

CHAPTER 7

JUDGING THE QUALITY OF CHRISTY'S TEACHING

Like other junior faculty members, Christy Evans was evaluated formally every year following the campus and college policies described in earlier chapters. Other assessments of her work took place during her third year evaluation and during her tenure review. The evaluation process focused not only on Christy's teaching but also included her research and service activity.

As described in the last chapter, Christy's Department used more than one source of information to make evaluative decisions about teaching but gave more emphasis to student ratings and peer observations. Each evaluation source provided valuable information about her teaching.

Student Ratings of Instruction

Student ratings were collected through the Teacher Evaluation System (TES) and through Departmental forms filled by students attending Christy's courses. As mentioned before, TES ratings had administrative priority over the other ratings because of their important role in the tenure process. A member of the executive committee said:

We have to pay attention to the TES ratings when we evaluate junior instructors like Christy because these ratings are weighted heavily by the campus. Peer review is considered a weaker perspective, as well as subjective data, at the campus level. The college of Arts and Sciences also requires TES ratings when granting tenure. So, we use the TES ratings for annual and tenure reviews. It is better to know what they mean and how the instructor performs in the TES evaluations from the beginning.

The TES ratings obtained by Christy varied during the years that she worked at Midwestern. Her undergraduate “required courses” received lower scores than her graduate courses. Her rating scores were increasing over time, as can be seen in the following table:

Table 4

TES Ratings for Christy's Courses During a Period of Five Years

Semester	Course – Level	TES Scores	Course Ratings
Spring 1996	100 – undergraduate	3.1 (low)	3.2 (low)
Spring 1997	100 – undergraduate	3.7 (low average)	3.4 (low)
Spring 1997	400 - graduate	4.8 (high average)	4.7 (high average)
Spring 1998	400 - graduate	5.0 (high)	4.8 (high average)
Fall 1998	100 - undergraduate	3.6 (low average)	3.5 (low average)
Spring 1999	200 - undergraduate	3.8 (average)	3.6 (low average)
Spring 1999	300 - undergraduate	4.0 (average)	3.8 (low average)
Fall 1999	100 - undergraduate	4.4 (average)	4.3 (average)
Fall 1999	400 - graduate	4.8 (high average)	4.5 (high average)
Spring 2000	200 - undergraduate	4.5 (high average)	4.4 (average)
Spring 2000	300 - undergraduate	4.4 (average)	4.3 (average)

During her first semester at Midwestern, Christy was assigned to teach an undergraduate course with a class size of 400 students. Her TES ratings were low. The size of the class and Christy's limited teaching experience may have negatively influenced her TES scores. Her Department Chair said:

When Christie first came to our Department, she was very young. When she first started teaching, she right away took a course with one of the largest class sizes. She didn't teach a lot before that. Predictably, her TES ratings were a bit low for that semester.

A departmental decision to increase grading standards for Christy's course may have also negatively influenced her scores that semester. Because this course was used to fulfill the undergraduate requirement for History, the Department's Administrators advised Christy to hold high teaching and grading standards for students enrolled in the course. The Department decision responded to the large number of undergraduate students trying to transfer to the Department, some of them on probation. Christy said:

When I began to teach in the Department I had to teach a very large class. That semester I was advised to hold high standards for my teaching. Because of this, I paid attention to the amount of reading and was strict in the grading. My ratings were low that semester. It was my first semester at Midwestern, it was hard to teach a large class. I think all of this and my decision of not decreasing the amount of reading, as my students wanted, influenced my scores.

The Department's Administrators were sensitive to Christy's limited teaching experience at the beginning of her practice, and to the characteristics of her courses when using the TES ratings for making decisions about the quality of her teaching. They also paid attention to how other instructors teaching the same course performed in the past, and were sensitive to the possible influence of Christy's age and gender in the evaluation.¹ They also took into account the class size of her courses. There were, however, other important aspects of the ratings that were not considered by the Department but deserve further attention here.

¹ As mentioned in chapter six, the Executive Committee of the department found that students tend to rate women lower than men in the TES evaluations and that students tend to give even lower ratings to younger women.

Limited Student Response Rate

An important aspect to take into account when interpreting the results from student

ratings of teaching is the response rate of the students completing the forms. As Theall and Franklin (1991) reported, good evaluation practice suggests caution in interpreting results from smaller samples of students. The smaller the number of students attending a course, the larger should be the minimum (percent) usable response-rate. According to Theall and Franklin (1991), for courses with a class size of more than one hundred students, the minimum acceptable response rate is fifty percent.

The rate of students completing the TES forms varied across Christy's courses. Her large "required" undergraduate courses had a smaller response rate than her graduate "elective" courses. In some of the undergraduate courses that she taught, the response rate was insufficient to be considered a representative sample of all students taking the course. For example, in 1996, only 37 % of the students taking Christy's beginning level course completed the evaluation forms. A similar situation took place in the Spring semester of 1997, when 40% of the 399 students registered in the course completed the evaluation forms.

Although these response rates were insufficient for making a decision about the perceptions of all students enrolled in the course, the Department did not consider this restriction when looking at Christy's ratings during her third year review. The Office of Instructional Support and Development (OISI), in charge of the TES evaluation system, consistently disregarded the response rate when selecting instructors to be included in the campus list of effective teaching, a list based on TES scores.

Problems with the Items of the TES Evaluation Forms

The evaluation form used for assessing the quality of Christy's teaching included two kinds of items. The first two items of the form were global, to be used by all Departments.

Rate the instructor's overall teaching effectiveness.	Exceptionally High	5	4	3	2	1	Exceptionally Low
Rate the overall quality of this course.	Exceptionally High	5	4	3	2	1	Exceptionally Low

There was also a set of 25 specific items (See Appendix D for a copy of the forms). These items were selected by the Department from a catalog provided by the Office of Instructional Support and Improvement (OISI). These items asked students to answer different kinds of questions regarding the course, the instructor's behavior and attitudes, and about how the respondent felt about the course.

