May 2020

Student Response Rate and Its Impact on Quantitative Evaluation of Faculty Teaching

Mingchu Neal Luo
Emporia State University, mluo@emporia.edu

Student Response Rate and Its Impact on Quantitative Evaluation of Faculty Teaching

Abstract

Student evaluations of teaching (SET) generate essential information for university administrators and faculty in assessing instruction quality. Lower response rates in student surveys have remained an important technical issue hurting the credibility of SET. This study examined the student response rate and its impact on the results of student quantitative evaluations of faculty teaching in a college of education. It analyzed the quantitative data of course evaluations collected by using the IDEA (Individual Development and Educational Assessment, 2016) survey at a teaching-orientated Midwest state university. Results indicate that 1) the average student response rate of all these courses is 63.6%, 2) there is no significant relationship between the enrollment size of the classes and response rate of SET for these classes, and 3) response rates to the course evaluation survey significantly associates with the overall mean scores of students' ratings, showing that classes evaluated by students with lower response rates tend to have lower evaluation mean scores. Implications for instruction evaluation policy and faculty practices for valid use of SET and increasing response rate are discussed.

Student Response Rate and its Impact on Quantitative Evaluation of Faculty Teaching

Neal Luo, Ed.D., Emporia State University, Emporia, KS

Dr. Neal Luo is an associate professor in the Department of School Leadership and Middle/Secondary Teacher Education, the Teachers College, Emporia State University. His research interests include data-driven decision-making, principalship, college student persistence, and TESOL. He can be reached at mluo@emporia.edu.

Introduction

Student evaluation of teaching (SET) is one of the most commonly used measures in colleges and universities that generate essential information for instructional decision makings of administrators and faculty. However, due to the controversial reliability and validity of SET, various studies question the properness and legitimacy of the persistent practice of using SET as a summative measure particularly for decision makings of retention, promotion, and pay for faculty (e.g., Boring, Ottoboni, & Stark, 2016; Hornstein & Law, 2017). Among the methodological criticism, the student response rate has remained a critical issue that negatively impacts the credibility of SET.

With the major shift away from paper-based to online SET for the benefits from enhanced data integrity, and savings of time and cost, researchers have been examining and are concerned about the reliability and validity of SET due to low response rates (e.g., Capa-Aydin, 2016; Nulty, 2008). One of the most important challenges is the low student response rate to online evaluations when compared to paper-based SET (Goodman, Anson, & Belcheir, 2015; Guder & Malliaris, 2010). Rates are substantially lower for online surveys while gaining more detailed feedback on qualitative responses (Liu & Armatas, 2016). The problem of low response rate has become obvious with lower than 30% for online evaluation surveys (Ling, Phillips, & Weihrich, 2012; Ogier, 2005), and the average response rate of SET generally falling between 30% and 60% (Anderson, Brown, & Spaeth, 2006; Nulty, 2008).

Reasons for lower response rates with online administration can be caused by various factors including privacy, time and social pressure, lack of engagement, incentives, communication, and demographic variables (Hatfield & Coyle, 2013; Mau & Opengart, 2012). Adams and Umbach (2012) suggest that the lower level of importance of the course perceived can yield low response rates. Different response rates also reflect actual differences in course content, course type, and student and teacher autonomy in the programs (Al-Maamari, 2015). Students taking lecture-based courses were more likely to complete SET surveys than any other course types (Macfadyen, Dawson, Prest, & Gasevic, 2016). Al-Maamari (2015) examined the variables that predict the SET survey response rate and revealed instructor gender contributes little to response rate while course type (i.e., whether a certain course was for credit or for non-credit) is a better predictor. The nature and the context of the programs, from which the relevant courses are evaluated, can also contribute to the response rate.

