An Interpretation Guide for the

[Rev Fall 2019]

Table of Contents

Closed-Ended Questions

Topic	**Item**	**Old** *(Fall 2003 – Spring 2017)*	**New** *(Fall 2017 – present)*
Relevance	Q1	Demonstrated relevance of the course content.	*[no change]*
Learning Environment	Q2	Used assignments that enhanced learning.	*[no change]*
Helping Students Think	Q3	Summarized/emphasized important points.	*[no change]*
Learning Environment	Q4	Was responsive to questions and comments from students.	*[no change]*
Learning Environment	Q5	Established an atmosphere that facilitated learning.	*[no change]*
Responsiveness to Students	Q6	Was approachable for assistance.	*[no change]*
Responsiveness to Students	Q7	Was responsive to the diversity of students in class.	Was **respectful of**

Informational Questions

Item	**Old** *(Fall 2003 – Spring 2017)*	**New** *(Fall 2017 – present)*
Q14	What is your current estimate of your expected overall grade in this course? *(A; B; C; D or F; Other)*	*[no change]*
Q15	You are a: *(Freshman; Sophomore; Junior; Senior; Graduate Student; Credential Student; Other)*	*[no change]*
Q16	Did you complete this form without undue influence from other students? *(Yes; No)*	*[no change]*

Q17) p(ue)-1hu form unh3.2 (or)324 r2 (o)-24ega (or)324dTjEM3unouruor

Interpretation of the SOTE Ratings

SOTE Reporting

To aid in interpretation, official SOTE reports provide data (means, standard deviations, and medians) for the instructor's department, college, and the university as a whole.

Mean: This is the arithmetic average of student responses. Note, however, that most student rating distributions are skewed (that is, the ratings bunch up toward one end, typically the right end), in which case the mean does not represent the typical or most frequently occurring rating.

Standard Deviation: This statistic measures the variability among the responses (i.e., how much, on the average, student responses vary from the mean). Like the mean, the standard deviation is an inappropriate measure of variability when the distribution is skewed.

Median: This is the middle ranking. A median of 3.5 indicates that half the students gave ratings higher and half lower than 3.5. The median is helpful in cases where outliers might influence the mean and standard deviation (e.g. cases in which a few extremely high or extremely low ratings push the mean

