

2.9 Teaching Evaluations

Mention teaching evaluations to most faculty members and you will get reactions ranging from horror to nausea: “Teaching evaluations are just a popularity contest.” or “Kids these days don’t want to be taught; they want to be entertained.” or “I know that I’m a good/competent teacher, and I certainly don’t need to be evaluated. I only distribute the damned forms because the administration makes me do it.”

In this section I am going to tell you that, yes, individual student teaching evaluations can be irritating. They can reflect nothing more than the students’ immaturity, or pique, or the sorry state of their digestion. But, taken as an aggregate, student teaching evaluations contain valuable information that will help you to become a better teacher.

The simple fact is that most people in most lines of work are regularly evaluated. The evaluation of physicians and lawyers is perhaps rather distant and painless. But it is done. It is a relatively recent development that tenured professors are evaluated—*by their students*—for teaching competence. How did this come about?

Part of the root of the teaching evaluation process comes from the student unrest—especially at U. C. Berkeley—in the 1960s. Students were partly distressed by free speech issues, and partly because they could never find their professors outside the classroom. In those days, professors were not required to hold office hours! The then chancellor at Berkeley, Clark Kerr, recalled wistfully that in the old days the duties of a chancellor were to provide “parking for the faculty, football for the alumni, and sex for the students.” He could see the tides of change coming, however, and he helped to usher them in.

Today, there are manifold reasons for teaching evaluations. University administrations are taking a hard look at tenured university faculty and demanding accountability (in some contexts, this sort of examination operates under the rubric “Total Quality Management” or TQM—see Section 2.14 for more on TQM); a system of post-tenure review is being put in place at more and more universities; taxpayers are pressuring universities to hire faculty who can teach (and who can speak English); if a department chair wants to get a faculty member tenured or promoted, then he must provide ample evidence that the candidate is a good teacher.

You may find that reading your teaching evaluations is a gut-wrenching experience. As an instance, James R. Martino of Johns Hopkins University compiled the following pairs of quotations (as cited in [MAR]). Each pair comes from different students evaluating the same instructor for the same course.

1. Does a good job of explaining new concepts and doing examples to demonstrate how it’s actually done.
2. Hopefully, he can explain the concepts at a more fundamental level; he assumes that some of the concepts are too trivial to be explained fully in class.

1. Great in all aspects of communication . . . is extremely intelligent when it comes to knowing [the] course.
2. The man is a horrible teacher. I'm sorry to say that he reflects the math department.

1. [The professor] is a giant. The man is humorous and very intelligent. He actually makes math interesting.
2. I don't go to lecture anymore because I never learn anything from the instructor. He seems to be talking about irrelevant topics and phrases questions in ways which are hard to understand. I see no benefit and reason to go to lecture except to obtain the homework assignment.

Not logically consistent, are they?⁵ Do these examples prove that teaching evaluations are worthless? That they are the puerile rantings of unformed minds? I don't think so. The extremes of opinion that you see here are no more bizarre than those that you would see in any public opinion survey. Any statistician will tell you that a sample of two is far too small for obtaining useful information.

When you read your teaching evaluations, don't let the outliers upset you. On the one hand, the isolated opinions may be those of thoughtful iconoclasts who really have something to say. You may indeed learn something from the musings of those independent thinkers. On the other hand, the outliers could be people who have been struggling all semester, who are too timid to get help, who are having personal problems, who have been ill, or who just like to complain. As award-winning teacher Tom Banchoff [BAN] says, "good teaching is not identical with perfect ratings." Instead, if you want to learn something about your teaching and improve it, look for trends in the student evaluations. If ten students say that you talk too softly or too rapidly, or if many of the evaluations say that you are a poor communicator, or if a plurality indicate that you cannot explain ideas at the students' level, then you should consider these criticisms carefully. Examine your conscience and determine whether you can learn something from your teaching evaluations and improve your technique.

If I have succeeded in convincing you that there may be something of value in teaching evaluations (taken as a group, not necessarily individually), now let me send chills up your spine. The article [AMR] describes the following experiment (which I present here in slightly simplified form), that was performed at an American college. Ten instructors were chosen from ten different departments. A random group of thirty undergraduates was assembled. Over the course of several days, each instructor gave a fifty minute lecture to this same group. At the end of each lecture, each student completed a standardized written teaching evaluation form for that instructor. The results were tabulated. Now each lecture was also filmed, and a ten second slice was taken from the beginning, from the middle, and from the end of each lecture. The sound was removed, and the thirty excerpts were spliced together in random order. The result was a five

⁵Note that we are unlikely to see pairs of teaching evaluations that differ over whether the professor wore plaid socks. What they will differ on is *subjective matters*, like clarity or level.

minute silent film showing randomly arranged ten second clips of each of the ten instructors.