A low response rate suggests that the samples involved can generate unrepresentative or even biased results that lead to undermined or erroneous generalization in using SET for assessing faculty teaching quality (Schutt, 2016). The sampling method used in SET is non-random, in which students in the population of a class participate in the surveys at their will or when available. If a low percentage of students respond, there is a high probability that the respondents differ systematically from the non-respondents producing a biased sample, from which a simple extrapolation of findings to the student population of the whole class is not valid (Nulty, 2008). This issue is even more serious if the student population of a class is small. Delicate statistical theories on sampling error and confidence intervals were constructed in guiding the adequacy of the response rate for non-random samples, which is beyond the review of this study. However, some researchers developed and/or suggested concrete indicators of adequate response rate for a representative sample in spite of its complexity. Babbie (2007) stated that 50% could be regarded as an acceptable response rate in social research surveys, while Richardson (2005) suggested that the desirable response rate should be 60% or more.

Prior studies relevant to survey response rates of student evaluation of teaching (SET) have primarily focused on two streams. The first stream is the concern of low response to online surveys (e.g., Liu & Armatas, 2016; Nulty, 2008) and the factors contribute to lower response rates (e.g., Al-Maamari, 2015; Hatfield & Coyle, 2013). The second stream of explorations focuses on the methodological negative impacts of low response on the credibility of SET (e.g., Boring et al., 2016; Hornstein & Law, 2017), and what response rates should be regarded as adequate for SET (e.g., Babbie, 2007; Schutt, 2016). There has been a striking lack of SET studies incorporating student response rate into the analysis on correlating various factors with the results of SET, which allows us to determine other potential impacts of response rate in addition to its well-informed negative effect of unrepresentativeness due to its low response. This study intends to fill this gap by selecting a unique lens in examining the pattern between the response rate and the SET mean score of a course, and further addressing whether different survey response rates correlate with the mean scores of the course-based SET. It aims to contribute to the paucity literature on the problem of SET response rates (Al-Maamari, 2015), enabling us to better understand the rarely researched correlation. It is also practically important to have further evidence and knowledge in better understanding the role of response rate and its implications in the proper use of SET for decision making on faculty teaching quality and competence.

Specifically, this study investigates student response rate and its impact on the results of student quantitative evaluation of faculty teaching in a college of education by analyzing the quantitative data of 259 student course evaluations collected by using the IDEA (Individual Development and Educational Assessment, 2016) survey at a Midwest state university. Specifically, this study addresses the following research questions: 1. What is the overall average student response rate of all the courses evaluated using online administration in a college of education? 2. Is there a significant relationship between the enrollment size of a class and its response rate of SET? 3. Is there a significant relationship between course response rate and the mean score of course-based SET?

Methodology

The Instrument of Student Evaluations

Faculty teaching quality in this study was assessed by conceptualizing the course teaching objectives selected by the instructors. It was measured with the assessment outcome of instruction by using the course-based Individual Development and Educational Assessment (IDEA, 2016), a cross-sectional student rating survey of instruction. IDEA is a nonprofit organization dedicated to improving student learning in higher education through analytics, resources, and advice. The IDEA survey is a course evaluation instrument for students to rate their learning experience and satisfaction from various types of classes.

There are a total of 13 learning objectives measured in the IDEA (2016) survey (see the Appendix). Examples of these learning objectives are: Learning to apply course material (to improve thinking, problem-solving, and decisions); Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course, and Acquiring skills in working with others as a member of a team (see the Appendix). Each course instructor is required to select at least three objectives that are deemed to be important or essential for students' learning during the course before students complete the survey. Students completing the course taught by the instructor evaluate on the same three or more learning objectives selected. The IDEA response choice using Likert-scales ranges from 1 representing "no apparent progress", through 3 representing "moderate progress" to 5 representing "exceptional progress".

Evidence of the measurement validity of the IDEA (2016) self-reported ratings of learning is supported by the significant and positive correlation between average exam scores measuring students' actual learning and student ratings on the learning objectives. Students' self-ratings of learning progress also has a significant relationship with their ratings of how frequently the instructor emphasized the specific teaching methods. Average reliability coefficients ranging from .78 to .94 for individual survey items support a sufficient degree of reliability among individual items (Benton, Duchon, & Pallett, 2013).

