

6-10-2015

Student Perceptions of Online Learning: An Analysis of Online Course Evaluations

Patrick Lowenthal
Boise State University

Christine Bauer
Boise State University

Ken-Zen Chen
Boise State University

Student Perceptions of Online Learning: An Analysis of Online Course Evaluations

Patrick Lowenthal
Boise State University

Christine Bauer
Boise State University

and

Ken-Zen Chen
Boise State University

Student evaluations of teaching provide a wealth of information about students' experiences in higher education. Colleges and universities, though, as a whole, need to spend more time mining these evaluations to better understand student perceptions of their college coursework. These evaluations are especially helpful to better understand students' experiences in online courses, which, despite continued growth, are still relatively new for most faculty and students. The analysis of seven years of student evaluations at a metropolitan research university is presented in the following article. The purpose of the analysis was to better discern students' experiences online as well as to address commonly held assumptions about online learning (e.g., that teaching evaluations are lower for online courses). Results indicate that students in this sample actually do rate online courses lower than face-to-face courses. The article concludes with situating these results in the larger student evaluation literature as well as addressing larger implications of these results for practice.

Colleges and universities have been conducting student evaluations of teaching for nearly a hundred years (Anderson, Cain, and Bird 2005; also see Remmers 1928; Smalzried and Remmers 1943). Faculty originally collected these evaluations simply for informational purposes about their teaching (Anderson, Cain, and Bird 2005). However, over time these evaluations became viewed as a way for administrators to evaluate teaching effectiveness (Kogan 2014). Today, nearly every college and university conducts some type of student evaluation of teaching (Anderson, Cain, and Bird 2005; Marlin and Niss 1980; Marsh 1984). While student evaluations of teaching began as a way for faculty to improve their teaching, over time they became used for much more—including support for promotion and tenure as well as information to help students make informed decisions about which courses to take (Donovan, Mader, and Shinsky 2010; Kogan 2014; Marsh 1984). Despite the potential value of getting feedback about one's teaching (see Donovan, Mader, and Shinsky 2010), faculty remain critical of student evaluations (Donovan, Mader, and Shinsky 2010; Marlin and Niss 1980) because these evaluations are often the primary or the only way that institutions evaluate teaching effectiveness, which, in turn, can influence such things as yearly evaluations, promotions, and salary decisions (Dziuban and Moskal 2011). Many faculty specifically question the degree to which these evaluations actually measure teaching effectiveness in the first place (Marlin and Niss 1980), as well as how these evaluations are influenced by situational variables (e.g., workload, class size, ease of grading), or, more realistically, how these evaluations are commonly misused (Clayson and Haley 2011). This skepticism has helped fuel one of the most popular and longest ongoing lines of research: the validity of student evaluations of teaching.

Unfortunately, much of the research on student evaluations contradicts other research. For instance, faculty often suspect that inflated grading or requiring less work can improve their teaching evaluations (Heckert et al. 2006). However, while some research has shown that reducing workload and giving higher grades is related to higher teaching evaluations, other research has found just the opposite (Marsh and Roche 2000). Similarly, some research suggests that class size can influence these evaluations. For instance, Grant (1971) found larger classes received better evaluations, whereas Kogan (2014) later found that courses with fewer enrollments had more positive

evaluations. To complicate matters further, research has shown that teaching evaluations can vary across disciplines (Felton et al. 2004; Kogan 2014) as well as gender (Kogan 2014; Sprague and Massoni 2005). In light of how teaching evaluations are used to shape the careers of faculty every day, inconsistent and contradictory results such as these call for continued inquiry. Given this, we set forth to investigate student evaluations of online learning, and specifically to compare evaluations of face-to-face vs. online courses, in order to gain a better understanding of what students think of online learning as well as to answer some common questions faculty have about online education.

Student Evaluations Move Online

Student evaluations of teaching traditionally were conducted using paper forms on one of the last days of class. However, increased use of the Internet prompted many institutions to begin conducting student evaluations online more than ten years ago (Stowell, Addison, and Smith 2012). Completing these evaluations online, instead of in person using paper forms, has the potential to save time, save money, and speed up reporting, while possibly even improving the depth and breadth of student comments (Donovan, Mader, and Shinsky 2010). Despite these potential benefits, faculty initially resisted conducting student evaluations online. Faculty feared this new practice could lower response rates and even increase negative feedback (Donovan, Mader, and Shinsky 2010; Hardy 2003; Norris and Conn 2005).