Next, a hand-picked group of thirty sophomore women was assembled, and they viewed the five minute silent movie. Each young woman completed the same standardized written teaching evaluation form for each instructor—based only on what she saw in the silent movie. The results were tabulated. The startling fact is that the data from the first group of students—who saw the original lectures—correlated extraordinarily well with the data from the second group—who only saw the silent film. What conclusions may we draw from these observations?

It is easy to focus on the flaws of this experiment. You may well ask, "How can a ten second slice show what a professor can really do?" If the professor is in the middle of the proof of the fundamental theorem of calculus, or explaining the more technical aspects of Marx's theory of commodity valuation, then he will not show well in a ten second slice. Perhaps, as in the review of a good restaurant, one should view the professor over several different days. Keeping these limitations in mind, let us see what we might learn from the study.

I think there is a nugget of insight buried in this exercise, and it may be this: Young people do not generally have the intellectual equipment to determine whether their calculus teacher or their genetics teacher really knows his stuff, or is doing an optimal job, or knows how to communicate the material most cogently. But young people have a lifetime of experience evaluating body language and nonverbal communication. The second group of students (the sophomore women) *only had* body language and nonverbal communication on which to base their evaluations.⁶ Since their evaluations correlated very closely with those of the first group (the students who experienced *all* aspects of the lectures), one is tempted to conclude that the first group based its assessments also on the body language and nonverbal communication aspects of the lectures.

Cynics like to conclude from this study that students simply cannot evaluate your teaching. They are just reacting to the "vibes" that you give off. If you are hellbent on writing off teaching evaluations as worthless, this reasoning may appeal to you.

An informal study conducted at Cornell University [CEW] performed the following experiment. A single professor taught the same class, to similar audiences, two semesters in a row. The first time around the lecturer evinced a monotone, staid teaching style. The second time around, he showed great enthusiasm for his task, using many hand gestures and vocal diversity. The students reacted far more positively in the second semester than in the first; in fact in the second class they even gave the textbook more enthusiastic praise. What may we conclude from this information? One arguable conclusion is that it is almost impossible for a person to evaluate his own teaching. While the Cornell professor was being monotone and staid, he may also have discouraged student questions and done a poor job of answering those that he did receive. In the second semester, while being enthusiastic, he also may have interacted much more vigorously with the

⁶Sophomore women were chosen because, in the opinion of the experimenters, young women are more sensitive to nonverbal communication than are young men.

students. Even so, the experiment is worth considering. Studies show that students react more to the instructor's enthusiasm than to anything else. Perhaps the Cornell experiment only serves to affirm that finding.

Students are not scholars. But they are people. They probably react to human input more than they react to intellectual input. That is just a given, and it is not likely to change. Even so, what they have to say has value. If a freshman says, "This instructor doesn't know what he is talking about." then you have reason to be skeptical. If instead that freshman says, "This instructor cannot present the material in a manner that is accessible to freshmen." then perhaps he is giving you information that you could not obtain in other ways. A student certainly has something of value to say about whether he feels he is learning something; about whether the instructor can explain the material. You should give the matter careful thought so that you can put your teaching evaluations into context and learn something from them.

There are objective studies that suggest that student evaluations of a given instructor are consistent over time, that they correlate well with administrative and peer review, that (taken as a whole) they are independent of extraneous student characteristics, and that they correlate significantly with how much students actually learn (see [CEN], [COH], [FEL1], [FEL2], [HCM], and further references in [NAR]). The paper [KULM] offers statistical evidence that student teaching evaluations measure (*accurately*) (1) instructional skill, (2) respect and rapport, and (3) instructional organization (remember that these are *statistics*; they do not speak of any one particular teaching evaluation form, but of the overall content of the aggregate of teaching evaluation forms). This information may or may not appeal to you. But there it is.

I repeat: It is a noteworthy observation that student teaching evaluations correlate well with peer evaluation (see [DAV]). They do *not*, however, correlate well with self-evaluation. On days when you are sitting around badmouthing your teaching evaluations as immature and worthless, bear this thought in mind.

In my experience, teaching evaluations (again taken as a group) have empirical value. Obviously, what information you glean from your teaching evaluations is up to you. If a significant number of your students say that you make too many mistakes in class, or you ridicule people who ask questions, or you don't prepare, or your tests are unfair, then I think it's a cop-out to just claim that they are reacting to your vibes. After all, this is your audience giving you feedback. Why not try to learn from it?

