


# Inspired College Teaching

A Career-Long Resource for  
Professional Growth

Maryellen Weimer

 **JOSSEY-BASS**  
A Wiley Imprint  
[www.josseybass.com](http://www.josseybass.com)

2010

3

## Rewriting the End-of-Course Ratings Story

What teachers learn from the reflection proposed in Chapter Two needs to be validated with feedback from students and colleagues. Faculty do get feedback from students. It's collected via surveys, mostly administered at the end of the course via processes now mandated by virtually all colleges and universities. The results are taken seriously. They inform the personnel decision-making process, at some places influencing who gets hired, tenured, promoted, and, in the case of part-timers, who gets to continue. Rating results are regularly included in annual performance evaluations, at some places determining the size of merit increases, although at most places even stellar ratings still garner modest pay increases. But how these summative assessments inform teachers' understandings of themselves and their efforts to affect learning is a different story.

The use of ratings as measures of instructional effectiveness does mark progress in the long journey to gain respect for teaching. Prior to ratings, teaching quality was assessed by hearsay, what a department head might overhear in the hall outside a classroom, or what colleagues might glean from student comments during an office visit. The now systematic assessment of instruction has been accompanied by a massive empirical enterprise, causing one prominent and prolific ratings researcher to observe that "probably students' evaluations of teaching effectiveness are the most thoroughly

studied of all forms of personnel evaluation, and one of the best in terms of being supported by empirical research” (Marsh, 1987, p. 254). Still more research on ratings has been completed since Marsh made that statement. Virtually anything and everything anyone might want to know about ratings has been studied, including whether distribution of chocolates positively affects results (Youmans and Jee, 2007). Within the research community there is widespread agreement that ratings can be reliable and valid measures of instructional effectiveness.

It should be a story with a happy ending, but more often than not the use of ratings ends up being a tale of woe. Despite research that clearly describes how to generate credible data, at many institutions, ratings are used in ways that compromise their effectiveness. Instruments end up being camels created by committees. They contain a hodgepodge of items, many of which are unrelated to those components of instruction that research has linked to learning outcomes. Results from very small classes are considered, despite cautions about reliability when the number of completed ratings is low. Faculty are ranked and raises determined when very small numerical differences separate individuals, even though researchers repeatedly warn that small differences in scores should not be equated with observable differences in classroom performance. Perhaps this woeful negligence of research could be excused if the findings only appeared in difficult-to-decipher journals, but ratings research has been integrated, distilled, and described specifically in terms of good policies and practices, making this story not only sad but ironic as well. There are many sources that could be recommended here. I regularly refer people to Braskamp and Ory (1994) and Centra (1993)—both thorough, well-organized, and easy-to-read treatments by researchers with impeccable credentials. More recent is a comprehensive volume by Arreola (2007).

For many faculty, the end-of-course rating experience has not confirmed what they believe about themselves as teachers or con-

tributed constructively to their growth and development. The results don’t tell them what they need to know, or they don’t make any sense, which causes frustration, sometimes anger, even despair. If the instruments aren’t valid, reliable measures, then the results may be bogus, which explains why they don’t make sense. Furthermore, if questionable results are informing personnel decision making, then the reasons for faculty anger and frustration are all the more justified.

Like a lot of administrators, many faculty (dare I say most) are woefully uninformed about ratings. Certain beliefs—that ratings measure popularity, for one—justify ignoring the results, which does diffuse the anger and frustration. But ratings aren’t measures of instructor popularity (Aleamoni, 1999, Myth 3). That’s a myth, like too many other beliefs about ratings. Because they are so uninformed, faculty perpetuate these myths, half-truths, and urban legends when ratings are discussed over lunch, in the mail-room, or in more formal venues like department meetings or the faculty senate.

The saddest part of the story involves how much teaching needs feedback from learners. As important as the view from the front of the room is, the experience in the desk is just as important, and the front of the room looks different when viewed from the desk. Students are in a good position to offer feedback. They are there for the course from start to finish; their experience is first-hand and fresh. They can say better than anyone else whether the course design and teacher actions motivated and expedited their learning. The potential for faculty to grow and develop as teachers based on feedback provided by students is enormous.

