Comparison of Student Evaluations of Teaching With Online and Paper-Based Administration

Claudia J. Stanny[1] and James E. Arruda[2]

[1] Center for University Teaching, Learning, and Assessment, University of West Florida

[2] Department of Psychology, University of West Florida

Author Note

Data collection and preliminary analysis were sponsored by the Office of the Provost and the Student Assessment of Instruction Task Force. Portions of these findings were presented as a poster at the 2016 National Institute on the Teaching of Psychology, St. Pete Beach, Florida, United States. We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Claudia J. Stanny, Center for University Teaching, Learning, and Assessment, University of West Florida, Building 53, 11000 University Parkway, Pensacola, FL 32514, United States. Email: cstanny@institution.edu

Abstract

When institutions administer student evaluations of teaching (SETs) online, response rates are lower relative to paper-based administration. We analyzed average SET scores from 364 courses taught during the fall term in 3 consecutive years to determine whether administering SET forms online for all courses in the 3rd year changed the response rate or the average SET score. To control for instructor characteristics, we based the data analysis on courses for which the same instructor taught the course in each of three successive fall terms. Response rates for face-to-face classes declined when SET administration occurred only online. Although average SET scores were reliably lower in Year 3 than in the previous 2 years, the magnitude of this change was minimal (0.11 on a five-item Likert-like scale). We discuss practical implications of these findings for interpretation of SETs and the role of SETs in the evaluation of teaching quality.

Keywords: college teaching, student evaluations of teaching, online administration, response rate, assessment

Comparison of Student Evaluations of Teaching With Online and Paper-Based Administration

Student ratings and evaluations of instruction have a long history as sources of information about teaching quality (Berk, 2013). Student evaluations of teaching (SETs) often play a significant role in high-stakes decisions about hiring, promotion, tenure, and teaching awards. As a result, researchers have examined the psychometric properties of SETs and the possible impact of variables such as race, gender, age, course difficulty, and grading practices on average student ratings (Griffin et al., 2014; Nulty, 2008; Spooren et al., 2013). They have also examined how decision makers evaluate SET scores (Boysen, 2015a, 2015b; Boysen et al., 2014; Dewar, 2011). In the last 20 years, considerable attention has been directed toward the consequences of administering SETs online (Morrison, 2011; Stowell et al., 2012) because low response rates may have implications for how decision makers should interpret SETs.

Online Administration of Student Evaluations

Administering SETs online creates multiple benefits. Online administration enables instructors to devote more class time to instruction (vs. administering paper-based forms) and can improve the integrity of the process. Students who are not pressed for time in class are more likely to reflect on their answers and write more detailed comments (Morrison, 2011; Stowell et al., 2012; Venette et al., 2010). Because electronic aggregation of responses bypasses the time-consuming task of transcribing comments (sometimes written in challenging handwriting), instructors can receive summary data and verbatim comments shortly after the close of the term instead of weeks or months into the following term.

Despite the many benefits of online administration, instructors and students have expressed concerns about online administration of SETs. Students have expressed concern that their responses are not confidential when they must use their student identification number to log into the system (Dommeyer et al., 2002). However, breaches of confidentiality can occur even with paper-based administration. For example, an instructor might recognize student handwriting (one reason some

students do not write comments on paper-based forms), or an instructor might remain present during SET administration (Avery et al., 2006).

In-class, paper-based administration creates social expectations that might motivate students to complete SETs. In contrast, students who are concerned about confidentiality or do not understand how instructors and institutions use SET findings to improve teaching might ignore requests to complete an online SET (Dommeyer et al., 2002). Instructors in turn worry that low response rates will reduce the validity of the findings if students who do not complete an SET differ in significant ways from students who do (Stowell et al., 2012). For example, students who do not attend class regularly often miss class the day that SETs are administered. However, all students (including nonattending students) can complete the forms when they are administered online. Faculty also fear that SET findings based on a low-response sample will be dominated by students in extreme categories (e.g., students with grudges, students with extremely favorable attitudes), who may be particularly motivated to complete online SETs, and therefore that SET findings will inadequately represent the voice of average students (Reiner & Arnold, 2010).

