

Center for Teaching and Learning

Student evaluations

[Close Menu](#)

[A guide to best practice for evaluating teaching](#)

[Self-assessment](#)[Peer review](#)[Student evaluations](#)

Student ratings and comments provide one source of data for review and promotion committees: The student perspective at a particular moment in time, at the end of the course. While students don't know at this point how they'll use what they learned or how they'll view the course after they've graduated, course evaluation data are still useful, because students are experts in evaluating their experience and perceptions as learners. It is important also to provide a means for students to have a voice and for faculty to remain accountable to their students.

As with other types of data, contextualizing data gathered from students is key for fairly assessing teaching practices. In this section we outline important factors to take into account when interpreting student evaluation data and provide further reading on discussions and debates in the research literature on this data. We also strongly recommend that review and promotion committees value student ratings and feedback data in light of self-assessment and peer review; taken together, these three areas offer meaningful data and analysis for evaluation. Reviewing the instructor's interpretation of the student data—along with the instructor's discussion of instructional contexts, innovations, and classroom evidence—reflects robust best practice.

[In practice](#)[In review & promotion](#)[Learn more](#)

Research

I. Instructional context

One of the most common evidence-based practices in evaluating teaching includes a recognition of the instructional context; when reviewing student evaluative data, note that:

Student ratings and comments:

- Are affected by students' reasons for enrolling in a course, such as: Whether it is required or elective, their expected grade relative to other courses, and class size. To account for these factors, UW's course evaluation IASystem computes Adjusted Medians for the first four items (the Summative Items) and for the combined median of the four items (the Overall Summative Rating).
- Can be biased by student perceptions of instructor identity, including the instructor's race, ethnicity, gender, disability, age, and other characteristics, including those for whom English is an additional language (1). See "[For Review and Promotion Committees: Countering Bias in Teaching Evaluation](#)" in this report.
- Overall, online courses are rated somewhat lower than face-to-face courses, and online evaluations are also slightly lower than paper evaluations (2). We recommend that review and promotion committees compare online and hybrid courses to similar courses. Online evaluations can become bimodal with students who rank very high or very low.
- Piloting teaching innovations may initially produce more negative student comments and a decrease in ratings due to student resistance to changes in roles and expectations. Thus, we recommend that review and promotion committees explicitly support teaching innovation by stating in writing that innovation in teaching is valued and, like innovation in research, not expected to succeed completely at first. Look for explanations of pilots and innovations in the teaching statement, letter, or philosophy and seek evidence that the instructor is using pilot data to improve the course in its next iteration.
- Course evaluations completed online tend to have a lower response rate and may produce bimodal results, with high scores, low scores, and few in the middle. Look closely at the response rate and the distribution of responses. If you are consistently getting low rates, think about ways to improve them going forward such as providing time in class (on a day other than the last day) for students to fill out forms and emphasizing with students the importance of their feedback and how it will be used.
- And, of course, response rates and response variability matter. Data from low response rates should not be considered representative of all students. In small classes, even with high

response rates, averages of small samples can be more susceptible to “the luck of the draw” than averages of larger samples, producing more extreme evaluations than in larger classes. Students in small classes might also imagine their anonymity to be more tenuous, perhaps reducing their willingness to respond or to do so truthfully (3).

Formative vs. summative

Formative and summative assessment differ in purpose — formative assessment aims to improve while summative assessment aims to evaluate. Yet, summative assessment, such as the work done by review and promotion committees, offers opportunities to reward the appropriate use of formative assessment.

Those involved in evaluation can seek evidence that a faculty member has systematically gathered student feedback on teaching effectiveness throughout a course and has made appropriate changes in response to these data. Examples of systematic collection of student data include: surveys, WebQs, and open-ended feedback forms. (For more more about methods for gathering and effectively using student feedback, see the CTL’s online resource page, [Gathering Student Feedback](#).)

Why might students give low ratings?

While low ratings usually indicate instruction that genuinely needs improvement, sometimes other factors are at play. Other sources of data may help to identify some of these other factors, such as student’s comments, observation by peers, and the instructor’s experience.

For example, students give low ratings in the end-of-term course evaluations to the question: “The course content was...”. However, the vast majority of students written comments indicate that students attribute their dissatisfaction with the sequence of the course required of the major, and not how the content was chosen or taught by the instructor.

Similarly, if a strong theme in student comments indicates student bias because of the faculty member’s gender, race, nationality, or other aspects of their identity, the faculty member may choose to share additional data to contextualize student ratings and/or other challenges in teaching the course.

Collecting midterm feedback from students

Midterm feedback allows instructors to check in with students with enough time to tweak their teaching before the end of the quarter. Effective methods for doing so include gathering written feedback through an anonymous survey or WebQ, or offering a whole class interview process, such as Small Group Instructional Diagnosis (SGID).