That’s the point where this chapter begins. The chapter rewrites the ratings story by showing faculty how institutionally mandated ratings can be used to fuel growth and development for teachers. Cynical faculty will think that support for the ideas in this chapter is provided covertly by administrators who want faculty perspectives on ratings changed. Not so. This chapter deals honestly with

rating realities, including ways they are misused and ways they are used that limit or become barriers to growth. It starts by proposing what lessons can be learned from end-of-course ratings and then offers a set of conclusions not justified by ratings results. It explores how faculty should deal with less than impressive ratings, really negative feedback, and how they need to avoid ratings “addictions.” It characterizes successful (as in positive and constructive) conversations with administrators about ratings and concludes by summarizing what actions to take and not to take based on end-of-course rating feedback. All together, it’s a discussion that sets the stage for Chapter Four, which showcases even more ways that feedback from students can be used to expand and deepen the understandings and insights that emerge from reflective practice.

### Lessons to Be Learned from End-of-Course Ratings

Even the most basic understanding of ratings research can help faculty understand this student feedback mechanism and the how the results it generates can be used to promote growth. By the same token, a little knowledge can be a dangerous thing. The random reading of a few isolated studies will not develop a good working knowledge of the research on ratings. Many studies address the same issue. Given the nature of social science research, seldom do all the studies agree. So, it’s possible to read one study that comes to a conclusion not supported by the bulk of the research.

Fortunately, a number of sources provide excellent, accessible, and nontechnical summaries of the research. (See Cashin, 1995; Wachtel, 1998; Aleamoni, 1999; Hobson and Talbot, 2001; Arreola, 2007, Chapter Twelve; and Felder and Brent, 2008, for examples.) These summaries can be used to develop the kind of working knowledge faculty need to better understand ratings results. In my perfect university, a copy of one of these summaries would be stapled to every set of ratings distributed. Untenured

faculty would be required to pass a quiz on the summaries before receiving ratings results. At my ideal university, the myths, half-truths, and old wives’ tales would quietly fade way as faculty and administrators learn to draw accurate conclusions and take appropriate actions based on ratings results. But this is a whole new ratings story, not a rewrite of the current one.

A basic knowledge of ratings research uncovers their most salient feature: end-of-course ratings offer a “big picture” view of instruction. They are summative, meaning they present data that summarize. They provide overview information: “How does this course or this instructor compare with other courses or instructors?” The course and instructor are judged but from several miles up, which means the view of how the course went is more general than specific. These comprehensive judgments of the course and its instructor are one of the reasons this feedback is so potent and why faculty react to it so viscerally.

Besides providing this general, overall assessment, summative data are a good source of information on trends, provided they are looked at across time. Ratings from different kinds of courses taught can be compared, revealing what courses and content generate the highest and lowest ratings. A collection of ratings from the same course can become a database against which current ratings can be compared. Ratings do tend to be stable across mid-career, but an occasional section may rate significantly above or below the norm. A single or even a couple of sections so rated do not constitute a trend; four or five sections across several different semesters probably do. The last section of this chapter and all of Chapter Four address what can be done in this situation. Noting what happens to ratings the first time through a course or when innovations are tried means fewer rating results surprises. Then teachers expect that ratings will be lower the first time through a course, or that a course revamp normally causes a two-semester dip in ratings.

Next there are lessons to be learned about comparisons. At many institutions, ratings results are tabulated so that faculty can see where they rank compared with colleagues. Even at places where this is not done formally, most faculty cannot resist at least a few informal comparisons. “Above average” scores can be gratifying, “average” scores comforting and not very impressive scores motivating. Two caveats do limit conclusions justifiably drawn from comparisons. First, despite extensive research and widespread use, standards specifying acceptable rating levels are all but nonexistent. This makes it very difficult to know whether an overall rating is good, despite how it compares with the ratings of colleagues. Couple the absence of standards with the fact that on many instruments the midpoint of the scale is well below the average score at an institution. This means that a score above the midpoint can be below average, but that still leaves the question of competence (or the lack of it) unanswered. These variables make comparisons with colleagues of little value beyond ego building or bashing.

Second, as already mentioned, a small difference between scores does not equate with observable differences in teaching performance. End-of-course ratings instruments do not differentiate meaningfully between almost equal teachers. They do separate the very good from the not so very good, but for the big group in the middle, ratings do not come with bragging rights. A teacher with a 6.12 overall score has no business gloating because a colleague teaching the same course has scored only 6.00.

