

Senate Task Force to Overhaul Student Opinion Surveys on Teaching Instruments (SOST)  
 Final Report  
 Executive Summary

The task force’s charge, timeline, and work process are detailed on pages 1-3 of the final report. All are encouraged to read the report in its entirety.

The following provides a brief summary of major findings and recommendations.

I. There is abundant research that the current practice of relying on an end-of-semester questionnaire not only provides little useful information about teaching effectiveness, but can harm faculty, particularly those from historically marginalized/excluded populations in higher education. At CSUSM, there is reason to believe that the reliance on an end-of-semester questionnaire is even more harmful to lecturer faculty for whom SOST and teaching-related materials form most of their Working Personnel Action Files (WPAFs). The current University RTP document requires reviewers to have “undergone anti-bias training, in include materials on bias in Student Opinion surveys on Teaching” thus the fundings from this task force further reinforce and provide greater detail on the already documented concern about bias in SOST. SOST can provide insight on the student experience which is different from assessment of student learning or even assessment of effective teaching by the instructor.

II. There are multiple approaches to the creation, administration, and application of student evaluation of instruction/student opinion surveys (SOS)\*. Common characteristics include:

- Multiple tools used at different times during the semester.
- Framework and tools that are discipline/pedagogy appropriate.
- Approaches that explicitly address DEI and mitigate bias.
  - This is an area where it is especially important the Colleges/Depts/Programs identify what is appropriate to their faculty and students.

III. Without the adoption and implementation of new approaches (likely by Colleges/Departments/Programs), the current, single end-of-semester questionnaire is to be used to fulfill the CBA’s requirement at detailed in 15.15. The task force provides some suggested revisions for the current, single end-of-semester questionnaire that may provide an intermediary step to mitigate potential for SOST bias while Colleges/Departments/Programs work on designing and implementing new approaches.

IV. While there are multiple recommendations in the report, there is no sequence to be followed for these recommendations. The recommendations are briefly noted below.

Recommendation	Within Senate purview	Not in Senate purview
<b><i>The Committee recommends that CSUSM fundamentally change institutional evaluation policies and processes, to implement structured and specific mechanisms to include multiple dimensions to evaluate instruction and</i></b>	Revise University RTP Revise University Lecturer Evaluation policy	Revise College/Department/Program RTP/Lecturer Evaluation policies (these are vetted by Senate but created at within Colleges/Departments/Programs)

<i>faculty. (page 9)</i>		
<b><i>Policy revisions: Appropriate referrals to Standing Senate Committees, Colleges and Departments (page 11)</i></b>	See above	See above
<b><i>Developing a framework that directly contributes to student learning and experience. (page 12)</i></b>	Not clear if this is something that could be achieved University-wide.	Colleges/Department/Programs could commence work on developing framework(s) appropriate to their fields.
<b><i>Recommendation: Mid-semester learning dialogues (page 15)</i></b>	Possibly within purview of FAC/LC/SAC.	Colleges/Departments/Programs can partner with the Faculty Center to develop approaches/toolkits for implementation.
<b><i>Revising the current SOS questionnaire (page 16)</i></b>	The questionnaire itself does not appear to be 'governed' by any existing Senate policy.	IPA/PAR** can act to respond to recommended revisions.

V. While the Senate, Colleges, and Departments engage in next steps, the CSUSM Faculty Center is poised to provide timely and ongoing support for faculty for discussing SOST and teaching in their WPAFs. Additionally, the Faculty Center is committed to programming that supports classroom best practices and a culture of continuous improvement. More specifically, the Faculty Center can provide routine workshops on:

- Best practices for candidates to discuss their teaching and contextualize SOST in their WPAF
  - Including alternative methods to discuss and demonstrate evidence of teaching success outside of SOST
- Training and/or guidance for faculty conducting peer classroom observations
- Mid Semester Learning Dialogues

Additionally, the Faculty Center can develop support resources and/or templates for:

- Discussing SOST as a candidate under review
- Discussing SOST as a reviewer
- Customizable mid-semester dialogue tools for use in the classroom

\*SOS is used in this executive summary of the task force report and includes the various terms used for this practice.

\*\*Student evaluations at CSUSM have been previously administered by Institutional Planning and Analysis but this process is transitioning to Planning and Academic Resources.

**SOST Task Force final report  
Submitted to the Executive Committee of the Academic Senate  
on February 10, 2023**

**Table of Contents**

I.	Task Force Charge, Composition, and Process.....	5
II.	Work process.....	6
III.	SOS in faculty evaluation requirement in the California State University (CSU) system .....	6
IV.	Background.....	7
IVa.	History and Validity of Student Opinion Surveys.....	7
IVb.	Bias Could Substantially Affect Responses to Student Opinion Surveys .....	9
IVc.	New Frameworks and Processes for Instructor Evaluation .....	11
V.	Recommendation: Fundamental changes to instructional and faculty evaluation .....	12
VI.	Recommendation: Policy revisions .....	14
VII.	Recommendation: Developing a framework that directly contributes to student learning and experience.....	15
1)	WHO will the framework benefit?.....	16
2)	WHAT are the specific objectives of the framework? .....	16
3)	HOW can activities provide immediate benefits to students? .....	18
VIII.	Recommendation: Revising the current SOS questionnaire.....	20
a.	Background .....	20
b.	Introduction to recommended changes .....	22
c.	Questions that should be deleted .....	22
d.	Questions that should be reordered in sequential appearance.....	25
e.	Questions that must be rephrased to reduce bias .....	25
f.	Questions specific to the online form.....	29
g.	The anti-bias statement .....	30
h.	What the revised SOS tool would look like with the above changes implemented.....	30
i.	Recommendations for administration of existing SOS tool .....	31
IX.	Next set of recommendations.....	<b>Error! Bookmark not defined.</b>
X.	Appendices .....	34
Appendix 1:	Ongoing and Historical Effort Related to SOS at CSUSM and CSU System-wide CSUSM – SOST .....	34
Appendix 2:	FAC Literature Review on the Biases of Student Opinion Survey (SOS) Spring 2022 .....	36
Appendix 3:	Efforts to reform Student Opinions of Teaching at other CSU campuses .....	42

**I. Task Force Charge, Composition, and Process**

The SOST Task Force was approved by the Academic Senate on April 20, 2022.

**RATIONALE:** Based on Academic Senate records, the last substantial overhaul of CSUSM’s Student Opinion Surveys on Teaching Instruments took place in 2004. Ongoing challenges exist with engaging students in this process, particularly with electronic distribution and collection of instruments. Additionally, a significant growing body of research literature examines bias in Student Opinion Surveys on Teaching. For these reasons, the evaluation of these instruments is timely and necessary.

**TASK FORCE CHARGE:** This task force is charged with evaluating and, as needed, recommending changes to CSUSM's Student Opinion Surveys on Teaching Instruments (CBA, Article 15.15) to reduce bias and increase the effectiveness of student evaluations based on an examination of relevant research literature and effective practices. The task force's work product shall be a recommended set of new instruments, to be approved by the Academic Senate, designed with the following goals in mind:

- Capture meaningful student input on course instruction and instructor effectiveness that can be utilized to improve pedagogy and evaluate performance;
  - Eliminate/minimize/mitigate bias against instructors;
  - Maximize student participation (particularly in online-only distribution and collection);
- and
- Develop questions relevant to varying modalities (online and hybrid instruction, for example).

**TIMELINE:** The work of the Task Force begins May, 2022 and runs through Summer 22 and Fall 22. Once the task force convenes, the task force members will select a chair and determine the modality of their meetings. Consultation with FAC and SAC, which have been investigating these instruments in AY 2021-22, should take place upon convening. We anticipate that the group will research and draft new instruments in Summer 22.

Draft instruments will be circulated and revised based on input across Colleges and faculty ranks during Fall 22, with final Senate approval of new instruments presumably by the end of that term. Consultation with Office of Inclusive Excellence, IITS, and the Office of Disability Support Services will occur as appropriate.

Faculty Affairs Committee (FC) Member (Fall 22)	Ruoxi Li, Political Science, CHABSS
Student Affairs Committee (SAC) Member-SAC (Fall 22) and Student Representative	Stephania Rey, Student
Lecturer Committee (LC) Member (Fall 22) and Lecturer Representative	Erika Shuh, Modern Language Studies, CHABSS

Anti-Racism, Anti-Colonialism, and Social Justice Committee Member (Fall 22) and Lecturer Representative	Lucia Gordon, Communication, CHABSS (summer 22)
CEHHS Representative	Devin Jindrich, Kinesiology, CEHHS
CoBA Representative	Qi Sun, Finance, CoBA
CSTEM Representative	Kambiz Hamadani, Chemistry and Biochemistry, CSTEM
Additional Lecturer Representative (total 3)	Janette Larson, Literature and Writing Studies, CHABSS
CFA Representative	Lori Walkington, Sociology, CHABSS
Institutional Planning & Analysis Representative	Cameron Stevenson, Interim Director, Institutional Planning & Analysis
Faculty Center Representative	Rebecca Lush, Faculty Center Director
Faculty Affairs Representative	Dmitri Ranieri
Dean Representative (Provost's Designee)	Jackie Trischman, Dean, CSTEM
Guest	Jimmy Young, Faculty Center Associate Director
Task Force Lead	Yvonne Nalani Meulemans, Academic Senate Chair

## II. Work process

Meetings began in June 2022 and continued through the Fall 2022 semester and into January 2023 on as-needed basis. In addition to task force meetings, asynchronous work was also done. Meeting agendas, minutes, and meeting recordings are available upon request.

Each member of the task force had the opportunity to closely review each draft and any substantive and/or copy edits throughout the process of creating this report. Of the twelve elected/appointed members that remained when the task force concluded their work in early February 2023 (excluding the task force lead and any guests), nine voted to endorse that the report accurately reflects the work and discussions of the task force. There was one vote of non-endorsement of the report. The remaining two task force members did not vote at all. (There was no abstention option.) In addition, task force members were given the opportunity to document any specific concerns/insights they had about the report, whether they voted to endorse the report or not. The following were noted:

- Concern about the general nature of the critique of the TEval approach from the Anti-racism, Anti-Colonialism, and Social Justice Committee.
- That the report did not include more ways to include student voices in the evaluation process.
- A desire to keep items on the current evaluation form regarding course organization/preparation, student learning outcomes, the instructor providing feedback on student work.

## III. SOS in faculty evaluation requirement in the California State University (CSU) system

The Collective Bargaining Agreement for Unit 3 CSU employees requires SOS in faculty evaluation. Specifics [are provided in Article 15.15-17](#) which are also included below.

[Process for Student Course Evaluations of Teaching Faculty Instructional Effectiveness](#)

15.15 Written or electronic student course evaluations of faculty instructional effectiveness, also called “student opinion survey” and “student perception of teaching effectiveness” on some campuses, shall be required for all faculty unit employees who teach. All classes taught by each faculty unit employee shall have such student course evaluations unless the President has approved a requirement to evaluate fewer classes after consideration of the recommendations of appropriate faculty committee(s). In cases where student course evaluations are not required for all classes, classes chosen for evaluation shall be representative of the faculty unit employee's teaching assignment, and shall be jointly determined in consultation between the faculty unit employee being evaluated and their department chair. In the event of disagreement, each party shall select 50% of the courses to be evaluated. The results of these evaluations shall be placed in the faculty unit employee's Personnel Action File. Results of course evaluations may be stored in electronic format and incorporated by extension into the Personnel Action File provided that individuals involved in evaluations and personnel recommendations or decisions are provided secure access for these purposes.

15.16 Students may, with the concurrence of the department and administrator, be provided an opportunity to consult with the department peer review committee.

15.17

a. Student course evaluations collected as part of the regular student evaluation process shall be anonymous and identified only by course and/or section. The format of student course evaluations shall be quantitative (e.g., “Scantron” form, etc.) or a combination of quantitative and qualitative (e.g., space provided on the quantitative form for student comments).

b. Any student communications or evaluations provided outside of the regular evaluation process must be identified by name to be included in a Personnel or Working Personnel Action File.

c. Faculty unit employees may submit written rebuttals to student course evaluations pursuant to Provision 11.2 when it is believed that additional information is needed or in the case of student bias. Evaluators must review such written rebuttals when reviewing underlying student course evaluations.

## **IV. Background**

### **IVa. History and Validity of Student Opinion Surveys**

Student Opinion Surveys (SOSs) have been used in instruction for at least 100 years ([Kulik 2001](#); [Wachtel, 1998](#)). The original purpose of SOSs was to provide feedback to instructors, so that instructors could improve the educational experiences of the students. Since the 1970s, SOSs have taken on a different role: a means for students to provide information about instructors to the university institution ([Ory, 2000](#)). Currently, SOSs are widely used for personnel decisions such as retention, tenure, and promotion (RTP; [Spooren, 2013](#)). Therefore, the purpose of SOSs has shifted, from their original purpose to help instructors, to their current role as primarily helping institutions to *evaluate* instructors.

The use of SOSs as tools to help institutions make personnel decisions has led research and development of assessment instruments to focus on questions of *validity*: on whether instruments accurately enable institutions to measure student learning ([Onwuegbuzie, 2009](#)). Many small studies, large-scale studies, and meta-analyses have been conducted to determine the validity of the SOS instruments ([Benton and Cashin, 2011](#), [Wright and Jenkins-Guarnieri, 2012](#)). Although student learning and performance is extremely difficult to measure in a controlled way given the large number of variables (different students, instructors, courses, etc.), early meta-analyses of the relationship between SOSs and measures of student performance suggested moderate correlations between student academic performance (assumed to be related to long-term retention and learning) and SET scores (Cohen, [1980](#), [1981](#)). Moreover, proponents of SET instruments have used sophisticated data analysis techniques such as factor analysis to try to account for the large number of covarying factors that could influence SOS responses, arguing that finding independent factors identified by SOSs indicate that SOSs successfully measure distinct aspects of learning instead of general biases that might undermine the validity of SOSs as measures of student learning ([Marsh, 1984](#)). The apparent validity of SOSs for evaluating instructor performance supported the utility of SOSs for instructors and institutions and contributed to their wide adoption ([Clayson, 2009](#)).