The items of the TES evaluation forms present problems regarding construct representation. Both global and specific items present problems of content relevance and construct representativeness.

The global items require students to make a judgment of the "instructor's effectiveness." This construct is not defined on the form. These items are generally used to give the respondents the opportunity to reflect on their overall experience with the instructor's teaching. Most researchers in the field of faculty evaluation who support the use of these items believe that global items are better than specific items because there is no agreement about the specific criteria of good teaching. By including global items, people have the opportunity to think about all relevant criteria. It was expected that the students filling out the forms would not all have the same understanding of the construct "good teaching" when evaluating their instructor. However, when the TES ratings are reported without an

understanding of the context of teaching, or of how motivated were students towards the course, or how students define the construct “good teaching,” there is a good possibility that the respondent may be assessing something other than teaching. The analysis of Christy’s evaluation forms over a period of two years provides evidence that some students, especially those taking the course as a requirement outside their major and who rate her teaching low, were not motivated in taking the course. Some of these students explicitly stated that they did not consider the course content to be worthy. Students outside the major also had more negative attitudes towards class assignments and the type of exams, partly because they expected to be assessed using multiple-choice exams similar to those used by their instructors in their own Departments. Some of these students complained about the grading procedures and said that classroom assessment was unfair because the instructor did not grade on the curve. There were also students who felt the grading was unfair because they fulfilled the requirements but did not get an “A.”

The “specific” items included in the evaluation forms included teacher traits and behaviors that were taken from other evaluation forms and from suggestions made to the OISI staff for inclusion in the forms. In some studies, these traits and behaviors have been correlated with effective teaching. But, correlation is not causation. It is also unknown what process the respondents followed when answering the items, or if they responded truthfully. Ory and Ryan (2001) claim the lack of this information can raise issues of validity for the items if they are used for administrative decisions (Ory and Ryan, 2000). There is little assurance of valid representation of the construct. Are the items valid for evaluating the goodness of teaching? Are all important criteria represented in the items included in the form?

Are the items appropriate for that particular instructor teaching under those conditions of teaching, using his or her teaching style? Some of the conditions are not ones that s/he can control.

Even if there is consistency among student responses, it is unclear what the evaluation forms are measuring. The meaning of the scores remains vague. If the ratings are representation of good instruction, there is no problem. But, if they are not measures of instructional quality, and if they are used to make decisions that affect the lives of faculty members and result in decisions affecting the quality of teaching on campus, then validity and fairness become serious issues.

Some of the “specific” items ask students questions outside their areas of expertise. For example, one of the items asks students to judge the quality of the instructional materials used in the course in relation to the course objectives. Students have little knowledge of instructional design. Their ability to evaluate the quality of instructional materials and to determine their relevance to the course objectives is subject to challenge. Other items ask students to determine the amount of knowledge that their instructor has about the subject matter. And students have to assess their instructor being conscientious and well prepared for their course. Braskamp and Ory (1994) claim that these items are inappropriate for evaluating instructional quality because they ask students to make a judgment of aspects that directly or indirectly judge "the quality of the course content or the knowledge and scholarship of the instructor" (p.99).

Other “specific” items included in the evaluation form were used for evaluating aspects of the course, rather than of the instructor. Although it can be argued that the

instructor designs the course and that these items provide evidence of the professor's abilities as instructional designer, it may be unfair to question the goodness of an instructor because the course did not increase the interest of students. There should be concern about the assumption that all the "specific" items are appropriate for all courses in the Department. Instructors who teach a course based on their research have an advantage over instructors teaching a course needed to fulfill a requirement. In addition, instructors who teach advanced seminar courses may have an advantage over those teaching general introductory courses in accomplishing the goal of having their students identify the "main points and central issues in the field."

The Office of Instructional Support and Improvement selects effective instructors by considering instructors who do well in both the item that points to teaching effectiveness and the one that points to course effectiveness. This may result in punishing professors teaching required courses. The problem of using the global course items, to select effective instructors is that the assessment of the course includes different aspects which students may not be in the best position to judge, such as relevance of course materials, appropriateness of course objectives, etc.

Use of Averages for Making Decisions About Teaching

The use of averages of the responses across students completing the TES forms also presented problems. As several members of the Executive Committee and the Department Chair commented in the previous chapter, they found the averages of the ratings to be one of the less useful parts of the evaluation. Averages are calculated from the responses of students to a checklist of items, such as those "specific" items included for evaluating Christy's

teaching. But, as Stake and Cisneros-Cohernour (2000) state, those aggregates have seldom been validated empirically.

Another problem with the use of averages and other aggregates of data is that they rely on the assumption that a single statistic is better than a distribution of views. Doyle (1992) states that the purpose of the evaluation of teaching should not be to obtain a descriptor. Selecting an average can hide differences in opinions among groups within the classroom. A review of the TES evaluation forms of Christy's courses over a period of five years provides evidence of such differences, especially among students attending her undergraduate “required” courses. Students attending those courses were not always History majors, nor always having a positive attitude toward the content. A member of her Executive Committee said:

Christy's 100 level course is used to fulfill College requirements for History. Many of the students taking this course won't take another History course again. Others will become History majors but not necessarily interested in Western European History. There are many students who just take these courses because they need credits to fulfill their General Education requirements. In these courses it is common to get engineering or business majors who need to take a course and don't want to be there or don't see the value of the content.

Christy's students were at different stages in their studies, from freshmen to seniors. They brought different knowledge backgrounds to these courses. With these differences in student prerequisite knowledge and interest in the course, and the “required” nature of the course, it was not surprising that the students who gave Christy lower ratings were freshmen or students with a major other than History.

