

About End-Of-Course Evaluations

Samuel A. Rebelsky

Version of 12 May 2006

This document can be found on the World Wide Web at

<http://www.cs.grinnell.edu/~rebelsky/about-eoc.html> or

<http://www.cs.grinnell.edu/~rebelsky/about-eoc.pdf>.

Summary: At the end of each semester, you will find that most Grinnell faculty ask you to fill out at least one end-of-course evaluation, perhaps a standardized form, perhaps a form specialized to the course, perhaps both. Unfortunately, faculty do not always discuss the context or purpose of those forms. This document attempts to provide some of that background. It also reflects on some related issues, such as merit raises and how Grinnell evaluates faculty.

Contents:

- A Warning
- Why Have End-of-Course Evaluations
- End-of-Course Evaluations at Grinnell
 - Merit Raises
 - Evaluating Teaching, Phase 1
 - Evaluating Teaching, Phase 2
 - What Does This Mean?
- Related Issues
 - Do We Need Merit Raises?
 - Alternative Mechanisms for Evaluating Faculty
 - Other Perspectives
- Evaluating My Courses
 - The Subject Matter of The Course

A Warning

Some students (and some faculty members) seem to be under the impression that end-of-course evaluations are retyped. Tutorial evaluations are, in most cases. Other end-of-course evaluations are not. That means that if you have recognizable handwriting, your faculty member can probably tell you wrote the evaluation.

Should that matter? It should not. I believe most faculty are responsible enough that they do not let either negative or positive comments impact how they relate to students.

Why Have End-of-Course Evaluations

There are two reasons that faculty members ask students to fill out end-of-course evaluation forms. The first is that faculty use the forms to reflect on our courses. While we have a perspective on what worked and what failed to work in courses, we also expect to gain insight from our students' comments.

At some institutions, like Grinnell, in which the quality of teaching is emphasized, we use end-of-course evaluations for a second reason: as a mechanism for evaluating teaching. At such schools, faculty members' forms are reviewed and used in raise and promotion decisions.

Note that these two purposes are somewhat in conflict. If our primary goal is to improve our courses, we want you to emphasize the things that went wrong (with some notes as to what went right so that we keep doing those things). If our primary goal is to stay at the institution and get paid well for doing so, we want you to write good things.

How do we resolve this conflict? The best way is to give you two forms. It is also important that you know the purpose of each form.

End-of-Course Evaluations at Grinnell

Grinnell now has a standard end-of-course form which is used primarily for promotion decisions. How did we get to this state? End-of-course evaluations have a long and twisted history at Grinnell.

Merit Raises

A few years before I came to Grinnell, the faculty (with some prompting from the board of trustees) voted to institute a merit raise system. That system was intended to emphasize teaching, but also to incorporate scholarship and service. Although the percentages vary from year to year, the basic idea was to make teaching about 50% of the merit score, scholarship about 33%, and service about 17%.

How are these components evaluated? Each faculty member writes an annual *Faculty Activities Report* (FAR). Department chairs read those reports and some associated documents and write recommendations to the budget committee. In the past, these FARs were reviewed annually.

Now, for tenured faculty (and, in essence, for junior faculty), every three years, each faculty member writes three reflective statements, one on teaching, one on scholarship, and one on service. Again, the department chair reviews and comments on these documents.

Afterwards, the Faculty Budget Committee reads a packet for each faculty member--including FARs, curriculum vitae (CV), statements, and chair's letter--and develops a numeric ranking between 0 and 5 for each of the three components. (A future version of this document will include the claimed "rules of thumb" used in determining the numbers.) A weighted sum of those components is then taken and rounded to the nearest integer. These merit scores are translated into raises. In the past few years, each point of merit has resulted in \$500 of raise. Most faculty end up with a merit score of 2 or 3.

Who is on the Budget committee? The chair of each division and the chair of the faculty. Do they enjoy spending their winter break reading all of these documents for approximately one-third of the faculty? More than they enjoyed reading them from all the faculty, I'm sure.

Evaluating Teaching, Phase 1

Since teaching is factored into raises, the college needs a way to evaluate such merit. When the merit raises were instituted, each department was asked to develop an end-of-course evaluation form.

That period was “interesting”, to say the least. Almost every form was different. Some forms were numeric, some were not. (As you might guess, the Math/CS form was not numeric.) Forms that were numeric used different scales. How did the budget committee come up with a number for each faculty member? With difficulty. Did anything else help? Each faculty member was asked to write a one-page reflection on each course they taught. (I think the reflection was helpful for me. I am unsure whether anyone else read them or whether my readers found the reflections helpful.) We no longer do so.

Evaluating Teaching, Phase 2

A few years ago, the Executive Council reacted to the related problems of (1) the dual nature of end-of-course evaluations and (2) the inconsistency of the end-of-course data and suggested that Grinnell come up with a standardized end-of-course form whose primary purpose was evaluative. Some very smart people (who I respect very much) worked hard to develop a form that could be used in many courses and to “validate” the form (that is, to show that it is consistent, not to show that the score it gives really says anything about our teaching).

The faculty were then asked to vote to have the form used in the raise system. Someone asked “How will you deal with confidence intervals?” (For those of you who do not understand the question, the result from a survey like this is consistent, but only somewhat consistent. Hence, a score of, say 5.5, could really indicate a real rating in a range, say between 5.3 and 5.7. If two confidence intervals overlap, it is inappropriate to say that the two values are significantly different. (Well, that is my understanding, which is not completely informed.)) Responses from the budget committee varied. One was something like “I don’t know what confidence intervals are. I don’t really understand math or statistics. But I can tell what the numbers mean.” The faculty, sensibly, voted not to use the end-of-course forms for annual merit raise evaluation. Since the every-three-year process was introduced, we have not revisited that vote.

