

# Student Evaluation Task Force Report

<b>Acknowledgments</b>	<b>2</b>
<b>Introduction</b>	<b>2</b>
<b>Literature Review</b>	<b>3</b>
Bias in Evaluations	3
Measuring Teaching Effectiveness	4
Best Practices in Teaching Evaluations	6
Instructions	6
Survey Design and Construction	6
Non-Response Bias	7
<b>Primary Research</b>	<b>7</b>
Findings from IFS	7
MBA Research Team Findings	9
<b>Task Force Recommendations</b>	<b>12</b>
Self-Evaluation and Reflection on Teaching	13
Mid-term Student Survey	14
Peer Observation	14
Summary Evaluations	15
<b>Conclusion</b>	<b>17</b>
<b>References</b>	<b>18</b>
<b>Appendix</b>	<b>21</b>
Additional Resources for Best Practices	21

## Acknowledgments

This report was prepared by Melissa Anderson, Melissa Birkett, Kristin Hocevar, and Donna Lane based on the research and meetings of the entire task force. Special thanks are owed to the two MBA research teams who conducted campus surveys and provided primary data about evaluations at SOU. Those teams included students Mike Monroe, Molly Pascale, Shu Yang, Zachary Noeker, Jake Riggs, and Murtadha Alhafufi. Additional thanks are owed to Donna Lane, who surveyed members of the Interinstitutional Faculty Senate to provide additional primary data for consideration.

## Introduction

On October 29, 2018, the Southern Oregon University Faculty Senate voted to create a task force to develop a proposal for student evaluations of teaching “that are both effective and equitable.” The specific charge of the task force was threefold: to research best practices in teaching evaluations that focus on student learning; to review the use of evaluations for promotion and tenure; and to develop a proposal for student evaluation procedures that are equitable and provide meaningful feedback to faculty. To these ends, a call for volunteers was sent to all faculty and a task force was convened that included Melissa Anderson (chair), Donna Lane, Kristin Hocevar, Melissa Birkett, Enrique Chacón, Robert Strahan, Erin Wilder, Alison Burke, and student member Britney Sharp. The task force met bi-weekly winter term to discuss research, and then a smaller subset of the group continued to meet spring term to work on this report. In addition to a review of the extensive literature on student evaluations of teaching in higher education, the task force also reviewed primary research resulting from a survey distributed to Interinstitutional Faculty Senate members and from survey research conducted by two MBA student research teams. A full report on the findings of the literature review and the primary research follows, but the overwhelming conclusion reached by the task force was that the current use of quantitative student evaluations of teaching in promotion and tenure decisions

is subject to substantial bias and should not be continued. Based on their research, the task force members identified a number of potential alternatives for the current quantitative evaluations that could be used in promotion and tenure decisions more equitably while still recognizing the importance of student feedback. Those alternatives are discussed below. In addition, the task force identified ways to improve electronic student evaluations so that they would provide more useful information to instructors, and those suggestions are also discussed below.

## Literature Review

### Bias in Evaluations

Significant biases in student evaluations of teaching (SETs) have been noted in numerous academic studies. Indeed, bias has become so recognized a feature of SETs that a few high-profile universities, such as the University of Southern California, have eliminated traditional teaching evaluations and have opted for alternative means of gathering information about student learning in hopes of improving teaching without unfairly penalizing instructors (Flaherty, 2018; Supiano, 2018). Other institutions, including the University of Oregon, Colorado State University, Fort Collins; the University of Colorado, Boulder, Ryerson University, and a division of the University of California, Berkeley, are changing the way SETs are used, particularly with regard to promotion and tenure (Doerer, 2019). Whether or not bias is an inherent feature in SETs, or if SETs can be constructed in a way that makes them still a useful tool for promotion and tenure decisions without bias remains to be seen, but SETs as they have been used in recent years have shown significant bias with regard to gender, race, age, sexual orientation, political stance, and other factors.