In sum, three lessons can be learned from end-of-course ratings. First, ratings provide a broad, overall view of the course and its instruction. Second, they can be looked at for trends, provided data from previous courses is kept and can be used for comparative purposes. Third, they offer feedback that allows instructors to make comparisons: with those teaching the same course, in the same department, across the institution, or with faculty teaching the same course or in the same discipline at different institutions,

provided a standardized instrument is being used. Those comparisons may provide some emotional support. They may motivate change, but comparisons do not tell teachers anything about how to grow and develop teaching skills. Unfortunately, faculty draw a lot of other unwarranted conclusions from end-of-course rating results beyond these three legitimate lessons.

### What Can't Be Learned from Summative Rating Results

Generally, end-of-course ratings do not provide much in the way of details. Typically items on these instruments describe aspects of teaching abstractly. For example, an item may inquire whether the instructor is organized. Organization is not a concrete entity. You cannot go to the teaching excellence center, get some organization, and bring it back to class. Rather, the presence or absence of organization is conveyed by behaviors (some verbal, some nonverbal) that have come to be associated with organization. We covered this relationship between abstract entities and the behaviors indicative of them in Chapter Two. With ratings, students decide on a high or low score based on the presence or absence of behaviors they associate with organization. From summative results based on abstract aspects of teaching (like enthusiasm, fairness, concern for students), instructors cannot discern what policies, practices, or behaviors students used to determine the rating. This may be why research has shown that the regular delivery of this feedback to teachers does not improve instruction as measured by subsequent scores (Kember, Leung, and Kwan, 2002).

The absence of specifics makes implementing changes based on summative feedback alone pretty much a hit-or-miss affair, especially when teachers haven't done the kind of reflection proposed in Chapter Two and end-of-course ratings are their main source of feedback. Targeting what to change is much easier when the behaviors, policies, and practices students are equating with the

abstraction are known. There are “low-inference” rating instruments that make those connections; they are discussed with examples provided in Chapter Four.

The spotty record of improvements based on end-of-course rating feedback is further compromised by how teachers respond to abstractions that describe personal attributes: organized, enthusiastic, fair, knowledgeable. If a teacher wants to improve scores on the organization item, but fails to get to the level of specifics, then that teacher ends up saying to himself or herself, “I have to be more organized.” We’re back to another point made previously. To be more organized, that teacher must change the way she is. That is much more difficult than changing what is done. “I’m going to post an outline at the beginning of class.” “I’m going to devote the last five minutes of every class to activities that summarize and structure content.”

Next, end-of-course summative feedback provides an incomplete picture of student learning. At this juncture, one of the myths about ratings really confuses the issue. Many faculty believe that easy courses (in which students aren’t challenged, don’t work especially hard, and don’t learn a lot but get good grades) get high ratings. For example in 1987, 72 percent of faculty at a research university reported believing that course difficulty biased ratings, 68 percent held the same belief about grading leniency, and 60 percent thought that courses that required more work received lower ratings (Marsh, 1987). In a 2002 study (Sojka, Gupta, and Deeter-Schmelz), 53 percent of faculty agreed or strongly agreed that students give better ratings to instructors who teach less demanding courses. I suppose that’s progress, but those beliefs about a leniency grading connection have been repudiated by research and lots of it. Take research by Centra (2003) and Marsh and Roche (2000), for example. Both of these are large, comprehensive studies—Centra analyzed data from 50,000 individual courses. Or, look at some smaller individual studies like Martin, Hands, Lancaster, Trytten, and Murphy (2008), who found that

students preferred harder courses that challenged them; Dee (2007), who documented that students don’t give low evaluations to hard courses; or Jansen and Bruinsma (2005), who reported that students had a higher positive perception of faculty who taught difficult classes than those who taught easy ones. Bottom line: Faculty with high ratings do not necessarily have students who are learning *less* because their course content is easy, they don’t require much work and give high grades. But that doesn’t establish whether those with high ratings have students who are learning *more*.

What the research offers here is a large family of studies known as multisection validity studies. (Cohen’s 1981 meta-analysis is the landmark review of this work.) In these sections of the same course, the content was held constant. Students heard the same lectures, they used the same textbook, they had the same homework assignments, they completed the same labs, they were graded on the same grading scale, and they all took the same comprehensive final. (This explains why math and science courses have been studied most often in this line of research.) What the studies showed, with some consistency, was that those sections for which students rated the instructor highly were the sections where students scored higher on the comprehensive final. To the extent then that a comprehensive final measures learning, students learned more in those sections taught by highly rated instructors.

However, there are caveats that limit what should be concluded about the ratings-learning connection. First, multisection validity studies establish correlational, not causal links. The results assume that with all other variables held constant, it is the teacher who made a difference in how much students learned. Chances are good that the finding is true, but that is different than being absolutely certain that the relationship is cause and effect. However, that is not the important caveat.