In comparison to the vast amount of literature on traditional student evaluations of teaching, surprisingly few studies have focused on the mode of delivery of evaluations (Avery et al. 2006). But like most of the research on student evaluations of teaching, the results on mode of delivery are mixed. For instance, while some research showed that evaluations completed online had a lower response rate (Avery et al. 2006; Donovan, Mader, and Shinsky 2010; Layne, DeCristoforo, and McGinty 1999; Stowell, Addison, and Smith 2012), other research found a higher response rate when evaluations were completed online (Carini et al. 2003). Some of this early research also suggested that response rates would increase over time (Donovan, Mader, and Shinsky 2010)—which some suggested happened simply from students becoming more comfortable with the evaluation system (Avery et al. 2006). Oliver and Sautter (2005) as well as Donovan, Mader, and Shinsky (2010) even questioned how using an online learning management system to administer evaluations could increase response rates. Faced with the benefits of online delivery but the risk of lower response rates, some researchers recommended using incentives to increase student response rate. These incentives range from basic things like simply reminding students to complete the evaluations and stressing the importance of these evaluations (Stowell, Addison, and Smith 2012) to more aggressive strategies like withholding the final grade or even giving extra credit to encourage higher completion rates (Avery et al. 2006).

Faculty were also concerned that completing evaluations online might lead to more negative feedback. Early research found that teaching evaluations completed online did not differ from evaluations completed in a traditional format (Avery et al. 2006; Donovan, Mader, and Shinsky 2010; Stowell, Addison, and Smith 2012). In fact, some research suggests that students leave even more feedback (Donovan, Mader, and Shinsky 2010; Hardy 2003; Hmieleski and Champagne 2000; Kasiar, Schroeder, and Holstaad 2002; Morrison 2011; Stowell, Addison, and Smith 2012) and at times even more positive feedback (Carini et al. 2003) when completing evaluations online. Some research even revealed that students prefer completing online evaluations (Hardy 2003).

Teaching Evaluations of Online Courses

Only a handful of studies have looked at student evaluations of online courses and how these evaluations compare to face-to-face courses (Carle 2009; McGhee and Lowell 2003). There are, however, a few notable findings. First and arguably most important, students overall do not rate instructors significantly different in online courses than in face-to-face courses (Carle 2009; Liu 2006; Spooner et al. 1999; Waschull 2001). This finding does not mean that students necessarily prefer one format over the other (see Dziuban and Moskal 2011); simply, this result suggests that—either as a whole or when tracking the same instructor in both formats—there is not an overall statistically significant difference between evaluations in the two formats. Dziuban and Moskal (2011) explain that results like these “suggest that students do not consider course mode an important element when defining the dimensions by which they evaluate their educational experience” (239). However, McGhee and Lowell (2003) did find that “students in online classes had much greater diversity of opinion regarding the effectiveness of the online delivery format than their on-campus peers” (46). Liu (2012) found that students rated elective and major courses

higher than required courses and that first-year students gave the lowest ratings. Liu (2012) also found that students ranked tenured and tenure-track faculty lower than all other types of instructors (e.g., adjuncts and non-tenure-track faculty). As McGhee and Lowell (2003) and Liu (2012) suggest, results like these might simply mean that individual characteristics of the teacher or the student influence evaluations more than whether the course was taken online or not. Interestingly though, other research suggests that some of these faculty teaching skills and dispositions might hold constant over time. For instance, Carle (2009) found that student evaluations of specific instructors showed little change over time; in other words, faculty “tend to receive similar evaluations across their classes” (433). Finally, some of the research on online courses does suggest, in fact, that students might spend more time on online courses than face-to-face courses (McGhee and Lowell 2003) and that class size might not influence evaluations like they once may have (Liu 2012).

While the results from the handful of studies comparing student evaluations of teaching in face-to-face and online courses appear to align with each other, the results from student evaluation research as a whole suggest that additional research is needed to better understand and support the high-stakes use of teaching evaluations.

METHOD

Student evaluations of teaching remain a rich source of data about the classroom experience. As such, some have argued that these evaluations should be used to establish credibility of online learning (Liu 2012). We did not set out in this study to look to these evaluations as a way to document teaching effectiveness or student learning, or to establish the credibility of online learning. We already strongly believe that online learning is a credible format for formal education. We also think student evaluations of teaching are not by themselves adequate measures of teaching effectiveness or student learning. But we do believe that student evaluations are prodigious sources of data about what students think of a specific course and instructor, opinions that are extremely important in today’s competitive higher education market.

