

CHAPTER 9

CONCLUSIONS

This study examined the validity of the evaluation of teaching at a research university, particularly for representing the quality of an instructor's teaching. In the second part of the study issues of substantive and consequential validity were studied by examining the meanings given to the construct "good teaching" and the processes followed by undergraduate students when rating their instructors, as well as to how race, gender and ethnicity affected the evaluation ratings of a sample of women and minority professors on campus.

The first part of the research was developed around Christy Evans, a junior professor with a scholarly approach to teaching. Evans was a member of the History Department at Midwestern University, a department well known for the quality of its teaching, and with a strong research reputation. The Department was also known for its progressive orientation and for support for a scholarly and cultural orientation, as well as for stressing the importance for instructors to adjust their teaching to contextual needs. Evans's teaching philosophy fit well within this approach.

As described in Chapter Seven, Christy Evans 's teaching was evaluated according to campus and college policies. Teaching was evaluated at the department using two main sources of evidence, student ratings and peer reviews. Student ratings of instruction took priority over other evaluation data. These ratings were collected through the campus Teaching Evaluation System (TES) at the end of each academic semester. Evans's Department had also developed its own peer evaluation system, using it to complement the ratings and to provide

more information about the context of teaching. The peer evaluation system provided all junior professors, during their tenure-track years, with classroom observations conducted by members of the Department's Executive Committee. The observers conducted these observations every semester and wrote a common report for both the administration and the instructor. Other evidence collected for evaluating teaching included student projects, course materials and artifacts, examinations, and the instructor's teaching philosophy, all part of the instructor's teaching portfolios. The Department had assigned junior faculty members to a mentor who shared similar research interests. According to the Department Chair, mentors were expected to provide evaluative advice on teaching, research, service, and the overall tenuring process.

The case study of Christy Evans showed that different evaluation sources provided different portrayals of her instruction. While colleagues rated Christy's teaching high, students had a wide range of perceptions, sometimes conflicting, about her teaching. Undergraduate students taking the course as a requirement outside their major, and thus were not motivated in taking the course, had more negative perceptions of her teaching. Some of these students explicitly stated that they did not consider the course content to be worthy. Graduate students, on the other hand, had more positive views of Evans's teaching, as did the undergraduate students taking her advanced undergraduate courses. Differences in perception appeared not only between graduate and undergraduate students, but also among students taking Christy's required and elective courses. The non-History students complained more about grading procedures and the kind of work she assigned. An analysis of 2000 TES evaluation forms of Christy's teaching collected over a period of 5 years revealed that

students who preferred a teacher-centered style of teaching had a more negative attitude perception of Christy's teaching style than those preferring a student-centered teaching style.

Peer evaluations of teaching provided insight into the contexts of Christy's teaching. Since ratings took place only once a semester, little student information about the continuity of teaching was available. Clearly, supervisors lacked a fully 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 faculty 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 contributed to incomplete reports as to quality of her teaching.

Both student ratings and peer faculty observations for evaluating teaching had strengths and weaknesses. For Christy Evans, the most important value of the ratings was their shock effect, bringing realization that there were problems in her teaching that she needed to address. For Evans and most professors and administrators in the Department, the greatest value of the student ratings were the comments made by students about their experience in the course. In a follow-up of the ratings over a period of ten years, Department Administrators concluded that student ratings were more discriminating when used in courses with a medium-class size. They were persuaded that almost every professor in the Department could obtain high ratings in courses with a class size of ten, and that almost anyone teaching a class with a hundred students or more would receive low student ratings. So, they paid more attention to ratings obtained from courses with a class size of thirty-to-forty. But class size was only one of the contextual aspects that they considered when

reviewing the ratings. The less useful part of the evaluation was said to be the score and the forms used in reporting the results. The administrators considered these forms burdensome. The members of the Department Executive Committee found that student ratings were often consistent with peer evaluations of teaching, but not always. According to the members of the Committee, sometimes student ratings reflected student lack of motivation toward the course content, or preferences toward the use of technology. Sometimes they signaled student preference for simple content. The members of the Executive Committee and the Department Chair also were persuaded that the ratings were drawn down by heavy assignment load, expectations of hard grading, and unwieldy class size. And race, gender and ethnicity, even the instructor's accent apparently influenced the perception. The faculty also was confident many students did not intend to be objective evaluators and expressed their dislike for the person. For some it was "pay back time." (The Department Administrators investigated claims made by students, trying to ensure that the claim was based on a teaching problem and not an issue of student bias against the instructor). The only variable that the Department Administrators found consistently influencing the student ratings of their instructors was the gender of the instructor, with men getting higher scores. The negative effect of gender was said to be highest for the youngest women faculty members.

