

CHAPTER 1

INTRODUCTION

The evaluation of teaching in higher education is an area of strong interest for different stakeholders, and a subject that goes beyond the world of researchers, given its social, political and legal implications. In recent years, pressures for accountability, changes in enrollment trends and financial retrenchment have forced universities to pay more attention to formal procedures for evaluating and improving the quality of their teaching. As Cuban (1999) says, when the cost of tuition rises, and "stories of large undergraduate classes taught by doctoral students have angered parents, pressures on university presidents and faculties to improve teaching have produced urges of new initiatives for teaching improvement." (p. 189)

Research universities have not escaped this trend if only because faculty evaluation is considered an essential element of institutional quality. Indeed, the evaluation of teaching is receiving more attention in research institutions because in the past, teaching has been under-evaluated and minimized, and research has been highly valued (Boyer, 1987; Nyquist and Staton-Spicer, 1987; Cross, 1987, 1983; Bailiff and Khan, 1987; Miller, 1979). As Atkinson and Tuzin (1992) states "only by reforming its teaching mission can the research university recover lost moral ground and restore credibility to its research mission" (p. 25). The growing concerns about the quality of teaching in research institutions have led to current discussions about post-tenure reviews, and to reports of requiring university professors in some universities "to be available in their offices at least four-days a week" (Wilson, 2001).

In addressing the pressures for accountability, research universities are supporting initiatives for improving and rewarding teaching at the college level. These efforts have also

resulted in using the results of the evaluation for making administrative decisions such as tenure, promotion or salary increases for faculty, and as indicators of unit or departmental accountability.

But, while the efforts for improving teaching have been increasing, the methods for assessing the quality of teaching on campus remain the same. Even though culture on the current campus has become more diverse, and the research on teaching and learning has expanded, research on the evaluation of teaching in higher education remains concentrated on the study of student ratings of instruction.

Major literature reviews during the last thirty years (Marsh, 1987; Centra, 1993; Costin, Greenough, and Menges, 1971; McKeachie and Lin, 1979; Cashin, 1988, 1995) support the reliability of student ratings when used for evaluating instruction. But little research has been conducted about the meaning of the ratings, the processes followed by students when filling out the evaluation forms, and on the consequences that arise from the interpretations or misinterpretations of the ratings. Among these consequences, issues of equity have special importance because they could result in a university creating a non-welcoming climate for a diverse faculty or could even result in civil rights violations (Bureau of National Affairs, 1964).

The present study addresses this gap in the literature. This study used holistic methods to examine the meanings, trade-offs, and the social consequences of the evaluation of teaching in the context of a research university.

Need for Conducting the Study

A broad body of research has been conducted on the evaluation of college teaching. This research usually has been undertaken using a positivistic approach that applies the criteria of objectivity, reliability, internal and external validity to disciplined instrumentation. Following these criteria, researchers have focused on determining the meaning of the construct "good teaching" with emphasis on teacher behaviors and/or teacher characteristics, the dimensionality¹ of the construct, and the sources for evaluating college instruction, primarily student ratings². Indeed, the research is practically synonymous with studies of student ratings since only a few studies have used other sources for evaluating campus teaching, and even fewer have used more than one source simultaneously to assess the quality of instruction.

According to Marsh and Bailey (1993), findings from the research show that student ratings of teaching effectiveness are "multidimensional, reliable, stable and relatively valid against a variety of indicators of effective teaching." According to Marsh (1987), the ratings are "relatively unaffected by multiple variables identified as potential biases to the evaluation, and are useful for faculty as feedback about their teaching" (p. 255).

Although Marsh's research supports a certain statistical validity of the ratings, it says little about the validity of the evaluation for representing the target construct. If classroom

¹ While some researchers affirm that teaching should be evaluated as a single global concept, others prefer to think of the construct as having multidimensional attributes of teaching (Abrami, 1989).

² Student ratings are the main source used in colleges and universities in the United States, England, Australia and Canada to evaluate teaching (Bess, 1982; Filkenstein, 1984). And Seldin (1993) mentions, nearly all colleges and universities in North America use student ratings.

teaching is misrepresented, the validity of the scores for decision making is to be questioned. In addition, if the evaluation is not soundly evaluating teaching and evaluation results are misinterpreted and misused, then serious consequences could arise affecting the professional careers of faculty members, or diminishing the quality of teaching on campus.