Effects of Format on Response Rates and Student Evaluation Scores

The potential for biased SET findings associated with low response rates has been examined in the published literature. In findings that run contrary to faculty fears that online SETs might be dominated by low-performing students, Avery et al. (2006) found that students with higher grade-point averages (GPAs) were more likely to complete online evaluations. Likewise, Jaquett et al. (2017) reported that students who had positive experiences in their classes (including receiving the grade they expected to earn) were more likely to submit course evaluations.

Institutions can expect lower response rates when they administer SETs online (Avery et al., 2006; Dommeyer et al., 2002; Morrison, 2011; Nulty, 2008; Reiner & Arnold, 2010; Stowell et al., 2012; Venette et al., 2010). However, most researchers have found that the mean SET rating does not change

significantly when they compare SETs administered on paper with those completed online. These findings have been replicated in multiple settings using a variety of research methods (Avery et al., 2006; Dommeyer et al., 2004; Morrison, 2011; Stowell et al., 2012; Venette et al., 2010).

Exceptions to this pattern of minimal or nonsignificant differences in average SET scores appeared in Nowell et al. (2010) and Morrison (2011), who examined a sample of 29 business courses. Both studies reported lower average scores when SETs were administered online. However, they also found that SET scores for individual items varied more within an instructor when SETs were administered online versus on paper. Students who completed SETs on paper tended to record the same response for all questions, whereas students who completed the forms online tended to respond differently to different questions. Both research groups argued that scores obtained online might not be directly comparable to scores obtained through paper-based forms. They advised that institutions administer SETs entirely online or entirely on paper to ensure consistent, comparable evaluations across faculty.

Each university presents a unique environment and culture that could influence how seriously students take SETs and how they respond to decisions to administer SETs online. Although a few large-scale studies of the impact of online administration exist (Reiner & Arnold, 2010; Risquez et al., 2015), a local replication answers questions about characteristics unique to that institution and generates evidence about the generalizability of existing findings.

Purpose of the Present Study

In the present study we examined patterns of responses for online and paper-based SET scores at a midsized, regional, comprehensive university in the United States. We posed two questions: First, does the response rate or the average SET score change when an institution administers SET forms online instead of on paper? Second, what is the minimal response rate required to produce stable average SET scores for an instructor? Whereas much earlier research relied on small samples often

limited to a single academic department, we gathered SET data on a large sample of courses (*N* = 364) that included instructors from all colleges and all course levels over 3 years. We controlled for individual differences in instructors by limiting the sample to courses taught by the same instructor in all 3 years. The university offers nearly 30% of course sections online in any given term, and these courses have always administered online SETs. This allowed us to examine the combined effects of changing the method of delivery for SETs (paper-based to online) for traditional classes and changing from a mixed method of administering SETs (paper for traditional classes and online for online classes in the first 2 years of data gathered) to uniform use of online forms for all classes in the final year of data collection.

Method

Sample

Response rates and evaluation ratings were retrieved from archived course evaluation data. The archive of SET data did not include information about personal characteristics of the instructor (gender, age, or years of teaching experience), and students were not provided with any systematic incentive to complete the paper or online versions of the SET. We extracted data on response rates and evaluation ratings for 364 courses that had been taught by the same instructor during three consecutive fall terms (2012, 2013, and 2014).

The sample included faculty who taught in each of the five colleges at the university: 109 instructors (30%) taught in the College of Social Science and Humanities, 82 (23%) taught in the College of Science and Engineering, 75 (21%) taught in the College of Education and Professional Studies, 58 (16%) taught in the College of Health, and 40 (11%) taught in the College of Business. Each instructor provided data on one course. Approximately 259 instructors (71%) provided ratings for face-to-face courses, and 105 (29%) provided ratings for online courses, which accurately reflects the proportion of face-to-face and online courses offered at the university. The sample included 107 courses (29%) at the

beginning undergraduate level (1st- and 2nd-year students), 205 courses (56%) at the advanced undergraduate level (3rd- and 4th-year students), and 52 courses (14%) at the graduate level.