Peer evaluations provided different views including an opportunity for confirming the student ratings. These faculty observations were useful for obtaining information about the context of teaching, important to teachers and less important to students. They presented some problems of objectivity. The dual role of the observers as members of the Executive Committee and colleagues were acknowledged to cause omissions of critical information.

Some observers seemed disposed to write positive reports that could be used for counteracting the low student ratings and helping the instructor make a good case for tenure. In some occasions, peer observations were influenced by the instructor's reputation. The observations usually took place only once per semester.

There was no single "departmental view" when integrating different evaluative data for making the judgment about the quality of the instructor's teaching. Each of the members of the Executive Committee drew upon common and personal information. The view was a composite of the common and personal information. For each instructor, it was an informal synthesis, not worked out formally by the Executive Committee and not informed by specialists in data aggregation and scaling. The view was shaped by student ratings, peer review, teaching portfolio, review for awards and benefits, reputation, and informal complaints. In most departments, only the student ratings were quantified. The aggregate was important information but a confounded expression of teaching quality. Almost no validity research on those inputs was conducted.

Student ratings were defined and constrained by the items and procedures the department used. Peer evaluation by members of the executive committee when shared with the instructor, were constrained by collegiality, but also shaped by a need to maintain a collective image of good teaching across the Department, partly to support the research mission of the Department.

The view of the instructor's teaching quality emerged differently for different purposes. For facilitating efforts to improve the teaching, the view was one thing. For

assessing the instructor's contributions to the Department and the instructor's readiness for promotion and for assuming advanced responsibilities, the view was different.

The validity (the quality of the Department perception of teaching quality) was not the same for the perception of teaching in order to facilitate ways of improving it and when considering reward for good teaching. (There may be other uses of the ratings and perceptions that require separate questioning of validity).

Using these and other evaluative procedures affected the instructor morale, department integrity, student credibility, and other aspects of the well-being of the instructor. These consequences need to be considered as part of the validity of the Department procedures and perception of rating teaching effectiveness.

The Christy Evans' case study raised five issues deserving serious review. The first was the pressure for accountability, especially the emphasis from outside agencies for measuring student outcomes even when they are difficult or impossible to measure. This pressure appeared to cause more teaching of "facts", less of historical interpretation. The second issue was a growing tendency to use staff from the Office of Instructional Support and Improvement (OISI) to support initiatives for improving (and thus defining) the quality of teaching at Northwestern. Because the staff from that office followed an approach for teaching improvement that stressed pedagogical principles (and not content and project development), campus-wide adoption could result in excessive emphasis on the technical principles of teaching. Defining teaching purpose as technical principles of teaching may not always be appropriate for different course contexts and conditions across campus. This too

could affect OISI staff load, given that growing demands for support would be in addition to the services they were already providing with a relatively fixed budget and staff.

The third issue was Midwestern University's growing emphasis on using technology-based teaching, with complex implications for the evaluation of teaching and for faculty professional development. The fourth and fifth issues were about the equity protections when students rate lower those instructors of different race, gender or ethnicity, and about the level of good faith and seriousness of students filling the TES evaluation forms. These two issues had special importance invoking the substantive and consequential validity of the evaluation.

Because of this, and because issues four and five could result in civil rights violations and contributed largely to the cultural integrity of the campus community, the second part of the research focused on studying these issues in depth. In this part, the problems were studied by collecting data from faculty members and undergraduate students of the College. Invitations to participate in the study were sent to all women and minority professors at the College, and to a random sample of 1523 undergraduate students. Of the 200 women and minority instructors invited to participate, twenty provided their Teaching Evaluation System (TES) Summary Profiles for analysis. Four of these instructors were interviewed plus another group of ten minority instructors who did not provide TES profiles. They all were asked about the influence of several variables on the TES ratings. Of the 1523 students invited, only 200 students responded to the online survey on the issues.