Study after study has shown gender bias in SETs in both online and in-person courses. MacNeill, Driscoll, and Hunt (2015) found that students in an online course rated instructors they perceived to be female lower than instructors they perceived to be male, regardless of actual gender. Burke, Head-Sturgess, and Siders (2017) found that students assign different characteristics to their professors based on gender, such as “nice” for women and “smart” for

men. Mitchell and Martin (2018) have shown that the differences in the language students use in evaluating male and female professors leads to discriminatory employment practices. Peterson, Biederman, Andersen, Ditonto, and Roe (2019) found that when students were given evaluations with language intended to mitigate bias, they tended to score female professors more highly than when given standard evaluations.

The bias in student evaluations is not limited to gender bias. Racial bias has also been shown to be a factor in student evaluations of teaching (Smith & Hawkins, 2011). For example, a 2003 study of 17,000 students showed clear bias against minority faculty (Hamermesh & Parker, 2005). In 2017, a court case showed that SETs had led to bias against minorities in personnel decisions (Schmidt, 2017). Huston (2006) suggested that SETs could be a factor impeding efforts to diversify university faculty.

Age and physical attractiveness have also proven to be factors that affect the validity of SETs. Joye and Wilson (2015) found that in addition to having different expectations for faculty instructors based on gender, students were also influenced by both age and physical attractiveness when evaluating teaching effectiveness. Hamermesh and Parker (2005) also found that instructor attractiveness had a direct effect on their ratings as effective teachers.

## Measuring Teaching Effectiveness

Research on measurement of teaching effectiveness – i.e., what student course or teaching evaluations are actually measuring – was particularly active in the 1980s and 1990s. Some research from that period suggests that teaching evaluations generally measure instructional skill, which in turn measures a combination of instructional delivery, interaction facilitation, and an instructor's evaluations of student learning (d'Apollonia & Abrami, 1997). These measures are generally reliable, stable, and multidimensional, when developed and used correctly (i.e., multiple items are used), are relatively valid when compared with other measures of effective teaching, and generally represent an evaluation of the instructor rather than the course (Marsh, 1984; Marsh & Roche, 1997). Other research from this period suggests that

teaching evaluations are significantly positively associated with student success measures (Cohen, 1981).

However, questionnaires developed by individual institutions for assessments of teaching should be tested for validity, and multiple criteria should be used in assessing teaching effectiveness (Marsh, 1984). Much of the research from this era that generated such findings examined specific validated questionnaires, such as the Student Evaluation of Educational Quality (Marsh, 1987). While the SEEQ is still in use today, not all institutions use this or another validated questionnaire to assess teaching effectiveness, calling into question the applicability of these findings at Southern Oregon University and other institutions that do not use the specific questionnaires from which these findings were derived. Additionally, more recent research notes that the validity and reliability of student evaluations of teaching effectiveness (SETs) is mixed, and can be influenced by a number of biases and factors such as the gender of the instructor (Boring, Ottoboni, & Stark, 2016; Spooren, Brockx, & Mortelmans, 2013).

Other more recent research that re-analyzes some of the original studies via meta-analysis finds that the link between SETs and student learning is small to nonexistent (Uttl, White, & Gonzalez, 2017). Finally, some researchers have suggested that SETs should be used only in combination with other measures of teaching effectiveness, such as peer ratings, self-evaluation, student interviews, videos, teaching scholarship, and teaching portfolios, in order to form a complete picture of the instructor (Berk, 2005). With this method, each institution should form a “triangulation” of these different components of strong teaching that fits the specific educational goals of the institution, and assess instructor effectiveness via a combination of these measures (Berk, 2005).

Regardless, using a single item to assess any core variable measured by questionnaire items (such as teaching effectiveness) is considered unacceptable by measurement scholars and general practice in social scientific measurement, because reliability and validity cannot be established from a single measure (Babbie, 2007; Brown, 2006; Crocker & Algina, 2008; Shadish, Cook, & Campbell, 2002; Tabachnik & Fidell, 2007). Thus, if Southern Oregon University chooses to retain SETs, more than a single item should be used to assess teaching

effectiveness for tenure and promotion, and the University should evaluate the reliability and validity of any current or new items used in SETs particularly if they are used for promotion and tenure purposes. Without further research and validation of the current questionnaire used at SOU, it is very difficult to assess what SOU's student teaching evaluations are actually measuring. Thus, it is unclear whether SOU's student teaching evaluations - particularly the single questionnaire item measuring overall teaching effectiveness - is truly a valid and acceptable measure of the desired constructs.