The comments made by Christy's students on the back of the TES forms indicated some differences in thinking as to how they rated the instructor. Although the Campus and the College Officers did not support the use of student comments for making administrative decisions about teaching, the comments were helpful identifying differences in the process followed by students when completing the forms, especially in courses with a highly heterogeneous student population. For example, in some cases, on the evaluation forms, Christy's undergraduate "required course" students wrote very positive comments, but rated her 3 points on a scale from 1-5. Other students who gave her a 3 on her teaching wrote both positive and negative comments about the teaching. In a few cases, they wrote mostly negative comments. So, as expected, the meaning of 3 points appeared to be different for some students than for others. For some students a rating of 3 was be something positive; for other students, it indicated that the instructor had both strengths and weaknesses, or mostly weaknesses in her teaching.

There were some contradictory comments on the back of the TES evaluation forms from these undergraduates. These comments also were useful for illustrating differences among students enrolled in these required courses. For example, while some students complained about the difficulty of the assignments and the amount of reading, other students saw the assignments as appropriate for the course and requested additional readings.

Differences were found not only in student perceptions about the goodness of Christy's teaching, but also in student understandings of what took place in the course, and about Christy's knowledge of the subject-matter. For example, some students taking Christy's undergraduate "required courses" saw her as not knowledgeable and uninterested in the

subject, whereas others saw her as enthusiastic and "knowledgeable of everything in the course and in the field." Some students described her teaching as boring whereas others attending the same class described it as exciting. A few comments on the TES evaluation forms referred to the way Christy was dressed, and her accent. A few undergraduates stated that Christy's accent was one of her weaknesses as an instructor. Other students complained of Christy's pronouncing the name of French sources in French. Some students enrolled in Christy's undergraduate "required course" seemed unhappy with Christy's efforts at getting students interested in the course, saying that she treated them as children. One of them stated that Christy was "power hungry" and that he laughed at her.

It is important to note that not all students wrote comments. Most of the comments were made by undergraduate students enrolled in Christy's required courses. The students attending Christy's 300 - 400 level courses rarely made comments about her teaching. For example, during Spring 2000, when this study was conducted, 13 of the 23 students attending Christy's 300 level course made comments on the back of the TES evaluation forms. All were positive, describing her as enthusiastic, engaging, thought provoking. Only one of the students stated that a weakness of Christy's teaching was her thematic approach, because he preferred to learn "the facts". Students who took Christy's 200 level seminar made no comments on the evaluation forms about Christy's teaching even as they gave her very high ratings, helping her to be included in the campus List of Effective Instructors.

The university policy of disregarding student comments on the evaluation form when making administrative decisions about teaching was established because, according

to the Provost, there was "no basis for estimating the quality of those views," and because in order for those comments to be considered, they needed to be "based on a legitimate sample of student views." However, in required courses and courses with a large class size and high heterogeneity among students attending the course, student comments provide useful information for understanding these differences among students. By eliminating the comments or not including other ways for capturing student understandings about teaching, valuable information is missed. Differences in student views are important because they raise the issue of how validly the "average" of the scores represents the perceptions of all subgroups of students within the course.

Focus Group Interview with Undergraduate Students

A focus group interview was conducted with Christy's students at the end of the Spring semester of 2000. The purpose of the interview was to understand the process followed by her students when assessing her teaching and the meaning that students gave to the construct "good teaching." Twenty-three students participated in the focus-group interview. As described in Appendix (F), the students decided in different ways how to assign ratings. Some students said that they assigned a score by comparing the instructor with other instructors they had had. Other students assigned the scores by comparing their instructor to an ideal of what they believed a teacher should be. A small number of students stated that they reached their decision by reflecting on their own experience in the course. A few other students stated that they did not know how they assigned the scores, saying they had not thought about it.

When asked about the meaning of good teaching, many students identified teacher behaviors associated with good teaching. According to the students, good instructors provide them with opportunities to acquire knowledge that they can take with them after the course (21 of 23 students), make the course enjoyable and interesting (all 23 students); grade their students fairly (all 23 students); are available after class (11 students), and respect their students (12 students).

When asked if one or several of these aspects would be so important that it would influence their ratings regardless of other aspects, all the twenty-three students participating in the interview mentioned fair grading. One student stated: "If I am not graded fairly, it would negatively affect my overall view of the teacher."

The students stated that they did not believe an instructor should be "culturally sensitive" when framing his or her course, nor frame the course depending on the kind of students, or connect the class materials with life experiences outside the classroom. One student said, "At this level, if you aren't smart enough to connect it yourself, there's something wrong. But I do think that History becomes more meaningful when connected with other History."

All students participating in the session stated that they believed the university valued the evaluation of teaching and used it to make decisions about firing instructors.

To understand further what students meant by fair grading and to obtain more information on findings from the focus group interview, a follow-up e-mail survey was sent to all the participants of the focus-group session. Only three students responded to the survey. One of them stated that she made the decision of assigning a score by comparing her

instructor to an ideal of what she thought was the best possible teacher that she could have. Another student stated that he made the decision by reflecting on his own experience in the course. A third compared Christy to other instructors that he had had. These last two students had taken more than one course with Christy, and made additional comments about her teaching:

I took another course with Dr. Evans before. I remember enjoying that class. Her lectures were very informative and our discussions were helpful. A number of readings we had were quite difficult but we worked in small groups on questions that she handed out and were able to think things through ourselves. Her teaching approach was useful for me. Most of my classes don't take that type of view. It is much more challenging and it has helped me to develop my analytical skills to a great extent. I think that knowing the methodology and being able to apply it to the reading and to ones own research is an invaluable skill. I get a lot more out of my readings now.

