

PEER EVALUATION OF TEACHING

Thor A. Hansen
Western Washington University

There has been a lot of discussion at Western recently about peer evaluation of faculty teaching, which, as a balance and supplement to student course evaluations, most faculty recognize the need for. Indeed, a great deal has been written about such evaluation in the assessment literature, although I do not intend to review that literature here. Instead I will describe the system that the Geology Department has recently implemented—and the thinking behind it—as a kind of nuts and bolts example of one way that peer evaluation could work at Western.

Goals for Peer Evaluation: It is important to recognize that there are two purposes behind faculty teaching evaluation:

- 1) *Accountability*, in which a judgment is made about the teaching ability of the professor, which in turn informs some authority (e.g. T&P committee, Dean, Provost, state legislature), that then acts on this information; and
- 2) *Assessment*, in which the professor's abilities are gauged for the purpose of self-correction and improvement.¹

Though differing in purpose, accountability and assessment should have a similar ultimate goal, that of improving teaching ability; after all, if not to insure we have excellent teachers, why have such measures? Unfortunately, gathering information for the purposes of accountability and assessment can have very different effects on the person being evaluated. Let's consider, for example, this scenario (somewhat extreme, though based on fact) where accountability, not assessment, is the primary goal:

Professor Smith is up for promotion next year. His colleague, Professor Bones, visits his class for the purpose of forming a judgment about Smith's teaching abilities. Smith sees Bones enter the classroom

and knows that he is about to be "graded" on his teaching. Professor Bones has never visited Smith's class before and will likely never visit it again. Smith's teaching abilities are on the line now; it is make or break. He had planned to try a new discussion technique in class today, one he read about in a teaching journal. Flustered by this sudden change in events, Smith fumbles his way through the discussion. The students sense his apprehension and say virtually nothing during the discussion. Bones scribbles a few notes and leaves halfway through. At the conclusion of the dismal class, Smith returns to his office and scans the want ads in the Chronicle.

In this example, the visitor's primary purpose is to make a judgment. The relationship between Smith and Bones is somewhat like that of jobseeker and interviewer, except in this case Smith already has the job, and the purpose of the visitor is to see if he keeps it. The person being reviewed has little reason to welcome the visitor and will probably be nervous during the class. The fact that Professor Bones makes only one visit means that he will get a small and possibly biased sample of Smith's abilities.

While there is the potential that Bones' bias might be mitigated by having made more classroom visits, I make the suggestion that the fundamental relationship between the observer and the observed be changed: from one of *accountability* to one of *assessment*. Rather than a culture where teaching is a private endeavor, we should strive to make it public; rather than an environment where members of a department sit in judgment, we should strive to create an atmosphere where faculty avidly seek out colleagues for pedagogical discussion and advice, where classroom visits by fellow faculty are frequent and welcome. Yes, accountability will always be present; we still need to make tenure decisions and at some point each faculty member must pass judgment on others. But there are ways to ameliorate this

adversarial component while encouraging support. Such a culture rests on two pillars: 1) non-judgmental feedback, and 2) frequent and multiple modes of assessment.

A good parallel for this model can be found in what we consider good practice in the teaching of writing. It is well known that student writing improves quickest if students are given *frequent* assignments which receive *ungraded* comments on drafts. The grade is given only when the final report is handed in. In this type of class it behooves the student to write the first draft as soon as possible and go through as many revisions as possible, with input from the instructor, in order to turn in the best final product on which the grade is based.

We can mimic this practice for faculty development by creating a system of regular evaluations of “drafts”—for example, visits to classes by reviewers who make observations, take notes, then review their observations with the instructor. The reviewer would give a copy of their comments to the instructor only (because this evaluation is primarily for self-review) and, at their discretion, keep a copy for themselves. A reviewer’s primary questions would be “What can this person do to improve?” and “Is this person making progress?” Ideally, over the course of a year several classroom visits would be made. The year-end and tenure evaluations would be *independent* and *separate* from the classroom evaluations but would be *informed* by the observations made during the year(s). At the point when judgments must be made, the questions informing the case would be: “How good a teacher is this person *now*?”, “What is the potential for this person in terms of teaching?”, and ultimately, “Is this person good enough to tenure?” In this system, there is an incentive for the probationary professor to encourage faculty to attend their classes and get feedback.