Comprehensive finals in most courses measure a particular kind of learning: content knowledge. Usually, they do not measure whether students learned anything about working in groups, even

in lab courses where they worked with others throughout the course. Usually they do not measure whether students have an explicit understanding of problem-solving processes, even in courses where students solved problems daily in class and on homework. Usually they do not measure the development and appreciation of learning skills, even in courses that remediate absent skills. True, finals could test outcomes like these, but most do not. The pertinent question, then, is how representative the final is of all learning that occurred in the course. Multisection validity studies that link learning to ratings generally say something about content learning but nothing much at all about other kinds of learning and whether the teacher influenced learning in those domains.

Furthermore, a comprehensive final provides little feedback on how well students learned something, how long they will retain it, and whether they can apply it subsequently. In fact the evidence here is rather bleak, as most college teachers' experiences in the classroom confirm. Shortly after exams, indeed after courses, students quickly forget content knowledge learned for exams, even when they are taking upper-division, major courses in which ostensibly they have interest in the content and professional reasons to retain it. (For two compelling examples, see Bacon and Stewart, 2006, and McIntyre and Munson, 2008).

To summarize this long discussion: from end-of-course ratings, teachers can learn little about the nature or lasting effects of student learning experiences in a course. Higher ratings are likely indicative of more content learning than are lower ratings, but that kind of conclusion is like pretending a half-written chapter tells the whole story.

Finally, end-of-course summative ratings are not definitive measures of a teacher's competence or worth. They offer a valuable view of the teaching and course, but they do not provide the be-all, end-all perspective. Researchers repeatedly caution against giving ratings too much credence. The director of the Kansas State IDEA (Individual Development and Educational Assessment) Center, an

organization that distributes one of the best-known and most widely respected student rating instruments, recommends that ratings account for somewhere between 30 and 50 percent of an overall evaluation of teaching effectiveness (Pallett, 2006). Rating results do not contain a detailed description of the effects of teaching on learning, and they say nothing about a teacher's intrinsic worth. Just like students, teachers too must learn to separate the person from the performance. Ratings results are about what happened in one class with one group of students at one point in a career. They do not measure capabilities or potential. Chapter Four describes a variety of ways to enlarge, enrich, and deepen the understanding of how an individual's teaching affects learning. Developing that larger view enables faculty to see that the story told by summative ratings results is decidedly not the whole or best part of the story.

### Dealing with Less Than Impressive Ratings

Most faculty overreact to negative feedback on ratings. They draw comprehensive conclusions laced together with emotion. Overall scores a few points down for a course or semester do not spell the end of teaching careers. Really low scores from one or two students do not mean that the teacher has no redeeming social value. Even a course with low ratings does not warrant crisis mental health counseling for the teacher. The first question to ask when end-of-course ratings are assuming these larger-than-life dimensions is whether they are really all that unimpressive. If they are regularly below institutional norms, that's one thing. If they are not as good as the instructor expected (or hoped), that's quite another thing.

Some teachers do have a history of lackluster ratings. Most are motivated (whether the pressure is external or internal) to do something when they first get those scores. But when they make changes based on summative data alone (as discussed in the previous section), and when their efforts garner no change in ratings,

there's usually some attempt to explain away the results. The feedback is from students—students who don't study, don't come to class, don't care, and don't take more than 45 seconds to fill out the form. They wouldn't know a good teacher if they had one. Or, the low ratings are the result of a crummy rating system. The instrument doesn't measure the really important aspects of teaching—those that matter to this teacher or with this content. No background variables (like class size, time of day, degree of difficulty) are considered. And the whole process favors the popular teachers, the ones who entertain and teach without standards.

If your scores are consistently low and you're blaming the instrument, I offer this challenge: if the end-of-course data presents a skewed or inaccurate portrait of your teaching, then collect other data that establish a more accurate picture of what's happening in your class. Chapter Four offers many alternatives. Blaming the instrument is more justified and looks less like an excuse when you have data that disagree with those results generated by the institution's rating instrument.

Many teachers do anguish over low scores (a lot of time more than they need to) because they don't know what to do about them. Happily this part of the ratings story can be rewritten. The situation may feel desperate, but it is far from hopeless. Here's what a teacher with a history (it can be long or short) of not very impressive ratings can do (in a nutshell; indeed the rest of the book offers an entire nut tree of options). First, you need diagnostic, descriptive details, not large, overall comparisons. So, if scores on organization are low, you need to solicit feedback on those aspects of instructional practice that relate to organization:

Do students know with some degree of confidence what to write in their notes when listening to a lecture?