As described in Chapter 8, the analysis of the TES Summary Profiles did not provide evidence of a pattern of low scores for minority and women instructors as compared to

campus norms. Uncertainty here was partly due to the changing nature of the norms each semester, and to the many differences found among the instructors. Differences associated with the course, students, and personal characteristics in some cases seemed to influence TES ratings, but not in all. Interviews with the fourteen minority and women professors, clearly supported the perceptions held by Christy's colleagues and History Department Administrators about the sometime negative influence of gender, race and ethnicity on student ratings of instruction. According to some of the interviewees, the influence of race, gender and ethnicity was difficult to pin down in the TES evaluation forms because students rarely made verbal comments of this nature in the forms, although they may have expressed a bias numerically. The presence of these biases, they said, could be observed during instructor-student interactions and in the students' attitudes towards the special knowledge and expertise of the instructor. One international instructor provided an example about how race relations could influence the classroom climate and thus could influence the instructor's teaching and the ratings. Some instructors spoke of these variables being moderated by instructor experience, the nature of the course being taught, class size, instructor teaching style, and teaching philosophy.

Another variable identified by the interviewees also identified by Christy's peers was the influence of course content, especially in undergraduate "required courses" that focused on a non-Western culture. The interviewees shared a belief that low ratings for instructors of these courses were partly attributable to having different kinds of students (freshmen, junior, sophomores, seniors; major and non-major students) with different prerequisite knowledge and expectations. Especially when the course enrolled over a hundred students and had more

than a single teaching assistant, student complaints about fair grading and poor teaching were seen to increase.

The findings from the online survey indicated that students, as expected, paid attention to different characteristics and behaviors of the instructor. They of course differed in defining good teaching. For some of them good teaching was expressed in terms of characteristics or behaviors of the instructor, such as: being open-minded, caring, confident, enthusiastic, patient, responsible, creative, with a good sense of humor; sensitive to student needs, fairness and with good presentation skills. Other students defined good teaching more in terms of their own like or dislike of the course content and the amount of the course workload. For another group of students good teaching existed when the instructor created an environment in which there was no sexism or racism and where learning was fun. A few students said that for them good teaching was reflected the instructor challenged them to learn. There were also students, who judged the quality of teaching based on the quality of the class notes provided by the instructor; the way in which the instructor covered the course content, his or her teaching style, and the clarity of the instructor expectations. In spite of the variety of opinions of the students, most saw good teaching as teacher-centered, that what the teacher had concluded what was more right and important than what the students were concluding. Most expected teaching to focus on knowledge reproduction, indicating that responsibility for student learning was a responsibility of the instructor. One student said: "If the instructor doesn't care, why should I?"

The second part of the study also indicated that these Arts and Science undergraduate students followed different processes when assigning their ratings. Some said

they compared their instructor to an ideal of teaching; some to instructors they had before. Some students said that they reflected mostly on their own experience in the course when rating their teachers. There were differences in the ways in which the students assigned the highest, lower and middle scores on the TES forms. Some did not feel comfortable assigning extreme scores, where as others said that they would never give a "middle" score, only assigning very high or very low ratings.

Although these students did not always agree, of course, among themselves about the most important criteria for good teaching, they consistently referred to fair grading, clarity, and good communication skills as influences on their evaluations. Good communication was explicitly defined by some students as speaking without an accent. Indeed, some students said that they would not mind having an instructor of another race or gender unless that affected the instructor's ability to communicate in English.

Fair grading was identified by the students as one of the most important elements of good teaching. They defined fair in different ways. Most said grading would be unfair if there were inconsistencies with pre-established grading rules, or the workload or level of difficulty of this course in relation to other courses. There were students who said that not obtaining an "A" or not being rewarded with a high grade for their hard work was unfair.

In general, the findings of the student online survey supported faculty perceptions that on some occasions, students were not focused on evaluating the quality of teaching. Also, that student ratings could be influenced by their perception of the worth of the course content and their grade expectations. In addition, the study again found that student motivation towards the course could influence the ratings. Although most students said that

having an instructor of their own gender, race and ethnicity was not important, some granted that this could negatively influence their ratings, confirming the position of several professors interviewed in the two parts of the study.