The other student commented:

I would have to say that in the classes that I have had with Dr. Evans, I do not believe I have noticed a significant real change in her teaching or attitude. I do believe that she changes or adapts to her class environment or course content, however. Dr. Evans's survey class, for example, was much more impersonal since there was a huge number of students in the class. In her smaller classes, she taught in much the same way, but in a much more personal manner. Dr. Evans is also very interested in finding new ways to teach things or at least better ways. She seems to genuinely enjoy what she does, which adds to her personality in front of the classroom.

An issue raised by one of these students was the lack of importance students give to the TES evaluation forms. This student said that, in his experience, most undergraduate students did not take the TES evaluation seriously:

Quite honestly, evaluation forms do not work well. Students do not always respond honestly to the questions. They often fill out the forms as quick as possible. Every undergrad student will probably tell you this...

The standardization of the forms makes the entire process seem meaningless to many students, and the way it is presented to the students is often in a sterile, chore-like way.

Although this was the opinion of a single student, these comments were consistent with statements made by an instructional specialist working for OISI. In a public interview he encouraged students to take the evaluation forms seriously. An article published by a former History student in the local student newspaper also supported the perception that students do not take the forms seriously. Some might use evaluation "to get even with teachers they have grown to dislike" (p. 1). The author of the article also stated that "the (TES) forms more closely resemble a cola taste test than a measure of the quality of teaching." And that teachers of required courses may "had a difficult time attracting favorable ratings, because dissatisfied students are compelled to take the course."

Strengths and Weaknesses of the Departmental Evaluation Forms

The History Department administered a short evaluation survey to students in the Department. This evaluation form (See Appendix E) was used to complement the information from the TES evaluation forms. These forms included five open-ended items:

1. Which class readings did you find most valuable? Why?
2. Which class readings did you consider to be a waste of time? Why?
3. What does the teacher in this class do that you'd recommend to other teachers?
4. What does the teacher in this class do that other teachers should avoid?
5. How has this class changed your view of the subject matter of this course or of historical study in general?

With the exception of Item Five, all these items can provide the instructor with more valuable information about some of the items included in the TES forms. The Departmental items are highly appropriate because they allow students to describe instructor behavior and to provide information regarding what students value about the teaching. Item Five relies on the assumption that the single responsibility for changing student views is the instructor's. As Tagamory and Bishop (1994) report, this kind of item asks for information that may not be associated with effective teaching performance.

Peer Evaluation of Teaching

For the last five years, six selected members of her Executive Committee observed Christy's teaching². An observer would usually contact Christy via telephone or e-mail and arrange a convenient time for the classroom observation to take place. Some observers got a copy of the course syllabus prior to the observations while others did not. For most, there was no prior meeting. Sometimes, there was a brief meeting afterwards, but in most cases the observer just wrote a report for the Administration, and gave Christy a copy. One of the observers illustrated the process:

Before I observed Christy's class, I arranged with her when to meet. She suggested a date convenient for the observation. We all choose a lecture that we think it will be better for observation. For me it was important to conduct the observation at a convenient time, I did not want my visit to be an intrusion into her class. After the observation of the class, I wrote the letter reporting on the observation. I spoke with her briefly but we did not meet.

²Of the six observers, only five agreed to be interviewed. Observation reports were available for all the six observers.

Each of the observers attended different course sessions (Christy does not teach the same course every semester). The observers differed in what they perceived to be their roles and the aspects to which they paid attention.

The first observer was especially attentive to Christy's ability for integrating the lecture to an overall picture of the course and historical time:

I consider that having different levels of delivery is part of good teaching. So, I expect an instructor to show comprehensiveness of events, to be clear when using evidence for stating a thesis and when making illustrations. When I observed Christy, I was impressed with the clarity, precision, and overall coherence of her lecture. She has an engaged and fluent lecture style that provides anecdotes and stories framed by a larger conceptual organization. My critique of her class was that, in her lecture, she lost the big picture. She got immersed in using illustrations and anecdotes to help students understanding but she needed to go back to the big picture to remind the students why they were there.

The second observer was also interested in delivery of content but his main focus was on style. He/She gave some attention to contextual aspects of the course, such as class size and the characteristics of the large auditorium where the class took place:

When I go to observe a class, I have an idea about how a good lecture should be. A good lecture is organized, very informative, it has clearly explained themes and illustration. A good presenter also uses a variety of media whenever necessary, and should use anecdotes because this helps to keep students attentive. These are the kinds of things I paid attention to when I observed Christy. It was a large lecture class with about 300 - 400 students. When I was observing her class I wondered how difficult for an instructor it would be to keep the interest of the hundreds of students during the lecture, and to maintain a good attendance rate throughout the semester. When I looked around, I was surprised to find that the huge classroom was filled with students. I found myself sitting through a very well organized lecture on the early phase of European imperialism. My suggestion for Christy was to include a few more anecdotes and to make more pauses during the presentation to get a sense how students were following her during the lecture.

The third observer paid particular attention to the way in which Christy drew students into the discussion and to her ability for promoting a classroom climate favorable for learning:

When I observed Christy's class, I paid attention to how well she was communicating with students. I did not judge content coverage. Instead, I paid attention to her effectiveness in drawing students into a discussion--when this was applicable. The class that I observed was about Martin Guerre. Christy drew students attention to a highly complex subject - matter with admirable simplicity. She also maintained a relaxed and cordial environment in the classroom, to which the students seemed to respond positively. I suggested to her in my report that she ask more questions about the topic. She could integrate some additional thematic connections to show students that History is multi-linear.

The fourth observer saw his role as searching for positive aspects of the teaching, as well as making sense of what was taking place in the classroom, and identifying areas for improvement:

When I observed Christy, I saw my role as helping her to gather information to counteract the TES ratings. So, in a way the primary purpose of my observation was to get something positive about the teaching. I also tried to get a sense of what the instructor was doing and what she can do for improvement. During that session, she was teaching a very large class. I was impressed, especially that somebody just beginning her teaching career already had the sovereignty and easy to speak freely over long stretches of the lecture.

