An Interpretation Guide for the

Student Opinion of Teaching Effectiveness Surveys (SOTES)

The SJSU Student Evaluation Review Board (SERB)

The information presented here includes a description of the SOTE instrument, and overview of the statistics included in the SOTE report, and a brief review of factors that influence SOTE information ( ).

Table of Contents

SOTE Interpretation Quick Guide

Background and Administration

The Student Evaluation of Teaching Effectiveness (SOTE) instrument was created to assess student perceptions of teaching effectiveness (the current version was revised in Fall 2019).

The survey begins with a brief introduction and overview, followed by 13 closed-ended items, 4 informational items, and 3 open-ended questions.

SOTE surveys are administered by the SJSU Office of Institutional Research and Strategic Analytics (IRSA) through CourseEval (online software integrated with Canvas).

Results, Reporting, and Interpretation

Results are provided to individual instructors and department chairs. Results are also included in

A recent memo from Provost Del Casino also required that RTP committee members interpret SOTEs from Fall 2020 with care" due to the challenges of converting instruction modality.

Instructors and department chairs may request a report of responses to questions asking about 'undue influence' from the IR Office. Typically, such requests occur when students make independent allegations of improprieties and an investigation is conducted.

Questions? For an up-to-date listing of Student Evaluation Review Board members (which includes one representative per college), visit

History and Policy

The Student Evaluation Review Board is an Operating Committee of the Academic Senate that reports to the Professional Standards Committee. The board includes one faculty member from each of the seven colleges on campus as well as one student representative (at-large). The directors of the Office of Institutional Research and Strategic Analytics (IRSA) and the Center for Faculty Development serve as ex officio members on the committee.

to correct

Closed-Ended Questions

Topic	Item	Old	New
Relev			

| Overall Effectiveness | Q13 | Overall, this instructors teaching was: | Overall, this instructor's teaching was **effective**. |

Notes: Items and instructions that were revised in Fall 2017 are in **bold font**. Response options for Questions 1-12 on the old instrument (Fall 2003 - Spring 2017) used the following scale:

. The new instrument (Fall 2017 -   ) adopts a slightly modified scale (for all questions):

Informational Questions

| **Item** | **Old** | **New** |

Q14   What is younb

Interpretation of the SOTE Ratings

SOTE Ree T

repeatedly observed for the same instructor, then RTP committees should consider further evaluations for that instructor.

Q12												1	0.790
Q13													1

The Pearson product moment correlation measures the strength of linear dependence between two variables, and varies between -1 and 1. As a rule of thumb, correlations between .00 and .50 are considered weak; correlations between .50 and .70 are moderate, and correlations over .70 are relatively strong. The correlations presented in the table above are all statistically significant at the p<.01 level.

In Fall 2022, 5.2% of students (n=3,229) responded 'no' to Question 16 ("Did you complete this form without undue influence from other **students**?") and 5.4% of students (n=3,316) responded 'no' to Question 17 ("Did you complete this form without undue influence from the **instructor**?"). Of these students, most (n=3,017) responded 'no' to both questions indicating that they may have misunderstood the question. In Spring 2023, 5.2% of students (n=2,860) responded 'no' to Question 16 ("Did you complete this form without undue influence from other **students**?") and 5.3% of students (n=2,945) responded 'no' to Question 17 ("Did you complete this form without undue influence from the **instructor**?"). Of these students, most (n=2,691) responded 'no' to both questions indicating that they may have misunderstood the question.

We also note that *several factors are known to systematically influence SOTE ratings* This is demonstrated below using Fall 2022-Spring 2023 data with references to similar findings from research conducted elsewhere. These factors should be considered in any RTP evaluation of SOTE data and we encourage faculty members to include additional information and explanation in their dossiers as necessary.

Course Characteristics

College and Content

From the following figure, there appear to be some differences in the averages of the ratings of overall teaching effectiveness (Q13) across colleges at San Jose State. This is a common trend, Stroebe 2020 contains a review of articles studying the correlation between disciplines and faculty ratings, with faculty in science and engineering obtaining less positive ratings (see also Gravestock & Gregor-Greenleaf, 2008).

Mean Response of Instructor's Teaching Effectiveness by College

Error Bars = +/- 1 SD

There are also differences in average ratings between departments within colleges. It is therefore important that RTP committees

Gregor-Greenleaf, 2008). Kreitzer et al. 2022 identify all the mentioned characteristics as measurement bias, when variables unrelated to teaching effectiveness systematically influence the results, and provide a wide literature review. Note, however, Beran et al. (2009) argue that these effects may be mediated by varying levels of student engagement.