The overall findings of this study were similar to those obtained from research conducted by the OISI staff showing that students define good teaching by combining both course and instructor variables. Findings also were consistent with those of Shu-Wan Chiu (2000) and other researchers, (such as Aleamoni & Graham, 1974; Feldman, 1978; Kulik and McKeachie, 1975; Marsh, 1984, 1987 Wigington and others, 1989), who found evidence that elective courses receive higher ratings than do required courses, and that graduate courses receive higher ratings than undergraduate courses. In addition, the findings of the study were consistent with the findings from the researchers named above regarding class size influencing the ratings. This study also found support for the findings of Cuban (1999), who identified "role and purpose" conflicts in research universities facing pressures for accountability. The study, however, did not support Cuban's perception that scholarly teaching was appreciated only by graduate students. Students taking Christy Evans undergraduate courses during Spring 2000 rated this kind of teaching highly. Some of these students stated that although challenging, scholarly teaching helped them develop their analytical skills. Skills they could apply to other courses and different learning situations.

Overall Conclusions and Implications

Midwestern University was an institution under public pressure to produce high quality teaching and learning outcomes. Its leaders needed evidence of the quality of its

programs. The pressures for measuring outcomes created a difficult situation for the University, given that many of the outcomes produced are, as for most social service institutions, intangibles, difficult to quantify. Selecting pre and post measures of institutional effectiveness has been problematic because of the complex multiplicity of variables that relate to University accomplishment. An example of the problems of measuring student outcomes taking place in Christy Evans's Department was an Ad-Hoc committee struggling to compare non-equivalent student papers from two different periods of time. As some of the interviewees noted, some outcomes may happen before the next quiz but most take months and years.

Although almost all of the professors interviewed said that pressures for accountability did not interfere with their priority of research over teaching, the increasing importance of TES ratings as teaching outcomes was evident in the case of junior instructors, identified in the second part of the study. Under pressure to improve their ratings, some tried different remedies, even if only to manipulate the system. Such faculty efforts were identified by the interviewees, such as inviting students for dinner a few days before administration of the TES forms and remaining in the classroom to tell students how to fill the TES forms in a way favorable to the instructor. Obviously these raise ethical issues and deserve further study.

This College's practice of rewarding teaching heavily on the basis of TES ratings was not seen to need research validation. Almost all the professors interviewed indicated that if a professor did not receive a TES score lower than 2.5 points (using a scale of 5 points) nor caused the Department to receive any student complaints, he or she had no reason to worry

about teaching quality. But, emphasis on the student ratings is having instructors and administrators become more concerned with increasing the "numbers" than with improving the quality of teaching. This is especially serious for departments that were not only using the TES ratings as the only source for evaluating teaching, but also publish the ratings of their faculty and use TES as the only information for making decisions about teaching.

The existence of a variety of approaches to professional development at Midwestern was foresighted and beneficial, given the many different orientations of the faculty interviewed. The support for different types of professional development was consistent with the belief of campus, college and department administrators interviewed that different teaching styles and philosophies should coexist. However, the growing trend of using the "best practices" approach to support new professional development initiatives could be problematic. Any trend leading to the selection of a particular model for professional development will serve poorly a comprehensive and diverse institution of Higher Education.

The professional development provided by OISI was consistent with many of the aspects identified by students as important when evaluating teaching. However, OISI did not address some important areas for professional development that deserve more attention, including plagiarism and student diversity in the classroom. OISI followed a professional development approach that stressed many aspects of teaching important to students. But the OISI approach was not a good fit for instructors like Christy Evans with a different teaching orientation. It could result in instructors paying attention more to technical aspects of teaching than to scholarly aspects of teaching. For example, after participating in the professional development provided by OISI, Evans started worrying more about the number

of lines in her overhead transparencies than on how to engage her students in meaningful learning experiences. In the end, Evans was able to keep a balance using both technical and non-technical suggestions for improving her teaching, but this was because she was exposed to different kinds of professional development and because she found other groups of support in her peers inside and outside her Department. Departments less invested in the complexities of teaching, ones that rely only on one kind of professional development, may press for changes that will not result in better instruction.

Campus administrators need to be sensitive to the existence of a variety of evaluative perceptions of what good teaching is, not only among different stakeholders but across different groups of stakeholders. Creating standardized systems for evaluating teaching and for providing professional development to faculty on campus is not without problems. A system can be practical and convenient, but a single approach for teaching evaluation and development can be insensitive to the needs of different students and the different teaching orientations of the faculty. Enforcing a "best practices approach" can result in discouraging other approaches that are needed for special situations and that should coexist within an institution of higher learning.