The positivistic thinking that has guided the research has ignored some of the complexities of teaching. Such evaluation has consistently failed to acknowledge the diversity of teaching and its situationality. In addition, the traits or categories used for defining good teaching lack empirical support and are not consistent with approaches to validity that stress multiple perceptions, situationality and complexity (Messick, 1989; Lincoln and Guba, 1989; Stake, 1995). Because research has disregarded the issues of diversity in courses that may require a different role for students and the instructor, more research is needed to address this important gap in the literature.

Purpose of the Study

This study used a holistic approach for examining the meanings, trade-offs, and social consequences of the evaluation of teaching in the context of a research university. The first part of the research was developed around Christy Evans, a junior professor of History who was trying to improve her teaching. As described in Chapter Four, Miss Evans had been a professor at Midwestern University for six years and had a reputation of being a good instructor among the graduate students in her Department. But she was disappointed with the low student ratings that she received from undergraduate students in a “required course.” Last Spring, when she obtained lower ratings than she expected in one of her advanced

courses, she decided to find ways for improving her teaching. After engaging in professional development organized by the instructional specialists of the Office of Instructional Support and Improvement (OISI), she began to implement some changes in her teaching. During spring 2000, while participating in professional development activities provided by another campus office, she learned about this study and voluntarily offered to participate. She wanted to understand how students perceived her teaching and why on some occasions, even when everything seemed to go well, she did not obtain the ratings she felt warranted.

A case study was used as the instrument for understanding how evaluation was conducted, and for identifying the meanings given to the evaluation, the way in which different stakeholders interpret and use evaluation results, and the trade-offs faced by professors trying to improve their teaching. In addition, attention was paid to how issues identified in the case were engaged by campus, college and departmental policies.

The second part of the research focused on how gender, race, ethnicity, and other variables affect the evaluation ratings of women and minority professors on the College of Arts and Sciences at Midwestern University. In addition, the second part of the study gathered information about the meanings that students give to the construct, "good teaching," the process that they follow when they rate their instructors, and the trade-offs that result from using student ratings as measures of instructional quality. By focusing on these issues, the study provides an examination of the content, substantive and consequential validity of the evaluation and addresses issues related to faculty diversity, important gaps of the research on faculty evaluation and diversity in higher education. In addition, the study provides a better understanding of how decisions about teaching are made. This is critical

because misuse of evaluation data can result in diminishing instructional quality, limiting academic freedom or creating a hostile environment for women and minorities.

Foreshadowing Issue / Questions³

The issues/questions that guided the observations, interviews, and document analyses are described next.

Case Study Issues

1. What was Christy's teaching philosophy? Why did she become a professor? Was her philosophy of teaching compatible or conflicting with the accepted philosophy of teaching supported at the Department level? What happens if a professor had a different teaching philosophy than the one adopted by the Department?
2. Were Christy's responsibilities as a researcher conflicting or supporting her teaching responsibilities?
3. Was the Department using more than one source of information to evaluate instruction, had the different sources and methods provided Christy with similar or conflicting evidence of the quality of her teaching?
4. Has the feedback provided by student ratings and other sources influenced Christy's decisions regarding classroom activities? In what ways?
5. What were the meanings of "good instruction" for Christy's students? Are some elements of good teaching more important than others? Did the meanings of good teaching for Christy's students emphasize style over quality?
6. What were the processes followed by Christy's students when evaluating her teaching? Are aspects of cultural sensitivity important to students when rating the instructor?
7. Was there pressure for obtaining higher student ratings in Christy's Department? Was this pressure putting Christy in the dilemma of eliminating important content

³ Issues are "research questions that emphasize trade-offs and contexts" (Stake, 1995, p. 171)

or making other decisions that could diminish the quality of her teaching in order to obtain higher ratings?

8. Had the evaluation provided Christy with valuable information to become more reflective about her teaching?
9. Was the evaluation able to identify the merit and shortcomings of Christy's teaching as relevant to the context in which it takes place?
10. Was the evaluation process inattentive to information about teaching quality that Christy could use for instructional improvement?
11. What could happen if Christy was not able to increase her student ratings? Was she at risk of failing to be granted tenure and/or losing her job?