## Best Practices in Teaching Evaluations

Best practices to deploy teaching evaluations help ensure that accurate data are collected. Best practices often include recommendations for *instructions*, *survey design and construction*, and *consideration for non-response bias*. The appendix to this report includes links to additional resources for best practices in student evaluations of teaching.

### Instructions

Clear instructions that are repeated with each page of an evaluation help ensure accurate data collection. Instructions for open response items should be clear. Each question should ask about one measurable quality at a time. Questions should be neutrally worded (non-leading). The purpose and use of the evaluation should be clear from the instructions. For example, students should be clearly informed of the purpose of the evaluation and how the results will be used (see [BRUSO](#) model). It should be clear which response(s) (if any) will be anonymous and how anonymity will be maintained.

### Survey Design and Construction

Collecting unnecessary demographic information should be avoided. Students might feel that their responses could be identified through their demographic information (for example, as the only graduate student in an undergraduate course). Considering survey fatigue and eliminating repetitive or unnecessary items may increase completion rates and generate more accurate data (Adams and Umbach, 2012). Using psychometric procedures may allow for the

isolation and inclusion of only evaluation items with adequate validity. Tabachnick and Fidell (2007) have published additional guidelines on survey design and analysis.

### Non-Response Bias

It is also important to consider the timing of opening or deploying the evaluation. The data from late responders differs from early responders (Estelami, 2015). Outreach to accommodate late responders could generate more accurate or useful data. It may be helpful to consider or track the characteristics of non-responders (Porter and Whitcom, 2005) and majors/non-majors (Adams and Umbach, 2012).

## Primary Research

### Findings from IFS

An online survey was distributed Winter term 2019 to the faculty senators of the Interinstitutional Faculty Senate (IFS), which includes members from the eight four-year public universities in Oregon, to obtain information about the student evaluation process from each university. Ten responses were collected which included representation from each of the universities with two of the universities completing two surveys. Respondents represented Eastern Oregon University, Oregon Health and Sciences University, Oregon State University, Oregon Technology, Portland State University, Southern Oregon University, University of Oregon, and Western Oregon University.

IFS member institutions (100%) reported “online” as the distribution method used to distribute formal student evaluations, and of those members, 60% reported that 100% of their courses are evaluated. IFS member institutions (70%) reported using “holding grades over a period of time” and “sending email reminders” as a method to increase the response rate. Response rates varied across institutions. The response rate for online student evaluations ranged from 10% to 75%, while the response rate for paper student evaluations ranged from 50% to

71%. The timing of formal student evaluations was very similar across institutions, with 80% reporting that the formal student evaluation is distributed by administration and they have no control over the timing, and 100% reported the formal student evaluation is distributed between weeks nine and ten. Member institutions (90%) reported that they prefer student evaluations be distributed between weeks nine and ten and ten weeks plus. The nature of the evaluations was fairly similar across institutions as well, with 90% reporting that the formal student evaluations are a mix of quantitative and qualitative questions.

One difference among IFS institutions was seen in the use of evaluations other than the formal online evaluations. Of the members, 50% reported that they do not implement an additional evaluation process on their own, 30% reported they do implement an additional evaluation process on their own, and 20% reported they sometimes implement an additional evaluation process on their own. Those that reported implementing an additional evaluation process on their own specified using midterm check-ins and inviting a randomly selected group of students to a one-hour conversation to address immediate needs of students. Those that reported sometimes implementing an additional evaluation process on their own specified occasionally giving a midterm feedback survey, particularly if they are trying a new instructional method.

Similarly to what will be seen below with SOU's own survey, IFS member responses varied somewhat on the usefulness, bias, and accuracy of the formal evaluations. The survey used a number of statements about evaluations with which members either said they agreed, sometimes agreed, or did not agree. Of those that responded, 70% sometimes agreed or agreed that they are useful to improve teaching effectiveness, whereas 30% did not agree.