Data Collection and Statistical Approach

The IDEA (2016) survey was administered online to all students who took the courses. Students were notified and reminded to complete the course evaluation surveys at the end of a semester. Data of the summary mean scores of each course measuring the average students' learning progress of the learning objectives selected by professors were used for analysis in this study. The mean score of the SET ratings in each course is used for evaluating faculty teaching by the university administrators. Logically, faculty teaching evaluation in this study was operationalized by using the SET ratings on the learning objectives selected by a professor who teaches the course.

Data compiled for analysis of this study are the 259 course SET collected using the IDEA (2016) survey in a college of education at a Midwest state university. These courses represent 95.5% those taught by faculty with ranks of assistant, associate, and full professors in both the spring and fall semesters in a calendar year at the college. The objective summary mean score of the

student-reported progress on all the instructor-chosen course objectives for each course was used as the unit of analysis measuring the dependent variable of faculty teaching quality in this study. The data of other variables including course evaluation response rate, course enrollment, course level (undergraduate vs graduate), course delivery method (face-to-face vs online), faculty gender (female vs male), faculty rank (assistant, associate, and full professor), and semester (spring vs fall) were also collected using the IDEA survey by the college and used for analysis in this study. The course evaluation response rate is the percentage of students who have validly responded to the evaluation out of all the students completing the course. Course enrollment refers to the class size indicating the total number of students who have completed the course.

Quantitative descriptive and correlational designs were adopted to address the research questions. The data collected using the IDEA (2016) instrument were analyzed using descriptive and inferential statistical tests. Mean scores, standard deviations, and percentages were used to address Research Question 1 on the overall average student response rate of all the courses evaluated and the mean SET by different categories of response rates. Pearson's correlation was used to address Research Question 2 on whether a significant relationship existed between the enrollment size of the classes and response rates of SET. For Research Question 3, the univariate test of ANOVA was conducted to determine whether there were significant differences in the course-based mean score of SET by response rates that were categorized into four groups of less than 40%, 40%-49%, 50%-75%, and more than 75%. Multiple regression was also performed to examine the correlation between response rate and SET score by controlling some factors of course features and instructor characteristics including course level, course enrollment, course delivery method, faculty gender, faculty rank, and semester. The purpose of controlling these extraneous variables was to achieve the higher nonspurious effect of the course survey response rate on SET (Schutt, 2016).

Results

The overall average student IDEA (2016) survey response rate of all the 259 courses evaluated using online administration is 63.6%. Twenty seven (10.4% of the total) courses received less than 40% of the student response rates. There are 38 (14.7%) of the courses examined receiving the survey response rates ranging from 40%-49%. The majority of the courses (194, 74.8%) received more than 50% student response rates. The mean course-based SET scores of all the 259 course is 4.39 (*SD*=.58), which represents the range from substantial progress to exceptional program on the learning objectives. Table 1 presents the overall and segregated means and standard deviations of SET by different course groups of response rates.

Pearson's correlation test indicated that there is no significant relationship between the enrollment size of the classes and the response rate of student evaluations of teaching (SET) for these classes, r =.085, p =.300. The ANOVA revealed significant differences in the course-based mean score of SET by response rate, $F(3, 255)$=2.345 p=.049, ω^2=.06, representing a medium effect size. Pairwise comparison follow-up tests indicated that the courses with a student response rate of lower than 40% have significantly lower SET than those with a 50-75% response rate, p=.039, r=.41 representing a medium effect size.

Table 1

Descriptive and ANOVA Statistics of SET on the Learning Objectives by Response Rate

Response Rate	Levels	n	Mean	SD	Test Statistics	p
	All Courses	259	4.39	.58		
	Less than 40%	27	4.10	.73	F=2.543	.049*
	40%-49%	38	4.40	.62		
	50%-75%	112	4.44	.61		
	More than 75%	82	4.42	.45		

Note. The IDEA (2016) response choice scales are 1=No apparent progress, 2=Slight progress, 3=Moderate progress, 4=Substantial progress, 5=Exceptional progress. * *p* is significant at .05.