However, despite arguments by proponents that SOSs are valid and reliable measures of instructor performance, the validity of SOSs has been the subject of intense controversy for as long as they have been in use ([Marsh, 1984](#); [Benton and Cashin, 2011](#); [Uttl et al., 2017](#)). Recent studies suggest that SOSs may not be as valid or useful as previously thought. For example, a recent re-analysis of the data from early meta-analyses suggests that SOS scores are not actually related to student learning ([Uttl et al., 2017](#)). SOS may measure short-term student satisfaction more than they measure long-term learning ([Kornell and Hausman, 2016](#)). Moreover, student opinions are subject to many systematic biases unrelated to learning that could be harmful to any instructor depending on context, but often impact instructors in marginalized groups ([Kreitzer & Sweet-Cushman, 2022](#); see section IVb below). The inability of surveys to accurately assess student preferences may, in part, reflect more general limitations to survey methodology for determining preferences ([Quaife et al., 2018](#)). Therefore, after over 50 years of research and thousands of research studies, whether SOSs can be considered valid indicators of student performance (and therefore instructor effectiveness), or are irredeemably corrupted by bias, remains unclear.

One potential reason for the difficulty of measuring instructor effectiveness with SOSs is the indirect nature of the measurement process. To evaluate instructor effectiveness, institutions employ SOSs to survey students. Student responses to surveys about the teaching methods used by instructors could be influenced by many factors. For example, students respond to surveys based on (unmeasured) preconceptions about what effective instruction is (and how effective instructors should look and act; [Ory and Ryan, 2001](#)). However, it is often unknown what students consider to be effective instruction, and student preconceptions about teaching quality may have been shaped by educational experiences that were not based in strong pedagogical practices ([Pallas et al., 2017](#)). The extent that survey questions themselves are consistent with students' evaluation of effective instruction could influence responses. Moreover, if an instructor uses teaching approaches that are unconventional, or different from students' previous experiences, it could adversely affect student responses even if the instructor's practices are effective for learning ([Simonson et al., 2022](#)). For example, using high-impact teaching practices that contribute to learning can result in lower teaching evaluations from students ([Carpenter et al., 2020](#)). Evaluation using SOSs could therefore be a *disincentive* to using unconventional but high-



impact teaching practices, and an impediment to adapting and improving educational practices ([Stroebe, 2016](#)).

Another factor that could reduce the validity of SOSs is student motivation to provide substantive and thoughtful information. SOS administration and score interpretation will be most valid and reliable when students are making good faith efforts to honestly evaluate teaching effectiveness ([Benton and Young, 2018](#)). However, students often provide *inaccurate* information on surveys ([Clayson and Haley, 2014](#)). Students have little incentive to make efforts to complete SOSs for their institutions, as evidenced by the low response rates observed when SOSs are voluntary (e.g. when using online surveys; [Luo, 2020](#)). Paradoxically, an optional survey with a lower response rate (where responses required intention from the students) could potentially contain more useful information than a coerced survey with a higher response rate (which could include many meaningless responses). In any case, whether the sampling of survey respondents reflects the students as a whole always has the potential to affect outcomes ([Wolbring, 2016](#)).

Most often, students do not directly benefit from completing SOS, but instead are requested to volunteer their time and thought during some of the most time-demanding and stressful periods of the semester (typically the final weeks). Even when students are offered incentives such as extra credit to complete SOSs, students have little incentive to spend time on the surveys themselves. On anonymous surveys, students skip questions to save time, or simply select a single score for all questions (e.g. “1” or “5” on a 5-point scale) based on a general satisfaction with the instructor, the course, their overall well-being, or other factors ([Uijtdehaage and O'Neal, 2015](#)). Therefore, a lack of incentives for student engagement and participation could decrease the validity of resulting data, in part by allowing biases to influence student responses.

Finally, although student experiences are perhaps the most important aspects of college courses, student experiences are inherently limited. In the best circumstances, students could provide valid information about their experiences and learning, and observations about the instructor’s management of a course. However, students can only be expected to evaluate the information and activities that were actually included in the class. Students cannot be expected to evaluate a course with respect to the overall department or university curricula, or with respect to the role or requirements of the course within a discipline. For example, students cannot reasonably be expected to identify content, concepts, or skills that may be common or necessary for a course to cover a topic but are missing from the particular course in which the students were enrolled. For example, courses with little content and high grades could be rewarded with positive responses to SOSs ([Krautmann and Sander, 1999](#)). The missed opportunities for learning in the short term, and long-term impacts when the students do not have the preparation or skills important for subsequent courses, will not be reflected in SOSs collected at the end of a given term ([Braga et al., 2014](#)). The fundamental inability of SOSs to provide critical information about course content necessitates other processes to evaluate courses in the context of curricular and disciplinary requirements and expectations.

#### **IVb. Bias Could Substantially Affects Responses to Student Opinion Surveys**

**FAC completed an initial literature review on the issues surrounding student evaluation of instruction in Spring 2022 (See Appendix 2).**

Human perceptions and decisions are influenced by many cognitive biases that are difficult or impossible to avoid ([Tversky and Kahneman, 1974](#)). People are biased towards information that

confirms their preconceptions or beliefs or impressions developed early in an experience rather than at later in time ([Nickerson, 1998](#)). Moreover, negative events affect perceptions and are remembered more than positive events of the same type ([Baumeister et al., 2001](#)). Surprisingly, there seem to have been few empirical investigations of the effects of general cognitive biases on Student Opinion Surveys, and the potential impacts on responses to SOSs. However, it remains likely that SOSs are systematically affected by the same general cognitive biases that affect perceptions and decision-making in many other contexts, which could contribute to decreasing the validity of SOSs as measures of teaching quality or learning.

Many specific factors can also bias student evaluations and decrease the validity of SOSs. External factors that have been shown to affect SOS evaluations include expected grades (grading leniency; [Brockx et al., 2011](#)), course difficulty and workload ([Marsh and Roche, 2000](#)), course type (e.g. lecture vs. Laboratory; [Langbein, 1994](#)), course modality (e.g. online vs. In-person; [Ayllon, 2022](#)), course level ([Marsh, 1980](#)), discipline (e.g. natural sciences, humanities; [Arroyo-Barrigüete et al., 2022](#)), class size ([Feldman, 1978](#)), whether the course is required or an elective ([Wachtel, 2006](#)), the timing of SOS administration ([Frey, 1976](#)), the specific design of survey instruments ([Courey and Lee, 2021](#)), survey modality (e.g. paper vs. online; [Loveland, 2007](#)), survey response rate ([Koh and Tan, 1997](#)), student interest in course subject ([Langbein, 1994](#)), overall student capability ([Culver et al., 2020](#)), student gender ([Boring et al., 2016](#)), instructor gender ([Boring, 2017](#)), instructor age and experience ([Arbuckle and Williams, 2003](#)), instructor race and/or ethnicity ([Wang and Gonzalez, 2020](#)), student race and/or ethnicity ([Arnold and Versluis, 2019](#)), instructor's native language ([Fan et al., 2019](#)), instructor appearance ([Felton et al., 2008](#)), instructor personality ([Clayson and Sheffet, 2006](#)), among other factors ([Stoesz et al., 2021](#)). Even activities such as providing chocolate ([Youmans and Jee, 2007](#)) can influence SOS scores, suggesting that many other un-measured factors may bias the results of SOSs. Clearly, many aspects of courses, SOS administration, and attributes of the students and instructors that are unrelated to learning can affect student evaluations of teaching.

Evaluating the potential for bias to affect SOSs is complicated because many effects are relatively small by themselves, and different studies report conflicting results ([Stoesz et al., 2021](#)). Moreover, many factors can interact ([McPherson and Jewell, 2007](#)). For example, gender is among the most-researched influences on student opinions, with many studies observing gender effects on teaching evaluations ([Mengel et al., 2019](#)). Whereas the *average* effects of gender may be small, the *interaction* of gender with other factors can be much more pronounced ([Basow and Martin, 2012](#)). For example, evaluations may also reflect the interactions between gender and student preconceptions, where instructors are punished for gender non-conformity with established stereotypes ([Freeman, 1994](#)). Student gender can also interact with instructor gender, resulting in students giving higher evaluations to instructors of the same gender ("affinity effects;" [Young et al., 2009](#)). In a particular situation, interactions such as stereotype nonconformity and affinity effects could affect evaluations, in additive or in opposite ways ([Anderson and Smith, 2005](#)).

The types of biases observed for gender are also likely to affect other marginalized groups such as BIPOC or LGBTQ instructors. Under-representation of BIPOC and LGBTQ faculty has limited research on bias against marginalized groups ([Kreitzer & Sweet-Cushman, 2022](#)). However, non-white instructors have been shown, on average, to be given scores lower than white instructors ([Reid 2010; Smith and Hawkins, 2011](#)). Similar to gender, affinity effects could also create interactions between student and instructor identities that affect SOS scores ([Gith, 2020](#)). Analyses of student written comments have revealed that student biases can manifest themselves as being cruel and abusive against marginalized faculty ([Heffernan, 2022; Wallace et al., 2018](#)). Therefore, the use of SOSs in faculty evaluation could

provide yet another mechanism whereby biases and stereotypes negatively impact underrepresented and marginalized faculty.

In summary, student opinion surveys are widely used by institutions (universities) to evaluate instructors but may not be valid measurements of instructor effectiveness. Among the reasons that SOSs may not be valid measurements of instructor performance are (1) that information from surveys of students reflect complex interactions between student preconceptions of quality instruction, the survey instrument itself, and the practices of the instructor (among other things); (2) students have little incentive to volunteer the time and effort necessary to thoughtfully complete SOSs, providing many openings for biases to influence responses; and (3) student opinion surveys have been demonstrated to be biased by many factors, including biases against marginalized groups. A common aspect of many sources of bias are that students often provide low ratings to aspects of courses that challenge them, in terms of course content, unfamiliar activities and/or assessments, and the extent to which courses or instructors challenge student preconceptions and stereotypes.

#### **IVc. New Frameworks and Processes for Instructor Evaluation**

Because Student Opinion Surveys are indirect measures of instructional effectiveness with questionable validity, and are biased by many factors, many institutions are currently re-evaluating the use of SOSs in instructional evaluation.

The task force gathered information about current practices for instructional evaluation at other CSU campuses, including some efforts to reform the process of instructional evaluation (Appendix 3). Most CSU campuses use SOSs like those used by CSUSM. However, some campuses are making efforts to reform the evaluation process. Examples include a proposal at San Francisco State University ([JEDI-TEA](#)) that seeks to use a “balanced approach, employing multiple instruments to gather data from diverse perspectives and for disparate uses.” (Cited with permission.)

Another prominent effort to reform faculty evaluation is the National Science Foundation (NSF)-funded [TEval collaboration](#), which “encourages the use of evaluation, incentive, and reward processes as a lever to promote greater use of evidence-based teaching within universities as complex systems.” The TEval approach seeks to change instructional evaluation and campus cultures to incentivize faculty to use evidence-based educational practices (EBEPs) that have been demonstrated to be associated with student learning. Among the efforts of TEval is to define quality teaching (e.g. the [Teaching Quality Framework \(TQF\) at University of Colorado](#)), which identifies seven core components — goals, content, and alignment; preparation for teaching; methods and teaching practices; presentation and student interaction; student outcomes; mentorship and advising; and reflection, development, and teaching service / scholarship. Other TEval institutions ([University of Kansas](#)) have used the core components to develop benchmarks for holistic evaluation of instructors and a detailed rubric for instructor evaluation. The [University of Massachusetts](#) has established principles using multiple dimensions to evaluate teaching, including using multiple lenses (multiple sources and types of data from instructors, students, and peers), triangulation, both formative and summative assessments, and including a balance between uniformity across units and customization to different disciplines. Different types of reviewers are necessary to contribute to each benchmark with different weights. For example, students could address class climate, but peers or other evaluators are necessary to address other benchmarks such as content and alignment.

The task force learned more about how the University of Colorado considered and implemented TEval through a conversation with Dr. Teresa E. Foley, faculty in Integrative Physiology. Dr. Foley generously shared the materials her institution has and are using. In this meeting, Dr. Foley was asked how TEval instruments mitigate student bias. Dr. Foley's response confirmed some committee members' concerns that controlling for bias was not adequately considered during the development of the TEval instrument.

Some members of the committee noted concerns that the institutions where TEval was developed are predominantly white and raised the question of whether TEval sufficiently addresses issues related to racial diversity, equity, and inclusion. For example, it was unclear to some task force members how the benchmarks/measures addressed student bias in SOS. Therefore, the Anti-racism, Anti-Colonialism, and Social Justice Committee of the Academic Senate were asked to review the materials with focus on how these materials address not only concerns about providing opportunities for biased responses from students but also reflecting the diversity of CSUSM's students and faculty. The Committee's response (see Appendix 4) expresses concern that the TEval materials do not adequately address these issues. With this noted, the TEval's general approach does provide another well-developed example of assessing multiple dimensions of teaching through multiple tools appropriate to discipline and pedagogy, it insufficiently addresses/mitigates bias in student evaluations of instructor effectiveness.

Other institutions like the [University of Oregon](#) have also revised their evaluation process to include multiple dimensions of teaching. Taken together, there are clear commonalities in contemporary approaches to student evaluation of instruction. The taskforce's recommendations are reflective of these commonalities.