A majority of IFS members felt that bias was a factor in evaluations. Specifically, 60% either agreed or sometimes agreed that evaluations are influenced by personal attributes (gender, ethnicity, age, etc.) rather than teaching effectiveness, and 40% did not agree. On whether or not evaluations are an accurate reflection of teaching effectiveness, 60% sometimes agreed and 40% did not agree. As far as whether or not the evaluations provided relevant information about teaching effectiveness, responses were more positive: 80% sometimes agreed or agreed, and 20% did not agree. Whether or not evaluations inhibit risk-taking is something that SOU is examining



as it considers ways to encourage innovative teaching methods, and 50% of IFS members either agreed or sometimes agreed that the formal student evaluation process negatively impacts their willingness to take risk or innovate their teaching. The additional 50% did not agree.

IFS members had a variety of responses concerning the beliefs of students. 50% reported they somewhat agree students believe they have an impact on their teaching effectiveness through the formal student evaluation process, 40% reported they do not agree, and 10% reported they agree.

As far as tenure and promotion processes are concerned, 60% reported they somewhat agree formal student evaluations are a major factor at their university in the promotion and tenure review process, and 40% reported they agree--totalling 100% who see it at least somewhat significant. Interestingly, 40% reported that not all questions are considered for promotion and tenure purposes, whereas 40% were unsure. Only 20% reported that all questions asked in the formal evaluations were considered for promotion and tenure purposes.

IFS members were also given the chance to provide qualitative information. Comments included, “the concept and the content of student evaluations is outdated and has very little effect in changing the overall outcome of our teaching” and “the use of faculty evaluations as the sole indicator of teacher effectiveness is flawed.” One IFS member noted that at their institution, faculty status and course type had an effect on how often evaluations were completed: “Evaluations are done every term for un-tenured tenure-track faculty and adjuncts. Tenured faculty are evaluated every other term. Some courses, such as seminars, are evaluated only if the faculty member requests. Online courses are evaluated every term, regardless of faculty status.”

Overall, the results of the IFS survey align well with the literature on bias and effectiveness in student evaluations of teaching. As shown below, they also align with the findings of SOU’s survey of faculty.

## MBA Research Team Findings

For BA 519 Applied Business Research, students form groups and conduct both primary and secondary research for a real-world client that needs information to solve a problem relating to business operations, growth, marketing, or other areas. In Winter 2019, the course instructors,

Donna Lane, Ph.D. and Mark Siders, Ph.D., reached out to the Senate task force on student evaluations of teaching to ask if the task force would be interested in being the “client” for two groups. The task force accepted, and the chair, Melissa Anderson, agreed to be the point person in contact with the students. Two three-person teams of MBA students then undertook research into student evaluations of teaching, one from the faculty perspective and one from the student perspective. The students then conducted literature reviews, met with the task force chair, and designed surveys to reveal both student and faculty perceptions of the effectiveness, accuracy, and potential bias of SOU’s current collection and use of student evaluations of teaching.

The team working on the faculty perspective on evaluations developed their survey based on conversations with the task force chair and a review of the literature positing inherent bias in student evaluations of teaching as well as a lack of useful information produced by typical evaluation methodologies. The team distributed the survey electronically to 234 faculty members and received 53 responses, a response rate of approximately 23%. Along with a few demographic questions, faculty were asked to rate the effectiveness of the current system and how accurately the current questions assess teaching ability. They were also asked to provide suggestions for more valuable questions. Faculty were also questioned about how seriously they believe their students take evaluations, how effective student responses are as a measure of quality, and whether or not they implement feedback from evaluations. The survey also posed several questions about bias in the current evaluation system, and what faculty would suggest to improve the current system.