How well do you establish direction for the day at the beginning of the period?

Do you provide clear signs indicating movement from one topic to the next? By doing what?

Do students see connections between what happens in class one day and the next?

Do they understand how what they are reading in the text relates to what's happening in class?

Next, after collecting and looking at the feedback (much more on this in Chapter Four), you need to identify a reasonable number of changes to implement (see Chapter Six for more information on how this works). As part of this process, explain what's happening to the class. "I want to make some aspects of my instruction easier for you to follow. Here's what I'm going to do and I will need feedback as to whether these changes make it easier for you to learn the material." After a period of implementation, feedback should be solicited and the results analyzed and discussed. There may be a need for further revision and adjustment. The process of soliciting, discussing, and responding to feedback improves ratings in and of itself, as is documented by a variety of studies (Marsh and Roche, 1993, for example).

Too often a history of less than impressive ratings ends up paralyzing teachers. The results are depressing and can rob teachers of their sense of efficacy and take away the joy that can be experienced in the classroom. Teachers feel badly about themselves, are disgruntled with students, and lose their commitment to the profession. Like any other self-defeating behavior, there comes a point when it is time to draw a line in the sand and determine whether the journey in this direction has gone far enough. Ratings can be improved, even ratings that have stayed the same or been unimpressive for some time. The place to begin is with a clear recognition that end-of-course ratings are the first word on instructional effectiveness. They are not the last word or anything like the whole story.

## Constructive Reactions to Those Really Negative Comments

I wonder if the most damaging part of the end-of-course rating story aren't those very negative responses to the regularly included open-ended queries. "What did you like most and least about the course?" "What should the instructor do to improve the course?" I don't know any instructor who hasn't received at least some truly negative, truly destructive, truly awful comments. Most have received some wonderful responses as well, but I don't think the goodwill they engender lingers nearly as long as the negative auras of those very hurtful statements. Hartz (2008, p. 4) describes personal and pedagogical consequences. Of the personal, he writes, "You wonder in the middle of a sleepless night where you went wrong. You dream of retirement. But the *pedagogical* consequences are even more dramatic. Such comments take aim at the very soul of teaching. They haunt you during the teaching day—make you hesitate to take risks in your interactions with students. You pull back from challenging them the way they need to be challenged if they are to learn how to think analytically and critically."

I once worked with a faculty member who was twenty years into his teaching career. He wanted to make some changes, and as we discussed what those might be I asked about feedback from students. He pulled a tattered file folder from his desk drawer and handed it to me. The first item in the file was a machine-scorable form with the following comment scrawled across the bottom: "Instructor should use his lectures for toilet paper." I didn't recognize the form and asked where it came from. It hailed from a previous job. Even more amazing, it was written during his second year of teaching and was made about a new course he was teaching for the very first time! What a burden to carry across the teaching career. I don't know a lot of faculty members who collect negative comments in file folders. I do know many who harbor them in their hearts.

At some level, those involved with the rating systems—faculty and administrators—must own at least some of the problems created when questions allow students to comment where they will. Doesn't the quality of a question play a role in determining the caliber of the response? Why in the world are students being asked what they liked most and least about an instructor or course? Since when did "like" become a relevant criteria for the assessment of an educational experience? Besides that, the questions are way too open-ended. They give students license to comment on everything, even things totally irrelevant to learning, like whether they liked an instructor's selection of ties or choice of earrings. Besides damaging teachers, questions like these do not teach students the principles of constructive feedback. Instead, under the cloak of anonymity, they are allowed to say whatever they want and without any consequences. Does that bear any resemblance to how critiques are conducted in professional venues? If I had the power, I would make these kind of open-ended questions illegal. Lacking any such authority, I can only encourage the collective raising of voices in protest and offer (in the next chapter) a collection of questions that more productively focus responses.

Faculty can protect themselves (more intellectually than emotionally) if they understand why these comments hurt so much. It is a case of reckoning with vulnerabilities. Teaching exposes personhood. It reveals to others areas of weakness, sensitivity, and places where pain can be inflicted. Some students (because they are immature or have problems of their own) take advantage of the situation. For teachers, the solution is not to pull back or fabricate some inauthentic teaching personae. That approach compromises effectiveness and denies authentic learning experiences to students who deserve them. The solution is knowing that we're vulnerable and then working to create classroom climates that respect the humanity of everyone.