Formative assessment—such as the collection of midterm student feedback to provide an instructor with data on the impact of his or her teaching—differs in purpose and audience from evaluative assessment typically shared in a performance review. For one thing, it’s voluntary. However, formative assessment data that the instructor has gathered might be usefully included in a performance review in one way: An instructor may choose to describe how and why they gather formative feedback and how it informs their teaching. Too, the instructor may choose to briefly highlight relevant examples of formative feedback that provided key insights into their teaching practice or development.

II. Interpreting data from student evaluations

Student evaluation data can be analyzed productively by seeking patterns and themes in the data as well as by considering it in the appropriate contexts.

Questions for identifying patterns and themes in data:

- What patterns, if any, are indicated in numerical ratings? Ratings for some items may help to make sense of ratings for others.
- What patterns or strong themes are indicated in qualitative comments? Themes in qualitative comments can be quantified to demonstrate the degree of student consensus on particular aspects of the course.
- Are there patterns across sources of data? Quantitative ratings may reflect points raised in students’ open-ended comments or vice versa. Both sets of data may inform an instructor’s own self-assessment, a peer review, or other forms of data.

Questions for contextualizing student data:

- What is the teaching context? Take into account course characteristics such as size of the course, whether an instructor is co-teaching, and whether the instructor supervises TAs. How do these ratings compare to: The instructor’s other courses; ratings for courses with similar sizes, levels, or content; or other courses with similar backgrounds and preparation?
- What changes have occurred over time? What has improved?
- What ratings stand out to the instructor and why? Instructors have the most context for the course, including their aims in teaching. What ratings do instructors find most useful for their

- own self-assessment in both strengths and areas for growth, and why?
- What information is available to help clarify specific issues? Qualitative data, particularly, may help provide a more nuanced understanding of the course context or a particular issue raised in other data under review.
- What information may help reconcile divergent ratings? Qualitative data may also help review and promotion committees interpret divergent student ratings—for example, when instructors receive very high and very low rankings from different groups of students in the same course, or across multiple quarters of teaching the same course.

For review and promotion committees: Countering bias in teaching evaluation

Given the research on bias in student course evaluations, we urge review and promotion committees to counter bias actively by considering student data through the lens of ongoing research on this topic and in light of additional sources of evidence, such as self-assessment, peer review data, and evidence of student learning.

Reviewers may also consider Merritt's (4) suggestions for evaluation practices:

- Use focus groups mediated by outsiders.
- Encourage instructors to do evaluations less often but more deeply (at least once per quarter).
- Develop instruments that encourage students to move past reacting intuitively but to spend more time thinking about their responses.
- Recognize teaching as an ongoing process rather than an end product.

For instructors: What should I do if I suspect bias in my teaching evaluation?(5)

1. Rule out that other factors are not in play. Student ratings should never be the sole source of data used in performance appraisal. Consider the following:
 - Do evaluations include criteria beyond basic presentation and delivery skills? These skills are associated with traditional lecturing.
 - Do evaluations include review of course plans, course design, and instructional development?
 - Do evaluations focus on learning outcomes and supporting assessment data?
 - Do evaluations consider faculty contact with and mentoring of students?
2. Be deliberate in your selection of student ratings forms or questions (if possible); including questions about teaching methods you use and reflect characteristics that are important to you.
3. Be responsible for interpreting your own student ratings. Do not leave it to anyone else to decide the meaning of your students' ratings and comments. You were in class with them, so you are more likely to be able to interpret what the scores and comments indicate.
4. Ask other faculty who may have experienced bias from students or peers for strategies to compensate for differential expectations.
5. Identify and align faculty and student expectations on the first day of class.
6. Consult with experts on instruction: E.g. consult with your campus teaching center, and seek information posted on the internet by other campus teaching centers or organizations.

Citations

- 1: [Huston, T. A. \(2005\). Race and gender bias in higher education: Could faculty course evaluations impede further progress toward parity? *Seattle Journal for Social Justice*, 4 \(2\), Article 34.](#)
- Ogier, J. (2005 October). Evaluating the effect of a lecturer's language background on a student rating of teaching form. *Assessment & Evaluation in Higher Education*, 30 (5), 477-488.
- 2: [McGhee, D., & Lowell, N. \(2015\). Effects of course delivery mode and course evaluation mode on student ratings of instruction \(Rep. No. 15-02\). Seattle, WA: University of Washington Office of Educational Assessment.](#)
- 3: [Stark, P.B., Freishtat, R. \(2014\). An evaluation of course evaluations. *ScienceOpen Research*.](#)
- 4: [Merritt, D.J. \(2008 winter\). Bias, the brain, and student evaluations of teaching. *St. John's Law Review*, 82 \(1\), 235-287.](#)
- 5: [Adapted from: Linse, A. \(2003\). Student ratings of women faculty: Research & strategies. Document created for an NSF ADVANCE program workshop.](#)

Advice for review and promotion

FOR CHAIRS AND COMMITTEES

See ratings data as suggestive, not definitive

FOR INSTRUCTORS

Follow best practices for administering evaluations in order to gather substantive feedback.