With this in mind, we set forth to investigate students’ perceptions of online learning by analyzing student evaluations of teaching (also known as end-of-course evaluations). Student evaluations of teaching, however, are not currently public at our institution as they are at many other institutions. Therefore, we located a comparable metropolitan research university in the region that posted its student evaluations publicly online. We will refer to this institution as Rocky Mountain University (RMU). We downloaded all of the student evaluations at RMU between 2007 and 2013. We then deleted evaluations from courses with less than a 10% response rate. We analyzed a total of 30,131 course evaluations; of those, 82.63% were from face-to-face courses and 17.37% were from online courses (see Table 1). The level of coursework and faculty status also varied across both formats (see Tables 2 and 3). Online courses at RMU also had lower completion rates compared to face-to-face courses. During this time period at RMU, student evaluations for face-to-face courses were paper based and typically administered during the last day of class, whereas student evaluations for online courses were completed online during the last week of class.

<insert table 1 about here>

<insert table 2 about here>

<insert table 3 about here>

The student evaluation instrument at RMU was updated in 2006 to include six universal questions regardless of course format. All of the ratings—except for one question about workload—are based on a scale of 1 = *lowest* to 6 = *highest* (see Figure 1). For the purpose of this study, we focused on analyzing students’ ratings of the course, the instructor, overall workload, availability, and course as a learning experience (see Figure 1).

<insert figure 1 about here>

More specifically, we were interested in answering the following research questions using corresponding inferential test statistics:

1. Is there a statistically significant difference between classroom and online course evaluations for (a) response rate, (b) course overall, (c) instructor overall, (d) grading fairness, (e) instructor accessibility, (f) workload, and (g) course as a learning experience?
2. How do course evaluations change over time?
3. How do course evaluations differ across lower division and upper division undergraduate vs. graduate courses?
4. How do course evaluations differ across faculty type (e.g., tenured/tenure track vs. others)?

RESULTS

Our analysis of the data revealed that students rate face-to-face courses at RMU higher on all measures than online courses. For instance, on a six-point scale, when asked to rate the “course overall,” students rated face-to-face courses as $M = 4.84$ but online courses as $M = 4.57$. Then when asked to rate the “instructor overall,” students rated face-to-face instructors as $M = 5.06$, but online instructors as $M = 4.75$ (see Table 4).

<insert table 4 about here>

Research Question 1: Is there a statistically significant difference between classroom and online course evaluations for the following: (a) response rate, (b) course overall, (c) instructor overall, (d), grading fairness, (e) instructor accessibility, (f) workload, and (g) course as a learning experience?

We found a statistically significant difference ($p < 0.000$) between face-to-face courses and online courses across all measures. However, the amount of variance due to delivery mode differed across each measure. Ultimately, the difference between sample sizes coupled with the significant difference in response rate between the two modes does not necessarily make the variance practically significant.

Research Question 2: How do course evaluations change over time?

While the overall student response rate failed to increase over time (see Table 5), the student ratings of the “overall course” and “overall instructors” have increased since 2007 (see Figure 2). This increase likely is due to a number of factors including students’ becoming more comfortable with the student evaluation instrument (which was changed in 2006), instructors getting better at teaching and designing fully online courses, and students’ becoming more comfortable with what it means to complete courses online.

<insert table 5 about here>

<insert figure 2 about here>

Research Question 3: How do course evaluations differ across lower division and upper division undergraduate vs. graduate courses?

Unlike previous research where first-year students were some of the most critical in their evaluations, at RMU graduate students had significant higher response rate to the course evaluations (54.96%) against lower (42.19%) and upper (42.93%) division students. Moreover, RMU graduate students seem to be less satisfied and/or simply more critical of their online courses and their instructors than undergraduate students (taking lower division or upper division courses), except instructor accessibility ($p = 0.158$). While the differences are relatively minor (e.g., graduate students rated their average course 4.50 on a six-point scale compared to 4.58 by lower division students), any differences should be investigated further in an effort to find reasons such differences might exist.

<insert table 6 about here>

Research Question 4: How do course evaluations differ across faculty type (e.g., tenured/tenure track vs. others)?

Finally, we looked at how end-of-course evaluations differ across faculty type. In other words, how do tenured and tenure-track faculty compare to instructors and adjuncts? Although there is no significant difference between instructors/adjuncts ($p = 0.146$), students rated tenured and tenure-track faculty lower than course instructors/adjuncts (see Table 7). This finding is similar to previous research (see Liu 2012) and supports the argument that tenured and tenure-track faculty might spend less time focusing on their teaching when their scholarship is ultimately what will get them promoted and tenure.