Faculty further protect themselves by learning how to deal with hurtful comments constructively. That begins with an upfront



decision as to whether or not the comment can simply be let go, as in ignored, disregarded, deep-sixed, or in any way that works dispensed with. It can be if the comment relates to something over which the teacher has no control. “Normal human beings do not teach at 8 A.M.” Most teachers are assigned teaching times or are expected to take less desirable class times on a rotating bases. A few of us actually sign up for the early classes. Students, not teachers, control when they learn best. The comment relates to a student issue over which the teacher has no control.

A comment that has nothing to do with how the teaching affects the learning can also be disregarded. Here’s one that came to me in a packet of evaluations from a faculty workshop. “Given your weight problem, you should not wear corduroy. And yours looked like you’d slept in it.” I suppose teaching in the buff or in truly unconventional garb might so distract students that efforts to learn could be compromised, but in general faculty fashions are irrelevant to learning and comments about them should be ignored. Even as I offer that very sensible advice, I must admit, I never wore that corduroy suit again. Learning to respond constructively, even when the comments are irrelevant, is not easy.

Comments that address student problems should not be transformed into professor problems with clever wording. “Professor harshly punished students for missing deadlines.” If the punishment is explained in the syllabus, discussed in class, meted out fairly, and is not excessive compared with how other professors respond, then it’s not likely the professor who has a problem but a student who missed a deadline and experienced the consequences of doing so.

For those comments a teacher cannot justifiably let go, it helps to rewrite them using language more conducive to analysis. This doesn’t mean watering down or otherwise diluting something about the instruction that may be a problem. It’s an attempt to use more neutral and less loaded language so that the teacher can

understand, explore the objection, and determine its legitimacy. So, a comment like “This instructor is so boring I couldn’t stay awake in class even after a good night’s sleep” becomes something more like “This instructor failed to keep my interest or connect me with content even when I was fresh and ready to listen.” Hodges and Stanton (2007) propose a similar tack. They assert that student objections may reflect challenges commonly experienced by novices. So when a student complains, “The problems on the test weren’t anything like the problems assigned as homework,” the student is really saying, “I haven’t yet learned how to transfer problem-solving skills from one type of problem to a related problem.” The Hodges and Stanton article presents a number of other examples and is a great illustration of how even negative student feedback—the kind that makes teachers defensive—may well have lessons to teach, but only if teachers can deal with the feedback in a more objective, intellectual way.

Responding constructively also involves dealing with the question of representativeness. If one student reports that the group work was a frustrating waste of time, was that experience reported by other students? How many? Out of how many in the class? Upon reflection, did it appear to you that a few, some, or many students were not working productively on the group task? If a comment is not representative of the majority response, that’s important. The experiences of a minority should not be ignored, but they should be addressed cognizant of what has been experienced by the majority.

Beyond cultivating constructive responses to hurtful feedback, teachers should also work to create climates of respect in which the potential damage from destructive feedback is recognized by everyone. A wise teacher taught me that the golden rule can be applied to feedback in the classroom. I ask students to give unto me feedback in the form they would have me give it unto them.

That doesn't mean they have to say only "nice things," but when critique is offered it should be presented to me as they would have me say negative things to them. I don't just propose this as a principle for the course on the day students do their course evaluations, I present it in the beginning of the class, and we revisit it through the semester. I encourage them to talk with me about any feedback I have given them that they felt was more hurtful than helpful. I even give them examples of comments students have made about my teaching that made me cringe and question my career choice. It is unethical for teachers to try to manipulate ratings results, but it is professionally appropriate to help students learn the principles of constructive feedback and justifiable for teachers to protect their vulnerabilities. Hurtful comments diminish commitments to instructional growth and development. They sap vitality and make it harder to keep teaching fresh and invigorated over the long haul. When a hurtful comment is received, teachers should have skills in place that enable them to respond constructively so that they can remain open and able to learn from all kinds of feedback.