This observer also mentioned that, in his observations, he usually gives special attention to the kind of course being taught because students tended to give negative ratings to non-Western courses. But he acknowledged that in Christy's case, this was not an issue because she was teaching a Western European History course:

When I conduct an observation, I also pay attention to the kind of course content being taught. There is a negative attitude and resistance of students towards some course content. If the instructor teaches U.S. History, s/he

will have less resistance from students because they don't question the value of the course content. However, if s/he teaches Latin American History s/he will confront the fact that most of the students don't consider this kind of course to be relevant for them. Student negative attitudes towards the content are reflected in the evaluations. I have seen student comments saying: "I hate Latin America." Fortunately for her, Christy teaches courses that are in the middle level of acceptance. They are Western European History courses that focus on women, race and reasoning during the Seventeenth Century. She still gets students who take her courses because they were not able to get into another course that they would had preferred. But there is less resistance because her courses are Western European History courses.

The fifth observer was more concerned with the kind of teaching that took place. For this particular observer, the scholarly orientation of the teaching was especially important:

In the Department, we expect the teaching not to focus on a series of facts. Instead we expect the instructor to have students get a sense of the questions a historian will ask. You can see this in Christy' s teaching. When I observed her teaching, Christy engaged her students in this way. I think her teaching is better than average, I did not find any problems with her teaching.

The sixth observer also paid attention to the delivery of the lecture with emphasis on the organization of the syllabus, and the ability of Christy to address a highly complex content not in her special area of expertise:

My overall impression was one of high professionalism in Christy's organization, lecturing, and leading of classroom discussion. I was struck by the unusually well organized syllabus that Christy had put together for the class. I was particularly impressed with Christy's ability to deal with the Scientific Revolution expertly, and explain difficult ideas clearly and patiently, even though it is far removed from her expertise in French Social History.

All the six observers prepared observation reports for the Department Chair and gave a copy to Christy. The reports followed a similar format. They included some brief description of the context of the classroom and of Christy's and her students' behavior during

the lecture and during the time her students were working in teams. A positive aspect of these reports was that, unlike the checklists used by the Teaching Academy to conduct peer observations, the reports developed by the History Department brought a glimpse into the context of teaching.

It is important to note that there were some differences between what the observers said during the interviews to what was described in the reports. When reading the reports, the differences between observers were subtle. Some comments from the interviews were omitted from the reports. For example, there were no comments from Observer Number Five regarding the emphasis that he put in the scholarly orientation of the teaching, or on Observer Number Four's focus on finding positive aspects of the instruction. The differences among observer expectations and observations were also less evident between what they wrote in their observation reports and what they said during the interviews. This could be because the Department Administration directed observers' attention to certain things when writing the reports, or because of the pressures that Campus and Department legislation puts on observers to minimize interpretation and judgment, and to emphasize description during the observation. As one of the observers said:

As observers we are encouraged by the Department and the Campus to follow a similar format when writing the report, and to write about what we saw in the classroom, more than expressing a value judgment on the teaching. As observers we are told to avoid focusing the report on "Do students like it?" but on different aspects such as: "Was the professor organized" Does s/he interact with students? Are handouts and different materials being used as well as different techniques? How do students react to instructor "efforts." People who obtain low ratings are sometimes encouraged to use technology because this is Campus policy.

During the interviews only one of the observers explicitly stated that he focused his observation on finding positive aspects in the teaching, but all the reports were very positive. When asked about this, all the interviewees but one stated that they had problems making negative comments about Christy's teaching. Some of the observers said that they did not feel comfortable being too critical because Christy was going to receive a copy of the report. One of the observers said:

Peer evaluations are usually positive, in different degrees. You don't usually get massively critical reviews. This is in part because you are sharing your observations with the instructor. You don't write super critical comments in the letter about a colleague if s/he will receive a copy. If the report were only for the Chair, maybe peer evaluations would include more criticisms than they actually do.

In a sense, using members of the Executive Committee for conducting the observations resulted in a trade-off. On one side these professors were well known for the quality of their teaching and were acquainted with the evaluation system and the problems of different sources of data when making administrative decisions. On the other side, having a dual role as observers and decision-makers influenced the observer decision towards writing positive reports. A Committee member stated:

You have to consider that those observing the instructor come from the Executive committee. This situation makes it harder to make criticisms. Whatever critical impulses you have are gone, because you have a dual role. You are expected to be supportive of your junior colleagues. It is a different perception when you are a Committee member than when you are making comments as an individual professor. As a Committee member you are concerned with your junior faculty members doing well.

The professors who observed Christy found consistency between TES ratings and peer evaluations. Even though the observers had not seen Christy's TES scores before they

observed her teaching, they had heard about her teaching reputation from her students. One of the observers said:

I did not know how Christy was doing in her TES evaluations when I observed her. But I knew that she was a good instructor. Some of my graduate students who took her course told me that she is a good instructor. Sometimes I have students who are taking her course and my course during the same semester. It is common for students to talk about how they feel about their other courses and their teachers.

Peer Mentoring and Its Influence on Christy's Teaching

As were other junior professors in the Department, Christy was working with a mentor. This professor was selected because she shared similar research interests with Christy. The mentor saw her primary role as helping Christy go through the tenuring process. She said:

I was assigned to work with Christy because of the compatibility of our research interests. I work with Christy mostly on the research side of her career management. We talk about how to spend time, how the University works with respect to research, how to get tenure, how to write grants, etc. When the decision about tenure is made, I am expected to be on her side, to be supportive. That is, when the Executive Committee reviews her records in her tenure year, I will be on her side as her mentor. I will support her record.