<insert table 7 about here>

DISCUSSION

Student evaluations of teaching provide a wealth of information about the student experience in higher education. Institutions as a whole need to spend more time mining these data to help make data-informed decisions about how to improve coursework and instruction rather than simply using them as a justification for promotion and tenure or dismissal of faculty.

Student evaluations currently are not publicly available at our institution. Therefore, in an effort to answer common faculty questions about online learning (e.g., Will one's teaching evaluations be rated lower if he or she teaches courses online? Do students spend less time online?), to better understand the student experience, and finally to be better prepared to analyze our own institution's student evaluations in the future, we decided to investigate a peer institution's end-of-course student evaluations.

The results of our inquiry—much like the larger body of literature on student evaluations—leave us with more questions than answers. First and foremost, we found that students at RMU actually do rate instructors lower in online courses than face-to-face courses (i.e., based on ratings of “overall instructor” and “overall course”). And while the differences are not practically significant, they might be enough to discourage some faculty from teaching online. The differences, however, are minor, and we suspect these differences would not keep someone from getting promoted and tenure. Further, we posit that faculty who care about their teaching and, thus, student learning will be successful in any learning environment. It might take a few semesters to learn the differences with teaching online, but good teachers will become competent in an online environment. So while claiming that “good teaching is good teaching” might oversimplify the differences with teaching face-to-face and online, we do strongly believe that good teachers are good teachers, regardless of the learning environment.

Perhaps most troubling but not surprising (given the history of research on teaching evaluations) is that these results contradict earlier research. In other words, while students at RMU rated their instructors lower across online courses than face-to-face courses, previous research (Carle 2009; Liu 2006; Spooner et al. 1999; Waschull 2001) found just the opposite: students do not rate instructors lower in online course. At the same time, students at RMU reported the appropriateness of workload lower in online vs. face-to-face courses, which could suggest they are spending more time on online courses, as McGhee and Lowell (2003) previously found in their study. The lack of alignment between different university student evaluations makes it difficult to form many firm comparisons in the end.

The number one limitation of this study is the difference in response rates between the two groups. During 2007–2013, students completed 75% of evaluations for face-to-face courses but only 43% for online courses. While evaluations conducted online historically have been a little lower than those completed face-to-face (Avery et al. 2006; Donovan, Mader, and Shinsky 2010; Layne, DeCristoforo, and McGinty 1999; Stowell, Addison, and Smith 2012), making any firm conclusions about student perceptions of online learning is nearly impossible when almost 60% of students are not completing the evaluations. As useful as student evaluations are, they remain one piece of data; ultimately, these results should be triangulated with additional data to help colleges and universities better understand what is happening in online courses.

The results of this study should not be generalized to the broader population. Despite certain similarities between RMU and Boise State, there are notable differences. Perhaps most important is that online development models essentially lie on a continuum from highly centralized to highly decentralized (Lowenthal and White 2009). While RMU has a centralized unit to administer the learning management system and provide tier 1 and tier 2 support for faculty and students, this centralized unit provides very little direct instructional design support to help faculty design and develop fully online courses. Therefore, even though RMU hosts multiple events and trainings throughout the year to support faculty as they design and develop their own courses (see Lowenthal and Thomas 2010), RMU ultimately leaves online course development up to individual faculty and, thus, lies on the decentralized end of the continuum. Boise State has a centralized unit for online learning called the eCampus Center. However, the eCampus Center provides more direct instructional design support to develop fully online courses as well as ongoing seminars to teach those who want to learn how to teach online. While faculty at Boise State can still design and develop their own online courses (and many still do), a growing number of faculty turn to the eCampus Center for help, which places it closer to the centralized end of the continuum. As such, these institutional and cultural differences suggest that faculty and students at these institutions might be working in very different environments. Differences between RMU and Boise State are not unusual. In fact, no two institutions of higher education are exactly alike. Therefore, as interesting and in many ways as useful as it was to investigate another institution's student evaluations of teaching, the real value of mining student evaluation data lies in investigating one's own institution's student evaluations and using these data to make informed improvements over time.

IMPLICATIONS FOR DISTANCE EDUCATION PRACTICE

Colleges and universities need to spend more time mining student evaluations of teaching. These evaluations can provide useful metrics to make data--driven decisions. We set forth to evaluate student evaluations at a comparable metropolitan research university, in part to help inform the future analysis of our own student evaluations. We learned a few things through this process that have implications for distance education practice.

