Evaluating Teaching in Promotion & Tenure Cases:
Guide to Best Practices

Table of Contents

Introduction: 1
Self-Assessment: 2-4
Peer Review: 4-6
Student Ratings 6-7
References 8-9
Appendices 9-10

Introduction

Nearly two years ago, some faculty members of color approached then Provost Cauce with a request that we think about a more sophisticated approach to evaluating teaching. Student evaluations, most admit, are limited in their value, and, at times, tell us more about the student evaluators than the faculty being evaluated. This can be especially true for faculty of color, and especially women faculty of color, who face a good deal of pushback from students. In response, the Center for Teaching and Learning was asked to draft a best practices guide to evaluating teaching for use by department and college groups that review tenure and promotion cases. The guide draws on research and on the work of peer institutions.

We hope this guide encourages tenure and promotion committees to seek evidence-based processes that reflect disciplinary norms and the "innovation imperative" that informs our research, leadership, and service at the UW. Given that teaching evaluation is a process--not a moment, a letter, a one-hour class observation, or a swift look at evaluation ratings--we also hope to spark disciplinary as well as interdisciplinary conversations about teaching as a scholarly practice conducted in a community.

The guide has three sections: Self-Assessment, Peer Review, and Student Ratings. Each section focuses on evaluating teaching as part of tenure and promotion committee work, rather than on instructional or formative assessment intended for the faculty member’s use. Yet despite differing goals and audiences, formative assessment aligns constructively with summative evaluation, both contributing to an ongoing culture of conversation, innovation, and excellence in teaching.

Many colleagues shaped this guide. Marcia Killien, Secretary of the Faculty; Nana Lowell, Cathy Beyer, and the research team of the Office for Educational Assessment; Jan Spyridakis of HCDE, the Faculty Council on Teaching and Learning; the Teaching and Learning Group; and the Campus Fellows on Teaching Assessment. We look forward to gathering additional feedback and comments from other faculty groups. The guide will be finalized and ready for use by June 15, 2016.

Calla Chancellor, Karen Freisem, Beth Kalikoff, Katie Malcolm, and Theresa Ronquillo
UW Center for Teaching & Learning
Section 1: Self-Assessment

“Self-evaluation is one of the most overlooked forms of explicit evaluation. Ideally and logically, this should precede all other forms of the evaluation of teaching effectiveness.”
- Warwick University Learning and Development Centre (2012) Self Evaluation

“The individual instructor is in the best position to evaluate the strengths and weaknesses of an individual offering of a course, and to suggest approaches to improving future offerings.”
- UW College of Engineering, Promotion and Tenure Toolkit

As one element of an evidence-based approach to evaluating teaching, self-assessment is an important tool for understanding teaching effectiveness. According to Berk (2005), self-assessment provides “systematic, ongoing reflection on your teaching and courses” and is used for evaluative purposes in most four-year colleges and universities (p.51).

Faculty members use self-assessment to evaluate their teaching and advance their growth as instructors. Departments decide which kinds of faculty self-assessment to use in promotion and tenure decisions and how often their faculty members should reflect on their teaching--weekly? monthly? quarterly? at the end of the year? In what forms? We recommend that tenure and promotion committees use faculty self-assessment, in the context of articulated department norms around self-assessment, as core data in their evaluation of the candidate’s teaching.

This section focuses on three aspects of instructor self-assessment:

I. Course reviews
II. Teaching portfolios
III. Philosophy of teaching statements

1. Course reviews

Systematic course reviews allow instructors to reflect on what is working well in each course and make appropriate changes at key points before, during, and after the quarter. Strategies may include:

1. End-of-term summaries. Many UW departments encourage faculty members to use end-of-term course summaries to offer assessment data to promotion and tenure committees. The College of Engineering advises candidates to “prepare short (less than one page) written evaluations for each course taught...immediately after each course is offered. A suggested model for the reports is that the faculty member summarize what was done in terms of improvement, innovation, updating, and so on, along with an evaluation of what was effective and what was less effective, and how the course could be improved in the future.” End-of-term summaries may also include instructors’ reflections on their end-of-term student evaluations. See Appendix A for samples of guiding questions.