The mentor had informal conversations with Christy about teaching--but this was not the main focus of her mentoring. As mentioned above, the mentoring was centered on Christy's research and on the tenure process:

Christy and I have talked about teaching in an informal way. We have talked about our experiences and about teaching methods. We talk more about research because this is a research university where knowledge is generated. Research counts more than teaching--although I think there is a connection between being a good researcher and a good teacher. Our Department pays attention to teaching but you have to be really delinquent to be denied tenure because of poor teaching. In general, research is harder than teaching because

in teaching you get an immediate reward. When you do good work in teaching, the rewards come from your students. Working in research takes time and it is harder to get recognition. The seduction of teaching is an "every day satisfaction." This is why it is kind of easy to feel interested in teaching. You have to work harder in research to receive the recognition.

Christy worked well with her mentor but did not see her as a source of feedback for her teaching:

My mentor and I meet during the semester. We talk about how my research is going, or about publications. She also makes suggestions and shares resources with me. We also talk about our teaching experiences but her main focus is on helping me to go through the tenure process.

How the Evaluation Influenced Christy's Views on Teaching

As mentioned before, when Christy began work at Midwestern University, she started by teaching a very large class. She felt overwhelmed with the requests from her students and began to feel that the demands from teaching were interfering with her research.

She said:

When I started teaching it wasn't easy for me. I was surprised the Department let me teach a class of 400 students. It was a hard class to teach. Students were always complaining about little things. It wasn't enough to prepare a package with the reading materials organized by session. They wanted the date and session number written in each reading of the package. I thought they were just whining to get every little detail. Everything seemed to be burdensome to them. I felt overwhelmed by student demand, but I tried to address their requests. I began to feel resentful for the time that teaching was taking from my research. I also felt isolated during that year. At that time, there were very few junior faculty members with whom I could talk. I talked to one of my peers who is known for his good teaching; otherwise there was no advice. There was informal support but there was no forum for talking about teaching.

When Christy received low ratings for her teaching at the end of that semester, she began to pay more attention to teaching. She started worrying about how to improve her

TES scores. Her Department Chair assured her that most junior professors obtained low ratings at the beginning of their teaching and told her that most of her peers teaching undergraduate “required courses” such as one she taught did not do well in the evaluations either. After that conversation, Christy felt more confident about improving her teaching. But she still felt the need for support:

At the end of my first semester, my TES ratings were low. The comments of some of the students on the back of the TES forms were very bold and harsh. My Department Chair talked to me and told me that it was common for junior faculty to receive low ratings at the beginning of their teaching. He added that there were other professors teaching similar courses who received low scores. He was confident I was going to improve over time because my scores were not very low and he knew I was working hard at teaching. I felt better after that conversation and kept trying to improve. Improving my teaching was important for me. I knew that in our Department research is first, but I also knew that they didn't like low evaluations of teaching. We are encouraged to obtain high ratings. It is a sense of duty and pride.

In the following semesters, my TES scores increased and the professors who observed my teaching wrote very nice comments in their observation reports about my class. I felt more confident, but I still needed to have someone with whom I could talk about my teaching. I felt left alone in finding how to improve. What happened is that the Department Administration did not want to interfere with my academic freedom but I would have welcomed the interference.

In the Spring of 1999, Christy taught an advanced undergraduate course. Although she had the perception that everything went well with her course, at the end of the semester, her TES ratings were not as high as she expected. This motivated her to search for help to improve both her teaching and her scores. As Christy commented:

In 1999, I taught one of my advanced undergraduate courses. At that time, I had a high number of education students taking the course. In general, the course went fine. Some Education students commented that the readings were a little hard for them but the overall course seemed to be going well.

At the end of the semester my TES ratings went down. I was very surprised. I did not expect this to happen. That semester, I decided to do something about my teaching. I couldn't go to the Teaching Academy at that time because they had just started the project and were working only with faculty just hired. So, I went to OISI for help.

The OISI staff asked me to bring my TES ratings to our first meeting. During the meeting, they looked at my TES ratings over two years and read the comments made by my students on the back of the forms. After this, the OISI staff decided that my low ratings weren't because of a personality problem or because my students didn't like me. They felt that that my problem was that I was failing to engage my students. So, they asked me to prepare a short lecture on film. The OISI staff sat and listened to my lecture as if they were my students. A few days later, they talked to me and gave me verbal feedback about my performance. After that, they made some suggestions for improvement. They told me to break the lecture into small parts and advised me to deliver the content in a "packaged way", because this will engage my students. They told me to reduce the time I lectured, to modulate my voice, to use an overhead projector and to present the content gradually. They said that I should show students the points of the lecture one by one. They believed that, if I showed all the points of the lecture at the same time, this was going to be distracting for my students. They also advised me to use signs during the lecture, and to summarize the main points at the end of the session. The staff member who worked with me suggested I avoid doing all the changes at once, saying I should change no more than two things every semester.

The following semester, Christy began to make some changes in her teaching. She started preparing overhead transparencies and a list of questions that students could use when doing the course readings. She made these changes but wasn't fully convinced to adopt all the teaching suggestions made by OISI staff. It was at that time when she learned about a seminar offered by another campus office that provided professional development. This was a Center for Writing Studies that offered seminars about writing across the curriculum:

During the following semesters, I started following some of the suggestions of the OISI staff. I began to use more media and to prepare discussion outlines for my students and started summarizing the main points of my lecture. I also began to be concerned about things I didn't do before. I

started paying attention if the equipment was on time, if the font of the transparencies was big enough or if there were too many sentences in my transparencies. I also began to worry if the overhead projector did not work well. I was OK with the changes but I still wasn't fully content with the prepackaged approach that the OISI staff suggested to me. I introduced some of their tips but I still wanted to do more to help my students understanding of History. I wanted them to identify issues, to use and interpret primary sources and to be able to write good arguments in their assignments.