#### **V. Recommendation: Fundamental changes to instructional and faculty evaluation**

The Committee recommends that CSUSM fundamentally change institutional evaluation policies and processes, to **implement structured and specific mechanisms to include multiple dimensions to evaluate instruction and faculty.**

The current Working Personnel Action File (WPAF) process does provide some mechanisms for evaluating faculty in multiple dimensions (e.g. faculty include both SOS evaluations and often choose examples of course materials and student work). However, current processes are vague, unstructured, and confusing. For example, much faculty and review committee effort is devoted to defining what constitutes an "item" in WPAFs. However, once decisions are made about what the acceptable "items" of evidence are, the standards and processes for evaluating the items relative to curricular or disciplinary needs and expectations are not well defined. Over-reliance on SOSs for faculty evaluation is due, in part, to the fact that SOSs are the one part of WPAFs where the processes and data seem consistent and clear (although, as explained above, the apparent clarity is extremely deceptive).

The task force recommends that more structured and consistent processes be implemented to ensure that multiple dimensions are included in faculty evaluation. For example, the University could modify the TEval approach to create formal processes to include multiple dimensions in evaluation, create more specific definitions of quality teaching, and develop benchmarks for instructor effectiveness. Multiple dimensions include different "lenses" (sources and types of data from instructors, students, and peers), triangulation, and both formative and summative assessments. Like TeVAL, processes should reflect a balance between uniformity across units and customization to different disciplines. However, although approaches like the TEval approach includes many useful principles, the task force recognizes that any policies implemented need to account for the unique institutional and student composition of CSUSM.

Specifically, the task force identified three dimensions that are currently insufficiently structured in the CSUSM evaluation process:

- 1) Instructor training and peer evaluation. The University should support processes for instructor training to help instructors use effective, evidence-based teaching practices. Institutional evaluation processes should also include peer feedback and evaluation as an important perspective on effective instructional practices.
  - a. Peer evaluation holds particularly strong promise as it could be relatively quickly implemented because this is not a new practice in evaluating teaching effectiveness. Indeed, there are Departments at CSUSM that already require peer evaluation and provide significant training and guidance to all who participate. The Faculty Center can provide leadership on any campus-wide efforts to adopt this practice.
- 2) Course evaluation with respect to student needs. Consistent with the first two goals of [the CSUSM strategic plan \(Academic Excellence and Student Success\)](#), the University should support processes to ensure that students have opportunities to learn information, concepts, and skills that are consistent with curricular needs and disciplinary expectations, so that the long-term needs of the students from a particular course are met (e.g. success in subsequent coursework, career preparation, etc.).
  - a. Using SOSs for evaluation creates incentives to prioritize short-term expediency over long-term benefits. For example, using SOSs may incentivize course experiences that minimize challenges to students, in terms of course content, expectations for teaching practices and learning processes, or challenges to stereotypes or preconceptions. Instructor evaluation should instead incentivize the contribution of courses and instructors to defined, meaningful, and long-term learning objectives that contribute to student success through academic excellence.
- 3) Formative assessment to directly improve student experiences. Consistent with the third and fourth goals of [the CSUSM strategic plan \(Diversity, Equity & Inclusive Excellence and Culture of Care\)](#), student voices should be heard, and student time and thought should be respected and valued. The University should support processes that include student voices in evaluation in ways that can immediately improve learning experiences. For example, implementing Mid-Semester Learning Dialogues (MLDs) based on the SELF framework (see section VII below) could contribute to instruction.
  - a. Formative assessments designed to benefit students could also provide high-quality information about student experiences for instructional evaluation

Implementing more structured and consistent evaluation frameworks and processes is vital to ensuring that SOSs are only one part of a comprehensive, multi-dimensional approach to evaluating faculty, and preventing the fundamental limitations and biases of SOSs from disproportionately affecting faculty evaluation.

The proposed changes to faculty evaluation will clearly require substantial workload for faculty at the Department/College level. It is important to note that the institutions that have already adopted similar approaches (e.g. TEval) are far larger than CSUSM. This larger size meant there were simply more faculty that could participate in this sometimes discipline-specific work. To simply recommend that all Departments convene groups to consider and adopt possible tools and practices appropriate to their

discipline is not feasible at CSUSM. Some Departments at CSUSM have few faculty and even larger Departments may be hard pressed to convene a group for what is sure to be a complex initial implementation process that will require 'upkeep' in the form of regular review and possible revisions. Therefore, the task force strongly recommends the following:

- Commence this work with college level workgroups that ensure Department level representation and/or collaboration.
- Partner and closely collaborate with the Faculty Center to provide professional development and/or support to faculty engaged in this initial implementation.
  - Specifically, the Faculty Center can lead on providing templates that departments/programs could adopt and modify, in a fashion similar to how RTP Standards are based on similar templates.
  - The Faculty Center can lead on providing training and support for faculty on best practices for reviewing, reflecting on, and discussing their SOST when preparing a WPAF
- Ensure workgroups include faculty with some knowledge/expertise of how racial, gender, ethnic and other forms of bias inform student response to these practices.
- Consider revising program assessment efforts as a means to contribute to evaluation of instruction. For example, a Department/Program could consider adopting practices in order to gain insight into effective teaching practices in particular courses; such findings could also be included in instructors' WPAF's (if they so choose.)
- Determine a detailed timeline in each College for examination, implementation, codifying, and documenting the new practices.
- Use significant piloting and/or an early-adopter approach before making new evaluation practices required.
- Provide appropriate compensation to workgroup members and early adopters in the form of assigned time and/or other compensation.
- Ensure lecturers are full participants in this work, with consideration given to their evaluation practices.

## **VI. Recommendation: Policy revisions**

### **a. Appropriate referrals to Standing Senate Committees, Colleges and Departments**

Clearly, any changes to SOS practices will require appropriate policy revisions. The task force concluded that including specific changes to specific policies in this report is premature. However, as noted in IV. b., any workgroup convened for this effort must also identify and move in a timely fashion to update evaluation policies.

The task force also recommends that Colleges/Departments prioritize revising lecturer evaluation policies/practices given the rather narrow approach of these policies/practices vis-à-vis tenure-track evaluation. As the implementation effort advances and when appropriate, FAC and LC would receive a referral to revise the University-level evaluation policies to reflect the newly adopted approaches.

The task force also recommends a university-wide policy for whatever practices are adopted. While a primary value of this multi-faceted approach is that Departments/College practices are appropriate for



their field, a university-wide policy would provide additional assurance that all faculty are treated equitably no matter what their Department/College chooses to do.

There has been some effort already to increase awareness of and response to bias in SOST and faculty evaluation writ large. Since Fall 2022, the campus has adopted a new [University RTP document](#) (see sections III.B.3.a, III.D.1, III E.1., III.G.1. and III.H.4.) that requires reviewers to have undergone anti-bias training. The Office of Inclusive Excellence and the Faculty Center have partnered to create training and a resource guides; which [is available to all reviewers](#).

### **VII. Recommendation: Developing a framework that directly contributes to student learning and experience.**

Efforts to reform faculty evaluation at other CSU campuses, TEval, the University of Oregon, and others provide important examples of the fundamental changes to the processes for valid and equitable evaluation of faculty instruction at predominantly white institutions (PWIs). However, current approaches to revising instructor evaluation share two major limitations that limit the ability to implement similar approaches in the intermediate term at CSUSM.

- 1) Implementing [new institutional frameworks](#) for evaluating instructors is a long-term process that involves substantial commitments from the entire university: faculty, departments, and administration, within – and perhaps beyond - the current requirements of the CBA. Although the Committee does recommend making fundamental changes to the process of instructor evaluation, major structural changes will likely require many years to implement.
- 2) The primary objective of evaluation frameworks is commonly to serve institutions: to help universities evaluate instructors, and are not primarily designed to help instructors or students directly ([Bowman, 2017](#)). Some revised approaches shift some focus to serving instructors by providing guidance and information to improve teaching. For example, providing [rubrics](#) or [benchmarks](#) that identify effective teaching, or using [SOSs as formative assessments](#) to provide instructors with more immediate feedback that can be used to improve their approach during a term. However, one group of stakeholders still derives little immediate benefit from current or previously proposed evaluation processes: students.

Students are asked to volunteer their time and thought to evaluation in the hopes that *future* students will benefit from their effective evaluation of instructors. Even in the case of mid-semester evaluations, the opportunities for immediate benefits to students are often filtered through the need for instructors to beneficially adapt teaching practices based on student feedback. Therefore, instruments that survey students to evaluate instructors provide poor information about the instructional quality that they seek to measure (providing only indirect information that is based on poorly-defined assumptions and preconceptions, and influenced by many biases), but also do not provide any immediate benefit to the students who are expected to volunteer their time and thought.

Similar to the conclusions of [John Hattie](#): “We need to shift from focusing on the impact of talking to focusing on the power of listening.” Evaluation of teaching should, in part, involve a reciprocal contract between both students and instructors to achieve learning outcomes ([Ching, 2018](#)).

**The task force recommends developing a framework that *directly* assesses student learning and experiences, with the primary objective of providing immediate feedback and benefits to students.**

Developing a student-focused framework is a reasonable step in the intermediate-term (1-2 years) because the framework would be complementary, and not competitive, with institutional processes. Student-focused activities/assessments could be developed and implemented over time without requiring changes to university policies or the CBA. Moreover, we propose using a dynamic approach that can be modified and improved over time to include input from students, faculty, and administrators.

To develop a student-focused framework for activities/assessments, the task force proposes to address three questions:

- 1) **WHO will the framework benefit?**
- 2) **WHAT are the specific objectives of the framework?**
- 3) **HOW can activities provide immediate benefits to students?**

Addressing each question in turn:

### 1) **WHO will the framework benefit?**

The task force recommends the development and implementation of a framework that has the express objective of helping students gain immediate benefits.

How can students immediately benefit from assessment? We propose that if assessing student learning is the goal, then students can directly benefit in many ways. For example, there is substantial evidence that “**metacognition**,” or conscious attention to learning, can improve learning ([National Research Council, 1999](#); [Donkers et al., 2013](#)). Encouraging metacognition is one way to make learning more active, even in courses where other types of active learning strategies may be difficult to employ (e.g. larger lecture courses). Metacognitive skills can contribute to self-regulated learning, providing long-term benefits in many learning contexts ([Schraw et al., 2006](#)).

Students could also immediately benefit from activities that facilitate **communication** among course participants: students and instructors alike. Among the immediate benefits of communication are understanding instructor assumptions and expectations, establishing patterns of communication with other individuals, and contributing to a sense of community and shared purpose ([Xie and Derakhshan, 2021](#)).

Of course, direct benefits to students do not preclude indirect benefits to others. Instructors would also benefit from engaged students who are actively thinking about their own learning and communicating with each other and the instructor. If information from activities that benefit students are ultimately used by the university for assessment, the information will be likely to reflect judgments and opinions that faithfully reflect the students’ experiences and increase the validity of student contributions to multi-dimensional evaluations of faculty teaching. Therefore, student-centered approaches would be likely to provide unique and valuable information for others as well.

### 2) **WHAT are the specific objectives of the framework?**

Many SOS instruments currently in use were not developed through a systematic validation process ([Spooren, 2013](#); [Marsh, 1984](#)). However, SOS instruments have been developed using a variety of approaches intended to improve validity, including using particular theoretical frameworks




([Mortelmans and Spooren, 2009](#)), panels of experts ([Barnes et al., 2008](#); [Toland and Ayala, 2005](#)), surveys of instructors ([Keeley et al., 2010](#)), surveys of students ([Kember and Leung, 2008](#)), or hybrid approaches using literature, faculty, students and administrators ([Marsh, 1984](#)).

More recent efforts have called for defining content domains for evaluation instruments based on evidence-based teaching practices ([Finkelstein et al., 2017](#)). Clearly, evaluations of teaching that reward effective, evidence-based andragogical methods are an important goal. However, there are inherent difficulties in creating standards for effective teaching. What constitutes effective teaching may depend on discipline, course type and level, course medium (in-person, online, synchronous, asynchronous, etc.), student identities, and a host of other factors. Moreover, there is a multiplicity of ways for teachers to be effective, and students are likely to benefit from a diversity of teaching approaches that are each effective in their own way. Teaching practices must also evolve over time as evidence, technology, cultural, and economic opportunities (and constraints) change.

**The task force proposes that the hallmarks of positive learning experiences may be more consistent over discipline, individual, and time than teaching strategies are.** That is, while there has been much effort devoted to identifying the hallmarks of effective teaching as well as learning, the task force suggests that more consideration on effective learning may be more effective than trying to reduce teaching to a limited set of attributes. Moreover, students can directly evaluate the impact of a variety of teaching approaches on themselves, specifically on how they are personally affected in ways that influence learning. Therefore, defining a content domain focused on student learning (instead of instructor teaching) may help to develop instruments that provide information directly relevant to the ultimate goal of student learning, but minimize assumptions or prescriptions about particular teaching practices.

An example of a student- and learning-centered assessment domain is a “Student Experience of Learning Framework” (SELF), which contains five dimensions: the student’s engagement with course CONTENT AND SKILLS, the LEARNING PROCESS, the expected and realized GOALS, OUTCOMES and ASSESSMENTS of the course, the course CLIMATE, and the INSTRUCTOR. The proposed SELF framework addresses student engagement with [content, peers and instructors](#), and is consistent with frameworks developed for instructional and faculty evaluation ([Follmer Greenhoot et al., 2020](#)).