2. Course portfolio. A course portfolio may contain (1) course components--syllabus, teaching materials, support materials, and assignments; and (2) critical analysis of teaching and learning, including reflections on teaching. A course portfolio may contain annotations, midterm feedback,
3. **Syllabi, assignments, tests, and class plan/agenda annotation.** Regularly recording brief reflections on course materials help instructors understand what to keep and/or change when they teach the class again. Guiding questions for annotations might include:

- What did I want students to learn (in this lesson, activity, assignment)?
- How did it go (and how do I know how it went)?
- What would I do differently next time?

### II. Teaching Portfolios

Seldin (2005) describes the teaching portfolio as “a collection of materials that document teaching performance….It is flexible enough to be used for tenure and promotion decisions or to provide the stimulus and structure for self-reflection about areas in need of improvement” (p.3). Teaching portfolios are used for promotion and tenure decisions at a number of universities, including the University of Michigan.

Teaching portfolios used for promotion and tenure typically include the following:

1. Teaching Responsibilities
2. Teaching Philosophy
3. Teaching Objectives, Strategies, Methodologies
4. Student Evaluations for Multiple Courses Using Summative Questions
5. Classroom Observations by Faculty Peers or Administrators
6. Review of Teaching Materials by Colleagues Inside or Outside the Institution
7. Representative and Detailed Course Syllabi
8. Evidence of Student Learning (Cognitive or Affective)
9. Teaching Recognition and Rewards
10. Short-Term and Long-Term Teaching Goals
11. Appendices

(Seldin, 2005, p.16)

As a “vehicle for structured reflection about teaching” (p.18), the portfolio can provide instructors with comprehensive insight. For more information about teaching portfolios, including links to sample portfolios, see the University of Virginia’s “Teaching Portfolios.”

### III. Statements of Teaching Philosophy

As a “purposeful and reflective essay about the author’s teaching practices and beliefs” (Vanderbilt 2015), the teaching philosophy statement is another element of self-assessment. Teaching Statements (1-2 pages) often include three elements:

1. **Description:** What you do when you teach, types of activities or thinking in which you
engage your students.

2. **Analysis**: Why you teach in the ways that you do, and how your thinking about your teaching has changed over time.

3. **Empirical data**: Experiences or observations of student learning on which your decisions about teaching are based.

(UW Center for Instructional Development and Research, 2003)

Statements of Teaching Philosophy provide an opportunity to reflect on and communicate what individual instructors do and why. To learn more about Teaching Philosophy Statements, including examples, see [Ohio State](#) and [Vanderbilt](#) guides. For more on writing a Statement of Teaching Philosophy, see the Center for Instructional Development and Research Bulletin, “Writing a Teaching Statement.”

Section 2: Peer Review

“Teaching is perhaps the most privatized of all the public professions. Though we teach in front of students, we almost always teach solo, out of collegial sight – behind closed doors.”

- Parker Palmer, *The Courage to Teach*

As with scholarship, effective teaching takes place in a community, one that generates and analyzes data, draws on research, and develops in collegial, public, and private conversations. Peer review is commonly used for scholarly activities in generative, instructional, and evaluative ways. Given that teaching is a scholarly activity, there is ample support for peer review of teaching as an evidence-based best practice (Gosling, 2014; Kalish, 2015).

Peer review contributes to a culture of continuous improvement by sharing disciplinary norms and assumptions, research-based best practices, and innovations. At UW and its peer institutions, ongoing reciprocal mentoring is a peer review practice with considerable benefits for all participants.