That semester the English Department had a seminar that I decided to take. The seminar had faculty members bringing their syllabi and working together developing strategies to engage their students in scholarly writing. All of us had the opportunity to share our work and learn from each other. During the seminar, I learned how to work with my students by working with my peers and the instructor. I also learned how to create student portfolios. This was what I was looking for. The seminar helped me so much. It was consistent with the kind of learning that I wanted to promote in my students. It provided me with the most helpful and valuable feedback that I needed for improving my teaching.

In addition to implementing what she had learned in the writing seminar, Christy began to share her new ideas about teaching and ask her peers for other ideas about how to improve her teaching:

Our Department has an Early Modern European History Group integrated by faculty and graduate students. This group was set-up in 1999. During our meetings, we present our work in progress to this group. They read it, and give us feedback. After I took the seminar, I began to ask this group about teaching. Now, we even share teaching resources. I also belong to a regional group in French History that meets every six weeks. I also share teaching resources with this other group. Sometimes, we share syllabus and discuss ideas for teaching certain content or share information about our course materials and our course websites. Both the Early Modern European History group and the regional group in French History have been very useful to me. I finally found peers with whom I can talk about teaching and share ideas about how to evaluate learning.

Over time, Christy became more confident about her teaching ability and still looked for new ideas for improvement of her instruction. During the time this study was conducted

she implemented different writing activities that she learned in the writing seminar. She also adapted an essay exam that one of her peers from the Early Modern European History Group gave to her as a model for designing her own take-home test that semester.

Usefulness of the Different Evaluation Sources for Instructional Improvement

Christy stated she believes that the different sources of information used for evaluating her teaching were useful. In general, she felt that all of them had contributed to increasing her interest and confidence in teaching. She introduced changes in her teaching, using writing activities to engage her students in scholarly writing, increasing student active participation in the class, and using technology in teaching.

Reflecting over time, Christy felt that student ratings were beneficial for her because they made her aware that students were experiencing problems in her “required” undergraduate courses:

Students really express what they think. Their messages are bold. Sometimes they can be very hard to take. If they don't like the class they will let you know, sometimes in a very harsh way. If they have problems understanding something, you will know it by their comments on the forms. Having students letting you know that “you suck” is not too easy. But, this kind of comment makes you wake up. It reminds you that you have a responsibility to them. It is a wake-up call that makes you think more about how your students feel. The shock value is precisely the main value of the ratings. The score that we obtain in the evaluation is mostly a reflection of how students feel about our teaching. There are students who have problems with female authority, and the scores may be a reflection of sexism. Negative student attitudes can hurt you, especially if you are teaching a small class, because you only need a few comments to get low scores. But, many times the ratings are a way for students to tell you how they feel.

If you think about it, this is the only power that students have in the teaching process. The honesty and brutality of their comments comes from feeling sometimes oppressed. Students take more than one course, and

sometimes are overburdened, tired and frustrated. If students take a course in which the instructor doesn't care or doesn't do a good job at teaching, they feel very unhappy and let you know that. You need to put yourself in their shoes, and realize that you may not care too much about teaching but they have no choice. They have to do all the work and take the evaluations.

Christy felt that peer feedback also was useful, especially the feedback from those teaching other Western European courses, and the professors who participated in the English Department seminar that she took in 1999. The less useful feedback was from the Department observation reports because these reports were not critical enough of her teaching:

The problem with colleague feedback is that in some cases it is not critical of your teaching. It is nice to receive good teaching evaluation reports. But, if you are trying to use the comments for improvement, then, the feedback is not very useful. Most of the professors who observed my classes made some useful suggestions for improvement. But, in general, their reports didn't provide me with critical information that I would like to have received. I think this happens because they want to help me to do well and to get tenure.

Peer feedback, however, has been very valuable for me. The most useful feedback that I received came from peers who participated in the Early Modern European History Group at our Department, my group on French History, and especially the professors who participated in the Writing Across the Curriculum seminar. In that seminar, we not only learned from each other about how to improve our work, I also learned to connect my work with other courses and to use writing to engage students in the kind of learning that I believe is important for them. My peers in those groups were critical and did not feel uncomfortable sharing both positive and negative comments about teaching. They also shared the experiences that they had when teaching their courses. We all shared resources and provided each other with ideas for improvement. One of my peers who is part of the Early Modern European History group keeps saying to me, 'Did you see, at the beginning you felt that teaching was taking you away from your research? Now you are the one who wants to improve your teaching. You became interested; now you care.'

Christy also felt that evaluative feedback from the offices outside the Department was helpful to her.

I think all the feedback that I have received about my teaching has helped me. The feedback from the OISI staff was less helpful for the kind of teaching that I try to promote in my courses. But, I learned some tips from them that were useful. They made me realize that students pay attention to details and that ignoring this can make it harder for you and for them. Now I use technology as they suggested and I notice that students like it. But, I still remain truthful to my own teaching philosophy, one that values complexity and diversity, and that requires an active role of students in their own learning.

The English Department group who organized the Writing Across the Curriculum seminar was great, they were the ones who helped me the most. They were what I was looking for. They taught us how to use cooperative learning by having us learning in this way.

Other feedback useful to Christy was the feedback that she received from students who took her course in the past and from her teaching assistants.

I have some students who have taken more than one course with me. When I received lower ratings than those that I expected in my advanced undergraduate course, I talked to two of these students who took the course. They provided me with good feedback. They mentioned what they thought went wrong and why. My Teaching Assistants also shared comments about my courses. Since we usually meet after class, we use this opportunity for reflecting on the session.