Issues About the Evaluation of Teaching At Campus, College and Departmental Levels

The following issues were about the context of teaching at the university, college and departmental levels:

1. What was the approach used by the Department to teach History? Was this a course-specific approach or an all-courses approach? What were the trade-offs of choosing one approach over the other?
2. How was teaching being evaluated? What kind of information on teaching quality was obtained from the evaluation, and how it is integrated and interpreted at the department level? How did the departmental and the campus systems interact with respect to teaching quality? Did they support or conflict with each other?
3. What sources of information were being used to evaluate teaching quality? In the Department, were each value the same? What conditions facilitated or inhibited comprehensive evaluation of teaching at the department level?
4. What was the meaning of teaching quality espoused by the Department? Was the definition of teaching quality accepted by the Department flexible enough to allow different types of teaching? Was it responsive to the diversity of both instructors and students?
5. What was the meaning of teaching quality for the different stakeholders involved in the evaluation of teaching?

6. How were the evaluation data interpreted and used at the department level? What was the importance that the Department gave to teaching versus research? Was the emphasis on research productivity interfering with faculty interest in improving their teaching?
7. Were the results from the evaluation used for instructional improvement being kept apart from the promotion process? If not, what were the costs?
8. Was the evaluation of teaching focused only on classroom instruction or did it consider the context and the contribution of individual instructors to the improvement of the Department's teaching?
9. How were the results of the evaluation being used? Were they used to punish faculty members who obtain low ratings? How did administrators and faculty members use the data from the evaluation for instructional improvement? Was the evaluation helping faculty members to become more reflective of their teaching? Under pressures for using the ratings, what were the dilemmas faced by the faculty?
10. What were the consequences of the evaluation? Were there any trade-offs resulting from the approach used for evaluating teaching on campus?
11. How well did the evaluation address issues of equity? Was the evaluation process biased against certain instructors because of their gender, race, or ethnicity?
12. Were the issues identified in the case study pertinent to other professors in the College?

Importance of the Study

A study of the meanings, trade-offs, and social consequences of the evaluation of teaching at a research university is, in part, a study of the validity of the evaluation. Messick (1989; 1995) explains meaning and consequences are essential to validity. The study addresses an important gap in the evaluation of teaching in higher education. As Ory and Ryan (2001) observed, "In the areas of content, substantive and consequential validity very little evidence has been provided or even looked for." (p. 23).

In 2001, Ory and Ryan found that "many evaluation forms used today have been developed from other existing forms without much thought to theory or construct domains" (p. 9). There has been a need for understanding if the scores obtained from the evaluation are really a representation of the quality of the instructor's teaching. If not, what else? Issues of construct representation have special significance when related to issues of diversity (of faculty and students⁴), different types of teaching, and the context in which teaching takes place.

A study of this nature is important because "little is known about how decisions using the information about teaching are made, and about how teachers, colleagues and administrators use this information." (Menges, 1998, p. 3). This is critical because misuse of evaluation data could result in diminishing instructional quality, limiting academic freedom and creating a hostile environment for women and minorities (Haskell, 1997).

Most research on the evaluation of teaching has been conducted using quantitative methods, "conducted largely by persons, who have been trained in the traditions of psychology and sociology, seeking generalizations and controlling for variation, more research using multiple methods . . . and varied sources." Alternatives are needed (Menges, 1998, p. 4). The use of a holistic approach was expected to be especially valuable because it should facilitate an understanding of the process and events taking place and the complexities of the organizational, political, and cultural contexts in which teaching takes place (Patton, 1980;

⁴ An important aspect not addressed in the literature is how instructors deal with issues of equity and cultural diversity (Dwyer, 1993).

Stake, 1995). In addition, its use addressed a gap in the literature since little research has been conducted using this approach.

A study of this nature has implications for evaluation practice, curricular policy and professional development. Findings of the study may lead to the improvement of the evaluation system and the design of professional development programs for faculty and administrators, especially regarding the use and interpretation of evaluation data.

The real merit of the study, however, is in clarifying the importance of teaching within faculty careers. Universities have a serious responsibility for evaluating and improving the quality of their teaching not only for accountability purposes but also because of a responsibility to prepare professionals for the society they serve. Serious social and professional consequences could result if the evaluation misrepresents the quality of teaching. Issues of discrimination, the diminution of instructional quality and violation of academic freedom can increase when teaching is misrepresented and evaluation data are misused.