While the [UW Faculty Code](#) calls for peer review in service of effective teaching, the specific goals and practices that support these goals are left to academic departments to determine (see [Faculty Code 24-32 & 24-57](#)). We urge department and college groups responsible for evaluating teaching in tenure and promotion cases to seek out evidence in a candidate’s file that the department has engaged in thoughtful, ongoing peer review and has articulated practices that align with disciplinary norms and contexts.

I Department Recommendations
II A Protocol for Effective Peer Teaching Observation

We recommend that departments determine purposes and protocols around peer observation through department-wide discussion or the formation of a one-year teaching assessment
committee. We then recommend that departments articulate these purposes and protocols explicitly for tenure and promotion committees, both departmental and college-wide. Questions to address include:

1. What are departmental expectations and purposes for peer review? How might peer review be used for supporting the professional development and growth of faculty members? For example, is peer review used for formative as well as evaluative purposes?

2. Which faculty and instructors perform peer review and how do we define “peer” in this context? A faculty member in the same department? The same area of the field? At the same rank? Or?

3. Will faculty be reviewed by one or more colleagues? Some models of peer observation favor pairs, triads, or quartets of colleagues. If the department selects this model, a “norming” or reconciliation discussion often takes place among the observers before they meet with the instructor to insure more consistent and coherent feedback.

4. What criteria will reviewers use in this process? While many institutions suggest using a shared rubric or template for observations, such templates must still allow for diverse practices, recognizing that effective teaching is not limited to one specific approach or set of practices.

5. What is the expected protocol that reviewers and reviewees are to follow? Once departmental guidelines have been established, faculty should be provided with a standard protocol. Below is an example of one such protocol.

A Protocol for Effective Peer Teaching Observation:

1. The reviewee and observer meet with each other to clarify goals and context. The reviewee describes the course and may share course materials (such as syllabi and course websites) to provide context for the observation. The discussion includes what elements of the course helps students learn, what the challenges are, and what kind of feedback the faculty member will find most useful.

2. Agree on a protocol for teaching observation. How many times to observe? For what period of time? More importantly, what is the peer observer looking for? What are observable practices that departments consider most useful? What kind of protocol is broad enough to encourage diverse practices and innovations while precise enough so that colleagues are looking to answer the same questions about a peer’s teaching?

3. Follow up with a conversation. The observer gets together with the faculty member to say what she or he saw and speaks, if possible, to the faculty member’s specific questions. The focus is on observed effective practice, open-ended questions, and the faculty member’s goals.
4. Collaborate on writing a report. The faculty member and observer collaborate on a summary report that describes the conversation, which may include observable strengths of the class session and what could be improved or refined, why, and how.

The appendix offers some resources for and examples of effective peer review practice.

Section 3: Student Ratings

Student ratings provide one source of data for tenure and promotion committees: the student perspective at a particular moment in time, the end of the course. Even though students don’t yet know, at this point, how they’ll use what they learned or how they’ll view the course after they’ve graduated, the ratings data are still useful, because students are experts in evaluating their experience and perceptions as learners. For example, when we ask students to rate how confident they are in their instructor’s knowledge (Item #11 on many of the IASystem forms), they rate their perception of their instructor’s knowledge. What they can’t rate is their instructor’s actual knowledge.

This section, then, offers the best ways to use student ratings when evaluating the teaching of candidates for tenure and promotion. It focuses on two aspects of student ratings:

I. Reliable data
II. Instructional context

Reliable data:

- Consider the number and percentage of students who provide data. The larger the N, and the greater the percentage of students who provide data, the more reliable the data. Course data from 7-10 students are far more meaningful in a class of 25 than of 100.

- Analyze data from at least five courses to ensure inter-class reliability.

- Think in terms of general categories (“Excellent,” “Very Good,” etc.) A difference of less than 0.3 is not statistically significant. We recommend that faculty members, tenure and promotion committees, and departments value student ratings in light of self-assessment and peer review: taken together, these three areas offer meaningful data and analysis for judgment.