CSUSM STUDENT EVALUATION of LEARNING FRAMEWORK (SELF)



<p><b>CONTENT AND SKILLS</b> The content and skills of the course are interesting and relevant to me.</p> <p><i>Examples:</i> *Course content is new and <u>relevant</u> to me *The course content aligns with my previous <u>interests</u> *The course content made me interested in <u>new</u> things *The course helps me develop <u>new skills</u> *The course helps me to gain a broader and/or deeper <u>perspective</u> on the field *The course topics clearly <u>relate</u> to each other *The course material is <u>appropriately challenging</u></p>	<p><b>LEARNING PROCESS</b> I am interested and engaged in learning.</p> <p><i>Examples:</i> *Learning builds on things that I <u>already know</u> *I have many opportunities to make course content <u>meaningful</u> to me *There are <u>many ways</u> for me to learn *I receive enough timely <u>feedback</u> to improve my performance *Out-of-class activities are rewarding and provide valuable <u>practice</u> *I have opportunities to be <u>autonomous, independent, and creative.</u></p>	<p><b>GOALS, OUTCOMES &amp; ASSESSMENTS</b> Class goals are clear, valuable, and achievable.</p> <p><i>Examples:</i> *The content and skill learning goals of the course are <u>clear</u> *I am learning concepts and skills that I will be able to <u>remember</u> for a long time *I am learning concepts and skills that will be <u>useful</u> for me in the long term *I have opportunities to demonstrate <u>many types</u> of learning in assessments *The class experience has helped me <u>achieve</u> the course learning goals</p>	<p><b>CLIMATE</b> The class is structured to support my learning. I feel comfortable and respected for who I am.</p> <p><i>Examples:</i> *I feel <u>respected</u> as a person by the instructor and my peers * I have <u>opportunities</u> to contribute to class discussions and work * My particular strengths and perspective are <u>valued</u> *The class is a vibrant learning <u>community</u> *I can <u>safely</u> provide feedback to the instructor *The structure of the course is <u>fair</u> for all students</p>	<p><b>THE INSTRUCTOR</b> I enjoy learning with the instructor.</p> <p><i>Examples:</i> *The instructor helps me develop an <u>interest</u> in course topics *The instructor's <u>explanations</u> help me understand course material *The instructor treats me and others with <u>respect</u> *The <u>expectations</u> of the instructor are consistent over time and for different parts of the course *The <u>organization and communication</u> used by the instructor allows sufficient time for learning</p>
---	--	--	---	---

The SELF framework is based on the assumptions that students will learn effectively when they are motivated, course content is interesting and relevant to them, sufficient scaffolding is provided for meaningful learning, and the students feel safe, respected, and valued by the other members of the class (including the instructor).

Unlike frameworks that are expressed from the perspective of reviewing faculty (e.g. [Follmer Greenhoot et al., 2020](#)), the SELF framework is expressed from a student perspective to emphasize that the role of students is to directly assess (and improve) their own learning experiences, NOT for students to assess the effectiveness or attributes of the instructor.

It is important to note that the above framework is not formally endorsed by the task force but instead helps to illustrate that evaluation of teaching effectiveness can and should be student/learner centered.

The student perspective will undoubtedly be a helpful source of feedback to instructors. Moreover, student perspectives should remain an important part of institutional assessment. However, it will be important to avoid the type “mission creep” that contributed to SOSs shifting from providing instructor feedback to becoming a primary element of faculty evaluation. The task force recommends that direct assessment of student learning experiences be designed for, and limited to, understanding the student perspective.

**3) HOW can activities provide immediate benefits to students?**

To directly benefit students, potential activities need to be at a time when students have familiarity with a course and an instructor, but enough time remains in a course for meaningful changes to occur. The time constraints suggest using mid-semester activities as formative assessments ([Byrne and Donlan, 2020](#)).

**VIII. Recommendation: Mid-semester Learning Dialogues**

The task force recommends the creation of a system to facilitate “Mid-semester Learning Dialogues” (MLDs), brief pauses in courses that allow students to reflect on teaching and learning with the express goal of improving student learning experiences while a course is ongoing. MLD activities could be designed to explore the dimensions of the SELF framework, to characterize student experiences in a given course.

The Faculty Center can provide leadership via workshops, toolkits, templates, and further examples.

Learning dialogues could take different forms. Examples of the diversity of activities that could be used are:

a. Activities

- “Cogenerative dialogues” among instructors and students ([Hsu, 2018](#)).
- Card sort (analogous to a [values assessment card sort](#)) to stimulate a discussion about priorities and constraints for learning ([Bissonnette et al., 2017](#)).
- Debrief activities ([Sutherland et al., 2019](#)).
- “Student Teachers” – have students design a plan to teach future students a course topic.
- “Find the golden sentence:” as a team, find sentences in a short reading about learning that express a positive aspect of the process of learning in the course (and an aspect that could be improved) ([Greenleaf et al., 2012](#)).

b. Surveys

- Provide template examples (grab-n-go) and encourage editing/adding questions as appropriate for class/department/majors.
- Creating a minimum “off the shelf, ready to go” survey for instructors who wish for minimum hassle, with additional, categorized questions available to select from for those who wish to customize.

Relationship of the MLD framework to long-term recommendations:

This proposal is consistent with the charge of the Task Force, to:

- Capture meaningful student input on course instruction and instructor effectiveness that can be utilized to improve pedagogy and evaluate performance;
- Eliminate/minimize/mitigate bias against instructors;
- Maximize student participation (particularly in online-only distribution and collection);

In summary, the objective of the proposed SELF/MLD activities is to empower students to directly evaluate their class experiences in ways that immediately contribute to learning. The MLD framework is an expression of the values of CSUSM (as embodied in its mission and vision) and could potentially be a touchstone to encourage strategic thinking (metacognition) and communication among students and instructors.

## **IX. Recommendation: Revising the current SOS questionnaire**

### **a. Background**

The SOS currently used at CSUSM was first adopted in 2004. The survey was administered during the last two weeks of each semester, mostly using the in person, paper format before Spring 2020 (except for fully online courses), and mostly using the online, electronic format since Spring 2020 because of the COVID-19 pandemic. Data collected in SOS are required to be included in instructors' Personnel Action File (PAF) according to the Collective Bargaining Agreement (CBA).<sup>1</sup> Notably, the collection of SOS data must be done "anonymously" rather than confidentially. (Please see section III for the full passage of the CBA, articles 15.15-17.)

The task force wishes to emphasize that the CBA provides few details on what is required in the PAF. First, the current SOS questionnaire used at CSUSM is not required to be used. That is, the CBA requires that student evaluations be included, but does NOT specify the form of evaluations that need to be included. In addition, article 15.17 states "The format of student course evaluations shall be quantitative (e.g., "Scantron" form, etc.) or a combination of quantitative and qualitative (e.g., space provided on the quantitative form for student comments)." Taken together, the task force emphasizes that campuses have significant leeway to create and use a variety of tools to fulfill the requirements in 15.15-17.

The task force unanimously concluded that simply revising or attempting to improve the current questionnaire does not and will not even partially address the serious concerns already articulated in this report about student evaluation practices at CSUSM. In fact, the task force believes reliance on the current SOS questionnaire can and does, in some instances, harm faculty in the evaluation process (see section IV). As the recommendations below reveal, the task force took a 'harm reduction' approach in the proposed revisions to the questionnaire. A 'harm reduction' approach recognizes that until alternative tools and approaches are available to faculty, the current questionnaire is what will be used to fulfill 15.15-17. Ergo, the revisions aim to minimize the harm and opportunities for harm while the questionnaire is still in use.

In the past, the CSUSM Academic Senate had formed a subcommittee to investigate faculty perception of the SOS (2007) and of the RTP process as a whole (2020). The Faculty Affairs Committee had made policy recommendations to change the use of SOS data in the personnel evaluation process (2020 and then in 2022).<sup>2</sup> However, the SOS instrument remained largely unchanged since its original adoption in 2004.

#### Identifying problems

The task force began its work with an evaluation of the current SOS instrument and identified three main problems, listed below in no order of importance.

- First, the response rate of the online format was alarmingly low and trended downwards since the beginning of the COVID-19 pandemic.

---

<sup>1</sup> For the Collective Bargaining Agreement policy on student course evaluations, see Appendix A1.

<sup>2</sup> For a detailed review of ongoing and historical effort on SOS, see Appendix A2.

- Second, instructors found the data collected in the SOS were not particularly useful for improving instruction because most of the responses were summative and not formative, there was a lack of meaningful variance in the responses, etc.
- Third, task force members concluded that the use of SOS in the personnel evaluation process is problematic due to the impact of the following: student biases (gender, racial, cultural, political, religious, and even disciplinary); a lack of alternative measures for teaching effectiveness (e.g. peer observation); and the lack of a clearly-defined, objective process for analyzing and interpreting the results of the SOSs during the RTP or lecturer evaluation process.

#### Low response rate of the online format

According to the data obtained by the Office on Institutional Planning and Analysis (IP&A) via the survey vendor, Class Climate, the average response rate of SOS in spring 2022 was 39.7%. CEHHS had the highest response rate of 50% due to a larger proportion of graduate level courses.<sup>3</sup> The other colleges had response rates ranging between 35.6% to 37.5%. (Source available upon request.) In contrast, in fall 2019, the last semester before SOS were moved online, the average response rate was 67.1% for paper evaluations and 54.6% for online evaluations.

The low response rate was alarming for three reasons. First, a small sample is unlikely to be representative of the underlying population. With average response rates below 40% and some classes having single digit response rates, the SOS responses were unlikely to be valid representations of the opinions of the students taking the class. Second, the low response rate was not uniform among the students, which can further skew the data. These skewed data can have serious consequences because of SOS's role in personnel decisions. Third, since SOSs were moved to an online format at the beginning of the pandemic in Spring 2020, the response rate continued to trend lower, despite recent efforts to boost the response rate.<sup>4</sup> If this trend continues, it will further undermine the fairness and usefulness of SOS data. To be clear, SOS response do reflect the perceptions of individual students taking the class; individual students may provide valuable feedback. However, it is extremely problematic to interpret aggregate results as representative of the collective student experiences when response rates are so low.

#### Limited usefulness for instructors

The task force conducted a review of the current SOS instruments and recognized several issues associated with question choice, question wording, response scales, and question order that limit how useful SOS data are for instructors to improve teaching. The task force's study of the higher education literature suggested that these concerns were shared across campuses and confirmed by empirical research. (See sections IV a. and b.)

Separate from the problems listed above, the task force could not determine if the questionnaire was ever validated. The task force centered their efforts on proposing revisions to the existing questionnaire.

---

<sup>3</sup> Graduate level courses had a higher response rate (55.8% for Spring 2022) than upper division undergraduate course (41.6% for SP 2022), which had a higher response rate than lower division undergraduate courses (31.6% FOR spring 2022).

<sup>4</sup> In Spring 2022, students were given the incentive of a random drawing of one of ten \$100 Amazon Gift Card to complete the SOS. This incentive led to some increase, but still

While validation of the questionnaire would appear to be positive improvement, the task force did not want to recommend efforts that may appear to suggest that the single questionnaire is preferable to the long-term recommendations. The task of fully validating the existing questionnaire is complex, unfeasible given current resources, and not a long-term, substantive way to address evaluation of teaching effectiveness.

**b. Introduction to recommended changes**

**Revising the current SOS questionnaire is a first-step, interim solution.** The recommended revisions are by no means exhaustive, and the campus should and must continue to engage in robust discussion and work to overhaul and reimagine student evaluations of instruction to be in line with ongoing scholarship about best practices and address the poorly defined impact of SOS results on performance evaluations for faculty at all ranks. Before detailing recommended changes, the task force reiterates their recommendations (please see sections V, VI, and VII) that the campus (and each College and/or Department/Program as appropriate) should establish formal policies/guidance about the tools used for and analysis of data from SOS in faculty evaluation.

**Recommendations are made using a harm reduction approach.** To make short term recommended revisions to the current SOS that will reduce harm to faculty and improve feedback for teaching, the task force recommends the following:

1. Some questions should be removed from the current SOS if they do not offer useful information for improving teaching, have potential for bias, and/or ask something that is redundant in the context of the full SOS tool.
2. Some questions should be re-ordered in sequential appearance.
3. Some questions must be rephrased to reduce bias and elicit responses that will better assist faculty with information about improving teaching.
4. The adoption of an anti-bias statement to the SOS tool. This is a practice that other campuses follow, and some studies suggest can help reduce biased comments that focus on an instructor's identity or embodiment.

**Questions recommended for deletion could be reinserted** by colleges, departments, and/or individual faculty. However, the task force recommends the careful review of the rationales in the charts below to ensure any individual decisions to bring back questions include a consideration of potential for harmful bias. **Additionally, it will be important to assess resource needs to develop, maintain, and administer multiple forms with custom items by department/college.**

**c. Questions that should be deleted**

Question:	Rationale:
I took this course because it is a requirement for my major/degree program. <i>Yes/No</i>	Students may not always know how an individual class contributes to their overall graduation goal (unless they consult their Degree Planner which assumes an extra step for students if we want accurate self-response). Faculty and departments already have access to information about what role a class plays in the larger curriculum and faculty should contextualize special class distinctions in their WPAF narratives (i.e. intro

	<p>classes, capstone classes, GE, etc). This question could aid an instructor if the evaluation data can be parsed based on students’ answers to this question. IP&amp;A notes that in the future it may be possible for faculty to have interactive/filtering capability in SOS reports, but the availability and specifics are not yet known. Unless this filtering capability is possible the task force recommends deletion.</p>
<p>I took this course because it fulfills a GE requirement. <i>Yes/No</i></p>	<p>Students may not always know how an individual class contributes to their overall graduation goal (unless they consult their Degree Planner which assumes an extra step for students if we want accurate self-response). Faculty and departments already have access to information about what role a class plays in the larger curriculum and faculty should contextualize special class distinctions in their WPAF narratives (i.e. intro classes, capstone classes, GE, etc). This question aids an instructor if the evaluation data can be parsed based on students’ answers to this question. IP&amp;A notes that in the future it may be possible for faculty to have interactive/filtering capability in SOS reports, but the availability and specifics are not yet known. Unless this is possible the task force recommends deletion. If interactive filtering is possible, this question should be combined with the above question regarding major requirements to streamline.</p>
<p>In this class, how actively have you participated in all aspects of the learning process (e.g., completing required readings and assignments, participating in class activities)? <i>Very/Moderately/Somewhat/Hardly at all</i></p>	<p><b>Redundant</b> with the question about hours spent each week on class assignments. The phrase “actively...participated” may be confusing for students based on the structure of an individual class and be widely interpreted making the information less helpful to the instructor.</p>
<p>I learned a great deal in this course. (1-5)</p>	<p>Assessing students’ perception of self-learning is very subjective for each individual responding. Student self-reflection in SOS is a valuable practice that would require more substantial review and consideration to make it an intentional part of the SOS Tool. Because the SOS Tool as currently used measures faculty performance, the faculty can be unfairly assessed using a question such as this one where barriers and successes to student learning are beyond the purview and control of the faculty member (e.g. students define “learned a great deal” in terms of</p>

	traditional rote learning as opposed to active learning, or students have previously taken courses in the subject, etc.).
The instructor is an effective teacher. (1-5)	Biased question that has potential to reinforce traditional definitions of professorship and assess student perception of instructor's personality. This evaluation metric would be best assessed using standardized and validated assessment data together with peer evaluations of teaching.
The instructor is enthusiastic about communicating the subject matter. (1-5)	Biased question that suggests a particular way faculty should deliver instruction that is based on personality and asks students to speculate on an instructor's internal state.
The instructor showed genuine interest in students' learning. (1-5)	Biased question that asks students to assess if they feel instructor "liked" them and gauges rapport. Significant risk of disciplinary, gendered, and racial biases. Instructors' labor to prepare and optimize a course and its delivery for a target student audience is in interest in student learning but are tasks that remain largely invisible to students in a given class. Students never get a complete sense for all this work and so should not be asked to gauge an instructor's "genuine interest in" and dedication to their learning.
Required assignments (e.g., exams, papers, projects, etc.) contributed positively to my learning experiences in this course. (1-5)	Students associate grades earned with positive vs negative feelings and also their familiarity/comfort with assignment structure. Value of what is learned in a class is often something students only fully realize over time, not in the moment when thinking about their grades for the semester. This question is also somewhat redundant with other questions about assignments below.
The instructor was sensitive to student difficulties with the course material. (1-5)	Redundant with other questions that ask for similar information. Biased question because disciplinary, gendered, and racialized expectations about how faculty should accommodate student discomfort that could harm faculty teaching classes that are perceived as more difficult and/or covering sensitive topics.
Insofar as possible, the instructor was receptive to student questions. (1-5)	Redundant and vague question.
The instructor seemed well-prepared for each class. (1-5)	This is a question about organization that could lead to biased responses per student expectations about how a class should be designed--i.e. active learning exercises vs lecture.