It is also important to vary the *objects* of evaluation. There are many sources of information on teaching

abilities besides classroom visits. The standard source, of course, is student evaluations. These are fine as far as they go, but I have found them, as presently constructed, to be a relatively undiscerning tool. Using the standard evaluation, I can tell when my students like my class and when they don’t. But the machine-graded questions are far too general (and in some cases useless) to inform my teaching, while the questions that elicit comments (What did you like? What needs improvement?) do not provide the kind of reflective thought I need for feedback. In order to get comprehensive feedback on my specific learning outcomes and on the teaching techniques I employed, I append customized questions to the standard evaluation form. For instance, I hand out and discuss a list of course objectives and learning outcomes at the beginning of each course. When I evaluate the course, I attach this list and ask each student to rate their improvement on each item. I also ask specific questions about the efficacy of new teaching techniques. With these directed questions, I get much fuller and more thoughtful comments than the usual “Great (or lousy) course!” My experience with these sorts of evaluations has convinced me that students are very discerning and astute commentators if they are asked the right questions.

Other sources for information on teaching include such course materials as syllabi, exams, project assignments, etc. Online materials, too, especially those that include multimedia and interactive components,

can give us excellent insight into the effort that is put into teaching. In the Geology Department we have an irregular forum where one or more faculty demonstrate a teaching technique that they have developed that they find particularly useful. These presentations are an excellent low-pressure vehicle for demonstrating creativity in teaching methods. Interviews with students, particularly graduate students are also important, because they touch on aspects that may not be reported in standard evaluations. Moreover, graduate student interviews are useful for understanding a professor’s abilities as a mentor.

I would also like to comment on the value of student evaluations in general. There are those who think that student evaluations are basically worthless; that they are governed mainly by how “popular” or easy a professor is. Some of these people advocate eliminating student evaluations altogether. Personally I can’t imagine not asking students how they felt about a course. Even in my introductory courses, which most students, as science-phobes, would rather avoid, my primary objectives include that students like taking my course, that they learn how science works and to enjoy thinking in a scientific manner—perhaps even be inspired. If most of the students don’t like the course by the end of the quarter, then I have failed.

Many times I have heard the “statistic” that student evaluations are correlated with grades, the implication being that easy-grading instructors get better evaluations. Yet there have been over 1300 articles and books published which contain research on the topic of student ratings, and when the data is synthesized it clearly indicates that students “who receive higher course grades do not give higher course ratings.”²...

As a way of bringing all their expertise together, faculty could assemble a teaching portfolio, which would, obviously, provide a place for assembling their materials, but more importantly a context for explaining and/or demonstrating their teaching philosophy. Indeed, the power of teaching portfolios came to the attention of the Geology Department during job searches conducted in the last four years. Our position announcement demanded demonstration of both research and teaching expertise. Those applicants that submitted a teaching portfolio along with their student evaluations stood out from the crowd because 1) they cared enough about teaching to create a teaching portfolio, and 2) the portfolio assembled their teaching materials and philosophy into a coherent whole.

At this point, having presented some alternative modes of peer teaching evaluations, let's revisit the scenario presented earlier involving Professors Smith and Bones. This time, rather than an adversarial approach to peer evaluation, let's imagine that their department has embraced a peer evaluation model based on non-judgmental feedback and the improvement of teaching—on the idea of assessment rather than accountability:

Professor Bones enters Smith's classroom. Smith looks up and says "Ah, Bones! So happy you're joining us! Today, I'm trying a new discussion technique and I would be most interested in your feedback." At Smith's invitation, Bones has visited his classes twice before. (Moreover, Smith's classes have been visited by two other faculty at different times of the year). Bones has also read Smith's teaching portfolio and understands his interesting though sometimes unorthodox approach to teaching. Bones has already had one discussion with Smith regarding his observations. Smith has also given a short departmental presentation on an innovative classroom demonstration he developed. On this day,

Smith handles the class discussion moderately well. Afterward Smith and Bones confer about the class and both agree that the new technique has merit but could be improved by letting the students work in groups for a few minutes. Smith looks forward to trying this idea out. When Professor Bones writes the tenure evaluation for Smith, he has a file of observations from multiple sources from which to draw and is aware of Smith's progress and potential.

Clearly, this scenario is strikingly different from the first. For one thing, it follows the "best practices" outlined in this essay by having multiple modes of input and frequent observations. For another, it is rooted in an atmosphere of trust—the understanding that the visitor is there to help and not to judge.

Granted that the first example, representing the accountability model, is somewhat extreme, it has nevertheless been my experience that teaching and teaching evaluation in most departments at most schools tends more towards that end of the spectrum. In the accountability model, teaching is generally done in isolation with little outside feedback. Student evaluations, when performed, are confidential and read only by the professor until it is time for the annual evaluation. Student evaluations are generally the only means of assessment, so there is pressure to make sure they are high. If the professor is lucky enough to have had relatively high student evaluations, there is now a disincentive to try

new teaching techniques for fear of lowering those scores. Worse, if the professor has low scores, there is incentive to hide this fact and perhaps stop giving evaluations altogether.

On the other end of the spectrum, the assessment model, the emphasis is on improvement and self-correction, on collegiality and teaching creatively.