Instructional context:

- There is evidence that student ratings are affected by students’ reasons for enrolling in a class, expected grade relative to other courses, and class size. Therefore, UW’s course evaluation IASystem computes Adjusted Medians for the first four (global) items and the combined mean of the four items.
• There is evidence that other factors affect student ratings, including the faculty member’s gender, race, and ethnicity; online vs. face-to-face courses; and online vs. paper evaluations. At present, these factors are not accounted for in IASystem’s Adjusted Medians. OEA is currently considering them. In the meantime, we strongly recommend that departmental and college-level tenure and promotion committees bear in mind that gender, race, ethnicity, and country of origin inflect student ratings.

• There is evidence that faculty members who pilot teaching innovations may initially encounter student resistance, because student expectations for how the course should be taught have been stymied, particularly regarding any change in roles of faculty member and student. This resistance can lead to students feeling less positive about their experience in the course, and, as a result, faculty members might see student ratings decrease.

    We strongly recommend that departmental and college-level tenure and promotion committees bear in mind that (1) students, like other humans, do not embrace change immediately; and (2) the first time any course or innovation is offered yields data that leads to improvement the next time. (We also encourage tenure and promotion candidates to explain pilots and innovations in their teaching statement or philosophy.)

• The faculty member’s self-assessment. Reviewing the instructor’s interpretation of the student ratings data, along with their discussion of innovations and pilot projects, and teaching reflection, teaching statement, or teaching portfolio is a robust best practice.

• The course learning goals, content, and discipline. What constitutes effective teaching varies from discipline to discipline – depending on course learning goals and content. It’s important for college-level tenure and promotion committees to interpret student ratings data within a disciplinary perspective.

Conclusion

To paraphrase the late Speaker of the U.S. House of Representatives, Tip O’Neill of Massachusetts (“all politics is local”), all assessment is local, inflected by disciplinary, college, and institutional cultures and values. One size fits none. To evaluate teaching in a way that reflects research-based best practice and the UW’s innovation imperative, we recommend that tenure and promotion committees

• consider self-assessment as a source of reliable data
• seek evidence of thoughtful, ongoing departmental peer review practice that provides constructive information about a candidate’s teaching
• judge student ratings in light of instructional context and what constitutes reliable data
• use self-assessment, peer review, and student ratings data together to evaluate teaching


Warwick University Learning and Development Centre (2012). Self evaluation. Retrieved
Appendix A: Guiding questions for end-of-term course summaries
Guiding questions for these reflections include:

○ What did I do improve, update, or innovate in this course?
○ What went well this term? What do I think supported student learning?
○ What assignments, readings, etc. will I keep the next time I teach this class?
○ What resources do I need to support my teaching?

Appendix B: Sample faculty self-evaluation of teaching form
http://www.coe.fsu.edu/content/download/50739/352379/file/Self-Evaluation-of-Faculty-Performance.pdf

Appendix C: 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 may include gathering written feedback through an anonymous survey or WebQ, or a whole class interview process, such as Small Group Instructional Diagnosis (SGID) (see Bowden [2004], Diamond [2004], and Finelli [2008]). See the CTL’s Gathering Student Feedback web page for more information.

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 of the kind performed by a tenure and promotion review. For one thing, it’s voluntary. However, formative assessment might usefully become part of a tenure and promotion review in one way: a faculty member could describe the formative practices s/he chooses to employ without including the data from those practices.

Appendix D: Examples of peer review protocols and forms

Best practices for forming review committees and for peer reviewers:

Ohio State University, Office of University Senate (2015), Peer Review of Teaching:
http://senate.osu.edu/PeerEvalTeach.html

Protocols and examples, pages 12-22:

Central Michigan University, Office of the Provost (2013), A Review and Recommendations for Evaluations
https://www.cmich.edu/office_provost/facit/Documents/Teaching%20and%20Instructional%20D
Templates for peer review rubrics:


University of California, Berkeley (2015), Peer Review / [http://teaching.berkeley.edu-peer-review-course-instruction](http://teaching.berkeley.edu-peer-review-course-instruction)