	Preparation language privileges traditional lecture classroom. Research-based courses (e.g. CUREs) are inherently open-ended and so have different needs for preparation, yet they are a high-impact teaching practice.
The lab/discussion sessions clarified the lecture material. (Lab Form C)	Based on too many traditional assumptions about division between lab and lecture classes. Vague and redundant with question about “applying concepts.”
If relevant, describe one or two specific aspects of this course that lessened your interest in the materials presented or interfered with your learning.	Historically an open-ended question that elicits harmful and biased comments that disparage instructor on basis of identity or appearance or personality. The negative phrasing invites biased responses and is ultimately redundant with the relatively more neutrally phrased “What suggestions, if any, do you have for improving this course?”  Please see section below for additional discussion.

**d. Questions that should be reordered in sequential appearance**

<b>Question:</b>	<b>Rationale:</b>
Based on your performance in this course thus far, what grade do you expect to receive? <i>an A/a B/a B or a C/a C or a D/Credit/an F or no credit</i>	Place this question at the end of the SOS Tool. Students filling out questions with their assumed grade in mind at the start may bias their responses. This question could be more helpful to faculty if comparisons between what the student expects to receive and what grade the student earns could be comparable factors with interactive sorting capabilities. Other campuses maintain SOS results in a confidential manner while the CBA for the CSU has very vague language requiring “anonymity”. If the SOS analysis could be done in a manner which ensured student anonymity to the instructor who is grading them, then actual grades as well as expected grades could be used to parse all other SOS question results. For the moment we recommend simply moving this question to the last position.

**e. Questions that must be rephrased to reduce bias**

<b>Current Question Phrasing:</b>	<b>Suggested Rephrasing:</b>	<b>Rationale:</b>
-----------------------------------	------------------------------	-------------------

<p>When you first enrolled in this course, how interested were you in its subject matter? <i>Very/Moderately/Somewhat/Hardly at all</i></p> <p>Now that the course is nearly over, how interested are you in the subject matter? <i>Very/Moderately/Somewhat/Hardly at all</i></p>	<p>How interested are you in the class subject matter? <i>Very/Moderately/Somewhat/Hardly at all</i></p>	<p>Previous phrasing assumed a two-question sequence of changing interest before starting classes and at the end of the semester. We propose removing the first question and retaining a modified version of the second question to: 1. reduce questions to make the process for students more appealing as the current length is one barrier to student participation, and 2. to avoid creating biased responses that will harm faculty who teach required classes that may be a “harder sell” in terms of student interest due to subject matter and/or perceived difficulty. Student feedback indicates that these questions can be unclear to students, and they are unsure how to respond, with some anecdotal evidence from students that they can be used to purposely give an instructor a lower/negative if they simply did not like the class. Making this a more neutrally phrased question gives the instructor insight to interpret the results of other questions on the SOS, but in manner that does not place the responsibility on the</p>
--	--	---

		<p>instructor to persuade or convince students to feel intrinsically interested in the course subject, which is a subjective measure.</p> <p><i>Please see section f below for additional discussion.</i></p>
<p>The overall quality of this course was high. (1-5)</p>	<p>The course was a valuable learning experience for me. (1-5)</p>	<p>The previous question ostensibly evaluates “quality of course” but in practice mostly reflects “student satisfaction,” while also being too vague to help provide feedback to faculty for course improvement. The rephrase makes the purpose of the question clear and follows phrasing in use by CSU Channel Islands.</p> <p>The revised language while an improvement is still rather vague. SOS Task force strongly recommends future work in the overhaul and complete redesign of SOS Tool and Processes provides a more robust revision.</p>
<p>The course objectives and requirements were clearly outlined in the course syllabus. (1-5)</p>	<p>I knew where to find the learning outcomes for this course (1-5)</p> <p>I knew where to find the instructions for the individual class assignments, papers, and/or activities (1-5)</p>	<p>Campus syllabus policy requires the inclusion of course learning outcomes. Requirements (such as completion of specific assignments) can measure learning outcomes but are not interchangeable with</p>

		learning outcomes themselves. Breaking into two questions attends to this distinction.
The instructor responded when I asked for individual help.	When I sought help from the instructor, I received it. (1-5)	Current phrasing privileges response time considerations that may be unrealistic to place on faculty (i.e. culture of instant response no matter day or time) and/or expectation of particular kind of support that can be very discipline-specific, gendered, and racialized so high chance of harmful bias. Move question up sequentially so that it is part of a Likert scale to get more useful responses. Revision follows phrasing in use by CSU Channel Islands.
The instructor's presentations added to my understanding of the course material. (Form A Lecture Class)	Class sessions added to my understanding of the course material (1-5)	Bias possible regarding language, speakers' accent, style for giving presentations and could encourage traditional expectation of a lecture format. Sets expectation that classes should be mostly traditional lecture/presentation and does not give enough space for active learning and other approaches commonly used in contemporary pedagogy practices
The instructor asked students to demonstrate their understanding of the course material by applying concepts. (Form C Lab)	The course provided opportunities to apply concepts to demonstrate learning, understanding, and achievement of the course learning outcomes. (1-5)	Strange phrasing. Revision clarifies that applying concepts has a learning related purpose.

List one or two specific aspects of this course that were particularly effective in stimulating your interest in the materials presented or in fostering your learning. (open ended)	What course components helped you learn or increased your interest in this course? Please focus your comments on the course materials and not on irrelevant factors such as the instructor’s appearance, gender, accent, likability, etc. (open-ended)	Makes phrasing less difficult to parse. “Stimulating your interest” too formal and ignores the issue of learning at the heart of pedagogy.  <i>Please see section f below for additional discussion.</i>

**f. Questions for additional discussion**

The task force was unable to come to consensus on a recommendation for two statements in the existing questionnaire. Suggested rephrasing for the following questionnaire items are found in Table e above:

- How interested are you in the class subject matter?
- List one or two specific aspects of this course that were particularly effective in stimulating your interest in the materials presented in fostering your learning.
- What suggestions, if any, do you have for improving this course?

Several task force members shared that these items can provide valuable, actionable insight to faculty in order to improve their teaching effectiveness. Simultaneously, open-ended questions such as these can be spaces for students to share biased and derogatory comments about faculty.

The task force also debated the value of items that inquire about course quality and the instructor’s organization/preparation. Table e also proposes possible rephrasing to reduce opportunities for biased student responses. This aspect of teaching can also illuminate specific areas of improvement for instructors. The task force recommends additional attention be given to preserving the utility of these items.

**g. Questions specific to the online form**

The SOST Task force recommends all the questions specific to the online form (listed below) be evaluated and discussed by the Senate bodies engaged in rethinking online course issues at CSUSM — these questions are very specific in comparison to other evaluation questions and make very specific assumptions about the type of activities and formats used in an online course environment. That is, they make the assumption that a discussion forum must be employed and that student-to-student interactions are necessary. The existing SOS for all other course formats do not make this level of assumption in course design. While it is unfortunate that suggesting short-term suggestions for this specific sub-form is not possible given its particularities, the SOS Task force hopes that the

recommended suggestions for reducing harm in the questions used in all evaluations will provide some short-term relief. The potential for differing levels of harm-reduction in the short term is one reason why the Task force strongly encourages the campus to pursue a robust overhaul of the entire SOS tool(s) and process(es).

The questions specific to the online form:

- **The activities and assignments related to the course objectives. (1-5)**
- **The course provided ample opportunity for on-line interaction with other students. (1-5)**
- **On-line discussions enhanced my understanding of the course content. (1-5)**
- **The on-line course materials were easy for me to access. (1-5)**

#### **h. The anti-bias statement**

While the subgroup has limited its work on generating new content and language, when possible, given that the campus must engage in on-going work to overhaul and fully update the entire SOS process and tool, the addition of an anti-bias statement is a small change that we can take now that might reduce harm. The following statement is recommended:

“Student evaluations of teaching play an important role in the review of faculty. California State University San Marcos recognizes that student evaluations of teaching are often influenced by students’ unconscious and unintentional biases about the race, gender, and other identities of the instructor. Women, instructors of color, and members of other minority identity groups are systematically rated lower in their teaching evaluations, even when there are no actual differences in the instruction or in what students have learned. As you fill out the course evaluation, please keep this in mind and make an effort to resist stereotypes about professors. Focus on your opinions about the content of the course (the lectures, the assignments, the textbook) and not unrelated matters (the instructor’s appearance).”

#### **i. What the revised SOS tool would look like with the above changes implemented**

##### **Anti-bias statement in All Evaluation Forms:**

Student evaluations of teaching play an important role in the review of faculty. California State University San Marcos recognizes that student evaluations of teaching are often influenced by students’ unconscious and unintentional biases about the race, gender, and other identities of the instructor. Women, instructors of color, and members of other minority identity groups are systematically rated lower in their teaching evaluations, even when there are no actual differences in the instruction or in what students have learned. As you fill out the course evaluation, please keep this in mind and make an effort to resist stereotypes about professors. Focus on your opinions about the content of the course (e.g., the lectures, the assignments, the textbook) and not unrelated matters (e.g., the instructor’s appearance).

##### **General Questions in All Evaluation Forms:**

On average, approximately how many hours per week have you spent preparing for this class? *At least 10 hours/8-9 hours/6-7 hours/4-5 hours/2-3 hours/1 hour at most*

How interested are you in the class subject matter? *Very/Moderately/Somewhat/Hardly at all*

*Please see section f above for additional discussion.*

The course was a valuable learning experience for me. (1-5)

When I sought help from the instructor, I received it. (1-5)

I knew where to find to the learning outcomes for this course (1-5)

I knew where to find the instructions for the individual class assignments, papers, and/or activities (1-5)

**Form A Lecture Questions:**

Class sessions added to my understanding of the course material. (1-5)

**Form C Lab Questions:**

I had ample opportunity to ask questions during the lab/discussion sessions. (1-5)

The course provided opportunities to apply concepts to demonstrate learning, understanding, and achievement of the course learning outcomes. (1-5)  
(1-5)

**General Questions in All Evaluation Forms:**

What course components helped you learn or increased your interest in this course? Please focus your comments on the course materials and not on irrelevant factors such as the instructor's appearance, gender, accent, likability, etc. (open-ended) *Please see section f above for additional discussion.*

What suggestions, if any, do you have for improving this course? (open-ended) *Please see section f above for additional discussion.*

Based on your performance in this course thus far, what grade do you expect to receive? *an A/a B/a B or a C/a C or a D/Credit/an F or no credit*

**j. Recommendations for administration of existing SOS tool**

Potential solutions to the low response rate problem

To mitigate this problem, it is important that the SOS process is made as accessible as possible and to provide the students with external as well as internal motivations. In Spring 2022, IP&A reached out to instructors who had received high SOS response rates in the past, and the common responses include:

- Send students multiple reminders to complete SOS.
- Obtain real time response rate from IP&A and share it with students.
- Give students class time (about 15 minutes, and some faculty leave the class during that time) to complete SOS.
- Communicate the importance of the SOS and how responses will be used to the students, including
  - Instructors really value student feedback and use constructive feedback to improve the course.
  - SOS is the way for students to have a voice about curriculum and faculty.

On the issue of offering extra credit/course points for completing SOS

Offering extra credit or course points for completing the SOS appears to be one way of increasing response rate. There are several serious concerns in such a strategy. There is concern that completing student evaluations will be perceived as another chore that students are compelled to do to get a better grade. A holistic reframing of the entire process could mitigate the need to implement coerced compliance. Additionally, the practice opens up the potential for students to think they should only respond positively to get points; as noted earlier in this report, faculty are seeking useful and accurate input from the students. Not all faculty grade the same and such a practice may not fit with their grading philosophy. This approach to addressing the concern of low response rate may be replaced by another concern of dealing with students struggling to “redeem” their points, creating another “headache.” Lastly, this could open more grade complaint processes around an issue that is not related to the student's performance in learning the material of a given course.

