

An Interpretation Guide for the

Prepared by:

SOTE Interpretation Quick Guide	3
History and Policy	4
The SOTE Survey	5
<i>Instructions</i>	5
<i>Closed-Ended Questions</i>	7
<i>Informational Questions</i>	8
<i>Open-Ended Questions</i>	8
Interpretation of the SOTE Rating with SOTE Ra35	

The Student Evaluation of Teaching Effectiveness (SOTE)

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 Effectiveness and Strategic Analysis and the Center for Faculty Development serve as ex officio members on the committee.

The committee is charged with designing evaluation instruments to be used The

F12-6: Since student opinion surveys measure student satisfaction rather than student learning, they cannot be considered perfect indicators of teaching quality.... To guard against the limitations of the instrument, all those using SOTES as part of the SJSU evaluation process must consult the official interpretation guide... Information from SOTES is but one source of information for assessing teaching effectiveness.

F12-6: SOTES shall be administered in all classes [except those officially excluded for technical or ethical reasons] and the results placed in the faculty personnel file. Faculty, however, under some circumstances may exclude the results of an occasional course from their periodic

to correct any problems identified earlier in either direct observations or prior performance evaluations. Recent direct observations and surveys of student opinion of teaching effectiveness (SOTEs) are also supportive. SOTEs are considered supportive if they are either within appropriate norms, or if a preponderance of student opinion from objective and subjective questions indicates effective teaching.”

S20-4: Optional Exclusion of Student Opinion of Teaching Effectiveness surveys (SOTEs) Administered during Spring 2020. Faculty be permitted, at their option, to exclude any SOTE results obtained during Spring 2020 from future evaluations.

S20-7: Students were allowed to petition the Registrar to change a letter grade to Credit/No Credit for all classes. Spring 2020 SOTE results incorrectly excluded students who petitioned the Registrar to change a letter grade to Credit/No Credit.

The most recent version of the SOTE instrument was administered for the first time in Fall 2017. See below for a comparison across the old and new instruments. Note that both versions begin with a brief introduction and overview, followed by thirteen (13) closed-ended items that assess students’ perceptions on teaching effectiveness and their learning experiences. These are followed by four (4) informational items and three (3) open-ended questions. Items and instructions that were revised in Fall 2017 are in .

This instrument is designed of your instructor's teaching performance. It is NOT designed to measure your reaction to the subject, the facilities (such as the physical conditions of the classroom), . Your individual ratings will be anonymous and a summary of items 1-18 will be available to your instructor after grades are turned in. This summary may enhance your instructor's teaching. It will also be used in the evaluation of your instructor for personnel matters such as retention, tenure and promotion

Q20 Please provide any other comments you feel would be helpful to the instructor regarding his/her teaching performance/ability.

To aid in interpretation, official SOTE reports provide data

The norms and statistics reported in this Interpretation Guide were calculated from SOTE survey results from Fall 2021 and Spring 2022. All courses across all colleges were included in this analysis, resulting in a total of 133,658 student responses (Fall '21 = 64,122 responses; Spring '22 = 69,536 responses)

Cronbach's alpha (α) is 0.97 across all 13 questions, indicating a very strong level of internal consistency across questions. We also note that Question 13 is strongly correlated with all of the other items. While Question 13 is often used as an index of overall effectiveness, we recommend that evaluations of teaching effectiveness consider all 13 items.

	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13
Q1	1	0.759	0.778	0.685	0.737	0.668	0.652	0.723	0.733	0.648	0.740	0.655	0.762
Q2		1	0.779	0.668	0.748	0.662	0.594	0.668	0.795	0.673	0.759	0.699	0.797
Q3			1	0.717	0.780	0.692	0.631	0.721	0.802	0.684	0.799	0.706	0.813
Q4				1	0.775	0.818	0.667	0.710	0.699	0.658	0.726	0.705	0.751
Q5					1	0.774	0.683	0.755	0.808	0.687	0.799	0.736	0.834
Q6						1	0.703	0.714	0.702	0.662	0.730	0.713	0.755
Q7							1	0.695	0.614	0.594	0.644	0.602	0.650
Q8								1	0.730	0.632	0.735	0.678	0.754
Q9									1	0.710	0.838	0.748	0.866
Q10										1	0.727	0.711	0.745
Q11											1	0.776	0.855
Q12												1	0.803
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 2021, 5.3% of students (n=3,298) responded 'no' to Question 16 ("Did you complete this form without undue influence from other _____?") and 5.3% of students (n=3,343) responded 'no' to Question 17 ("Did you complete this form without undue influence from the _____?"). Of these students, most (n=3,087) responded 'no' to both questions indicating that they may have misunderstood the question. In Spring 2022, 5.7% of students (n=3,840) responded 'no' to Question 16 ("Did you complete this form without undue influence from other _____?") and 5.7% of students (n=3,855) responded to these 1 students,

We also note that . This is demonstrated below using Fall 2021-Spring 2022 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.

literature review. Note, however, Beran et al. (2009) argue that these effects may be mediated by varying levels of student engagement.

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 14% of the variability in student evaluations (e.g., Uttl, White & Gonzalez, 2017; Clayson, 2009; Cohen, 1980). In addition, student evaluations can also contribute to a teacher-centered style of instruction.



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).

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 (47.4% in Fall 2021; 57.2% in Spring 2022). We also note that there is no evidence for a significant difference in student responses to Question 13 across the Fall and Spring semesters ($M_{\text{fall}} = 4.35$, $SD_{\text{fall}} = .96$; $M_{\text{spring}} = 4.37$, $SD_{\text{spring}} = .94$).

A study by Guder and Maliaris (2013) showed that the response rate of online evaluation raised 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.

In a quasi-experimental design study by Chavez & Mitchell (2020), faculty members teaching identical online courses recorded welcome videos that were presented to students at the course onset, constituting the sole exposure to perceived gender and race/ethnicity. Results showed that

fare worse than their white and native English-speaking counterparts (Kreitzer & Sweet-Cushman, 2021). While Saunders (2001) did not find differences in evaluations of instructors whose native language is English compared to those for whom English is a second language, Gill (1994) found that students view teachers with “standard North American accents” more favorably.

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 & 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 to be a consistent or systematic difference among the ratings of full professors compared with associate professors or of junior versus senior lecturers (Spooren, 2010; Ting, 2000).

Faculty and Student Perceptions

Research has shown that student evaluations are influenced by whether students perceive the evaluation process as making a difference. Chen and Hoshower (2003) found that students are motivated to participate in student evaluations “by the expectation that they will be able to provide meaningful feedback” (p. 71). Furthermore, Worthington (2002) found that “students who perceive the evaluation process as a process for improving teaching in the future...have a higher probability of giving a more favourable ranking” (p.61).

Other research shows that students may not believe that the opinions they express on their evaluations are taken seriously by faculty or administrators (Spencer & Schmelkin, 2002). Richardson’s (2005) comprehensive review of literature on student evaluations concluded that “[m]any students and teachers believe that student feedback is useful and informative, but for a number of reasons many teachers and institutions do not take student feedback sufficiently seriously” (387).

Some studies find that information from student evaluations does not contribute to changes in teaching practices (Blair & Valdez Noel, 2014; Kember et al., 2002; Nasser & Fresko, 2002; Spencer & Flyr, 1992). Others, however, find that student evaluations are generally perceived as useful for “formative and summative” purposes (Schmelkin et al., 1997, p. 588) and may lead to changes in instruction (Beran et al., 2005; Chan et al., 2014; Gravestock & Gregor-Greenleaf, 2008; Parasuk & Lebaron, 1999). Arthur (2009) lists four reasons why
