reminder emails to students can also be destructive to the learning process. Students have directly commented that they receive too many emails every day from their different professors and do not realize what may or may not be important. One student stated that the number of emails was overwhelming, so she simply opted to go about her busy life without reading any of the documents, a notion also asserted by Merrillat and Scheibmeir (2016). Similarly, when assessing students work, comments should be deeply detailed with understandable remarks that directly relate to the content but are not in such high degrees of quantity and content jargon that the student does not understand what you are correcting. In short, the communication must be necessary but caring and also clear and concise.

Overall, by employing these ideas from the practices of online teaching as shown by direct example, hopefully a connection is built, and the students feel a sense of responsibility to complete voluntary evaluations. Then, to the best of our abilities through the employment of best practices, the online students’ evaluations of course quality and instructor rapport will match those from the traditional seated work. With all of that said, however, the evaluations for asynchronous coursework may show students still lack the connective bonds with the professor and the content. Even when all best practices are parsed out and employed, the scores may remain lower. At that point, one must separate the two entities by modality and move forward with a different mentality, which would mean embracing the differences without comparison of evaluation scores between seated and online educational environments.

CONCLUSION

In summary, the comparative results for the statistics course in question show a glaring difference with respect to the learning platform. The seated students felt more connected, as demonstrated in their comments. Despite all
continuous improvement efforts, professors simply may not have any influence in an asynchronous virtual environment to demonstrate significant change in evaluation scores. The online students missed the connecting relationships founded in trust and rapport with the professor. Many students did not take the time to complete the evaluations and those that did chose Likert-scale values that were consistently lower in mean value and were higher in degree of variation.

Nonetheless, readers may still find hope for future scores by virtue of the potential limitations of the study. For instance, the 316 students involved in this study represent only the population of students in one private university whose ACT test scores approximate the national average. The relatively small sample size might also negatively impact the results and generalizability. Since different universities employ a variety of marketing strategies to attract differing student groups, the student demographics of this study may not align with other universities’ virtual learners. Additionally, the reflective modifications to implement best practices in the online course may have confounded the results. Although the results demonstrated a medium effect size, some may still consider the results less than impactful to the general population. The content area may also limit generalizability as quantitative coursework is generally rated lower than other coursework (Bleske-Rechek & Michels, 2010; Constand & Pace, 2014; Royal & Stockdale, 2015). Finally, as is the case with most research, a longer study involving more students and a different collection of instructors or other evaluation survey instruments could possibly change the results.

A logical extension of this research is to call for further studies concerning evaluation scores in online and seated environments with a larger group of students with a larger range of course topics. Other avenues of related research would be to compare other leaning modalities such as hybrid learning and/or separation of evaluation results by matriculation status (i.e., certificate programs, postsecondary option programs, and other nonmatriculating coursework), full- or part-time instructor status, or content- versus skills-based courses. I would also posit that although the limitations of generalizability for school, sample size, and content exist in this study, it would be worth further exploring the rationale for the lack of connection and its direct relationship to potential differences by synchronicity. I echo the notion expressed by Callister and Love (2016) to encourage future researchers to explore the disconnect between content and outcomes associated with virtual learning.
REFERENCES


McDonald, J. (2002). Is “as good as face-to-face” as good as it gets? Journal of Asynchronous Learning Networks, 6(2), 10–23.