Innovative Pedagogy

Significant differences in student evaluations are observed due to course type and pedagogical structure, which can be daunting for faculty engaging in pedagogical innovation to improve student learning and lead to an entrenchment of traditional lecture-heavy, teacher-centered pedagogies. Numerous studies have cautioned against using student evaluations as an indicator of student learning, with student learning outcomes explaining only 1-14% of the variability in student evaluations (e.g., Uttl, White & Gonzalez, 2017; Clayson, 2009; Cohen, 1980). In addition, student evaluations were generally developed to assess a teacher-centered learning environment, with a knowledge transmission model, and in many cases fail to capture the benefits of desettling the classroom to a student-centered paradigm (Kolitch & Dean, 1999; Theall, 2010). Kember, Jenkins & Ng (2010) argue that student responses on evaluations depend largely on what students consider to be good teaching; this may align more with student previous experience or disciplinary conceptualizations rather than effective teaching practice.

While some studies have shown a positive correlation between innovative approaches like the flipped classroom model and student evaluations (Samuel, 2019; Lag & Saele, 2019), this is not universally the case with pedagogical innovation and teaching effectiveness. The fear of decreased evaluations and the consequences therein can lead faculty to be wary of negative consequences of pedagogical innovations that are research-based as these can lead to lower student evaluations (Henderson, Khan & Dancy, 2018). It should be emphasized that student evaluations are assessments of student experience, not of student learning, and many researchers have questioned whether students have the ability to assess the appropriateness and effectiveness of the pedagogies faculty employ as student evaluations and other metrics of teaching effectiveness are often anti-correlated (Braga, Paccagnella & Pellizzari, 2014; Kornell & Hausman, 2016). This leads both to the correlation between perceived easiness (and thus grades obtained, as discussed below) and student evaluations, as well as a decrease in student evaluations when students experience discomfort (Felton, Mitchell & Stinson, 2004; Walker et al., 2008).

In many innovative, student-centered approaches, the burden of knowledge construction is more clearly placed on the student, which can lead to student unease as they may be concerned about not having a clear-cut correct answer, of being negatively evaluated, or experiencing greater anxiety due an increasingly active role and responsibility in the class (Cooper, Downing & Brownell
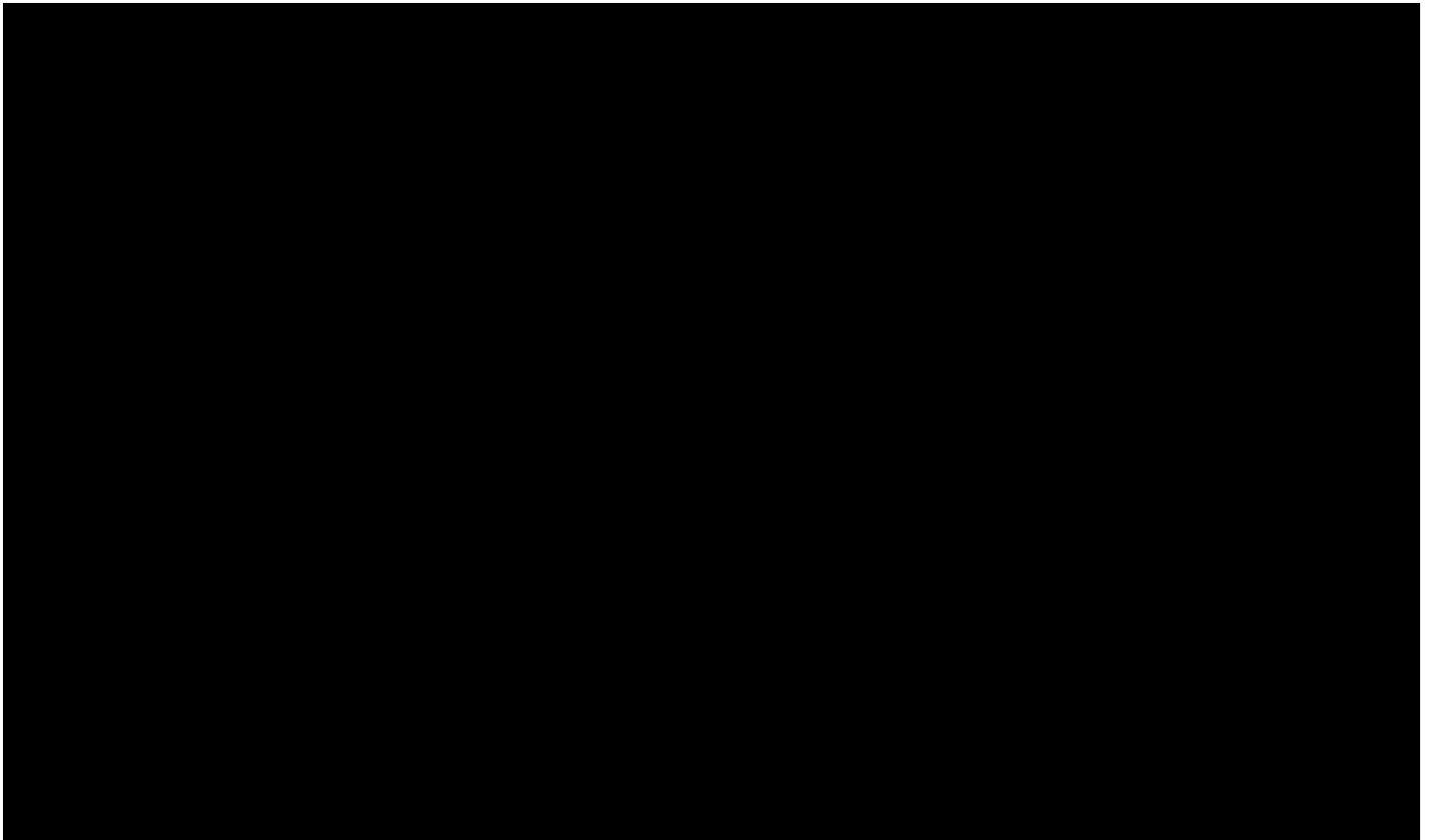
Course Level