On the other hand, if a large number of your students say that you are inspiring, well-informed, creative, and an excellent pedant, then don't just say, "Aw shucks" and forget about it. This is positive reinforcement. You must be doing something right!

Bear in mind that the evaluation of teaching need not take place only at the end of the semester. Mid-semester evaluations can be extremely useful. They tell you how the course is going, whether the students perceive that they are learning anything, what problems may have arisen.⁷ It is one thing to get your

⁷Note that, at many institutions, midterm teaching evaluations can be just between you and the students. Nobody else need see them after you have collected them.

evaluations at the end of the term and say to yourself, "Boy, there's something I did wrong. I'll try to fix it next time." It is quite another to realize after only half the course has gone by that corrections must be made *while the course is in progress*. For then you have an opportunity to make things right, and your end-of-term evaluations should improve as a result. The book [GOL] offers a number of insights on course evaluation techniques.

Some math faculty are uncomfortable with the standardized teaching evaluation forms that the university provides. Usually these forms ask the student to answer twenty or so questions with a rating of "1" to "5" to indicate "Poor", "Fair", "Good", "Very Good", "Excellent". While some of the questions are about reasonable issues (such as "Was the text useful?", "Was the instructor prepared?", etc.), others are rather vague ("Evaluation of the Instructor Overall" and "Evaluation of the Course Overall"). It is a fact that the dean is a busy person and wants a quick and dirty estimation of the teaching of any given faculty member. Often the numerical responses to these last two questions can give such a gauge, but it is not a gauge with which the faculty member being examined is comfortable.

You may feel, with some justification, that the evaluation process described in the last paragraph does not give the mathematics professor an adequate chance to show what he can do. If this describes you, then I encourage you to discuss the matter with your colleagues, your chairman, and your dean. The deans I have met are not averse to individual departments developing their own methods of teaching evaluation. Possibilities to consider are

- Videotaping of lectures
- Peer review
- Self-evaluation
- Consultation with other faculty experts
- Exit interviews by a third party professional

Let me give a brief explanation of each of these techniques.

(1) Few things are more honest, and sometimes demoralizing, than a videotape of a lecture. The videotape will show all your awkward mannerisms, your squeaky voice, your dandruff, your strange pauses and facial expressions. A lecture that you thought was ethereal will come across as peculiar when first viewed on videotape.

If you agree to have yourself videotaped, then view the tape with an experienced faculty member who can point out both what is good and what needs improvement. You will need some help in keeping the matter in perspective.

(2) Peer review consists of having some of one's colleagues attend some class lectures, and perhaps also review some class materials that were prepared, by the person being reviewed. Most intelligent people feel comfortable receiving cogent remarks from someone whom they respect and admire. Why not learn

about teaching from a fellow mathematician?

(3) Self-evaluation might consist of the candidate preparing a "Teaching Portfolio". This portfolio would contain an enunciation of the candidate's overall teaching goals, plus a list of particular goals connected with particular courses that the candidate is going to teach. A mentor would review the portfolio regularly with the candidate, and help him assess whether he is achieving his objectives.

(4) Some people are shy, or uncomfortable in front of groups, or are poor speakers. Notable success has been had by having such individuals work with experts from the Communications Studies or Speech Department. While it may be uncomfortable to have a fellow mathematician tell the candidate that his teaching is inadequate, it may be more natural (akin to going to a fitness trainer or a podiatrist) to have the candidate consult with a well-meaning faculty member from another department. And I can tell you that this is a method that works in practice.

(5) I rather like this last method for evaluating teaching, but it is expensive both in terms of time and in terms of money. The idea is that a professional interviewer, perhaps someone with a background in psychology, will interview each student at the end of the course. The interview can be brief—perhaps ten minutes. But the interviewer can ask questions that will draw out the student's concerns. He can also zero in on important points that the student is trying to articulate and help him to develop them. At the end of the interview process, the interviewer will write up an in-depth report on the class, and the instructor, in question.

Once you start thinking creatively about ways to evaluate teaching, you will certainly develop ideas of your own. Bear in mind, as you do so, that the dean has an affection for "Evaluation of the Course Overall" and "Evaluation of the Instructor Overall" because these simple questions give him two numbers. He can quickly assess whether a given candidate cuts the mustard or not. When you devise alternative assessment techniques, be sure that each one results in useful and accurate advice for the teacher being examined and also in a *quick and incisive* take on the candidate's teaching abilities. The dean does not have the time to view videos, or read long position papers. He may not insist on a *number*, but he needs the evaluation to be of the nutshell variety.

2.10 Exams

In this section I will discuss how to compose an exam, how to formulate questions on an exam, how to judge the length of an exam, how to grade the exam, and so forth.