In addition to the above examples, task force members also considered the following suggestions:

- Embed the link to SOS within the Canvas interface of individual courses to “make it appear to be required.” Discussion with Canvas expert on campus seems to confirm this possibility. It may even be possible for the link to operate within the Canvas interface, instead of opening up a new window/tab. It will be essential to assure students that their responses are indeed anonymous.
- To make SOS official by providing example syllabus language on SOS to be incorporated in course syllabus.
- Provide students with training on how to give effective feedback in surveys and make their voice heard, so that students feel more empowered with taking surveys. (In the current new student orientations, there is a brief component on survey culture and the importance of completing surveys, though the discussions did not focus on SOS specifically. There is also support for giving students training on giving feedback effectively.)

Most of the above suggestions focus on facilitating the SOS process and providing students with extrinsic incentives such as course credit in order to increase response rates. Task force members – and in particular our student representative - also thought it important to provide students with intrinsic incentives and opportunities for reflection on instruction and learning. Importantly, all felt it was important to ensure that students felt that their feedback and input was being valued and having an impact on pedagogy as the semester is still ongoing. It is difficult to provide such intrinsic motivation with an end-of-semester SOS. Therefore, task force members believed that an optional midterm survey, or more broadly, a midterm learning dialogue, should be created, so that when instructors choose to use the midterm survey/dialogue, it would give to students an opportunity to give timely feedback on



instruction and learning. The task force took the initiative of drafting a midterm instruction and learning dialogue for the campus community to consider. Please see section VIII for more details on midterm learning dialogues.

The task force also discussed reverting back to administering the SOS questionnaire via paper to address low response rate. This appears unfeasible given the challenge of administration and data analysis, in addition to the fundamental concern of continuing to expend limited human resources on a tool that, as the task force has made clear, may cause more harm than benefit for faculty and students.

## X. Appendices

### Appendix 1: Ongoing and Historical Effort Related to SOS at CSUSM and CSU System-wide CSUSM – SOST

#### 1. FAC recommendations for changes in University RTP policy pertaining to SOS in 2022

In spring 2022, FAC conducted [a literature review on the bias in SOS](#) and consequently recommended the following changes to the University RTP document (tenure-track):

1. Candidates may rebut SOS as part of their teaching narrative or as a separate addendum.
2. Departments are encouraged to require or suggest non-SOS, discipline-appropriate materials as part of the evidence of teaching success.
3. Department chairs, PRC members, Dean/Director, PTC, and the President or Designee are required to undergo “anti-bias training, to include materials on bias on Student Opinion Surveys on Teaching, within the last 36 months.”
4. “In the evaluation of teaching performance, Student Opinion Surveys on Teaching shall not constitute the sole or primary evidence of teaching quality.”

A full list of these changes [can be found in this document](#). The new University RTP document incorporating these changes [can be found here](#). At the conclusion of spring 2022, the proposed changes went through the second reading at the Academic Senate.

#### 2. FAC recommendations for explicit SOS policy in the RTP process in 2020

In Fall 2019, FAC conducted a faculty survey (tenure-track) inquiring about faculty experience with the RTP process. The raw survey results [can be found here](#). A summary of the survey results [can be found here](#). Based on the problems identified in the survey, FAC made several recommendations including the following pertaining to the use of SOS in the RTP process:

1. Revise Department and/or University RTP standards “with explicit statements as to how to use student evaluations.”
2. “Revise University RTP Standards to include language regarding how student evaluations are considered in context of WPAF contents.”

The entirety of FAC recommendation [can be found here](#). Recommendations on SOS are listed on page 6.

#### 3. (Failed) attempt to move all SOS online in 2014

In spring 2014, FAC was tasked to consider whether to move all SOS online. FAC recommended the move. The recommendation [can be found here](#). It was voted down at the Senate (?). (Note: official record of the Senate decision is not included in sharepoint.)

#### 4. CSU committee's recommendations regarding SOS in 2007

In spring and fall 2007 CFA and CSU tasked a joint committee to study “the best and most effective practices for the student evaluation of faculty teaching effectiveness.” Based on a review of the literature and surveys done with the Faculty Development Council of the CSU, the committee made numerous recommendations including:

1. “Acknowledging that most such instruments primarily measure student satisfaction” as opposed to student learning.
2. “Recognizing that evaluation results cannot be used in a linear manner to rank faculty or to place them in categories (‘excellent’, ‘below average’).”
3. “Student evaluations should never be the sole basis for evaluation of teaching effectiveness.”
4. “Student evaluations must be recognized as only one component of an evaluation of teaching effectiveness. Evaluation policies for all faculty (lecturers as well as tenure-track) should require that reviewers use multiple measures of teaching effectiveness. High Student ratings in isolation do not necessarily mean that an individual is an effective teacher, nor do lower ratings necessarily mean that an individual is an ineffective teacher.”
5. “Campuses should monitor the student evaluation process and be particularly sensitive to the potential for bias in evaluations.”
6. “Campuses should use a well-designed student evaluation instrument (with demonstrable validity and reliability) in providing diagnostic information and feedback to faculty, and those involved in evaluations should have an understanding of their formative as well as summative uses.”

The full report [can be found here](#). The recommendations are listed on page 9.

#### 5. Faculty perception of SOS in 2007

In fall 2007, FAC formed a subcommittee to review SOS based on a faculty survey (tenure-track and lecturer). Notable findings include:

1. The subcommittee believed that SOS measure teaching effectiveness.
2. Surveyed faculty were not confident “in the validity or summative value of the data.” “only 35% believe the instrument provides valid information about the quality of teaching for retention, tenure, and promotion.”
3. “Of the faculty surveyed, a majority (82%) also state the instructor’s entertainment value significantly impacts student evaluation ratings.”
4. “Interestingly, grading standards are also perceived by a large number of the faculty surveyed (85%) to impinge upon student ratings. In fact, more than half of the survey participants (59-61%) contend that if they were to raise their grading standards or increase course content their student evaluation ratings would suffer.”
5. “Consistent with what is reported in the literature, faculty perceive personal characteristics such as the gender of the instructor, religion, age, sexual orientation, and race to have little effect on student ratings. On the other hand, faculty contend grading standards, course rigor, and the amount of content covered, impinge upon student ratings.”

The full report [can be found here](#).

## Appendix 2: FAC Literature Review on the Biases of Student Opinion Survey (SOS) Spring 2022

### Introduction

In response to the referral: “*Work with SAC to review and revise University RTP Document to include language regarding how student evaluations are considered in context of WPAF contents,*” FAC investigated the issue of bias in Student Opinion Surveys (SOS) in depth. FAC concluded it was necessary to review and consider the existing literature on this issue as an initial step in their work and inform further discussions regarding this referral. The following is provided solely as context and background. It is also necessary to note that bias in SOS is but one of several issues identified in their administration and use at CSUSM. SOS are required by Article 15.15 of the CBA. Therefore, it is necessary to note that the conclusions noted in the following serve as foundation to improve their use in faculty evaluation.

### Terminology

Student Opinion Surveys (SOS) have many names. The CBA (2022-24) mentions the following three: student course evaluations of teaching faculty instructional effectiveness, student opinion survey, and student perception of teaching effectiveness. This document uses student opinion surveys (SOS) as it is the term that FAC considers to be the most accurate.

### Highlights

- SOS results are communicated in a way that summary statistics and summarized comments are emphasized, which encourages evaluators to focus on these few items, a seemingly efficient way of thinking: an average score of 4.7 out of 5 surely means better teaching than an average score of 3.7, and a set of SOS full of praises is surely a reflection of the instructor’s superior teaching skills. This document cautions against this kind of “Yelp-ification” approach of using SOS in faculty performance evaluation.
- SOS are biased at multiple fronts. First, SOS are myopic due to their timing and students’ limited experiences with higher education. Second, SOS discriminate against instructors who are from underrepresented groups<sup>1</sup>, paralleling the larger, societal discrimination against members of these groups.
- Skeptics may point to the fact that many faculty members at CSUSM, both tenure-track and lecturer faculty, despite being from underrepresented<sup>5</sup> groups, have achieved similarly high scores and positive comments in SOS compared with their white, cisgender male counterparts. This is evidence of the remarkable overachievements of our outstanding faculty who have compensated for some of the biases through hard work, not evidence of the absence of biases.

### Scope

This document focuses on the biases of SOS as documented in the higher education literature and how these biases render SOS an ineffective tool for faculty performance evaluation. It is not a comprehensive

---

<sup>5</sup> Within the term “underrepresented groups,” we include instructors who identify as persons of color, women, and/or LGBTQ+.

review of the overall usefulness of SOS, as SOS are used at multiple fronts including but not limited to faculty performance evaluation. For example, instructors themselves may choose to reflect on SOS at the end of a course of their own volition, and they can do so in the context of the overall classroom dynamics, student performances in assignments and tests, and student interactions with the instructor and other students, etc., a context with which the instructors are intimately familiar. In the best-case scenario, an instructor can critically assess responses in SOS, disregard the biased and nonsensical comments, and reflect on and learn from the useful ones. This nuanced approach towards SOS requires the instructors to be equipped with the awareness and the training to critically assess SOS, which is not always the case, especially for junior faculty.

This document focuses on the use of SOS by outside evaluators in the formal process of faculty performance evaluation. An evaluator, even a well-intentioned and hardworking one, when using SOS to assess candidates' teaching performance, is not positioned to take a highly detailed and nuanced approach due to the sheer volume of work required from both the candidates and the evaluators. SOS results are communicated in a way that emphasizes summary statistics and summarized comments, which encourages evaluators to focus on these few items, a seemingly efficient way of thinking: an average score of 4.7 out of 5 surely means better teaching than an average score of 3.7, and a SOS full of student praises is surely a reflection of the instructor's superior teaching skills.

This document cautions against this kind of "Yelp-ification" approach to using SOS in faculty performance evaluation based on an extensive body of literature on the biases of SOS. It is not to advocate the complete disregard of SOS. Well-designed and well-executed SOS can be informative and helpful to the instructors. There is, however, a distinction to be made between the instructors themselves critically reflecting on and learning from SOS with a thorough understanding of the instructors' own personal background and the overall course versus the evaluators using summary statistics and comments from SOS in faculty performance evaluation. When SOS are used in faculty performance evaluation without sufficient recognition of the biases, it pressures instructors, especially instructors from underrepresented groups, to singularly aim for certain scores and/or certain positive comments in SOS, which is unfair and counterproductive, as demonstrated in the higher education literature.

## Literature Review

CSUSM has been using SOS since its inception. SOS provide a venue for students to anonymously communicate their class experiences and are mandatorily included in the faculty evaluation process as a measure of teaching effectiveness. SOS are valuable in that they give students the opportunity to provide feedback without worrying about any negative repercussions; however, they are ill-suited for the purpose of evaluating faculty's teaching effectiveness due to a myriad of problems.

**First, SOS are myopic due to their timing and students' limited experiences with higher education. By using SOS to evaluate teaching, it incentivizes instructors to also focus on the short term, winning the "popularity contest," to the detriment of achieving long-term education goals.**

SOS are completed before courses are fully concluded, before students experience the impact of a course on their subsequent college education, not to mention the impact of a course on their work and life years after graduation. SOS tend to focus on the students' immediate reactions to their course experience, positive reactions when they experience successes (better grades) and negative reactions when they experience struggles. Multiple studies have found that SOS ratings are responsible for grade

inflation (Berezvai, Lukáts, and Molontay 2021; Langbein 2008; Eiszler 2002), but not responsible for improving student learning (Stroebe 2020; Uttl, White, and Gonzalez 2017).

Research also indicates that using SOS in evaluating teaching effectiveness incentivizes instructors to focus on increasing students' positive perception of the instructors personally, often through raising contemporaneous student achievement scores, and sometimes by resorting to practices that may harm long-term student learning outcomes. Braga, Paccagnella, and Pellizzari (2014) and Carrell and West (2010) demonstrated a negative relationship between SOS ratings and subsequent academic performance. In their experiments, instructors were randomly assigned to students in introductory courses. At the end of the courses, instructors who gave higher course grades to their students also received better SOS scores. Students of those instructors, however, performed *worse* in subsequent follow-on courses in mathematics, humanities, basic sciences, and engineering, despite receiving better grades in the introductory courses, compared to peers with less lenient introductory course instructors who gave lower grades and received worse SOS scores. When SOS are used to evaluate instructors' teaching quality and skills, it increases incentives to inflate grades, teach to the test, and/or to reduce rigor, harming both student learning outcomes and the quality of higher education.

Similarly, the use of SOS in the evaluation of teaching effectiveness reduces incentives for the instructors to innovate and to experiment with new pedagogical practices, to encourage critical thinking, and to teach controversial topics. Courses that are likely to be considered "difficult" by the students and therefore penalized in SOS are often courses in which the instructors challenge the students intellectually, courses that are likely to have a lasting impact on students long after graduation (e.g. Andersen and Harsell 2005; Murawski 2014). These long-term impacts cannot be detected by end-of-semester SOS, as researchers often relied on the alumni associations for follow-up surveys years after students departed campus to assess their college educations' long-term impacts. Considering that even consumer product companies nowadays give both short-term and long-term customer surveys for product satisfaction, it is rather inappropriate for higher education to rely on end-of-the-semester SOS as a major instrument to assess instructor teaching quality and skills (Hornstein 2017; Zabaleta 2007).

Even more alarmingly, in today's increasingly polarized political environment, instructors of political science, history, sociology, and other social science and humanities disciplines are penalized in SOS when their courses evaluated contemporaneous political issues critically and did not coddle students' personal political views. Instructors all over the country, including our CSUSM colleagues, noticed significant fluctuations in SOS scores in courses with politically relevant content, especially regarding the political behavior of prominent, polarizing political figures. In a particular instance, Dr. Kenneth Mayer, who teaches an undergraduate course on presidential politics at the University of Wisconsin-Madison, was publicly criticized by a state legislator because in a course syllabus, leaked to the public by a disgruntled student, Dr. Mayer included an introductory paragraph that expressed negative views of the Trump presidency, after a short paragraph expressing positive views of Mr. Trump.<sup>6</sup>

---

<sup>6</sup> According to the Associated Press:

The syllabus' first page includes a two-line paragraph saying Trump supporters "rejoice in his contempt for what they insist is a corrupt D.C. establishment."