The feedback from all the different sources helped Christy to reflect more about her teaching. She also made some changes based on her experience working with her students:

Over time, I have changed the way in which I teach some content. But, I haven't changed my teaching philosophy. What I have done is learn how to adapt what I learn without giving up the kind of learning that I want to promote in my students. Sometimes, I have made changes because of my own reflections and my experiences teaching. For example, when I began to teach my undergraduate "required course" I thought that providing my students with some background knowledge could help them in the course. But, this didn't work well. Some of students felt overburdened with the amount of reading. Now, I have learned to be more organized. The

sequence of my courses is better. I feel more relaxed. I think students can feel this.

Summary

The History Department at Midwestern University followed Campus and College policies when evaluating professors, like Christy, on tenure track. Teaching was evaluated every semester using multiple sources of information, giving special importance to student ratings and peer observations.

The different evaluative resources used for assessing the quality of Christy's teaching provided an image of an instructor improving her teaching. Christy's TES ratings improved over time in both her undergraduate and graduate courses. During Spring 2000, she was included in the campus "list of effective instructors." Peer observations of Christy's teaching were positive.

Because of their limitations, and the way in which they were used by the Department, the various resources did not fully reflect the quality of Christy's teaching and her inquiry teaching approach. The TES ratings were helpful in providing Christy with some information about how students felt in her course, especially their comments, on the backs of the evaluation forms. Not all students commented; in some cases only a small sample of students completed the forms. At times, the comments were a reflection of student negative attitude towards the course, or their limited background in the content. The items included in the evaluation forms were potentially problematic in that they asked students to make judgments about aspects that they were not in a good best position to judge, such as when an item asked about the scholarship of the instructor. Other items in the evaluation forms presented

construct relevance problems and measured student related behaviors. In other words, TES was broader than rating the effectiveness of the instructor's teaching. Yet for some purposes, it was used for just that.

The use of averages of the TES ratings could be problematic if one assumed that students share a single understanding of the construct being measured. But there are vital differences among students and groups of students. Data from a focus-group interview of students who took Christy's 300 level course, Spring 2000, and an analysis of comments made by Christy's students over the five years provided evidence that students taking her course have different, even contradictory perspectives of her teaching. When students assign their ratings in different ways, it indicates that a particular score has different meanings for different students. In the undergraduate "required courses" taught by Christy, there were groups of students, including freshmen and non-History majors, who had different perceptions about the course and the instructor. The campus-wide decision of not looking at student comments in the TES forms when making decisions about teaching quality resulted in omitting valuable information.

Peer evaluations of teaching provided insight about the context of Christy's teaching. Since ratings took place only once for a semester, information about the continuity of teaching was incomplete. As a result, supervisors lacked a developmental understanding of Christy's teaching style and scholarly orientation. Some of the negative attitudes towards her teaching needed follow-up but did not get it. In cases where observers glimpsed the context of teaching or Christy's scholarly orientation, the pressures on them for avoiding

interpretation and the format used for reporting on the teaching resulted in reports incomplete about the quality of her teaching.

The use of members of the Executive Committee as observers resulted in a trade-off because they had a dual role as observers and decision-makers. On one side they had a better understanding of the evaluation process and about the strengths and limitations of the evaluation sources and campus policies, which made them valuable observers. On the other side, their dual role resulted in mostly positive reports that focused on finding ways to underscore favorable TES ratings.

The feedback provided by offices outside the Department was useful to Christy. Feedback from the Office of Instructional Support and Improvement was somewhat useful but focused on teaching in a prepackaged way that was not fully compatible with Christy's teaching orientation and philosophy. She, however, found some of the tips suggested by that office to be useful and adopted them into her teaching. The most useful evaluative feedback that Christy received came from fellow professors in a writing seminar organized by the English Department. That seminar focused on having faculty members work cooperatively in designing different activities for engaging students in scholarly writing and analysis, and in assessing student learning through portfolios.

Christy did have a mentor, selected to work with her because of her compatibility with Christy and her research interests. This mentor met with her at different times during the semester but the mentoring focused more on helping Christy get familiar with the tenure process and on planning her research and publication agenda.

In addition to the peer observers and her mentor, Christy Evan also received evaluative feedback from her peers in a Modern European History group and a French History Group. These two groups provided an opportunity for her to obtain feedback and new ideas about research and teaching, information she was unable to get from the student ratings, the peer reports, and her mentor.

Partly because of the way in which the Department used the different resources, Christy was not negatively affected by the evaluation system. She found that the value of student ratings was that it made her aware that there were problems in her teaching that she needed to address. The lack of critical feedback from the peers observing her class or from her mentor motivated her to find help outside of her Department. She used the feedback that she received from the external sources and adapted what worked well for her without compromising her teaching orientation and philosophy. Because she was able to find help, Christy went from worrying about how to improve her scores to become more interested in finding how to improve her teaching in a way consistent with the learning that she wanted to promote in her students.

The evaluation of Christy's teaching raises important issues for Departments infrequently sensitive to the complexities of teaching and learning, especially ones that use a single source for evaluating teaching. Christy was in a special situation because her Department gave her opportunity to teach different kinds of courses to different classes of students. This provided her with the opportunity to demonstrate her teaching skills by teaching different courses. Instructors teaching one kind of classes, especially undergraduate “required courses” addressed to heterogeneous student populations, are in a difficult

situation. The Department may not be sensitive to the limitations of TES ratings and the importance of the contexts of teaching. Instructors working for Departments that are not highly sensitive to the limitations and the possible factors negatively influencing their scores may be in trouble if the evaluation misrepresents the quality of their teaching. More research is needed to learn if the negative influence of gender, race, ethnicity and instructor accent have also been identified elsewhere as negatively influencing TES scores.

There is also need for more research to learn how and when students take the TES evaluations seriously. What are their meanings of good teaching? What are their rating processes? Since students who took Christy's 300 level course in Spring 2000 stated that the most important aspect of good teaching is fair grading and that this aspect can make them disregard other merit in the teaching, there is also a need for learning more about their meanings of fair grading.