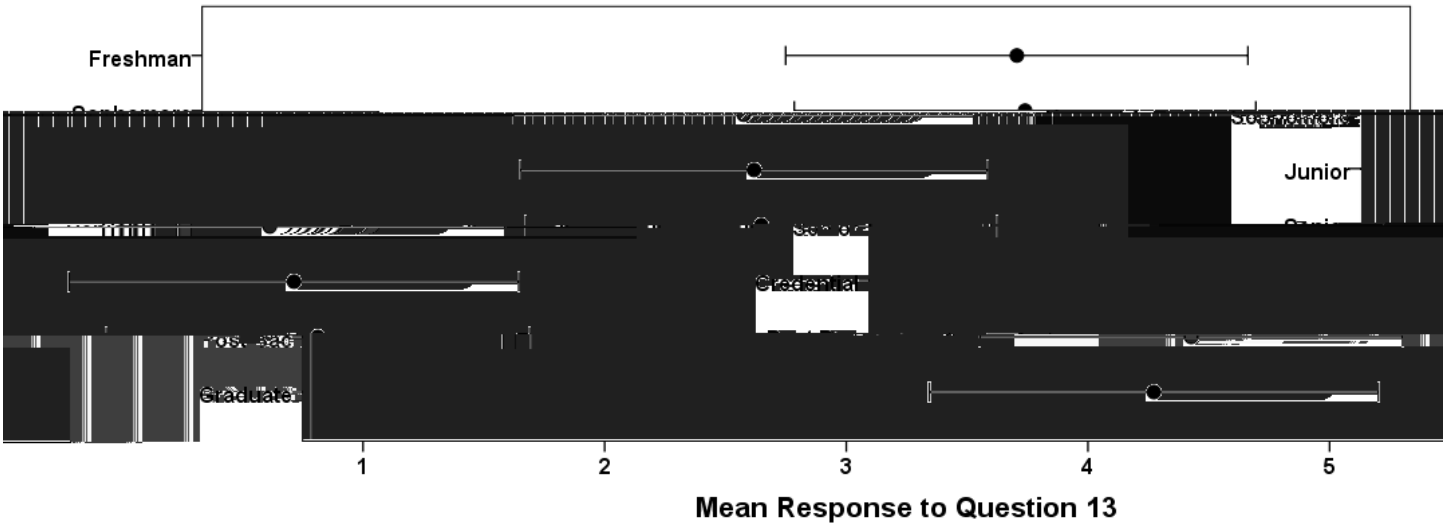
Course Characteristics

College and Content

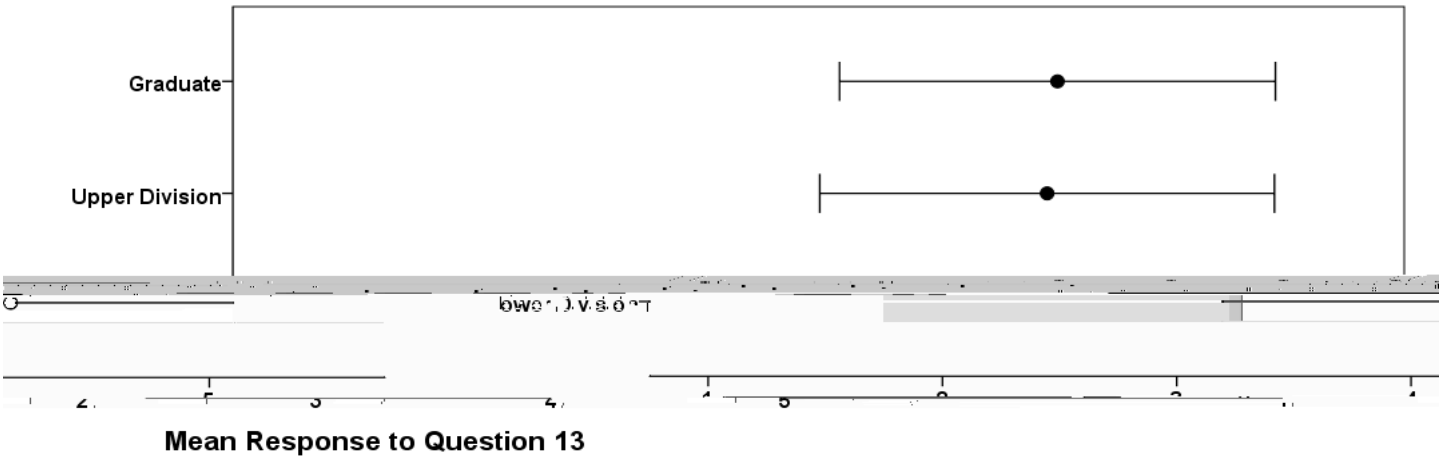
There appear to be some differences in average ratings of overall teaching effectiveness (Q13)

Course Level

There appears to be slight differences in the average ratings of overall teaching effectiveness (Q13) across student level (i.e., frosh, junior, graduate, etc.) as well as level of instruction (e.g., upper- vs. lower-division courses).



Error Bars = +/- 1 SD



Error Bars = +/- 1 SD

Research on student evaluations at other universities shows that ratings in graduate and credential classes tend to be higher than in undergraduate classes (see also Arreola, 2000; Marsh & Hocevar, 1991). However, ratings across lower and upper division courses tend to be relatively similar (Arreola, 2000).

Official and Expected Grades

Possibly the most notable impact on student ratings is their anticipated and official grade in the course.

Error Bars = +/- 1 SD

In fact, it is well established that student ratings are positively associated with both expected and actual course grades (e.g., Kulik, 2001). Greenwald & Gillmore (1997) further concluded that grading leniency exerts an important influence on ratings. However, another possible explanation for this result is that strong instructors teach courses in which students both learn a lot (therefore, they earn and deserve high grades) and give appropriately high ratings to the course and the instructor.

Nevertheless, when interpreting SOTE ratings, we encourage RTP committees to note the distribution of expected grades. Classes in which the majority of students expect either low or high grades should be fairly rare (exceptions to this would be graduate and credential classes in which a grade lower than a "B" is often considered equivalent to a failing grade). In addition, expected grades for a class should show some relationship to actual grades. In cases where there is a wide discrepancy (e.g. 80% of the class expects a grade of "A" while

Administration

Several studies have failed to detect a significant difference in ratings between online evaluations and paper evaluations (Donovan et al., 2006; Hardy, 2003; Heath et al., 2007; Laubsch, 2006; Spooner et al., 1999). At SJSU, a study by Sujitparapitaya and Briggs (2010) indicated that there was no significant difference for a majority of the responses between online evaluations and paper evaluations (similar to findings from a study conducted at Brigham Young University, Sorenson & Johnson, 2006). While some

Instructor Characteristics

Whereas analyses of SOTES responses in relation to various instructor characteristics is not reported here, the factors discussed below have been identified in existing literature as possible threats to the validity of student evaluations. Note that this is not intended to be a comprehensive review of such factors, but a brief review is presented here as a point of consideration.

Genderpon (ent)-6(ec8.918(ed hec -0.pon (ent)-8.86(c)-2 (ons5(f)-1(s)-2)0.40 Td-13(ed5(f)-120.820.15d)0

Rank and Tenure

Findings on the impact of student evaluations according to the faculty members' status, rank, and tenure are mixed. While some have found that non-tenured faculty receive lower ratings than tenured faculty (e.g., McPherson & Todd Jewell, 2007), others have found that adjunct and temporary faculty tend to receive higher ratings than tenure-track faculty (Figlio, Schapiro & Soter, 2015; McPherson et al., 2009). There does not appear

References

Andersen, K., & Miller, E. D. (1997). Gender and student evaluations of teaching. *Political Science & Politics, 30*, 216-219.

Arbuckle, J., & Williams, B. D. (2003). Students' perceptions of expressiveness: Age and gender effects on teacher evaluations. *Sex Roles, 49*, 507-516.