Teaching Evaluations Need to Be Fine--Tuned

Teaching evaluation instruments at colleges and universities throughout the country need to be fine--tuned. While some argue that the differences between teaching face--to--face and online warrant different instruments with specific questions focused on the differences of each learning environment, we believe that keeping a core set of questions to ask all students, regardless of the format of the course, is imperative. Consistent questions enable comparisons across formats, comparisons across sections of the same courses, and comparisons across individual instructors. As the boundaries or distinctions between online and face--to--face courses continue to blur, identifying key components of any learning experience and seeking students' feedback on those key components will become even more important.

At the same time, institutions should strive to give faculty the opportunity for input, not only on the core questions asked, but also to give departments as well as individual faculty the ability to add additional course or discipline--specific questions as needed to the evaluation. However, as important as it is to regularly review and improve one's instrument and to strive for consistency across formats, institutions also need to finalize their instruments as quickly as possible to enable the collection of longitudinal data and to increase students' familiarity with these instruments.

Response Rates Need to be Improved

Colleges and universities need to increase the response rate on *all* student evaluations, but especially on evaluations for online courses. Faculty can use a number of different incentives to help increase the response rate for online courses. For instance, research suggests that assigning points or giving extra credit for completing evaluations can increase response rates. Others suggest alternatives such as dropping the lowest grade for an assignment (e.g., weekly quizzes) if a certain percentage of students in the course complete the evaluation (e.g., 95%). Still others have found that threatening students that final grades will not be posted until everyone completes the evaluations can be an effective strategy to increase response rates. While incentives like

these might prove to be effective, research also suggests that simply reminding students to complete the evaluation can help increase response rates. One strategy used at Boise State is to ask students to acknowledge that they completed the evaluation in a specific discussion forum during the last week of class.

In addition to the aforementioned strategies faculty can use, there are things that colleges and universities can do at the institution level to improve response rates. For instance, the platform used to administer the student evaluations can also influence response rate. Boise State started using a product by CollegeNET to administer course evaluations. One key feature of this product (though other products on the market offer similar functionality) is that faculty can send reminder emails to only those students who have not completed the evaluation. The first author of this article found that his response rate improved by 10–20% once his department started using CollegeNET and he was able to send targeted reminders during the last few days of class. Similarly, institutions need to strive to have one consistent system across the university and then help establish a culture that publicly shares, values, and actually uses the results of these evaluations. Students complete evaluations on *ratemyprofessor.com* because they help other students make decisions about what instructors they should seek out or avoid. RMU publically shared student evaluations of teaching, but these results were not widely promoted or shared and in many ways remain hidden and unused. Unfortunately, many universities have no interest in sharing student evaluations publically (Hoffman 2003; Marsh 1984), which further prevents students from seeing the utility of these types of evaluations.

Improve Instruction

Student evaluations were originally collected to help instructors improve their instruction. The problem with most student evaluations is that faculty often might not get the results until weeks, if not months, after the course is over. This becomes especially problematic for faculty who might teach a specific course only once a year. Faculty not only need timely (i.e., weeks after a course, not months) summative feedback, but they also need formative feedback during the semester. Therefore, more and more institutions need to make midterm evaluations a standard operating procedure. Midterm evaluations or any type of feedback survey during a course enables faculty to make just--in--time adjustments to a course that can help improve student satisfaction and student learning (Dobrovolny and Lowenthal 2011).

At the same time, student evaluations offer only one perspective on teaching effectiveness. Ultimately student evaluations should be combined with other forms of evaluation, such as self--evaluation (discussed next) as well as peer evaluation. Multiple forms of evaluation become especially important for faculty who find themselves teaching difficult courses in which they might be unjustifiably rated lower simply due to the rigorous nature of the course.

Finally, these evaluations should be used by centers of teaching and learning to identify in what areas faculty might need additional support. In other words, if a specific course has high attrition rates and low student evaluations, faculty developers and/or instructional designers can work with faculty to improve the course and/or instruction to improve the learning experience for students.

Faculty Evaluations

Last, but not least, colleges and universities spend a great deal of time and money on conducting student evaluations of teaching, often under the guise that they are used to improve teaching and learning. However, one shortcoming of nearly all end--of--course evaluations is that they take only the students' perspectives into account. While some have argued that a course is a course is a course, not all courses are actually the same and, at times, even the students in specific sections of a course can differ drastically. Therefore, institutions need to spend more time collecting faculty feedback on courses. This factor becomes even more important as more and more faculty are teaching courses designed by other faculty. Whether at the department, school, or college level, institutions need to collect data formally from those in the trenches. What worked? What didn't? What should be changed? How much time did he or she spend each week teaching the course? What was the workload? Do you feel like students met the outcomes of the course? Questions like these should be asked every semester for every course. These types of data, in conjunction with student evaluations and peer evaluations, can be used to improve the faculty and student experience in all courses—but especially in fully online courses.