It then includes an 11-line paragraph that begins: "To others, he is a spectacularly unqualified and catastrophically unfit egomaniac who poses an overt threat to the Republic."

While it is inevitable that students would bring their own limits and personal biases into SOS, to then use SOS to determine instructors' teaching abilities and skills means allowing those limits and personal biases to influence faculty evaluations and life-changing decisions on hiring, instructional assignments, retention, tenure, and promotion. Faculty who teach those topics in their courses either have to suffer the negative consequences or choose not to take the risks to avoid being penalized in SOS, which is demoralizing and contradictory to CSUSM's core values.

**Second, SOS discriminate against instructors from underrepresented groups, paralleling the larger, societal discrimination against them. As a result, the use of SOS in the faculty evaluation process places an undue burden on these affected faculty, pressuring them to overcompensate for biases against them and harming their assignments, retention, tenure, and promotion prospects.**

The stereotype of a college instructor is an older white male. Because faculty from underrepresented groups do not conform to this stereotype, students in SOS tend to rate these faculty much more negatively. Both observational studies and controlled experiments in educational research on SOS demonstrated a statistically and substantially significant bias against women and persons of color, especially junior women (Mengel, Sauermann, and Zölitz 2019) and persons of color from immigrant backgrounds (Boring and Ottoboni 2016; Fan et al. 2019). MacNell, Driscoll, and Hunt (2015) demonstrated that women faculty are penalized by as little as having a female-sounding name. In a randomized and controlled experiment, a number of online courses were delivered by 1) actual female instructors with female-sounding names, 2) actual female instructors with male-sounding names, 3) actual male instructors with female-sounding name, and 4) actual male instructors with male-sounding name. The actual gender of the course instructors was unknown to the students, but the name gave a (sometimes intentionally misleading) clue. In the final SOS, while the courses delivered by actual female instructors were rated higher, courses delivered by perceived female instructors were rated lower. Students rated the perceived female instructors significantly lower across the board on items of consistency, enthusiasm, knowledge, professionalism, etc. than the perceived male identity, yet the same students rated the actual female instructors (unknown to the students) higher than the actual male instructors. This type of discrimination exists even on seemingly "objective" items. Boring and Ottoboni (2016) found with randomized and controlled experiments of online courses that even when assignments were graded and returned to the students at the same time, perceived male instructors received higher ratings on "promptness" than perceived female instructors. The effect of such distribution is large enough that a more effective women instructor may be rated lower in SOS compared to a less effective male instructor solely because of gender (Boring and Ottoboni 2016).

Because of the biases against faculty from underrepresented groups, when SOS are used to evaluate faculty teaching quality and skills, these faculty found themselves working twice as hard to get half as far, and often had to work themselves to the point of exhaustion, both physically and emotionally, in the hope of over-compensating for at least for some of the biases against them. Women instructors are expected to exhibit traditional feminine qualities such as being caring, helpful, responsive, etc., and

---

For additional information, see:

<https://apnews.com/article/6b67b0557d1944a4b9c0c27f698de602>

<https://www.jsonline.com/story/news/education/2019/02/01/uw-professors-description-trump-presidency-stokes-debate/2739701002/>

penalized by students in SOS when they were perceived to deviate from these gendered expectations (e.g. boring).

Women faculty commonly reported holding extra office hours, working late into evenings and during weekends to respond to student emails, and spending significant amounts of mental energy to provide students with emotional support on top of academic assistance, when the same is not required of their white, male counterparts. Instructors who are female and/or persons of color are also more likely to be targeted by harsh and hurtful student comments, yet have no choice other than becoming even more patient and available (Heffernan 2021; LeFebvre, Carmack, and Pederson 2020). Absent the ability to overcompensate because of family obligations, medical conditions, or other factors outside one's control, women faculty had no choice but to swallow the unfair penalty in SOS and therefore be deemed less capable and less skillful teachers in their personnel evaluations, suffering from career stagnation or even job loss across academic disciplines (Martin 2016; Russell, Brock, and Rudisill 2019; Shreffler, Shreffler, and Murfree 2019).<sup>7</sup>

The *Washington Post* recently reported an alarming case on gender bias in tenure decisions in the U.S. Naval Academy: in 2021, 11 men and 4 women applied for tenure in USNA, while 10 out of the 11 men received tenure, 0 out of 4 women did. Among the 4 women being denied tenure, one of the applicants, Dr. Carolyn Chun, a woman of Asian descent, was given the feedback that a direct cause for her tenure denial was because of SOS (called student opinion forms at USNA), that the tenure review committee felt that she did not establish "rapport with the class or a cohort within the class", despite the unanimous support from her colleagues at her department.<sup>8</sup> The USNA is not CSUSM; its student body and faculty are less diverse and is possibly less concerned with diversity and equity issues. The problem that underrepresented faculty at USNA face, however, is the same problem that such faculty at CSUSM face, the difference being a matter of degree. For underrepresented faculty, to receive biased SOS is demoralizing enough; to then have these SOS play a major role in their retention, tenure, and promotion decisions is unfair and unjust.

The literature on the bias of SOS, like all empirical research on higher education, has its limitations. It focuses more on the outwardly apparent traits of instructors, such as perceived gender and race, and pays less attention to other identities such as sexual orientation. It also focuses more on the easily measured, such as the numerical points students give on a few selected items and pays less attention to the harder-to-measure items such as student comments. It does not mean that biases do not exist outside the areas that are more thoroughly examined in the literature. If SOS scores are biased against women, persons of color, and members from other underrepresented groups, it is only logical to infer that SOS comments are similarly biased, despite a smaller number of published research focusing on SOS comments.

### **Concluding Remarks**

SOS are used to evaluate faculty teaching quality because of convenience, not because of effectiveness. SOS are convenient because they give the appearance of a standardized measure of teaching quality,

---

<sup>7</sup> A small but growing number of studies suggest that LGBTQ+ faculty also face discrimination in SOS, see (Heffernan 2021).

<sup>8</sup> <https://www.washingtonpost.com/magazine/2022/01/31/naval-academy-asian-bias-tenure-gender/>



applicable to every instructor and every course. The numbers are easily understood: an average score of 4.7 is higher than an average of 3.7, therefore a faculty member whose scores are consistently above 4.7 is surely a much better instructor than a faculty member whose scores hover around 3.7. What if the faculty who consistently receives 4.7 is charismatic, lenient, and teaches non-controversial topics? And the faculty member who struggles around 3.7 challenges the students more, does not shy away from controversial topics, and comes from a different racial and ethnic background compared to most of the students?

## References

- Andersen, Kristi, and Dana Harsell. 2005. "Assessing the Impact of a Quantitative Skills Course for Undergraduates." *Journal of Political Science Education* 1: 17–27.
- Berezvai, Zombor, Gergely Dániel Lukáts, and Roland Molontay. 2021. "Can Professors Buy Better Evaluation with Lenient Grading? The Effect of Grade Inflation on Student Evaluation of Teaching." *Assessment & Evaluation in Higher Education* 46(5): 793–808.
- Boring, Anne. 2017. "Gender Biases in Student Evaluations of Teaching." *Journal of Public Economics* 145: 27–41.
- Boring, Anne, and Kellie Ottoboni. 2016. "Student Evaluations of Teaching (Mostly) Do Not Measure Teaching Effectiveness." *ScienceOpen Research*. <https://www.scienceopen.com/hosted-document?doi=10.14293/S2199-1006.1.SOR-EDU.AETBZC.v1> (August 22, 2021).
- Braga, Michela, Marco Paccagnella, and Michele Pellizzari. 2014. "Evaluating Students' Evaluations of Professors." *Economics of Education Review* 41: 71–88.
- Carrell, Scott E., and James E. West. 2010. "Does Professor Quality Matter? Evidence from Random Assignment of Students to Professors." *Journal of Political Economy* 118(3): 409–32.
- Eiszler, Charles F. 2002. "College Students' Evaluations of Teaching and Grade Inflation." *Research in Higher Education* 43(4): 483–501.
- Fan, Y. et al. 2019. "Gender and Cultural Bias in Student Evaluations: Why Representation Matters." *PLOS ONE* 14(2): e0209749.
- Heffernan, Troy. 2021. "Sexism, Racism, Prejudice, and Bias: A Literature Review and Synthesis of Research Surrounding Student Evaluations of Courses and Teaching." *Assessment & Evaluation in Higher Education*: 1–11.
- Hornstein, Henry A. 2017. "Student Evaluations of Teaching Are an Inadequate Assessment Tool for Evaluating Faculty Performance" ed. Hau Fai Edmond Law. *Cogent Education* 4(1): 1304016.
- Langbein, Laura. 2008. "Management by Results: Student Evaluation of Faculty Teaching and the Mis-Measurement of Performance." *Economics of Education Review* 27(4): 417–28.

- LeFebvre, Leah E., Heather J. Carmack, and Joshua R. Pederson. 2020. “‘It’s Only One Negative Comment’: Women Instructors’ Perceptions of (Un)Helpful Support Messages Following Hurtful Course Evaluations.” *Communication Education* 69(1): 19–47.
- MacNell, Lillian, Adam Driscoll, and Andrea N. Hunt. 2015. “What’s in a Name: Exposing Gender Bias in Student Ratings of Teaching.” *Innovative Higher Education* 40(4): 291–303.
- Martin, Lisa. 2016. “Gender, Teaching Evaluations, and Professional Success in Political Science.” *PS: Political Science & Politics* 49: 313–19.
- Mengel, Friederike, Jan Sauermann, and Ulf Zöllitz. 2019. “Gender Bias in Teaching Evaluations.” *Journal of the European Economic Association* 17(2): 535–66.
- Murawski, Linda M. 2014. “Critical Thinking in the Classroom...and Beyond.” *Journal of Learning in Higher Education* 10(1): 25–30.
- Peterson, David A. M. et al. 2019. “Mitigating Gender Bias in Student Evaluations of Teaching.” *PLOS ONE* 14(5): e0216241.
- Russell, Jared A., Sheri Brock, and Mary E. Rudisill. 2019. “Recognizing the Impact of Bias in Faculty Recruitment, Retention, and Advancement Processes.” *Kinesiology Review* 8(4): 291–95.
- Shreffler, Megan B., Jacob R. Shreffler, and Jessica R. Murfree. 2019. “Barriers to Advancement of Female Faculty in Higher Education: An Examination of Student Evaluations of Teaching, Work-Family Conflict and Perceived Organizational Support.” *Journal of Higher Education Theory and Practice* 19(8).
- Stroebe, Wolfgang. 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): 276–94.
- Uttl, Bob, Carmela A. White, and Daniela Wong Gonzalez. 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.
- Zabaleta, Francisco. 2007. “The Use and Misuse of Student Evaluations of Teaching.” *Teaching in Higher Education* 12(1): 55–76.

### Appendix 3: Efforts to reform Student Opinions of Teaching at other CSU campuses

CSU	SOS links	Useful information	Concerns	Other notes
Bakersfield				
Channel Islands	<a href="#">student survey</a> <a href="#">Faculty report</a>	Student responses are confidential rather than anonymous. All Online. 12	Last updated 2015. No mention of bias	Updated every 5 years according to policy. Giving up

	<p><a href="#">2015 Policy</a> (defines how the SOS is periodically updated by FAC)</p> <p><a href="#">2003 Policy</a> (older)</p>	<p>Likert-scale Q's (6 on course + 6 on instructor) + 2 demographics Q's (expected grade+hours worked/wk) + 2 open-ended (what worked + what didn't work/why). FAC defines SOS instruments and Senate votes to approve them. Colleges can add a few questions. Departments can also add a few questions. FAC approves all surveys.</p>	<p>and mitigating bias. Lecturers and TTF have different requirements. Students don't give input on survey development. No formative assessment mechanism (all summative).</p>	<p>to 0.5% of class points is "encouraged" to promote high response rate. <a href="#">University RTP Policy</a> and <a href="#">Lecturer Evaluation policy</a> both require Peer observations of instruction.</p>
Chico		<p>At Chico, we have the University Student Feedback of Teaching (USFOT) committee, which is a subcommittee of the Senate standing committee Faculty and Student Policies committee. Every form for student feedback is approved through this committee. Departments can (but very rarely do) create their own feedback form but the form must be approved by the USFOT committee.</p>		
Dominguez Hills	<p><a href="#">survey form</a></p> <p><a href="#">Guidelines for online evaluation</a></p>	<p>Standard <a href="#">survey form</a> focused on general (vague) questions about instructor practices (very similar to CSUSM form). <a href="#">Extensive system</a> for online evaluation. "The Office of Faculty Affairs and Development strongly recommends not offering extra credit to students for participating in the PTE process [because] providing incentives may risk the anonymous status</p>		

		of the students providing feedback.”		
East Bay	<a href="#">Policy</a>  <a href="#">Student Evaluations of Learning Experience</a>	<p>Anonymous online survey through Blackboard LMS. Students are emailed when survey is ready. For 15-week courses the survey remains open for 21 days, with three reminder emails to complete survey.</p> <p>There is an option for up to two student evaluations per year administered in <a href="#">paper format</a>. This occurs in the last two weeks of semester.</p>		Last update 2018 Their subcommittee is through FAC
Fresno	<a href="#">instrument for mation and analysis</a>  <a href="#">main web page on our instrument</a>		Strong content domain based on evidence-based teaching practices. However, they used a Likert scale survey and ended up at the same place as IDEA (current instrument) with little variance.	
Fullerton	<a href="#">Fullerton RTP doc with SOS info</a> (See 1.c)			Seems pretty forward-thinking and holistic approach to RTP.
Humboldt		At Humboldt, we have two SETs, one for lectures and one for labs (so a course with lectures and a lab would have students fill out one for the lecture portion and another for the lab portion of their class). They were both created through the Faculty Affairs Committee, vetted broadly by departments and voted on		