Opinions differed somewhat on the effectiveness of the current system, with answers ranging from not effective at all (18%), slightly effective (43%), moderately effective (33%), and very effective (6%). Although given the option, no one rated the current system “extremely effective.” When asked whether the current questions assess teaching accurately, the answers were more dramatic—73% of the faculty surveyed responded negatively. Results for the question about how seriously students take evaluations were of a range similar to the results for overall effectiveness, with the majority of the responses falling into the “somewhat seriously” category and none at all for “very seriously.” The survey revealed that faculty do implement feedback from student evaluations—87% of the respondents said that they implement feedback

sometimes, most of the time, or always. When asked whether bias exists in the current evaluation system, 84% of the respondents replied that it does. When asked specifically about bias related to gender, ethnicity, age, and “other,” responses varied a bit across the scale of “no bias,” “some bias,” “bias,” and “extreme bias,” with “bias” being chosen a majority of the time across all categories. Suggestions for improvements that could be made to the current evaluation system varied and some are discussed more in the last section of this report.

The team working on the student perspective on evaluations of teaching also based their survey on conversations with the task force chair and a review of the literature. The survey was sent to a representative sample of 950 current students, and had a response rate of 16%. In addition to demographic questions, students were asked how often they complete evaluations and their motivation for doing so. They were also questioned about what might make them more likely to complete evaluations. Questions providing counterpoint to those asked in the faculty survey included whether or not students feel faculty value the evaluations, whether or not they believe changes are made based on evaluations, and if they would be more likely to complete evaluations if they knew they were used to make changes. Students were also asked if they were able to express themselves fully in evaluations, whether their evaluations were influenced by outside factors, and whether or not they answered evaluation questions honestly. Based on anecdotal data, the research team believed that fears about anonymity might be a factor, so students were also asked if they believed the evaluations were anonymous. Finally, students were also asked to make suggestions to improve the evaluation process and/or make additional comments relevant to evaluations.

Surprisingly, the majority of students said they never complete course evaluations, which does not align with completion rate information provided by the registrar’s office nor with answers to follow-up questions about completion and motivation. Motivations to complete evaluations varied, but included improving the course (56%), improving teaching effectiveness (63%), getting access to final grades (39%), and being unsatisfied with the course (19%). Many students (27%) said they would be more likely to complete evaluations if they could do them in class, if they would be entered in a drawing upon completion (36%), and/or if they could do them on a mobile device (24%). A number of students (28%) said that nothing would make them

more likely to complete them. Questions about the value of the evaluations themselves were revealing, and showed that 54% of students “somewhat agree” or “strongly agree” that instructors value evaluations but only 28% “agree” or “strongly agree” that changes are made based on feedback from evaluations. Of those that responded, 87% of students “somewhat agreed” or “strongly agreed” that they would be more likely to complete evaluations if they knew that changes were made based on them. Somewhat contradictorily, 62% of students “somewhat agreed” or “strongly agreed” that they are able to express all of their feelings on course evaluations, but 70% of student respondents also said that they never answer evaluations honestly. Of those that responded, 55% said that they “somewhat agreed” or “strongly agreed” that the evaluations really are anonymous, which leaves 45% who may be influenced by their lack of confidence in the anonymity of the process. Some of the suggestions made by students for improvements to the process are discussed further in the following section of this report.

## Task Force Recommendations

As stated above, research shows that due to the potential bias inherent in student evaluations of teaching, the use of such evaluations in promotion and tenure decisions is potentially discriminatory. Therefore, the task force recommends that student evaluations in their current quantitative form be removed from promotion and tenure consideration. Since the current form of student evaluations are used to determine teaching effectiveness, and a certain score must be received for promotion and/or tenure, the Senate would need to change Bylaws 5.224 and 5.260-5.263 if student evaluations were no longer used in this way. However, despite the bias identified in quantitative student evaluations of teaching, the task force recognizes the importance of including student feedback in evaluations of teaching. Therefore, the task force suggests more equitable options for both evidence of good teaching and the evaluation of teaching effectiveness that might replace the quantitative use student evaluations currently practiced. Interestingly, the University of Oregon, one of the IFS members included in the survey described above, has this term also removed standardized electronic student evaluations of teaching from promotion and tenure considerations, and has instituted a number of practices

similar to the ones the task force identified as viable replacements for our current evaluations (University of Oregon, 2019).