CONCLUSION

Institutions have been collecting student evaluations of teaching for decades. While student evaluations of teaching are possibly the most researched topic in education, the results of this line of research are inconclusive—in part because of the narrow focus of much of this research. Rather than focus solely on the validity of student evaluations to document teaching effectiveness, more researchers should investigate these evaluations simply as indicators of student satisfaction with their learning experience. As such, these evaluations can provide valuable metrics to make data-driven decisions about one's courses, programs, and institution.

We set forth to investigate student evaluations of online learning, and, specifically, to compare evaluations of face-to-face vs. online courses, to gain a better understanding of what students think of online courses, to answer some common questions faculty have about online education, and to help inform the future analysis of our own student evaluations. Through this process, we found that at RMU online courses are rated lower than face-to-face courses and while these differences were statistically significant, they were not practically significant. Further, the difference in response rate alone questions the usefulness of these comparisons. These results, coupled with previous research, call for the need for additional research on this topic. These results should not be generalized to the larger public; they are institution specific. Instead, every college and university needs to take time to mine its own student evaluation data to help make data-driven improvements to its online courses and programs, thus continuing to improve the quality of the online courses and programs offered. We are left optimistic about what we might learn as we begin the systematic analysis of students' evaluations at our institution.

REFERENCES

- Anderson, H. M., J. Cain, and E. Bird. 2005. Online student course evaluations: Review of literature and a pilot study. *American Journal of Pharmaceutical Education* 69 (1): 34–43.
- Avery, R. J., W. K. Bryant, A. Mathios, H. Kang, and D. Bell. 2006. Electronic course evaluations: Does an online delivery system influence student evaluations? *The Journal of Economic Education* 37(1): 21–37.
- Carini, R. M., J. C. Hayek, G. D. Kuh, J. M. Kennedy, and J. A. Ouimet. 2003. College student responses to web and paper surveys: Does mode matter? *Research in Higher Education* 44 (1): 1–19.
- Carle, A. C. 2009. Evaluating college students' evaluations of a professor's teaching effectiveness across time and instruction mode (online vs. face-to-face) using a multilevel growth modeling approach. *Computers & Education* 53 (2): 429–435.
- Clayson, D. E., and D. A. Haley. 2011. Are students telling us the truth? A critical look at the student evaluation of teaching. *Marketing Education Review* 21 (2): 101–112.
- Dobrovolny, J., and P. R. Lowenthal. 2011. Using reflection surveys to improve teaching and learning. *The CU online handbook*, ed. P. R. Lowenthal, D. Thomas, A. Thai, B. Yuhnke, M. Edwards, and C. Gasell, 97–113. Raleigh, NC: Lulu Enterprises.
- Donovan, J., C. E. Mader, and J. Shinsky. 2010. Constructive student feedback: Online vs. traditional course evaluations. *Journal of Interactive Online Learning* 9 (3): 283–296.
- Dziuban, C., and P. Moskal. 2011. A course is a course is a course: Factor invariance in student evaluation of online, blended and face-to-face learning environments. *The Internet and Higher Education* 14 (4): 236–241.
- Grant, C. W. 1971. Faculty allocation of effort and student course evaluations. *The Journal of Educational Research* 64 (9): 405–410.
- Hardy, N. 2003. Online ratings: Fact and fiction. *New Directions for Teaching and Learning* 2003 (96): 31–38.
- Heckert, T. M., A. Latier, A. Ringwald-Burton, and C. Drazen. 2006. Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to 'buy' better evaluations through lenient grading? *College Student Journal* 40 (3): 588–596.
- Hmieleski, K., and M. V. Champagne. 2000. Plugging in to course evaluation. *The Technology Source Archives at the University of North Carolina*. Available online at http://technologysource.org/article/plugging_in_to_course_evaluation/
- Hoffman, K. M. 2003. Online course evaluation and reporting in higher education. *New Directions for Teaching and Learning* 2003 (96): 25–29.