Analysis using a multiple regression indicate that the model significantly predicts the student ratings (*F*=2.001, *p*=.046), explaining 10.23% of the variances of the student ratings, representing a medium effect size. The course-based response rate of the SET surveys significantly predicts the course-based mean scores in a positive direction. Lower response rates tend to lead to lower course evaluation scores. Other variables including course level, course delivery method, class size, and faculty gender were not significantly correlated to SET ratings. Table 2 presents the specific statistics of the multiple regression by the independent variables.

Table 2
Multiple Regression Coefficients

Variables	Beta (unstd.)	Std. Error	Beta (std.)	t	p	VIF
(Constant)	4.218	.208		20.325	.000	
Response Rate	.003	.002	.116	1.819	.050*	1.070
Course enrollment	-.004	.005	-.052	-.755	.451	1.259
Course Level (Undergraduate vs Graduate)	.008	.024	.033	.353	.725	2.286
Course Delivery Method (Face-to-face vs Online)	.046	.101	.039	.456	.649	1.989
Semester	-.075	.072	-.065	-1.049	.295	1.010
Faculty Gender	.068	.078	.059	.877	.381	1.182
Faculty Rank Dummy 1 (Assistant vs Associate Professor)	.076	.082	.062	.928	.354	1.118
Faculty Rank Dummy 2 (Assistant vs Full Professor)	-.252	.103	-.176	-2.439	.015*	1.373

Note. The IDEA response choice scales were 1=No apparent progress, 2=Slight progress, 3=Moderate progress, 4=Substantial progress, 5=Exceptional progress. * *p* is significant at .05.

Discussion, Conclusions and Implications

The average student response rate of all the 259 courses using online administration is fairly high with 63.6%. This finding somewhat differs from the previous research showing much lower response rate for online administration surveys (Ling et al., 2012; Ogier, 2005), but also seems to

be consistent with the result that the average response rates are between 30% and 60% (Anderson et al., 2006; Nulty, 2008). However, considering the valid criteria of response rates proposed by the previous researchers (47-58%, Nulty, 2008, p.309; 80%, McMillan, 2016, p.134), the average of 63.6% is still at a low level. Moreover, there are 65 courses (25.1%) with their response rates of lower than 50%. This finding indicates that there remains a significant gap in the actual response rates and in those that are deemed representative of the student population of the classes evaluated. There is still a need for continuous improvement in student response rates in SET using online administration. Without adequate response rates maintained, the benefits of implementing online surveys including efficiencies in survey administration, data management, as well as richer written feedback are not very worthwhile to achieve (Liu & Armatas, 2016).

Statistical examination of the linkages between the enrollment size of the classes and the response rate of SET of these classes reveals non-significant findings. This finding seems to suggest that the student response rate of SET is not impacted by the differences in class sizes. However, there is a significant difference in the mean score of SET among class groups of different student response rates. The courses evaluated with a response rate of lower than 40% have significantly lower course-based mean scores of SET than those with 50-75% response rates. These results are consistent with the finding of significant association indicated by the analysis of multiple regression. Specifically, the course-based mean scores of SET are related to the survey response rates controlling the effects of some factors of course features and instructor characteristics including course level, course enrollment, course delivery method, faculty gender, faculty rank, and semester.

More importantly, this study further reveals that the low course evaluation score is associated with a low response rate. Faculty members teaching courses that are evaluated with low response rates seem to suffer from both the small sample bias and low rating scores. If such information is used as a summative assessment to judge a faculty's teaching performance leading to retention, promotion, and pay, it can be an erroneous decision. This study confirms the critical issue of the low student response rate of SET that negatively impacts its credibility of SET. It affects the representativeness of SET due to the questionable credibility of course evaluation in terms of valid generalizability with unreliable extrapolation (e.g, Nulty, 2008; Hornstein & Law, 2017). Therefore, it is advocated that university administrators and faculty attend to the issue of low response rate and avoid using results of SET generated by low response rates in important decision-making. It is suggested that college administrators and faculty collaboratively build and maintain a feasible mechanism to encourage more students to respond to the course evaluation survey. Some specific strategies are effective to improve student response rates of SET including faculty emphasis on the importance of SET to the students, clearly indicating the course learning objectives evaluated in the course syllabus, making the first survey page simple, sending reminder emails, and considering timing when sending the survey and reminder emails.