It may be useful to provide a list of tips or recommendations to faculty along with any summative evaluation to build rapport, increase participation and solicit useful feedback as suggested in the “best practices” above and in the appendix to this report. This might also extend to sharing research with faculty and explaining the intent of formative and summative evaluation processes, as well as reminders that faculty can use their own evaluation processes to obtain feedback throughout the term. It may also be useful to provide faculty with a summary/list of research related to non-response bias and bias in feedback more generally.

## Self-Evaluation and Reflection on Teaching

If the goal of student evaluations of teaching is to provide feedback to instructors to improve both teaching and student learning outcomes, then it seems appropriate for reflections on teaching to become part of the evaluation process. In addition, by using a self-evaluation with a reflection on teaching in promotion and tenure packets in place of a number derived from potentially biased quantitative student evaluations, the task force hopes that concerns about a lack of equity in the evaluation process would be alleviated. Because qualitative student feedback is highly valuable, the task force recommends that new anonymous student evaluations be administered. The new evaluations would ask for qualitative feedback on teaching and learning from students; some suggestions of specific questions are provided below. Once the instructor has received this feedback, they would use it to write a self-evaluation and reflection on the course. By allowing the instructor to reflect on the feedback and self-evaluate, the task force hopes to avoid a problem identified in our research, that of inhibiting innovation and experimentation in the classroom by placing so much weight on student evaluations of teaching in promotion and tenure decisions.

In 2018, the University of Oregon began including centrally-administered “instructor reflections” in their teaching evaluations. These reflections are archived for the personal use of the instructor, and are also available for evaluators to read so that “the instructor’s voice can inform evaluators’ interpretation of student feedback” (University of Oregon, n.d.). Although the

task force is not at this time recommending that reflections be collected automatically for each course, we do see the value in allowing an instructor to contextualize student feedback and we also recognize how this type of reflective practice contributes to the professional development of the instructor.

## Mid-term Student Survey

Many SOU faculty currently use mid-term surveys or “check-ins” as a formative assessment of student learning, and the task force is recommending that the mid-term survey could serve as additional information for the self-evaluation and reflection on teaching. As part of its overhaul of student evaluations of teaching, the University of Oregon has developed a campus-wide “Mid-term Student Experience Survey” for instructors to use for formative assessment. According to the Provost’s Office at the University of Oregon, “Instructors can use this tool to check student experience against instructor expectations, acquire a sense of class climate, and consider adjustments to class plans if appropriate” (University of Oregon, 2019). Although the task force is not recommending a campus-wide mid-term survey, we are in favor of optionally including formative assessment such as this as part of a reflective teaching practice that informs self-evaluation.

## Peer Observation

Teaching observation is currently included in formal colleague evaluations, but a more flexible form of peer observation including reflective feedback could be chosen by faculty who wish to use feedback from observations to improve teaching (Cantillon & Sargent, 2008). This peer observation could supplement the self-assessment based on student evaluation feedback. The University of Toronto has compiled a number of processes, tools, and instruments for effective peer observation of teaching that includes reflection on teaching goals and practices and provides benefits to both the observer and the observed (University of Toronto, 2017).

## Summary Evaluations

On March 8, 2019, the task force met with the University Registrar, Matt Stillman, to discuss how summary electronic evaluations might be customized to align better with learning outcomes. Stillman assured the task force that programs could provide their own customized questions that would evaluate how well individual courses met the learning goals of programs, and that responses to these customized evaluations could still be distributed electronically and anonymized. Since the task force sees the value in more general questions that apply to all or most programs as well, the recommendation for student ratings going forward is to have a small set number of campus-wide qualitative questions that provide useful feedback about the course structure and learning environment and an option to include an additional small set of questions linked to program learning goals. The second set of questions would be designed by programs to provide useful information about specific program directions and objectives. All questions would be focused on student learning, and not on individual personalities or characteristics of the instructor in order to minimize bias. These recommendations align with feedback from IFS members who made suggestions for improvement to evaluation questions such as, “we should also look at whether students met the learning objectives in the class, whether they’re able to succeed in future courses that rely on the original class,” and “I would like to see some way of tying course evaluations...to course objectives, and to students’ own objectives in taking a class.” Faculty developing the questions may want to review those developed by Peterson, Biederman, Andersen, Ditonto, & Roe (2019) for possible ways of mitigating bias in question formulation. Some examples of possible campus-wide questions follow. The task force additionally recommends that the evaluation form itself includes directions reminding students to focus on the learning experience, not on personal aspects of the instructor. Additionally, since many students indicated in their suggestions that they wanted to know that course changes had been made based on evaluations, we feel that including a short explanation of how evaluations are used for course improvement might increase student motivation to complete evaluations. Finally, we recommend that the directions include an explanation of the anonymity of the evaluation