- Kasiar, J. B., S. L. Schroeder, and S. G. Holstad. 2002. Comparison of traditional and web--- based course evaluation processes in a required, team---taught pharmacotherapy course. *American Journal of Pharmaceutical Education* 66 (3): 268–270.
- Kogan, J. 2014. Student course evaluation: Class size, class level, discipline, and gender bias. Paper presented at the 6th International Conference on Computer Supported Education, Barcelona, Spain, April 1–3. Available online at https://files.ifi.uzh.ch/stiller/CLOSER%202014/CSEDU/CSEDU/Learning/Teaching%20Methodologies%20and%20Assessment/Short%20Papers/CSEDU_2014_140_C R.pdf
- Layne, B. H., J. R. DeCristoforo, and D. McGinty. 1999. Electronic versus traditional student ratings of instruction. *Research in Higher Education* 40 (2): 221–232.
- Liu, O. L. 2012. Student evaluation of instruction: In the new paradigm of distance education. *Research in Higher Education* 53 (4): 471–486.
- Liu, Y. 2006. A comparison study of online versus traditional student evaluation of instruction. *International Journal of Instructional Technology & Distance Learning* 3 (4): 15–29.
- Lowenthal, P. R., and D. Thomas. 2010. Digital campfires: Innovations in helping faculty explore the online learning wildness. *Journal of Online Learning and Teaching* 6 (3): 665–672.
- Lowenthal, P. R., and J. W. White. 2009. Enterprise model. In *Encyclopedia of distance and online learning*, ed. P. Rogers, G. Berg, J. Boettcher, C. Howard, L. Justice, and K. Schenk, 932–936. Hershey, PA: IGI Global.
- Marlin, Jr., J. W., and J. F. Niss. 1980. End---of---course evaluations as indicators of student learning and instructor effectiveness. *Journal of Economic Education* 11 (2): 16–27.
- Marsh, H. W. 1984. Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of Educational Psychology* 76 (5): 707.
- McGhee, D. E., and N. Lowell. 2003. Psychometric properties of student ratings of instruction in online and on---campus courses. *New Directions for Teaching and Learning* 2003 (96): 39–48.
- Morrison, R. 2011. A comparison of online versus traditional student end---of---course critiques in resident courses. *Assessment & Evaluation in Higher Education* 36 (6): 627–641.
- Norris, J., and C. Conn. 2005. Investigating strategies for increasing student response rates to online---delivered course evaluations. *Quarterly Review of Distance Education* 6 (1): 13–29.
- Oliver, R. L., and E. P. Sautter. 2005. Using course management systems to enhance the value of student evaluations of teaching. *Journal of Education for Business* 80 (4): 231–234.
- Remmers, H. H. 1928. The relationship between students' marks and student attitude toward instructors. *School & Society* 28:759–760.
- Smalzried, N. T., and H. H. Remmers. 1943. A factor analysis of the Purdue Rating Scale for Instructors. *Journal of Educational Psychology* 34 (6): 363.
- Spooner, F., L. Jordan, B. Algozzine, and M. Spooner. 1999. Student ratings of instruction in distance learning and on---campus classes. *The Journal of Educational Research* 92 (3): 132–140.
- Sprague, J., and K. Massoni. 2005. Student evaluations and gendered expectations: What we can't count can hurt us. *Sex Roles* 53 (11–12): 779–793.
- Stowell, J. R., W. E. Addison, and J. L. Smith. 2012. Comparison of online and classroom---based student evaluations of instruction. *Assessment & Evaluation in Higher Education* 37 (4): 465–473.
- Waschull, S. B. 2001. The online delivery of psychology courses: Attrition, performance, and evaluation. *Teaching of Psychology* 28 (2): 143–147.

Course overall: Rate the course compared to all others [Scale 1 (lowest) to 6 (highest)]

Instructor overall: Rate the instructor compared to all others [Scale 1 (lowest) to 6 (highest)]

Workload: Rate the appropriateness of the workload relative to the credit given.

Availability: Rate the instructor's availability for course-related assistance such as e-mail, office hours, individual appointments, phone contact, etc. [Scale 1 (lowest) to 6 (highest)]

Course as a learning experience: Rate how much you have learned in this course. [Scale 1 (lowest) to 6 (highest)]

FIGURE 1 End-Of-Course Evaluation Questions Analyzed In This Study.