Instrument

The course evaluation instrument was a set of 18 items developed by the state university system. The first eight items were designed to measure the quality of the instructor, concluding with a global rating of instructor quality (Item 8: "Overall assessment of instructor"). The remaining items asked students to evaluate components of the course, concluding with a global rating of course organization (Item 18: "Overall, I would rate the course organization"). No formal data on the psychometric properties of the items are available, although all items have obvious face validity.

Students were asked to rate each instructor as *poor* (0), *fair* (1), *good* (2), *very good* (3), or *excellent* (4) in response to each item. Evaluation ratings were subsequently calculated for each course and instructor. A median rating was computed when an instructor taught more than one section of a course during a term.

The institution limited our access to SET data for the 3 years of data requested. We obtained scores for Item 8 ("Overall assessment of instructor") for all 3 years but could obtain scores for Item 18 ("Overall, I would rate the course organization") only for Year 3. We computed the correlation between scores on Item 8 and Item 18 (from course data recorded in the 3rd year only) to estimate the internal consistency of the evaluation instrument. These two items, which serve as composite summaries of preceding items (Item 8 for Items 1–7 and Item 18 for Items 9–17), were strongly related, $r(362) = .92$. Feistauer and Richter (2016) also reported strong correlations between global items in a large analysis of SET responses.

Design

This study took advantage of a natural experiment created when the university decided to administer all course evaluations online. We requested SET data for the fall semesters for 2 years

preceding the change, when students completed paper-based SET forms for face-to-face courses and online SET forms for online courses, and data for the fall semester of the implementation year, when students completed online SET forms for all courses. We used a 2 × 3 × 3 factorial design in which course delivery method (face to face and online) and course level (beginning undergraduate, advanced undergraduate, and graduate) were between-subjects factors and evaluation year (Year 1: 2012, Year 2: 2013, and Year 3: 2014) was a repeated-measures factor. The dependent measures were the response rate (measured as a percentage of class enrollment) and the rating for Item 8 ("Overall assessment of instructor").

Data analysis was limited to scores on Item 8 because the institution agreed to release data on this one item only. Data for scores on Item 18 were made available for SET forms administered in Year 3 to address questions about variation in responses across items. The strong correlation between scores on Item 8 and scores on Item 18 suggested that Item 8 could be used as a surrogate for all the items. These two items were of particular interest because faculty, department chairs, and review committees frequently rely on these two items as stand-alone indicators of teaching quality for annual evaluations and tenure and promotion reviews.

Results

Response Rates

Response rates are presented in Table 1. The findings indicate that response rates for face-to-face courses were much higher than for online courses, but only when face-to-face course evaluations were administered in the classroom. In the Year 3 administration, when all course evaluations were administered online, response rates for face-to-face courses declined (M = 47.18%, SD = 20.11), but were still slightly higher than for online courses (M = 41.60%, SD = 18.23). These findings produced a statistically significant interaction between course delivery method and evaluation year, $F(1.78, 716)$ =

101.34, *MSE* = 210.61, $p < .001$.[1] The strength of the overall interaction effect was .22 (η_p^2). Simple main-effects tests revealed statistically significant differences in the response rates for face-to-face courses and online courses for each of the 3 observation years.[2] The greatest differences occurred during Year 1 ($p < .001$) and Year 2 ($p < .001$), when evaluations were administered on paper in the classroom for all face-to-face courses and online for all online courses. Although the difference in response rate between face-to-face and online courses during the Year 3 administration was statistically reliable (when both face-to-to-face and online courses were evaluated with online surveys), the effect was small ($\eta_p^2 = .02$). Thus, there was minimal difference in response rate between face-to-face and online courses when evaluations were administered online for all courses. No other factors or interactions included in the analysis were statistically reliable.