It is also important to have a systematic way to calculate the sample size required for a specified level of confidence concerning a class population of a specified size (Schutt, 2016). For instance, "for class sizes below 20, the response rate required needs to be above 58%", and the decent response rate of "47% is only adequate" to achieve validity "when class sizes are above 30" (Nulty, 2008, p.309). Generally, a representative sample from the popular class sizes of 10-40

students needs almost all the students. For a class population of 50-100 students, at least 80% need to be in the sample (McMillan, 2016). On the other hand, the characteristics of the student population in a class also matters in assessing the representativeness of a sample. Samples of more homogeneous student populations in classes can be smaller than samples of more diverse student populations. If the class size is large or the students in a class are more homogenous, the response rate of 40%-50% constitutes to fairly good representation (Schutt, 2016).

There are a few limitations to this study. Theoretically, although this study shows a low mean score of SET is associated with a low response rate, it is unclear that a low response rate can be a causal indication of student dissatisfaction on faculty teaching. The quantitative design of this present study with the limited variables included for analysis does not have the capacity to well address why the low scores of SET is coupled with low response rates. It is worthwhile to conduct expanded and further research with more contextual analysis and in-depth qualitative explorations to test or validate the unclear assumption. This present study did not dig into the in-depth analysis of the characteristics and class performance of students who responded and those who did not respond to the course evaluation survey within the classes of lower response rates. It is beneficial to understand the information patterns of these students for the purposes of increasing the response rate and improving teaching quality. Finally, the data of SET in a college of education at a Midwestern teaching-oriented state university can possibly make it untrue to generalize the research findings to other institutional settings since SET and response rates vary significantly across institutions and classes (Nulty, 2008). It is suggested that universities of different contexts conduct studies analyzing their own data in seeking valid information for administrative and policy decision-making.

References

Adams, M. J. D., & Umbach, P. D. (2012). Non-response and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education, 53*(5), 576-591. http://dx.doi.org/10.1007/s11162-011-9240-5

Al-Maamari, F. (2015). Response rate and teaching effectiveness in institutional student evaluation of teaching: A multiple linear regression study. *Higher Education Studies, 5*(6), 9-20. http://dx.doi.org/10.5539/hes.v5n6p9

Anderson, J., Brown, G., & Spaeth, S. (2006). Online student evaluations and response rates reconsidered. *Innovate: Journal of Online Education*, *2*(6). Retrieved from https://pdfs.semanticscholar.org/6396/8a476c8c073112baee4582fc4c2e8279df07.pdf

Babbie, E. (2007). *The practice of social research* (11th ed.). Belmont, CA: Thompson Wadsworth.

Benton, S. L., Duchon, D., & Pallett, W. H. (2013). Validity of self-reported student ratings of instruction. *Assessment & Evaluation in Higher Education, 38*(4), 377-389.

Boring, A., Ottoboni, K., & Stark, P. B. (2016). *Student evaluations of teaching (mostly) do not measure teaching effectiveness.* Retrieved from Science Open Research. http://doi.org/10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1

Capa-Aydin, Y. (2016). Student evaluation of instruction: comparison between in-class and online methods. *Assessment & Evaluation in Higher Education, 41*(1), 112-116. http://doi.org/10.1080/02602938.2014.987106.

Galbraith, C., Merrill, G., & Kline, D. (2012). Are student evaluations of teaching effectiveness valid for measuring student outcomes in business related classes? A neural network and Bayesian analyses. *Research in Higher Education, 53*, 353–374.

Goodman, J., Anson, R., & Belcheir, M. (2015). The effect of incentives and other faculty-driven strategies to increase online student evaluation response rates. *Assessment & Evaluation in Higher Education, 40*(7), 958-970, http://doi.org/ 10.1080/02602938.2014.960364.

