



Debunking a Few Myths about Course Evaluations

Miller, W., & Rinker, T. (2019). Myths and realities: An examination of course evaluations in higher education. *Data in Action Series, Campus Labs*. Retrieved from: <https://www.campuslabs.com/data-in-action/myths-and-realities-an-examination-of-course-evaluations-in-higher-education/>

Have you ever wondered whether students who complete course evaluations outside of class time are more likely to be critical in their comments and ratings? Or, do you believe that student written comments only reflect the extremes and consequently are not constructive? Recently, researchers at Campus Labs sought to answer these and other questions by analyzing course evaluations collected from 12 U.S. institutions of higher education. The schools were geographically dispersed, represented various accrediting bodies, and included both two- and four-year institutions. More than 216,000 respondents provided 2.3 million class ratings.

Specifically, the authors examined six commonly held beliefs about course evaluations in an attempt to assess the veracity of each. Their findings are described in the following paragraphs:

Belief #1: Students who complete course evaluations outside of class time are more likely to be critical in their comments and ratings. Prior to analyzing the data, all time zones were adjusted to Eastern Standard Time. Then, the authors arbitrarily set scheduled class time hours from 8 a.m. to 4 p.m. Contrary to popular belief, plots of average class ratings by time of day revealed that evaluations completed during the 8-to-4 timeframe were actually more critical in average ratings than were those outside of it. However, the effect was limited to four-year institutions. No meaningful time-of-day differences were found within two-year institutions.

Belief #2: Comments on course evaluations only reflect the extremes and consequently are not constructive. To examine this belief, the authors plotted histograms of students who left comments and those who did not by their ratings of the instructor (1 to 5 scale). For both two- and four-year institutions, ratings by both those who left comments and those who did not were negatively skewed. Thus, the notion of bimodal extremism is a myth.

Belief #3: Course evaluation instruments do not accurately measure what faculty and administrators want. To investigate the validity of this statement, the authors classified course evaluation questions into 23 separate categories, such as student growth, assessment, instructor behaviors, course design, and facilities. Survey instruments at two-year institutions tended to include questions about the effectiveness of the instructor, whereas those four-year schools focused more on courses, assessment, and students. If these differences reflect what faculty and administrators at those institutions want to know, then the instruments are measuring what they intend to measure. If not, then institutions need to be more intentional about the questions they ask.

Belief #4: Low response rates skew course evaluation results. To address this belief, the authors looked at discrepancies in response rates for the same instructor teaching the same course on two occasions. They arbitrarily considered a response rate gap of 35 percent as a means for distinguishing between classes with high and low response rates. Slightly higher, although not meaningfully different, ratings were found in classes with high compared to low response rates. However, regression modeling revealed that response rate explained only 1.35 percent of variance in ratings.

Belief #5: Respondents have a consistent attitude across different evaluations. This belief holds that students are not very discerning in their evaluations. To examine if this is a myth, analyses were performed on ratings from 191,755 students who completed more than one course evaluation. A histogram of standard deviations of ratings within respondents was positively skewed, indicating that the overwhelming majority of students offer unique perspectives when evaluating different courses and instructors. Those at two-year institutions, however, were more likely to offer more consistent ratings across multiple courses.

Belief #6: Evaluation comments can be a predictor for average ratings on course evaluations. The authors began by classifying the sentiment level (positive, negative) of each comment, utilizing a dictionary-based approach to sentiment classification. They then randomly split the course ratings into training and test datasets, using an 80 percent and 20 percent split, respectively. Simple linear modeling was employed to predict whether or not comment sentiment predicted course evaluation ratings. Word count and hour of the day that the comment was written were also included as explanatory variables. Results indicated that more positive sentiments, fewer words, and comments written later in the day predicted higher ratings. The full model explained 10.75 percent of variance in average ratings.

In conclusion, the authors recommend that campuses have data-informed discussions about course evaluations—why they conduct them, how they use them, and what value they provide to students, faculty, and administrators. Moreover, instructor concerns

about course evaluations should be addressed. Together, faculty and administrators can examine those concerns and inform stakeholders about course evaluation myths and realities.