Arreola, R.A. (2000). *Developing a comprehensive faculty evaluation system*. Bolton, MA, Anker Publishing.

Arthur, L. (2009). From performativity to professionalism: Lecturers' responses to student feedback. *Teaching in Higher Education, 14*, 441-454.

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,* 21-37.

Bachen, C. M., McLoughlin, M. M., & Garcia, S. S. (1999). Assessing the role of gender in college students' evaluations of faculty. *Communication Education, 48*, 193-210.

Basow, S. A. (1995). Student evaluations of college professors: When gender matters. *Journal of Educational Psychology, 87*, 656.

Basow, S. A., & Montgomery, S. (2005). Student ratings and professor self-ratings of college teaching: Effects of gender - and divisional affiliation. *Journal of Personnel Evaluation in Education, 18*, 91-106.

Basow, S., Codos, S., & Martin, J. (2013). The effects of professors' race and gender on student evaluations and performance. *College Student Journal, 47*, 352-363.

Basow, S. A., Phelan, J. E., & Capotosto, L. (2006). Gender patterns in college students' choices of their best and worst professors. *Psychology of Women Quarterly, 30*, 25-35.

Beran, T. N., & Rokosh, J. L. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science, 37*, 171-184.

Beran, T., Violato, C., Kline, D., & Frideres, J. (2005). The utility of student ratings of instruction for students, faculty, and administrators: A" Consequential Validity" study. *Canadian Journal of Higher Education, 35*, 49-70.

Blair, E., & Valdez Noel, K. (2014). Improving higher education practice through student evaluation systems: Is the student voice being heard? *Assessment and Evaluation in Higher Education, 39*, 879-894.

Bosshardt, W., & Watts, M. (2001). Comparing student and instructor evaluations of teaching. *The Journal of Economic Education, 32*, 3-17.

Cao, Y., Clark, A., Schrimer, J., & Nelson, M. (2007). *Online and paper course evaluations: Are the response rates and results different?* Paper presented at the Association of Institutional Research Annual Forum, San Francisco, CA.

Centra, J. A., & Gaubatz, N. B. (2000). Is there gender bias in student evaluations of teaching? *The Journal of Higher Education, 71*, 17-33.

Chamberlin, M. S., & Hickey, J. S. -2.9 (hamf0 Tc 0 Tw (0-6.3S1 ( )0.0.5 (s)-8 (t)-13Td[(C)-2.9 .2 ( )0.5 (e)-12.3 (v)4 (a)-12.2 (l)3.

Finegan, T. A., & Siegfried, J. J. (2000). Are student ratings of teaching effectiveness influenced by instructors' English language proficiency? *The American Economist, 44*, 17-29.

Gill, M. M. (1994). Accent and stereotypes: Their effect on perceptions of teachers and lecture comprehension. *Journal of Applied Communication Research, 22*, 348-361.

Gravestock, P., & Gregor-Greenleaf, E. (2008). *Student course evaluations: Research, models and trends.* Toronto: Higher Education Quality Council of Ontario.

Greenwald, A. G., & Gillmore, G. M. (1997). Grading leniency is a removable contaminant of student ratings. *American Psychologist, 52*, 1209-1217.

Hardy, N. (2003). Online ratings: fact and fiction. *New Directions for Teaching and Learning, 96*, 31-41.

Heath, N. M., Lawyer, S. R., & Rasmussen, E. B. (2007). Web-based versus paper-and-pencil course evaluations. *Teaching of Psychology, 34*, 259-261.

Johnson, M. D., Narayanan, A., & Sawaya, W. J. (2013). Effects of course and instructor characteristics on student evaluation of teaching across a college of engineering. *Journal of Engineering Education, 102*, 289-318.

Kember, D., Leung, D. Y., & Kwan, K. (2002). Does the use of student feedback questionnaires improve the overall quality of teaching? *Assessment and Evaluation in Higher Education, 27*, 411-425.

Kohn, J., & Hatfield, L. (2006). The role of gender in teaching effectiveness ratings of faculty. *Academy of Educational Leadership Journal, 10*, 121.

Kulik, J. A. (2001). *Student ratings: Validity, utility, and controversy*. In M. Theall, P. C. Abrami, & Mets, L.A. (Eds), The student rating debate: Are they valid? How can we best use them? (pp. 9-25). San Francisco: Jossey-Bass.

Laubsch, P. (2006). Online and in-person evaluations: A literature review and exploratory comparison. *Journal of Online Learning and Teaching, 2*, 62-73.

Liu, Y, (2006). A comparison study of online versus traditional student evaluation of instruction. *International Journal of Instructional Technology & Distance Learning.*

MacNell, L., Driscoll, A., & Hunt, A. N. (2014). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education, 40*, 291-303.

Marsh, H. W., & Hocevar, D. (1991). The multidimensionality of students' evaluations of teaching effectiveness: The generality of factor structures across academic discipline, instructor level, and course level. *Teaching and Teacher Education, 7,* 9-18.

Martin, L. L. (2016). Gender, teaching evaluations, and professional success in political science. *Political Science & Politics, 49*, 313-319.