Guder, F., & Malliaris, M. (2010). Online and paper course evaluations. *American Journal of Business Education, 3*(2), 131-138.

Hammonds, F., Mariano, G.J., Ammons, G., & Chambers, S. (2017), Student evaluations of teaching: Improving teaching quality in higher education. *Perspectives: Policy & Practice in Higher Education, 21*(1), 26-33. http://doi.org/10.1080/13603108.2016.1227388

Hatfield, C. L., & Coyle, E. A. (2013). Factors that influence student completion of course and faculty evaluations. *American Journal of Pharmaceutical Education, 77*(2), 7-27.

Hornstein, H. A., & Law, H. F. E. (2017). Student evaluations of teaching are an inadequate assessment tool for evaluating faculty performance. *Cogent Education, 4*(1). Retrieved from https://www.tandfonline.com/doi/full/10.1080/2331186X.2017.1304016 http://doi.org/10.1080/2331186X.2017.1304016

The Individual Development and Educational Assessment (2016). *The IDEA Student Ratings System.* Manhattan, KS: The IDEA Center.

Ling, T., Phillips, J., & Weihrich, S. (2012). Online evaluations vs in-class paper teaching evaluations: A paired comparison. *Journal of the Academy of Business Education, 12*, 150–161.

Liu, D. C. S., & Armatas, C. (2016). Response rate and ratings for student evaluation of teaching: Does online administration matters? *Asian Journal of Educational Research, 4*(5), 1-13.

Macfadyen, L. P., Dawson, S., Prest, S., & Gašević, D. (2016). Whose feedback? A multilevel analysis of student completion of end-of-term teaching evaluations. *Assessment & Evaluation in Higher Education, 41*(6), 1-19. http://dx.doi.org/10.1080/02602938.2015.1044421

Mau, R. R., & Opengart, R. A. (2012). Comparing ratings: in-class (paper) vs out of class (online) student evaluations. *Higher Education Studies, 2*(3), 55-68. http://doi.org/10.5539/hes.v2n3p55

McMillan, J. H. (2016). *Fundamentals of educational research* (7th ed.). Upper Saddle River, NJ: Pearson Education.

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

Ogier, J. (November, 2005). *The response rates for online surveys: A hit and miss affair.* Paper presented at the 2005 Australasian Evaluations Forum: University Learning and Teaching: Evaluating and Enhancing the Experience, UNSW, Sydney.

Richardson, J. T. E. (2005). Instruments for obtaining student feedback: a review of the literature. *Assessment & Evaluation in Higher Education, 30*(4), 387–415.

Schutt, R. K., (2016). *Investigating the Social World: The Process and Practice of Research* (8th ed.). Thousand Oaks, CA: Sage Publications, Inc.

Appendix

The IDEA (2016) Learning Objectives

1. Gaining a basic understanding of the subject (e.g., factual knowledge, methods, principles, generalizations, theories)
2. Developing knowledge and understanding of diverse perspectives, global awareness, or other cultures
3. Learning to apply course material (to improve thinking, problem-solving, and decisions)
4. Developing specific skills, competencies, and points of view needed by professionals in the field most closely related to this course
5. Acquiring skills in working with others as a member of a team
6. Developing creative capacities (inventing; designing; writing; performing in art, music, drama, etc.)
7. Gaining a broader understanding and appreciation of intellectual/cultural activity (music, science, literature, etc.)
8. Developing skill in expressing oneself orally or in writing
9. Learning how to find, evaluate, and use resources to explore a topic in-depth
10. Developing ethical reasoning and/or ethical decision making
11. Learning to analyze and critically evaluate ideas, arguments, and points of view
12. Learning to apply knowledge and skills to benefit others or serve the public good
13. Learning appropriate methods for collecting, analyzing, and interpreting numerical information

Rating Choice Scales
1 = No apparent progress,
2 = Slight progress - I made small gains on this objective,
3 = Moderate progress - I made some gains on this objective,
4 = Substantial progress - I made large gains on this objective, and
5 = Exceptional progress - I made outstanding gains on this objective.