In addition, the campus practice of having courses with a class size of 100 or more is not only pedagogically questionable, inconsistent with the needs of many students. The overwhelming majority of the undergraduate students who responded to the online survey and those who completed the evaluation forms of Christy Evans' undergraduate courses asked for a closer relationship with the instructor. Many spoke of needing more instructor-

student interaction. The university policy of assigning teaching assistants to work with students in these large courses in small group discussion, added to the problem of fair grading.

Claims of campus administrators and faculty members that teaching and research fully support each other has little backing from research on the relationships between teaching and research. Indeed, as Fedman (1987) stated, “ the likelihood that research productivity actually benefits teaching is extremely small or that the two for all practical purposes, are essentially unrelated... . Productivity in research and scholarship does not seem to detract from being an effective teacher” (p. 225) and vice versa. As Hattie & Marsh (1996) said, “We must conclude that the common belief that teaching and research are inextricable entwined is an enduring myth.” (p. 529). An exception to this, could be the case of a scholarly orientation to teaching especially in the natural and medical sciences (Kyvik and Smeby (1994).

Professors teaching large undergraduate courses appear at a special disadvantage in establishing linkage a relation between teaching and research. Most undergraduate courses emphasize acquisition and use of knowledge through memory, making it difficult for the instructor link with the theory building, groping, hypothesis aspects of research. On the other hand, those instructors teaching graduate students or advanced undergraduate courses that ask for “the abstraction of meaning, and emphasize learning as an interpretive process aimed at understanding reality”(Hattie & Marsh, p. 531), are in a better position to link their teaching and research.

Implications and Issues at the Department Level

The History Department faced trade-offs similar to those reported by Cuban (1999) in his study of the History Department at Stanford University, trade-offs related to the teaching of History as a scholarly discipline. Stanford teachers were pressed to reflect on their own teaching methodology, instead of on preparing students for becoming good citizens and for addressing their post-graduation expectations. A scholarly approach to teaching will not fit the learning appetites of many undergraduate students taking History courses. In addition, the decision of the History Department at Midwestern to stress a cultural and thematic approach, promoting specialization rather than universalization, results in the trade-offs described in Chapter Six. This is problematic for students. They may be caught between expectations of the University to address History questions in a scholarly way and expectations of the State for future History teachers and their employers to focus on facts, simple structure rather than nuanced interpretation.

Preserving the role of research over teaching and, at the same time, satisfying external calls for accountability, is difficult for the University. It was an important issue for the History Department. For instructors like Christy Evans, whose teaching style is compatible with her Department teaching philosophy, there was little problem in addressing Department and campus expectations. She did disregard the longing of some students for providing the answers to tests in an interesting way. Some had problems with this teaching style. The overwhelming majority of the students who completed the online survey supported fact-based presentations. Some of the students who rated Evans's teaching low in her undergraduate "required courses" were students who valued this kind of approach to

teaching. A conflict of expectations between departments and students raises the question about what can be done. Both expectations are legitimate.

Implications for the Evaluators

The Office that designed and administered the campus-wide system for evaluating teaching at Midwestern worked hard at improving the system. Some members indicated that if there were problems with the rating system, it was partly because administrators did not use the data appropriately and because students and faculty did not take the evaluating seriously enough. Although, some of the interviewees identified cases of departments giving little attention to teaching and its evaluation, the overwhelming majority of the professors, administrators and students who participated in this study, affirmed their concern for good teaching and tried to make sense of the results of the evaluation provided. Many said that the problem with the evaluation system was that campus administrators did not take seriously its professional development purpose. They saw student ratings created to preserve the status quo rather than to improve teaching, for control rather than improvement.

OISI has highly qualified staff for conducting the evaluation of teaching on campus. Members had received training in both quantitative and qualitative research methodologies as well as on program and personnel evaluation. The staff was knowledgeable of the research on evaluation of teaching in higher education and the limitations of different data sources for evaluating that teaching.. However, their work as evaluators promoted both good and not so good practices.