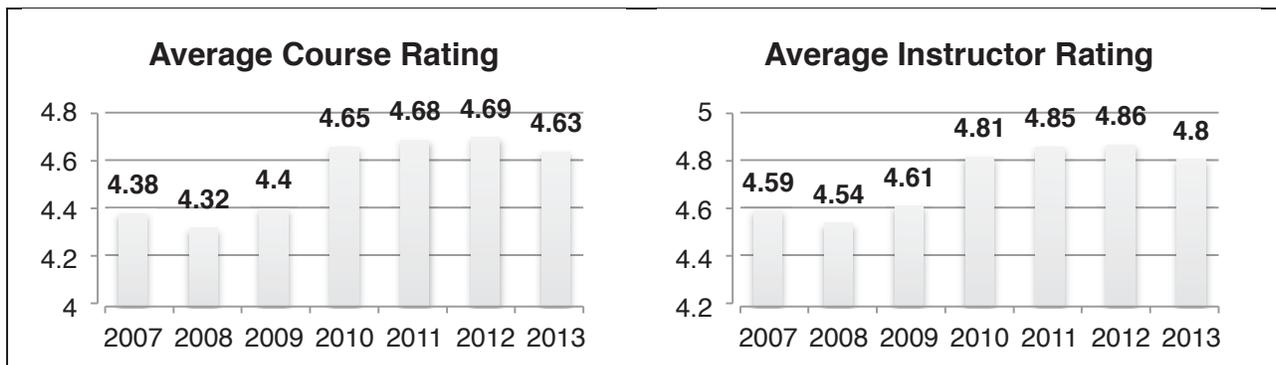


FIGURE 2. End-of-Course Evaluations Over Time.

TABLE 1
General Descriptive Data

	Average course size	Total completed evaluations	Percentage of evaluations completed
Face-to-face	25.74	24,896 (82.63%)	75
Online	26.03	5,235 (17.37%)	43

TABLE 2
Level of Coursework

	Lower division	Upper division	Graduate
Face-to-face	9,726 (39.07%)	8,788 (35.30%)	6,382 (25.63%)
Online	1,714 (32.74%)	1,695 (32.38%)	1,826 (34.88%)

TABLE 3
Faculty Status

	Tenure or tenure-track	Other*
Face-to-face	7,344 (29.50%)	17,552 (70.50%)
Online	1,180 (22.54%)	4,055 (77.46%)

*These include lecturers, adjuncts, and teaching assistants.

TABLE 4
Average Scores of All Face-to-Face vs. Online Course Evaluations

	Course overall	Instructor overall	Grading fairness	Instructor access	Workload	Course as learning experience
Face-to-face	4.84	5.06	5.19	5.27	4.99	5.09
Online	4.57	4.75	5.02	5.07	4.85	4.85

TABLE 5
Response Rate for Student Evaluations for Online Courses

Year	Average response rate
2007	42%
2008	39%
2009	47%
2010	47%
2011	46%
2012	42%
2013	39%

TABLE 6
End-of-Course Evaluations Across Student Level

Level	Course (p < 0.000)	Instructor (P = 0.001)	Grade fairness (p = 0.003)	Workload (p < 0.000)	Learning experience (P < 0.000)
Graduate	<i>M</i> = 4.50 (<i>SD</i> = 0.81)	<i>M</i> = 4.70 (<i>SD</i> = 0.82)	<i>M</i> = 4.99 (<i>SD</i> = 0.63)	<i>M</i> = 4.72 (<i>SD</i> = 0.67)	<i>M</i> = 4.78 (<i>SD</i> = 0.75)
Lower division	<i>M</i> = 4.58 (<i>SD</i> = 0.88)	<i>M</i> = 4.78 (<i>SD</i> = 0.86)	<i>M</i> = 5.06 (<i>SD</i> = 0.70)	<i>M</i> = 4.92 (<i>SD</i> = 0.74)	<i>M</i> = 4.89 (<i>SD</i> = 0.77)
Upper division	<i>M</i> = 4.63 (<i>SD</i> = 0.83)	<i>M</i> =4.79 (<i>SD</i> =0.82)	<i>M</i> = 5.02 (<i>SD</i> = 0.67)	<i>M</i> = 4.91 (<i>SD</i> = 0.68)	<i>M</i> = 4.90 (<i>SD</i> = 0.75)

TABLE 7
End-of-course Evaluations Across Faculty Type

Instructor type	Course (p = 0.007)	Instructor	Grade fairness (p = 0.015)	Workload (p < 0.000)	Learning experience (p = 0.009)
Instructors/adjuncts	<i>M</i> = 4.59 (<i>SD</i> = 0.85)		<i>M</i> = 5.04 (<i>SD</i> = 0.66)	<i>M</i> = 4.87 (<i>SD</i> = 0.71)	<i>M</i> = 4.87 (<i>SD</i> = 0.76)
Tenured & tenure- track	<i>M</i> = 4.50 (<i>SD</i> = 0.81)		<i>M</i> = 4.97 (<i>SD</i> = 0.64)	<i>M</i> = 4.77 (<i>SD</i> = 0.68)	<i>M</i> = 4.79 (<i>SD</i> = 0.75)