Evaluation Ratings

The same 2 × 3 × 3 analysis of variance model was used to evaluate mean SET ratings. This analysis produced two statistically significant main effects. The first main effect involved evaluation year, $F(1.86, 716) = 3.44$, *MSE* = 0.18, $p = .03$ ($\eta_p^2 = .01$; see Footnote 1). Evaluation ratings associated with the Year 3 administration (*M* = 3.26, *SD* = 0.60) were significantly lower than the evaluation ratings associated with both the Year 1 (*M* = 3.35, *SD* = 0.53) and Year 2 (*M* = 3.38, *SD* = 0.54) administrations. Thus, all courses received lower SET scores in Year 3, regardless of course delivery method and course level. However, the size of this effect was small (the largest difference in mean rating was 0.11 on a five-item scale).

---

[1] A Greenhouse–Geisser adjustment of the degrees of freedom was performed in anticipation of a sphericity assumption violation.

[2] A test of the homogeneity of variance assumption revealed no statistically significant difference in response rate variance between the two delivery modes for the 1st, 2nd, and 3rd years.

The second statistically significant main effect involved delivery mode, $F(1, 358) = 23.51$, $MSE = 0.52$, $p = .01$ ($\eta_p^2 = .06$; see Footnote 2). Face-to-face courses ($M = 3.41$, $SD = 0.50$) received significantly higher mean ratings than did online courses ($M = 3.13$, $SD = 0.63$), regardless of evaluation year and course level. No other factors or interactions included in the analysis were statistically reliable.

Stability of Ratings

The scatterplot presented in Figure 1 illustrates the relation between SET scores and response rate. Although the correlation between SET scores and response rate was small and not statistically significant, $r(362) = .07$, visual inspection of the plot of SET scores suggests that SET ratings became less variable as response rate increased. We conducted Levene's test to evaluate the variability of SET scores above and below the 60% response rate, which several researchers have recommended as an acceptable threshold for response rates (Berk, 2012, 2013; Nulty, 2008). The variability of scores above and below the 60% threshold was not statistically reliable, $F(1, 362) = 1.53$, $p = .22$.

Discussion

Online administration of SETs in this study was associated with lower response rates, yet it is curious that online courses experienced a 10% increase in response rate when all courses were evaluated with online forms in Year 3. Online courses had suffered from chronically low response rates in previous years, when face-to-face classes continued to use paper-based forms. The benefit to response rates observed for online courses when all SET forms were administered online might be attributed to increased communications that encouraged students to complete the online course evaluations. Despite this improvement, response rates for online courses continued to lag behind those for face-to-face courses. Differences in response rates for face-to-face and online courses might be attributed to characteristics of the students who enrolled or to differences in the quality of student engagement created in each learning modality. Avery et al. (2006) found that higher performing students (defined as students with higher GPAs) were more likely to complete online SETs.

Although the average SET rating was significantly lower in Year 3 than in the previous 2 years, the magnitude of the numeric difference was small (differences ranged from 0.08 to 0.11, based on a 0–4 Likert-like scale). This difference is similar to the differences Risquez et al. (2015) reported for SET scores after statistically adjusting for the influence of several potential confounding variables. A substantial literature has discussed the appropriate and inappropriate interpretation of SET ratings (Berk, 2013; Boysen, 2015a, 2015b; Boysen et al., 2014; Dewar, 2011; Stark & Freishtat, 2014).

Faculty have often raised concerns about the potential variability of SET scores due to low response rates and thus small sample sizes. However, our analysis indicated that classes with high response rates produced equally variable SET scores as did classes with low response rates. Reviewers should take extra care when they interpret SET scores. Decision makers often ignore questions about whether means derived from small samples accurately represent the population mean (Tversky & Kahneman, 1971). Reviewers frequently treat all numeric differences as if they were equally meaningful as measures of true differences and give them credibility even after receiving explicit warnings that these differences are not meaningful (Boysen, 2015a, 2015b).

Because low response rates produce small sample sizes, we expected that the SET scores based on smaller class samples (i.e., courses with low response rates) would be more variable than those based on larger class samples (i.e., courses with high response rates). Although researchers have recommended that response rates reach the criterion of 60%–80% when SET data will be used for high-stakes decisions (Berk, 2012, 2013; Nulty, 2008), our findings did not indicate a significant reduction in SET score variability with higher response rates.