itself, since task force research indicated that many students are concerned about whether the process is in fact anonymous, and one of the most common suggestions from the student survey was for the evaluations to be “truly anonymous.”

Instructions:

Thank you for filling out this evaluation. When answering the following questions, please avoid personal remarks. Instead, focus on your **learning experience** in order to help the instructor and program, who are thinking about how best to help students learn. Your responses will be anonymous, and will be collected by a third party. Your instructor as well as the program head(s) will also be able to read your anonymous responses, but your name will not be connected to your feedback.

Questions:

Please discuss how course materials and assignments affected your learning. For example, were you engaged by the course materials (e.g., lecture, readings)? Did you understand assignment guidelines? Was the course well-organized?  
(open-ended; text box)

Please discuss how your interactions with the instructor affected your learning. For example, did you feel welcomed and respected in the class? Did you feel supported by the instructor in your learning? Was communication with the instructor timely and meaningful? Did you receive helpful feedback on your work? (open-ended; text box)

What did **you** do, or could you have done, to support your success in this course?  
(open-ended; text box)



Do you have any additional feedback? For example, what were the most valuable aspects of the course? What were some of the most important or useful things you learned? (open-ended; text box)

The questions suggested above were adapted from various models encountered during our research and took into account the best practices described above as well. Our primary goal was to develop qualitative questions less prone to the bias seen in most quantitative evaluations. A future committee assigned to develop the actual questions to be implemented may want to consider other options as well.

## Conclusion

In conclusion, our research on best practices and effective use of teaching evaluations suggested that in order to provide meaningful feedback to faculty, as well as decrease bias in evaluation results, the current use of quantitative student evaluations of teaching in promotion and tenure decisions should not be continued, and the questions themselves should be redeveloped with a focus on meaningful responses about teaching quality from students, which we believe can best be achieved with qualitative and/or program-specific questions. We recommend that another task force or ad hoc committee work on developing these questions in the next academic year. These questions can then be used by faculty in development of a self-evaluation and reflection on the course. This process would allow instructors to gather more useful feedback from students than is allowed in the current system, would help limit bias, and would allow instructors who innovate or redevelop a course in a given quarter to reflect on that process and the outcome with the guide of student feedback.

## References

- Adams, M. J., & Umbach, P. D. (2012). Nonresponse and online student evaluations of teaching: Understanding the influence of salience, fatigue, and academic environments. *Research in Higher Education*, 53(5), 576-591.
- Babbie, E. (2007). *The practice of social research* (11th Ed.). Belmont, CA: Thomson Wadsworth.
- Berk, R. A. (2005). Survey of 12 strategies to measure teaching effectiveness. *International journal of teaching and learning in higher education*, 17(1), 48-62.
- Boring, A., Ottoboni, K., & Stark, P. (2016). Student evaluations of teaching (mostly) do not measure teaching effectiveness. *ScienceOpen Research*.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: The Guildford Press.
- Burke, A., Head-Sturgess, W., & Siders, M. (2017). He's smart and she's nice: Student perceptions of male and female faculty. *International Journal of Gender and Women's Studies*, 5(1), 1-8.
- Cantillon, P., & Sargeant, J. (2008). Giving feedback in clinical settings. *British Medical Journal*, 337(7681), 1292-1294.
- Cohen, P. A. (1981). Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of educational Research*, 51(3), 281-309.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage Learning.
- d'Apollonia, S., & Abrami, P. C. (1997). Navigating student ratings of instruction. *American psychologist*, 52(11), 1198.
- Doerer, K. (2019, January 13). Colleges are getting smarter about student evaluations. Here's how. *The Chronicle of Higher Education*. Retrieved from <https://www.chronicle.com/article/Colleges-Are-Getting-Smarter/245457>