Some comments members of the budget committee have made since then lead me to believe that we made the right decision. For example, one member of the committee, noting that most of the scores are between 5 and 6 on a six point scale, referred to “the Lake Wobegon Effect”. As you may know, in Lake Wobegon, “all the children are above average”. However, if 5 and 6 are “I mostly agree” and “I completely agree” to “I learned a lot in the course”, I think you could reasonably expect that most Grinnell courses would earn those numbers.

What Does This Mean?

We have a standardized form. The faculty voted that the budget committee cannot use it. So, how is the form used and how does the budget committee evaluate teaching? The answers are not encouraging.

Right now, the primary use of the standardized end-of-course form is for tenure and promotion decisions. The hope is that we will see evolution in scores over a faculty member’s time at Grinnell, which can then serve as a positive recommendation for tenure. For faculty not up for tenure, promotion, or renewal, the data are currently probably useful only in providing comparative data for those who are up for tenure,

promotion, or renewal.

The Office of the Academic Dean claims that we can use the forms for development, but their use for development is limited to barometer-type notes.

Should you still take the forms seriously? You should certainly take them seriously for junior faculty, since the ratings can have a significant effect on the careers of junior faculty. It is probably a good idea if you take them seriously for all faculty, as we need good comparative data.

Given that the budget committee has even less data than it did before this endeavor, how do they currently evaluate teaching? Well ... for a number of years, the committees decided that because they lacked sufficient data, they had to give each faculty member the same score. give every faculty member the same teaching score. Does that mean that quality of teaching was not factored into raises? Yes. Is that bad? If we have merit raises, it is bad.

The current Faculty Budget Committee is doing a much better job of setting criteria for determining scores. (I will admit that I do not necessarily agree with the criteria, but I appreciate that there are criteria.)

Related Issues

Do We Need Merit Raises?

Of course, there are faculty at Grinnell who believe that we don't really need merit raises. (I'm usually on that side of the fence, even though I usually get high merit scores.) Why? Some believe that merit raises encourage faculty to compete with each other instead of working cooperatively. Others believe that merit raises encourage faculty to work on things that get them bigger raises rather than things that are important. (For example, it is likely that my work to encourage more women to major in CS fits in neither scholarship nor service nor teaching. Similarly, my significant scholarly service for international organizations seems (e.g., reviewing) doesn't seem to count much. I don't care. Others might.) Still others believe that the process is insanely time-consuming for very little benefit. I believe that a system that tells half the faculty "you're worse than average" brings little benefit.

However, in 2002 or 2003, the faculty voted to continue with merit raises. So, I guess we're stuck with them.

Alternative Mechanisms for Evaluating Faculty

I will admit that I think end-of-course evaluations are a particularly bad mechanism for evaluating faculty comparatively. Why? Well, it is fairly easy to influence the results. It's also a silly time to ask students to evaluate courses: at the time they're most stressed and perhaps least reflective. There are currently no measures of their accuracy (in terms of measuring quality of teaching). Numbers are often hard to interpret. Students use the forms in interesting ways. (For example, I've had a number of students give me low rankings with a note that "I would have given him the highest ranking if he hadn't given us so much work!") Unfortunately, those comments don't appear when the data are communicated to others.)

A simple change I'd make would be to ask for the evaluations a year or so after the course is over. After that time, the students are more likely to have seen the impact of what they've learned and are more likely to be able to reflect on the course as a whole. The Dean also regularly sends out a survey to alums about teaching when we're deciding on tenure and promotion, and I feel that that survey should be used more regularly.

I've also suggested that we get comparative data directly, by asking students to rank all their faculty at the end of each year. Another possibility would be to ask each student to list the best course or faculty member they had over the past year. Both of these systems have some flaws and need the details worked out, but I think they'd provide better data than our current system.

If we had lots and lots of money (hey, this *is* Grinnell), I'd hire two or three people whose profession is educational evaluation and ask them to visit each faculty member's classes for a week each semester and to do a one-hour meeting with each faculty member. I expect that such visits would provide excellent data and also give us some success at development.

Other Perspectives

John Stone hates the college's reliance on "just numbers" in evaluation. If you search around his Web site, you're likely to find a nice essay on the topic at <http://www.cs.grinnell.edu/~stone/misc/scales.html>. If you talk to Mr. Stone more, you'll also learn things about the loss of information going from textual answers to numerical answers.

David Lopatto dislikes the college's piecemeal approach to raises and other forms of funding. He has written a very nice essay on a *humanistic* approach to evaluating faculty. I suppose if you asked him nicely, he might give you a copy.

Evaluating My Courses

My primary concern is making my courses better. I'd prefer that you put lots and lots of helpful comments wherever you find it appropriate to put those comments (on the course-specific form, on the standardized form, in an electronic mail message).

Do I care how you rank me on the standardized form? Certainly. Do I want you to be honest? Even more certainly. Think carefully about the different questions and give whatever answer you think best.

The Subject Matter of The Course

All of the questions on the standardized form as you about "the subject matter of the course". I used to give a long lecture about that topic. For awhile, I stopped giving the lecture because I worried that it biased the results, but that also seemed to bias results. I now give a shorter lecture.