Implications for Practice

Improving SET Response Rates

When decision makers use SET data to make high-stakes decisions (faculty hires, annual evaluations, tenure, promotions, teaching awards), institutions would be wise to take steps to ensure

that SETs have acceptable response rates. Researchers have discussed effective strategies to improve response rates for SETs (Nulty, 2008; see also Berk, 2013; Dommeyer et al., 2004; Jaquett et al., 2016). These strategies include offering empirically validated incentives, creating high-quality technical systems with good human factors characteristics, and promoting an institutional culture that clearly supports the use of SET data and other information to improve the quality of teaching and learning. Programs and instructors must discuss why information from SETs is important for decision-making and provide students with tangible evidence of how SET information guides decisions about curriculum improvement. The institution should provide students with compelling evidence that the administration system protects the confidentiality of their responses.

Evaluating SET Scores

In addition to ensuring adequate response rates on SETs, decision makers should demand multiple sources of evidence about teaching quality (Buller, 2012). High-stakes decisions should never rely exclusively on numeric data from SETs. Reviewers often treat SET ratings as a surrogate for a measure of the impact an instructor has on student learning. However, a recent meta-analysis (Uttl et al., 2017) questioned whether SET scores have any relation to student learning. Reviewers need evidence in addition to SET ratings to evaluate teaching, such as evidence of the instructor's disciplinary content expertise, skill with classroom management, ability to engage learners with lectures or other activities, impact on student learning, or success with efforts to modify and improve courses and teaching strategies (Berk, 2013; Stark & Freishtat, 2014). As with other forms of assessment, any one measure may be limited in terms of the quality of information it provides. Therefore, multiple measures are more informative than any single measure.

A portfolio of evidence can better inform high-stakes decisions (Berk, 2013). Portfolios might include summaries of class observations by senior faculty, the chair, and/or peers. Examples of assignments and exams can document the rigor of learning, especially if accompanied by redacted

samples of student work. Course syllabi can identify intended learning outcomes; describe instructional strategies that reflect the rigor of the course (required assignments and grading practices); and provide other information about course content, design, instructional strategies, and instructor interactions with students (Palmer et al., 2014; Stanny et al., 2015).

Conclusion

Psychology has a long history of devising creative strategies to measure the "unmeasurable," whether the targeted variable is a mental process, an attitude, or the quality of teaching (e.g., Webb et al., 1966). In addition, psychologists have documented various heuristics and biases that contribute to the misinterpretation of quantitative data (Gilovich et al., 2002), including SET scores (Boysen, 2015a, 2015b; Boysen et al., 2014). These skills enable psychologists to offer multiple solutions to the challenge posed by the need to objectively evaluate the quality of teaching and the impact of teaching on student learning.

Online administration of SET forms presents multiple desirable features, including rapid feedback to instructors, economy, and support for environmental sustainability. However, institutions should adopt implementation procedures that do not undermine the usefulness of the data gathered. Moreover, institutions should be wary of emphasizing procedures that produce high response rates only to lull faculty into believing that SET data can be the primary (or only) metric used for high-stakes decisions about the quality of faculty teaching. Instead, decision makers should expect to use multiple measures to evaluate the quality of faculty teaching.

References

Avery, R. J., Bryant, W. K., Mathios, A., Kang, H., & Bell, D. (2006). Electronic course evaluations: Does an online delivery system influence student evaluations? *The Journal of Economic Education*, *37*(1), 21–37. https://doi.org/10.3200/JECE.37.1.21-37

Berk, R. A. (2012). Top 20 strategies to increase the online response rates of student rating scales. *International Journal of Technology in Teaching and Learning*, *8*(2), 98–107.

Berk, R. A. (2013). *Top 10 flashpoints in student ratings and the evaluation of teaching*. Stylus.