There appear to be slight differences in the average ratings of overall teaching effectiveness

Official and Expected Grades

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

Mean Response of Instructor's Teaching Effectiveness by Students' Official Grade[6]

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 (Spooren and Mortelmans, 2006).

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 the actual average grade for the class is a "C") RTP committees may request further information from the instructor.

---

[6] Please note that although WUs appear on this graph, students with a WU are removed from the SOTE/SOLATE reports.

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 studies have found that specific questions may be answered more favorably in online evaluations (Liu, 2006; see also Avery et al., 2006; Cao et al., 2007), others have reported that paper evaluations produced higher scores for individual questions and total scores (Chang, 2003; Mau et al., 2012).

Importantly, the overall response rate at SJSU has remained the same, if not improved, since the university moved to online implementation in 2013 (48.1% in Fall 2022; 47.1% in Spring 2023). We also note that there is no evidence for a significant difference in student responses to Question 13 across the Fall and Spring semesters ($M_{fall} = 4.36$, $SD_{fall} = .95$; $M_{spring} = 4.39$, $SD_{spring} = .93$).

A study by Guder and Maliaris (2013) showed that the response rate of online evaluation increased when emails were sent and when faculty emphasized the importance of completing the evaluations in class. Van Mol (2017) suggested that sending extra reminders with specific reminder content is effective for increasing online evaluation response rates.

Instructor Characteristics

Whereas analyses of SOTES responses in relation to various instructor characteristics are 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

compared to those for whom English is a second language, Gill (1994) found that students view teachers with "standard North American accents" more favorably.

Rank and Tenu

References

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (1999). . Washington, DC: American Educational Research Association.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). .

Maul, A. (2017). Rethinking traditional methods of survey validation. , 15(2), 51–69.

McPherson, M. A., & Jewell, R. T. (2007). Leveling the playing field: Should student evaluation scores be adjusted? , 868–881.

McPherson, M. A., Jewell, R. T., & Kim, M. (2009). What determines student evaluation scores? A random effects analysis of undergraduate economics classes. 37–51.

Mengel, F., Sauermann, J. & Zölitz, U. (2019). Gender bias in teaching evaluations. (2), 535-566.

Mitchell, K. M., & Martin, J. (2018). Gender bias in student evaluations. , 1-5.

Nasser, F., & Fresko, B.

        hurt us.   , 779-793.
Stark. P. B. (2018 May-June).
                              , John W. Lawrence, American Association of University Professors.
        https://www.aaup.org/article/student-evaluations-teaching-are-not-valid
Stark, P. B. (2018 Oct).
              e?, Stanford-Berkeley Joint Colloquium, Department of Statistics, Stanford University, Stanford,
        CA. https://www.stat.berkeley.edu/~stark/Seminars/setStanford18.htm
Stark, P. B. (2019).   .
        https://www.stat.berkeley.edu/~stark/Preprints/setNotes19.pdf
Stark, P. B. (2021 March).
                                          , Berkeley Evaluation and Assessment Research (BEAR) Center,
        University of California, Berkeley, CA.
        https://www.stat.berkeley.edu/~stark/Seminars/setUCBED21.htm
Stroebe, W. (2020). Student evaluations of teaching encourages poor teaching and contributes to grade inflation:
        A theoretical and empirical analysis. Basic and Applied Social Psychology, 42(4), 276-294.
Subtirelu, N. C. (2015). "She does have an accent but…": Race and language ideology in students' evaluations
        of mathematics instructors on RateMyProfessors. com.   (1), 35–62
Sujitparapitaya, S. & Briggs, J. (2010).
                                        Paper presented at the Student Evaluation Review Board, San Jose, CA.
Theall, M. (2010). New resistance to student evaluations of teaching.
                  ,   (3), 44.
Theall, Michael, & Franklin, Jennifer. (2001). Looking for bias in all the wrong places: A search for truth or a
        witch hunt in student ratings of instruction?   , 45-56.
Ting, K. (2000). A multilevel perspective on student ratings of instruction: Lessons from the Chinese
        experience.   , 637–661.
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.
                  ,   , 22-42.
Van Mol, C. (2017). Improving web survey efficiency: the impact of an extra reminder and reminder content on
Ь   web survey response.   (4), 317-327.
Walker, J. D., Cotner, S. H., Baepler, P. M., & Decker, M. D. (2008). A delicate balance: integrating active
        learning into