Student evaluations should be considered in the context of other evaluation materials, such as peer review and self-assessment.

Account for variables

Consider the number and percentage of students who provide data (the larger the N, and the greater the percentage of students who respond, the more reliable the data); consider whether the instructor used best practices for administering evaluations (e.g. providing students with context and in-class time to do evaluations).

Seek evidence of responsiveness

Look for examples of changes in the course as a result of student feedback in the evaluations summary and other materials.

Read comprehensively

While focusing on the Overall Summative Rating and the four Summative Items (the Standard Formative Items are intended for diagnostic purposes in improving teaching), look at the distribution of student responses rather than the median; focus on general categories (such as "Excellent," "Very Good") rather than minor point differentials; if the instructor has offered multiple versions of the same course, aggregate data across courses.

Customize

Add customized questions that speak to specific aspects of the course, such as teaching methods, innovative practices, or course goals.

Contextualize

Tell your students how evaluation data will be used and by whom (reminding them that responses are anonymous). Emphasize that you value their honest, constructive feedback and will use it to improve teaching practices.

Make time in class

Research has shown that setting aside in-class time for evaluations yields higher response rates.

When presenting student evaluation data:

Summarize and contextualize the data

Do not assume the data speak for themselves. Summarizing student feedback within a narrative can contextualize the feedback in terms of your teaching approach and goals. Describe strong themes represented across student feedback, and clearly indicate how these comments were determined to be representative (e.g. what percentage of students provided comments? How many comments support each theme?)

Organize strategically

You could provide an overall summary of all data gathered over several years and/or a half- or full-page summary for individual courses. Feedback can also be summarized in a chart that may be structured in a number of ways: Chronologically; by courses taught; by themes in student feedback; or in another way that aligns with your approach to teaching.

Describe specific feedback

Include representative examples of student feedback that evidences and demonstrates your strengths and growth as a teacher. Cite student comments describing specific practices (e.g. how a particular activity helped student learning) rather than generalized comments.

Demonstrate growth and flexibility

Describe exactly how and why you have solicited student feedback and incorporated useful feedback into your teaching.

Frame summaries as student-

centered

Focus on the effectiveness of teaching practices on student learning rather than on other aspects of the course, such as your own knowledge.

Further reading

Description	Recommended Reading
Online resources for instructors:	<ul style="list-style-type: none">• Assessing and Improving Teaching (UW Center for Teaching & Learning)
On debates around student evaluations:	<ul style="list-style-type: none">• For a review of the literature on student evaluations, see: Spooren, P., Brockx, B., & Mortelmans, D. (2013). On the validity of student evaluation of teaching: The state of the art. <i>Review of Educational Research</i>, 83 (4), pp. 598-642.• Find a critique of the validity of course evaluations in: Stark, P.B., Freishtat, R. (2014 Sept. 29). An evaluation of course evaluations. <i>ScienceOpen Research</i>.• For a meta analysis of the research methods used to study student evaluations of teaching (SETS), see: Uttl, B., White C.A., Wong Gonzalez, D. (2017). Meta-analysis of faculty's teaching effectiveness: Student evaluation of teaching ratings and student learning are not related. <i>Studies in Educational Evaluation</i>.
On bias in student evaluations:	<ul style="list-style-type: none">• For a literature review and in-depth discussion of bias in student evaluations, read: Lazos, S. R. (2012). Are student teaching evaluations holding back women and minorities? The perils of 'doing' gender and race in the classroom. In G. Gutiérrez y Muhs, Y. Flores Niemann, C.G. Gonzalez, and A.P. Harris (Eds.), <i>Presumed Incompetent: The Intersections of Race and Class for Women in Academia</i> (pp. 164-185). Boulder: University Press of Colorado.• Find a recent (and growing) annotated bibliography on gender bias at: Savonick, D., Davidson, C. N. (2016). Gender Bias in Academe: An Annotated Bibliography of Important Recent Studies. Retrieved from HASTAC (Humanities, Arts, Science, and Technology Alliance and Collaboratory)

A guide to best practice for evaluating teaching

[Self-assessment](#)

[Peer review](#)

Student evaluations

COUNTERING BIAS IN TEACHING EVALUATION

[READ MORE](#)