### Addicted to Ratings

A few faculty get addicted to ratings, usually to good ones, although sometimes it can be an addiction to ratings in general. The addiction is evidenced by excessive concerns over ratings and their results. The addiction of very good teachers is especially ironic, since the ratings almost always confirm their excellence. However, their sense of accomplishment is very much tied up with ratings results. If those results decline, even go down just a little, these teachers can feel like failures. Unfortunately, being that dependent on high ratings, like other addictions, can compromise personal integrity. I once heard a truly outstanding teacher tell a class how much good ratings meant to him—how they made him try harder

and how students benefited when he felt his efforts were appreciated. I worked with another teacher, a recipient of teaching awards, who confessed that he went through the ratings and tossed all the really low ones. Good rating systems should be designed so that teachers don't have the opportunity to compromise the outcome, but that's a different issue. I was stunned that someone this competent in the classroom still felt the need to hide the assessments of those for whom the class was not an effective learning experience. I am also reminded of a group of new faculty who, after a few libations at a local watering hole, told me they wrote positive comments about each other and posted them on the Rateyourprofessor.com Web site.

Clearly there are ethical issues here, but more endemic to the discussion is the problem of being so identified with the teaching performance that feedback must be manipulated to prop up that sense of the teaching self. Yes, feedback from others (students and peers) is essential to validating perceptions of the teaching self. But external feedback should be used as building material, not the foundation on which understandings of self as a teacher rests.

### Constructive Conversations with Administrators About Ratings

Conversations with administrators about ratings are not always positive. Many of us have stories to tell or have heard stories worth telling. I have worked with faculty who've had administrators tell them, "Get those ratings up," with virtually no suggestions as to how that might be accomplished. Admonitions like this convey the sense that faculty should do whatever it takes (begging or passing out donuts). Comments like these motivate teachers to take action, but not in the interest of more and better learning for students. Their goal is to get those numbers up. I have worked with faculty who've had administrators fixate on one or two negative

comments and propose a whole series of instructional alterations based on comments that are not representative. I've worked with faculty whose administrators have asserted that ratings this high cannot be obtained without a sacrifice of academic standards. Not all administrators comment inappropriately, but it still behooves faculty to consider what they will say and how they will respond to those in academic leadership roles when meeting to discuss end-of-course ratings.

These conversations are always more construction when faculty have done the homework on ratings research proposed earlier in the chapter. Some administrators are informed about this research; others are not. Faculty members should know what the ratings research says about the effects of background variables like class size, course level, course difficulty, and workload. They should know the conditions under which ratings tend to be stable and when results from one course or instructor should be compared with another. Conversations with academic leaders should not be yet another venue for the exchange of misinformation.

Faculty also need to work to avoid defensive responses, especially in the face of negative comments. Defensiveness is natural given the vulnerabilities already discussed and given that the conversation is with a superior, but defensiveness slides into emotional protestations and argumentative exchanges. There are more effective ways to proceed, to wit, with questions. "So these ratings need to improve, what would be a reasonable increase by this time next year?" "Are you recommending this change when 38 out of 40 students don't mention it as a problem?" "If these ratings are not what you expect, what would be an acceptable overall course and instructor rating?"

It also helps to go into these conversations having collected additional information about areas in which the scores tend to be lower, items for which there is a wide range of student responses,

or items about which the results are confusing or contradictory. "I was concerned last semester about the scores on whether this course motivates best efforts. I asked students taking the course this semester a series of open-ended questions, and here's a summary of their responses." "I know why students rate the textbook low in this course. They discussed it in groups and generated a list of five things they think makes the text difficult. They also suggested some things I can do to support their efforts with the book." Not only does an approach like this show a certain amount of initiative, it enables an administrator (and certainly the faculty member) to interpret a score in a much more specific and detailed sense. Having developed an action plan and being in the process of implementing it sends the message that this teacher takes feedback seriously.

Relatedly, many instructional evaluation instruments focus on didactic teaching and do not provide feedback on learner-centered approaches like problem-based learning or cooperative learning. High scores cannot be expected on an item like "gives lectures that facilitate note taking" if that teacher uses group work and discussion more often than lectures. Any aspect of instruction used regularly but not included on the standardized rating instrument should be assessed even if the instructor must develop the items and administer a separate evaluation instrument. If the goal is to provide the administrator with a complete and accurate portrait of teaching and learning in the classroom, then there may be a need to supplement those data the institutionally mandated form supplies.

Faculty need to be able to deal objectively with feedback on classroom performance. Granted, that's easier said than done, but the ability to create an emotional distance enables the faculty member to put the feedback in perspective. Once a faculty member has that perspective, the chances of being able to communicate constructively with an administrator about them improves

exponentially. If those conversations are constructive, then there's another way that end-of-course ratings can contribute positively to long-term instructional growth and development.