		<p>by the Senate. Once approved they are used campus wide. I believe individual departments can add a question to their SETs, but as far as I know, most do not. I can share our SETs if folks are interested. I hope this is helpful.</p>		
Long Beach	<p><a href="#">Survey Questions (SPOT)</a></p> <p><a href="#">1. SOS help for students (including a training video)</a></p> <p><a href="#">2. SOS help for faculty</a></p> <p><a href="#">3. SOS policy</a></p>	<p>SOS is fully online, and each semester’s SOS administration information is also shared online.</p> <p>Faculty are encouraged to increase response rate by offering class time and emphasizing the importance of student feedback.</p> <p>Fairly succinct SOS policy. Departments and instructors seem to have some flexibility 1) deciding which courses to evaluate, 2) adding optional questions on top of common questions, not sure how much these options are exercised.</p>		
LA	<p><a href="#">You can see the questions and open-ended prompt here</a></p>	<p>Form for SOS mostly Likert Scale questions and one open-ended response with a prompt.</p> <p>Per the link in the left column, this campus has a clearly published time table of when evaluations get administered via an online portal that stays open until the end of the semester—students have a two-week window to complete SOS via their GET</p>	Standard SOS similar to existing form at CSUSM.	

		portal (which sounds like CSU LA’s version of the MyCSUSM portal our students use—for those curious GET is an acronym for “Golden Eagle Territory” and is the “personalized web page with services and information tailored to you”)		
Maritime				
Monterey Bay	<a href="#">More info</a>	<p><b>From the Website:</b></p> <ul style="list-style-type: none"> <li>• <b>The default method for evaluations is online, with the option open for paper. This is decided at the department level.</b></li> <li>• <b>If there are two or more instructors of record for a course, students will receive one eval per instructor.</b></li> <li>• <b>The default method for evaluations is online, with the option open for paper. This is decided at the department level. Contact your department chair or ASC for information.</b></li> <li>• <b>Students receive an email with a link to the site where they complete the evaluations.</b></li> </ul>	<p>It seems student evaluations are conducted exclusively through the school’s portal dashboard platform. No third party evaluation company involved. Faculty has direct access to the results.</p>	<p><b>Note: CSUMB uses this tactics to ensure better student evaluation participation.</b></p> <p>Suggestive tips listed below:</p> <ol style="list-style-type: none"> <li>1. Set up a threshold so if 90% of the class completes an evaluation, all students get 5 extra credit points.</li> <li>2. Create a quiz where students are asked to indicate if they have completed the evaluation on the honor system. If they indicate "yes" they get the 5 points.</li> <li>3. Provide class time to conduct evaluations.</li> <li>4. Have students "turn in" their digital receipt for their completed evaluation into an assignment. This</li> </ol>

		<ul style="list-style-type: none"> <li>• <b>Evaluations are available to students until the published close date in the email notification.</b></li> <li>• <b>Evaluation results are made available to faculty via email (link) in a web-based summary report as well as both quantitative and qualitative PDF 2 days after the grade due date published CMS that session. (Summer sessions and special sessions excluded)</b></li> </ul>		would be a screenshot indicating that they have completed the evaluation.
Northridge	<a href="#">SEF Resources Page for Faculty:</a>	From the website: <i>Student Evaluations of Faculty (SEFs) are one of the ways in which teaching effectiveness is measured at CSUN. The Office of Institutional Research administers the SEF, including the survey to students and distribution of summary reports, guided by the policies outlined in Sections 600 and 700. If you have questions about the SEF process or timeline, including questions about CoursEval, the platform used to distribute and collect SEF data, please contact the Office of Institutional Research (x3277). If you have</i>		The SEF resource page is easy to navigate and includes links to tips for instructors on adjusting the evaluation time period, retrieving course evaluation reports, the schedule for SEF and best practices. A separate page is available regarding setting up SEF for online courses with tips on how to add a link to the SEF inside Canvas and how to make it a “required” activity (it must be viewed to be

		<i>questions about SEF policies, including questions about how these data are used in the faculty review process, please contact the Office of Faculty Affairs (x2962).</i>		logged as completed). When set up this way, students have to click on the link to satisfy the requirement before moving on with other course assignments. A FAQ goes over basic questions that instructors might have about the process. There is a separate page on SEF for departmental administrative staff.
Pomona	<a href="#">Student evaluation policy</a>			
Sacramento		<p>Currently, our University Appointment, Retention, Tenure, and Promotion Policy states that departments are responsible for the development of evaluation questionnaires. The questionnaires do not get approved at any level beyond the department. I was hoping to learn from you:</p> <p>Do departments develop their own student evaluation of teaching questionnaires on your campus?</p>		



		Is there any oversight or approval process for the developed questionnaires?		
San Bernardino	<a href="#">SOTE website</a> <a href="#">SOTE policy &amp; forms</a> (2017) <a href="#">RTP</a> <a href="#">procedures</a>	<ul style="list-style-type: none"> <li>• Has required and optional components; required form is very simple</li> <li>• Policy seems to focus on in-person administration</li> <li>• Evaluations are collected by Depts/Colleges and forwarded to Academic Personnel Office for processing</li> <li>• Review procedures allow for evaluations to be excluded from personnel file for some circumstances dependent on faculty role</li> </ul>		
San Diego	<a href="#">More info</a>	<p>SDSU used Interfolio for performance review (see linked pasted left).</p> <p>Each college administrated survey distribution and analysis of results.</p> <p>Reached out to multiple departments for copies of survey questions and finally obtained the ones used by business school</p>	List of 17 questions seems to be too long	<ol style="list-style-type: none"> <li>1. Rating on scale of 5 with specification of what numbers mean, e.g. "5" is for "Excellent" and "1" for "poor".</li> <li>2. Ask for current GPA</li> <li>3. Ask for total # of units taken, including this course</li> <li>4. open-end questions are only for "strength" and "best part of the class", and how course can be improved. I like the removal of the</li> </ol>

				question on the weakness.
SF	<p><a href="#">Tool can be found here.</a></p> <p>Has the history of their work and a good website design.</p> <p>New approach: <a href="#">JEDI-TEA focuses on DEI</a></p>	<p>At SFSU, they have a very short survey, then departments can add their own open-ended questions.</p> <p>The process was developed by a task force that was led by an Associate VP, with the policy pieces being handled by a Senate committee in consultation with several administrators.</p> <p>It seems to be well-integrated with Canvas.</p>	Instrument looks to be similar to our existing forms at CSUSM.	
San José	<a href="#">More info</a>			
San Luis Obispo	<p><a href="#">Senate resolution and procedure for online evaluations (2016)</a></p> <p><a href="#">Evaluation timeline</a></p> <p><a href="#">Faculty FAQ</a></p> <p><a href="#">Student FAQ</a></p>	<p>Fully online, using Class Climate – similar procedurally to CSUSM</p> <p>Colleges and departments dictate the content – wasn’t able to find copies of the forms online. All evaluations required to offer opportunity for student comments, and two common “summative evaluation prompts” at the end of every evaluation.</p> <p>Managed by Academic Personnel office</p>		
Sonoma	<p><a href="#">Student Evaluations of Teaching Effectiveness (SETE)</a></p> <p><a href="#">University Policy on Student</a></p>	<p>SETE are anonymous deployed through the SETE Surveys Path in Canvas LMS. SETE opens for student input within the last three weeks of classes.</p> <p>Survey is run through The Office of Institutional</p>		Last updated 2014

	<p><a href="#">Evaluation of Teaching (2014-1)</a></p> <p><a href="#">Student Eval of Teaching Policy</a></p>	<p>Effectiveness/ a component of Academic Affairs</p> <p>This <a href="#">SETE Summary</a> form is completed by 3-year Lecturers for their WPAF and shows some of the closed questions on the SETE.</p>		
Stanislaus		<p>At Stanislaus, we have a single, <a href="#">university-wide Student Perspectives on Teaching (SPOT)</a> evaluation form that was locally created to replace the IDEA evaluations. Departments and individual faculty can and do create their own evaluations as well, but on an informal basis. They are not vetted by any governance or administrative body and are not a required part of RPT or any evaluation files. Faculty can, of course, opt to include such information in their files if they choose.</p>		

#### **Appendix 4: Evaluation of TEval from Anti-Racism, Anti-Colonialism, Social Justice Committee**

**TO:** Yvonne Meulmanns, Chair, Academic Senate  
**FROM:** **Anti-Racism, Anti-Colonialism, Social Justice Committee**  
**RE:** ARACSJC Response to SOST Workgroup on TEvals  
**DATE:** December 12, 2022

The ARACSJC met on 12/1/22 and we support the assessment that the student evaluations are structurally biased. Research indicates that they are marred with racist, sexist and biased comments.

We have observed that course content that challenges students' worldviews are likely to yield critical, biased evaluations. Courses that challenge deeply held conceptions that colonialism and imperialism are policies implemented only in distant lands are problematic when considering student evaluations. Many that take these courses are presented, for the first time, with nationalistic, and federally endorsed and enacted policies that systematically utilized violence, removal, relocation and dehumanization to secure land. This historic reality directly challenges traditional national rhetoric. Furthermore, when western morality reveals as vital mechanisms of settler colonialism, we see Eurocentric morality and value systems pervading our decision-making and judgment; the end result, are biased evaluations that ultimately damage the opportunity for improvement, tenure, employment continuation, and promotion.

Student evaluations tend to judge women and racial minorities more harshly than white instructors, and in courses that emphasizing diversity and inclusion topics, the performance evaluations tend to be especially negative in tone, particularly when diversity and inclusion topics are the focus of the course being evaluated.

The ARACSJC sees this issue as the center for the SOST workgroup focus, and we question the rationale of consider using TEval, an instrument that fails to completely address biases in the student evaluation process.

The ARACSJC recommends that either the TEvals are adapted to address and mitigate biased evaluations or adopt a new instrument.

## Appendix 5:

### Additional Points not mentioned in the Task Force Report but worthy of further discussion:

**Holistic Evaluation of Student Learning and Instructor Effectiveness using student performance in follow-on coursework and programmatic retention rates:** In a recent comprehensive book called “Cracks in the Ivory Tower” ([Brennan and Magness 2019](#)) there is an entire chapter dedicated to detailing the strengths and limitations of using student evaluations in assessing teaching effectiveness. Within this chapter [Braga et al., 2014](#) and [Carroll and West, 2010](#) (both cited in the Task Force Report) are highlighted as two of the most robustly controlled studies illuminating the relationship (ie. Negative correlation) between SOST-reported teaching effectiveness metrics and actual instructor teaching effectiveness. These two articles have been cited roughly 376 times (average impact factor of ~21). They argue clearly, decisively, and convincingly that students give instructors poor SOST evaluations when their instructors demand more from them and thereby get them to learn the material better – as measured by performance in subsequent post-requisite coursework. These papers argue that instructors can and should be evaluated for their teaching effectiveness by using the performance of their students in subsequent required coursework. While this is not universally applicable to all courses and programs it does offer some opportunities and insights that should be explored further. Such an approach would not be subject to poor student response rates, is completely free of implicit bias from students, and is also free from potential “teaching to the test” objections often raised against instructors who try to assess their own teaching effectiveness within their own course using standardized assessments. If given together with a formative evaluation tool focused on student affective experience and satisfaction with a given course such an objective measure of teaching effectiveness and student learning could help address the various concerns targeted by this Task Force.

While this approach wouldn’t be applicable for courses that don’t have any post-requisites, it offers a possible way to assess both teaching effectiveness in individual courses and the cohesiveness of courses within a given program – the later objective was again emphasized by the Task Force’s student representative as being particularly in need to attention. This approach is not a golden bullet for all courses and contexts and does have some important limitations that should be considered/remedied:

- **To prevent artificial improvement of teaching effectiveness metrics by instructors that set the bar too high and thus prevent too many students from being able to proceed to the post-requisite courses, it will be critical to also track not only the DFW rate for the course but also the retention rate or programmatic persistence rate of students taught by a given instructor. The best instructors should be detectable as both highly encouraging (as measured by maximal retention rates rather than simply minimal DFW rates) and optimally demanding (as measured by student performance in subsequent coursework).**
- **To address the issue of variability in the rigor of the potentially multiple post-requisite instructors that might eventually teach the different students taught in a given pre-requisite course it will be necessary to normalize student performance in each post-requisite course. This is the approach taken by the articles referenced above. While the assumption of a normal grade distribution is not always a good one, it’s not a bad place to start.**
- **The scalability of this approach requires...**
  - a. **A very detailed forensic analysis of student performance data across multiple courses in their particular programmatic sequence of study. This will require significant resources for those who will create and maintain the analytical tools to be used.**

- b. There will need to be multiple (at least 3) instructors teaching each pre-requisite course to be assessed over the course of a few years in order to be able to obtain a reasonable “profile line” for each pre-requisite course being taught. This will also encourage programs to maintain a healthy rotation of different faculty teaching each course in a programmatic sequence.**
- c. The pairwise pre-requisite relationships between courses in program need to be clearly laid out and fed into the analysis. When a programmatic change is made this approach may be able to rapidly and automatically quantify the impact of the change on student performance in subsequent coursework and on their persistence in a given program of study. The data analysis conducted would need to be integrated with PeopleSoft and updated yearly to account for changes in pre-requisite/post-requisite relationships.**
- d. The identities of the instructors who taught the pre-requisite course and the identities of the students that took pairs of pre/post-requisite courses will need to be correlated in ways that are not currently allowed in RADAR. It may be necessary to explicitly decide/negotiate who should have access to such non-anonymized data since the CBA does not explicitly elaborate on this point.**

Sincerely,

Kambiz M. Hamadani  
Associate Professor, Dept. of Chemistry and Biochemistry  
CSTEM representative on the SOST Task Force

Stephania Rey  
CSUSM Student Representative on the SOST Task Force