On one side, they worked very hard to increase faculty evaluation within partly by giving them opportunity to propose items to be included in the departmental evaluation

forms. They identified stakeholders that could be affected by the evaluation and encouraged department administrators to include stakeholders whose needs were not already addressed. They also provided timely reports and encouraged follow-up of the evaluation. In addition, the OISI staff used practical and cost-effective procedures for evaluating teaching and protected the anonymity of the students responding to the evaluation forms.

The evaluators, however, failed to follow some tenets of sound evaluation practice. First, this evaluation of teaching, using student ratings, designed as a survey administered at the end of the academic semester, failed to provide a broad array of information about the quality of teaching. Focus-group interviews with the professors' students were provided only on a request basis. Other sources were omitted. Second, there was little monitoring of the process of obtaining information from students. As the Personnel Evaluation Standards stress, the assessment gathering procedures should be chosen or developed and implemented in a way that ensures that "inferences concerning the evaluatee are valid and accurate" (p.1). Third, the evaluation reports provided by the OISI staff lacked information that decision makers needed for understanding the findings and for making appropriate decisions about teaching. For example, the TES reports and TES Summary Profiles lacked information about the limitations of the evaluation and the rationale for interpreting evaluation results, as well as about the evaluation context. Many of the interviewees complained about the lack of clarity of these reports. All of this raises issues about the completeness and fairness of the evaluation, as well as about its accuracy for representing the quality of teaching. These issues draw from *The Guiding Principles for Evaluators* (AEA, 1993).

Although evaluations conducted by the OISI staff took place alongside evaluations conducted by other offices on campus, the way in which the student rating system was implemented failed to fully address the Guiding Principles. By not monitoring carefully the process for collecting the ratings from students, technical standards for conducting their work were not adhered to, nor were the limitations of their methods, and the values, and assumptions of their practice were not set forth. The evaluators failed to adequately identify differences among participants, such as those related to culture, religion, gender, disability, age, sexual orientation and ethnicity. Members argued that since the literature on faculty evaluation had not found significant correlation in meta-analyses conducted in other settings, there was no need for conducting such analyses here. Limitations on budget and staff were said to prevent it. However, not addressing these issues in the conduct, analyses and reporting of the evaluation was not responsive to sound evaluation practice. The overwhelming majority of faculty members interviewed stated that they had raised claims about the capacity of the TES ratings system for representing the quality of teaching of women and minority instructors. The findings of this study included evidence that those variables may influence evaluation results, and concluded that their influence operates in conjunction with other variables. It was the evaluators' ethical and professional responsibility to publicize these cautions.

Campus administrators and evaluators should make a strong effort to ensure that the evaluation process is improved and does not result in unfairness or inequity, not only to minimize the legal vulnerability of the institution but also because it is in the best interest of the institution to properly assess its faculty work.

Research Limitations and Future Directions

This study provided evidence of the complexities of teaching and its evaluation within the context of a research university. It also provided a better understanding of the meaning of the construct “good teaching”, the evaluative process followed, and the social consequences that could result from the evaluation, particularly with regard to equity concerns. The study addressed an important gap in the literature of teaching in higher education by examining issues related to the content, substantive and consequential validities of the evaluation. It added to the literature on cultural diversity in higher education.

Because the small response rate, it is not possible to assert that the findings of the study are applicable to the collective faculty members, students, administrators, colleges and departments of Midwestern University.

The main value of the study is in the issues that it raises, especially those regarding the validity of the evaluation of teaching. Researchers in the field of faculty evaluation stress the importance of the reliability of student ratings of instruction --as found in their analyses of different sections of a course taught by the same instructor. But, as the findings of this study suggest, an over-reliance on consistency disregards essential aspects of validity. Consistency does not assure quality in the rating process. Consistency does not assure good interpretations nor avoid bad consequences. Good consequences of an evaluation system are affirmed by its validity.

A valid evaluation of teaching on campus represents the complexities of teaching, including a variety of evaluative perceptions as to what good teaching is within its context. Using a standardized procedure for evaluating teaching has administrative advantages (e.g.

simplicity), but standardization can result in misrepresentation of the construct being evaluated. Serious consequences can come from decisions based on the misinterpretation of evaluation results. A holistic approach of inquiry involving the use of different methods and sources of data collection, including participant observation, in-depth and focus group interviews, and document analysis, like the ones used in this study, can contribute to understanding the issues and to bringing contextual information vital to the support of high quality teaching on campus.