Boysen, G. A. (2015a). Preventing the overinterpretation of small mean differences in student evaluations of teaching: An evaluation of warning effectiveness. *Scholarship of Teaching and Learning in Psychology*, *1*(4), 269–282. https://doi.org/10.1037/stl0000042

Boysen, G. A. (2015b). Significant interpretation of small mean differences in student evaluations of teaching despite explicit warning to avoid overinterpretation. *Scholarship of Teaching and Learning in Psychology*, *1*(2), 150–162. https://doi.org/10.1037/stl0000017

Boysen, G. A., Kelly, T. J., Raesly, H. N., & Casner, R. W. (2014). The (mis)interpretation of teaching evaluations by college faculty and administrators. *Assessment & Evaluation in Higher Education*, *39*(6), 641–656. https://doi.org/10.1080/02602938.2013.860950

Buller, J. L. (2012). *Best practices in faculty evaluation: A practical guide for academic leaders*. Jossey-Bass.

Dewar, J. M. (2011). Helping stakeholders understand the limitations of SRT data: Are we doing enough? *Journal of Faculty Development*, *25*(3), 40–44.

Dommeyer, C. J., Baum, P., & Hanna, R. W. (2002). College students' attitudes toward methods of collecting teaching evaluations: In-class versus on-line. *Journal of Education for Business*, *78*(1), 11–15. https://doi.org/10.1080/08832320209599691

Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and online surveys: Their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education*, *29*(5), 611–623. https://doi.org/10.1080/02602930410001689171

Feistauer, D., & Richter, T. (2016). How reliable are students' evaluations of teaching quality? A variance components approach. *Assessment & Evaluation in Higher Education*, *42*(8), 1263–1279. https://doi.org/10.1080/02602938.2016.1261083

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment*. Cambridge University Press. https://doi.org/10.1017/CBO9780511808098

Griffin, T. J., Hilton, J., III, Plummer, K., & Barret, D. (2014). Correlation between grade point averages and student evaluation of teaching scores: Taking a closer look. *Assessment & Evaluation in Higher Education*, *39*(3), 339–348. https://doi.org/10.1080/02602938.2013.831809

Jaquett, C. M., VanMaaren, V. G., & Williams, R. L. (2016). The effect of extra-credit incentives on student submission of end-of-course evaluations. *Scholarship of Teaching and Learning in Psychology*, *2*(1), 49–61. https://doi.org/10.1037/stl0000052

Jaquett, C. M., VanMaaren, V. G., & Williams, R. L. (2017). Course factors that motivate students to submit end-of-course evaluations. *Innovative Higher Education*, *42*(1), 19–31. https://doi.org/10.1007/s10755-016-9368-5

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. https://doi.org/10.1080/02602931003632399

Nowell, C., Gale, L. R., & Handley, B. (2010). Assessing faculty performance using student evaluations of teaching in an uncontrolled setting. *Assessment & Evaluation in Higher Education*, *35*(4), 463–475. https://doi.org/10.1080/02602930902862875

Nulty, D. D. (2008). The adequacy of response rates to online and paper surveys: What can be done? *Assessment & Evaluation in Higher Education*, *33*(3), 301–314. https://doi.org/10.1080/02602930701293231

Palmer, M. S., Bach, D. J., & Streifer, A. C. (2014). Measuring the promise: A learning-focused syllabus rubric. *To Improve the Academy: A Journal of Educational Development*, *33*(1), 14–36. https://doi.org/10.1002/tia2.20004

Reiner, C. M., & Arnold, K. E. (2010). Online course evaluation: Student and instructor perspectives and assessment potential. *Assessment Update*, *22*(2), 8–10. https://doi.org/10.1002/au.222

Risquez, A., Vaughan, E., & Murphy, M. (2015). Online student evaluations of teaching: What are we sacrificing for the affordances of technology? *Assessment & Evaluation in Higher Education*, *40*(1), 210–234. https://doi.org/10.1080/02602938.2014.890695

Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, *83*(4), 598–642. https://doi.org/10.3102/0034654313496870

Stanny, C. J., Gonzalez, M., & McGowan, B. (2015). Assessing the culture of teaching and learning through a syllabus review. *Assessment & Evaluation in Higher Education*, *40*(7), 898–913. https://doi.org/10.1080/02602938.2014.956684

Stark, P. B., & Freishtat, R. (2014). An evaluation of course evaluations. *ScienceOpen Research*. https://doi.org/10.14293/S2199-1006.1.SOR-EDU.AOFRQA.v1