- Estelami, H. (2015). The effects of survey timing on student evaluation of teaching measures obtained using online surveys. *Journal of Marketing education*, 37(1), 54-64.
- Flaherty, C. (2018, May 22). Most institutions say they value teaching but how they assess it tells a different story. *Inside Higher Ed*. Retrieved from <https://www.insidehighered.com/news/2018/05/22/most-institutions-say-they-value-teaching-how-they-assess-it-tells-different-story>
- Joye, S.W. & Wilson, J.H. (2015). Professor age and gender affect student perceptions and grades. *Journal of the Scholarship of Teaching and Learning*, 15(4), 126-138.
- Hamermesh, D.S., & Parker, A. (2005). Beauty in the classroom: instructors' pulchritude and putative pedagogical productivity. *Economics of Education Review*, 24, 369-376.
- Huston, T.A. (2006). Race and gender bias in higher education: Could faculty course evaluations impede further progress toward parity? *Seattle Journal of Social Justice*, 4, 591-611.
- MacNell, L., Driscoll, A. & Hunt, A.N. (2015). What's in a name: Exposing gender bias in student ratings of teaching. *Innovative Higher Education* 40(4).40, 291-303.
- Marsh, H. W. (1984). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and utility. *Journal of educational psychology*, 76(5), 707.
- Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. *American psychologist*, 52(11), 1187.
- Mitchell, K., & Martin, J. (2018). Gender Bias in Student Evaluations. *PS: Political Science & Politics*, 51(3), 648-652.
- Peterson, D.A.M., Biederman, L.A., Andersen, D., Ditonto, T.M., & Roe, K. (2019). Mitigating gender bias in student evaluations of teaching. *PLoS ONE*, 14(5): e0216241.
- Porter, S. R., & Whitcomb, M. E. (2005). Non-response in student surveys: The role of demographics, engagement and personality. *Research in higher education*, 46(2), 127-152.
- Schmidt, P. (2017, January 13). When students' prejudices taint reviews of instructors. *The Chronicle of Higher Education*. Retrieved from <https://www.chronicle.com/article/When-Students-Prejudices/238892>

- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Independence, KY: Wadsworth Cengage Learning.
- Smith, B.P. & Hawkins, B. (2011). Examining student evaluations of Black college faculty: Does race matter? *The Journal of Negro Education*, 80 (2), 149-162.
- Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. *Review of Educational Research*, 83(4), 598-642.
- Supiano, B. (2018, June 29). A university overhauled its course evaluation to get better feedback. Here's what changed. *The Chronicle of Higher Education*. Retrieved from <https://www.chronicle.com/article/A-University-Overhauled-Its/243803>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Boston, MA: Pearson.
- University of Oregon. (2019). Revising UO's teaching evaluations. Retrieved from <https://provost.uoregon.edu/revising-uos-teaching-evaluations>
- University of Oregon. (n.d.). Teaching engagement program. Retrieved from <https://tep.uoregon.edu/self-presentation>
- University of Toronto. (2017). *Peer observation of teaching: Effective practices*. Retrieved from <https://teaching.utoronto.ca/wp-content/uploads/2017/01/Peer-Observation-of-Teaching-Guide.pdf>
- Uttl, B., White, C. A., & Gonzalez, D. W. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. *Studies in Educational Evaluation*, 54, 22-42.

# Appendix

## Additional Resources for Best Practices

[Iowa State University provides a concise list of best practices with literature review](#)

[UW Madison provides suggestions for best practices](#)

[2018 IDEA report on best practices](#)

[University of Washington evaluation best practices](#)

[More information about the psychometrics and caveats of data interpretation from SETs](#)

[Rice University Guidelines for Student Ratings](#)