... Another comment I often hear is that students didn't like a particular course because it was "too rigorous". Yet again, when the numerous data sources are synthesized students "do not give lower ratings to difficult or challenging courses that require a heavy work load."³ Actually, just the opposite is true: a recent study of Western course evaluations found that students responded positively to challenging courses.⁴ Moreover, data synthesized from national studies indicate that "students' overall ratings of course quality and teaching effectiveness correlate positively with how much they actually learn in the course (as measured by their performance on standardized final exams)."⁵

In my personal experience I have seen many very rigorous yet popular professors. For example, one professor in the Geology Department teaches a series of courses that are highly quantitative in nature and require copious amounts of difficult homework. I regularly see crowds of students in the lounge, calculators in hand, conferring over their problem sets for this course—if not exactly "enjoying" themselves, they are definitely fully engaged. Yet in spite of their level of difficulty (and the fact that grade averages for these courses are at or below the averages for the department), students flock to these courses and give this professor among the highest course evaluations in the department. Clearly, factors other than grades and ease of coursework are at work here.

Putting this model into place can transform the atmosphere for peer evaluations from one of wariness and skepticism to one of trust, can transform nerve-wracking stress into meaningful hard work.

CONCLUSION

Most importantly, we must accept the fact that there are many kinds of teaching with different audiences and that even if the ideal system for improving teaching were in place, not all faculty would excel in all modes of teaching. Large non-major introductory courses require different teaching skills than those needed for mentoring graduate students. I am an outstanding undergraduate classroom teacher, but I am only moderately successful as a graduate mentor. Likewise, members of my department display a wide variety of strengths. One, an only adequate large lecture teacher, attracts graduate students and upper division undergraduates like bees to honey, involving them in an endless variety of independent research projects. Another professor is particularly gifted in teaching field classes. It cannot be stated forcefully enough that teaching is not one size fits all; indeed, for a department to have real strength, we need all types of teaching expertise. When making course assignments, the trick lies in playing to an instructor's strengths while at the same time trying to improve areas of weakness. Importantly, a teacher's varied dimensions need to be recognized and appreciated by those who make tenure decisions. Otherwise we run the risk of selecting teachers who score well on standard student evaluations (such as those in large undergraduate classes) and neglecting those whose strength lies in the role of mentor.

Finally, all this talk about a less stressful and more meaningful peer evaluation model is well and good, but where do we find the time for it? Although Geology's teaching evaluation system contains all of these components, and generally occurs in a positive and supportive atmosphere, it is by no means clockwork. Our classroom visits tend to cluster in the quarter before evaluations are due, and the winter and

spring quarters prior to tenure applications see a flurry of professors, sometimes two or three at a time, visiting other's classrooms. But when duty calls we respond, and those probationary faculty who are up for review can be sure of having at least three faculty visits in the quarter prior to their evaluation. Clearly, however, for peer evaluation techniques to change universally, the challenge would indeed be one of incorporating changes systematically. Like anything new, there would be transitional issues to address, and certainly not all instructors would wildly embrace the changes.⁶ Yet there is clearly a need to make peer evaluations more meaningful if they are to have continuing influence on hiring and tenure decisions.



1 Frye, Richard (February, 1999). *Assessment, Accountability, and Student Learning Outcomes* (Dialogue, Issue No. 2). Office of Institutional Assessment and Testing, Western Washington University, Bellingham, WA.

2 Cuseo, Joe (October, 2000). *Evaluating New-Student Seminars & Other First-Year Courses via Course-Evaluation Surveys: Research-Based Recommendations Regarding Instrument Construction & Administration, Data Analysis, Data Summary, & Reporting Results*. www.brevard.edu/fyc/fya/CuseoLink.htm

3 Ibid.

4 Frye, Richard (April, 2001). *Teaching Evaluation Summary: An Analysis of Fall, 2000 Course Evaluations*. (Unpublished.) Office of Institutional Assessment and Testing, Western Washington University, Bellingham, WA.

5 Ibid, Cuseo, Joe, et. al.

6 As an example: when Western changed its phone system in the late 1980's, one professor, so accustomed to the old ways, refused to direct dial long-distance calls, but rather continued to have the department secretary make such connections for him. He retired having never made such a call from his office.

Thor Hansen, Ph.D., is a professor and the chair of the Geology Department at Western Washington University



published by:

Office of Institutional Assessment and Testing

Dr. Joseph E. Trimble, Director
Gary R. McKinney, General Editor

For copies of **Dialogue**, OIAT technical reports, Focus Research Summaries, or InfoFacts, please contact Gary McKinney, Western Washington University, MS: 9010, Bellingham, WA 98225. Telephone: (360) 650-3409. FAX: (360) 650-6893. E-mail: garyr@cc.wvu.edu. TTY: (800) 833-6388. Join in discussions of Dialogue issues on the web at <http://www.wvu.edu/~dialogue>