Stowell, J. R., Addison, W. E., & Smith, J. L. (2012). Comparison of online and classroom-based student evaluations of instruction. *Assessment & Evaluation in Higher Education*, *37*(4), 465–473. https://doi.org/10.1080/02602938.2010.545869

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110. https://doi.org/10.1037/h0031322

Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, *54*, 22–42. https://doi.org/10.1016/j.stueduc.2016.08.007

Venette, S., Sellnow, D., & McIntyre, K. (2010). Charting new territory: Assessing the online frontier of student ratings of instruction. *Assessment & Evaluation in Higher Education*, *35*(1), 101–115. https://doi.org/10.1080/02602930802618336

Webb, E. J., Campbell, D. T., Schwartz, R. D., & Sechrest, L. (1966). *Unobtrusive measures: Nonreactive research in the social sciences*. Rand McNally.

Table 1

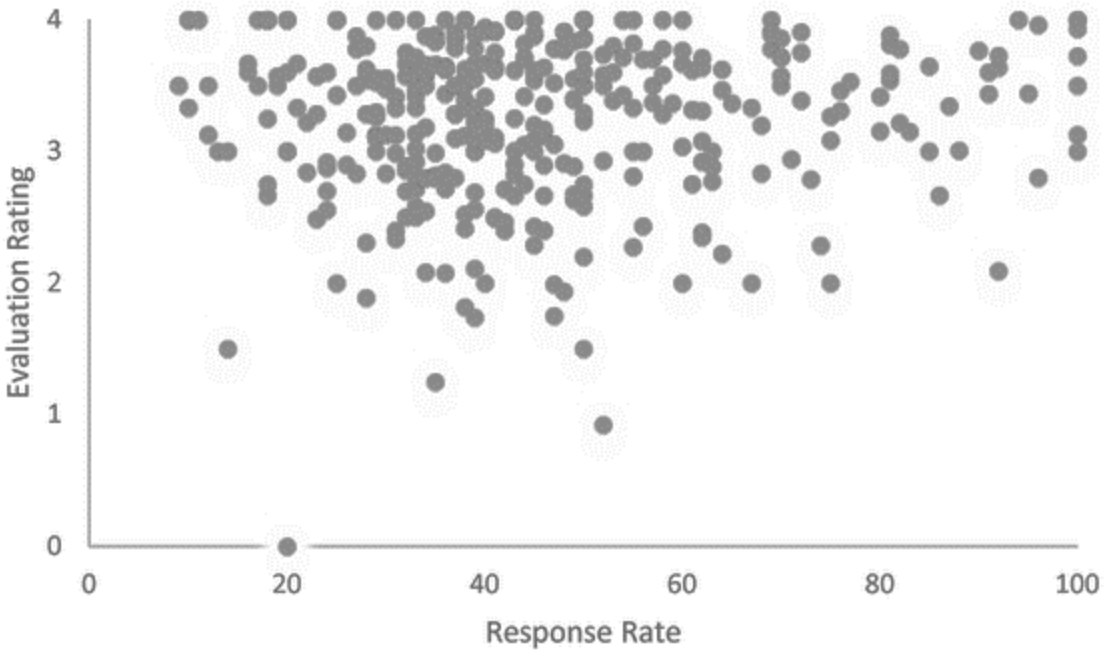
Means and Standard Deviations for Response Rates (Course Delivery Method by Evaluation Year)

Administration year	Face-to-face course		Online course	
	M	*SD*	*M*	*SD*
Year 1: 2012	71.72	16.42	32.93	15.73
Year 2: 2013	72.31	14.93	32.55	15.96
Year 3: 2014	47.18	20.11	41.60	18.23

Note. Student evaluations of teaching (SETs) were administered in two modalities in Years 1 and 2: paper based for face-to-face courses and online for online courses. SETs were administered online for all courses in Year 3.

Figure 1

Scatterplot Depicting the Correlation Between Response Rates and Evaluation Ratings



Note. Evaluation ratings were made during the 2014 fall academic term.