### **Actions: Those to Take and Not to Take**

In essence, end-of-course ratings results are a call to action, regardless of whether those results are high, low, or the same as they were last semester. They require action because these data offer one view—a distant, judgmental, nonspecific perspective—of teaching and learning in a classroom. That is not the only view; in most cases it is not even the best way to understand how teaching is affecting learning. It's the proverbial trying to understand the elephant blindfolded with one hand touching some part of the creature's anatomy. Taking no follow-up action pretty much ensures that next semester will produce the same incomplete chapter in the ratings story. It's a way to make the story short, but not very sweet.

What follow-up actions should be taken? There are lots of possibilities. For starters, if the instrument being used has validity or partial validity (meaning some questions provide feedback in important and relevant instructional areas), a database of results can be constructed. Simple spreadsheets expedite this task. Overall instructor and course scores for each different course taught should be collected. Scores from courses taught for the first time or courses significantly revised might be aggregated. High and low items for each course might be identified. In other words, a careful and systematic review of these data can reveal more information than what comes up during a cursory run-through.

Soliciting further feedback is another obvious follow-up action. The value of feedback is enhanced if it's collected with a particular goal in mind. Is there interest in or a need to confirm or deny end-of-course results? If so, another summative instrument, perhaps one with greater validity, can be administered. Is

the institutionally mandated instrument asking about those aspects of instruction relevant to this teacher, course, and instructional setting? If not, then the goal is to fill in the blanks, to obtain feedback on those aspects of instruction that may not be assessed on the required form.

Or, the goal of soliciting more feedback might be answers to questions raised by the summative results. Why are students evaluating an aspect of instruction like clarity so diversely? Why do some of the results seem contradictory? If there's a need to answer questions like these, then the follow-up feedback needs to focus on obtaining more information from students. Is there a question about whether a particular change (be it a policy, practice, assignment, or teaching behavior) will take care of a particular problem? If so, then the goal is feedback that asks students to anticipate how a proposed change (maybe a couple of different ones) might influence their learning experiences in the course. Clear thinking about the goal gives purpose and direction to the solicitation of follow-up feedback.

As for actions *not* justified by end-of-course ratings, I'd put drawing lots of conclusions first on the list. Wanting to figure out what the ratings mean is natural, but that pushes you to arrive at conclusions prematurely. It is far better to see summative results as one of several different information-generating activities, all of which contribute to a growing and evolving understanding of how one's teaching influences efforts to learn.

Talking with someone else about the end-of-course feedback forestalls conclusive interpretations of the results. This person needs to be a trusted colleague or a professional from the teaching center. You want to talk with someone who will discuss the feedback in ways that help you gain perspective, insights, and ideas about what to do next. Conversing with an experienced teacher is especially appropriate for new faculty. I can't tell you how often I have looked at rating feedback for a new faculty member, listed the three areas I'd like to discuss, and then had a conversation

during which none of those three areas were ever mentioned by the new teacher. In truth, any faculty member can benefit by sharing a set of ratings results with a colleague and offering no commentary when making this request: "I'd be interested in what you would say about this course and instructor after having reviewed these results."

Another action not justified by summative feedback alone: implementing changes before soliciting more feedback. This may seem like it flies right in the face of the whole ratings process. Institutions evaluate courses and instructors so that they can improve. However, faculty need to make good change choices, as established and explored in Chapter Six. They need to stop doing what isn't working and do more of what it is. They need to select those changes that fit comfortably with who they are, how they teach, what they teach, and the learning needs of their students. Changes implemented on these terms require careful planning, thoughtful implementation, and still more assessment. The generic, general feedback that hallmarks end-of-course ratings results does not provide the kind of detailed information needed to make change choices that are likely to affect learning and subsequent ratings in positive and significant ways. End-of-course ratings do a much better job of motivating change than they do informing it. For that reason, in most cases, the wise counsel is against implementing changes based on summative feedback alone.

Can the sad story of end-of-course ratings be rewritten? The history of how ratings have been used cannot be changed—it is part of our legacy. How they continue to be used is something that could be changed and is one of the advocacy roles for senior faculty recommended in Chapter Nine. Individual faculty do have the power to change how ratings influence their growth and development as teachers.

This chapter has tried to show how that can be accomplished. It begins with teachers being at least somewhat conversant with ratings research. It builds by proposing lessons that can and cannot

be learned from the results, constructive responses when rating feedback is not what it should be or contains destructive commentary. It suggests ways of talking about the results and proposes actions that should and should not be taken based on the results. In essence it's a rewrite in which teachers take charge of end-of-course ratings feedback and use the results to add positively to the story of their development as teachers.