

David Ackerman
Barbara L. Gross

and

Franck Vigneron
California State University, Northridge

Peer Observation Reports and Student Evaluations of Teaching: Who Are the Experts?

This study is an exploratory inquiry into the perceptions of university faculty regarding two forms of teaching evaluations, student evaluations of teaching (SET), and peer observation reports (POR). Which, if either, better assesses the quality of instruction? Who are the real experts in judging teaching quality: peers who are experts in their field or students who are the recipients of instruction? Results suggest that it is better not to rely on a single source of information as evidence of teaching effectiveness. SET and POR provide complementary information from differing perspectives. Advantages and disadvantages of both are discussed.

Ce projet de recherche est une enquête portant sur les perceptions des professeurs universitaires relatives à deux formes d'évaluations de l'enseignement : les évaluations par les étudiants et les rapports d'observation par les pairs. Quelle méthode évalue le mieux la qualité de l'enseignement : les pairs, experts dans leur domaine, ou les étudiants qui a qui on transmet l'instruction? Les résultats donnent à penser qu'il vaut mieux ne pas se fier à une seule source d'information comme preuves de l'efficacité de l'enseignement. L'information découlant des évaluations par les étudiants et celle que fournissent les rapports d'observation par les pairs proposent des perspectives différentes qui se complètent. Une discussion des avantages et des inconvénients des deux méthodes d'évaluation vient terminer l'article.

Introduction

Teaching, research, and service constitute the three primary factors considered in university faculty retention, tenure, and promotion (RTP) decisions. Teaching is typically assessed using student evaluations of teaching (SET), sometimes accompanied by embedded learning measures and/or teaching portfolios. There has also been increasing use of teaching assessments by faculty peers resulting in peer observation reports (POR). This trend raises the question of how these two assessment tools differ as well as the implications of the differences. Examples of typical questions asked in SET are provided in Table 1. Responses are usually indicated on a scale (e.g., a 5-point scale) ranging from *strongly agree* to *strongly disagree* or perhaps *excellent* to *poor*. SET also typically provide space for free-form comments, asking students to indi-

David Ackerman is an associate professor of marketing.

Barbara Gross is a professor of marketing and Chief of Staff, Office of the President.

Franck Vigneron is an associate professor of marketing.

Table 1
Example Questions Asked in SET¹

- The instructor's objectives for the course have been made clear.
- The instructor identifies what he/she considers important.
- The instructor provides an adequate syllabus that identifies course content and evaluation.
- The instructor informed students how they would be evaluated in the course.
- The course requirements and grading standards were made clear.
- Exams and/or graded assignments accurately measure student learning in the class.
- The instructor has students apply concepts to demonstrate understanding.
- The instructor emphasizes conceptual understanding.
- The assignments helped in learning the course material.
- The reading assignments helped me learn the course material.
- The writing assignments and/or homework exercises helped me learn the course material.
- The instructor provides constructive comments on students' work.
- The instructor gave helpful feedback on my work.
- Grading is fair and impartial.
- The instructor graded my work fairly.
- The instructor is well prepared.
- The instructor organizes the course well.
- The course is properly paced.
- The instructor explains clearly.
- The instructor explained difficult material clearly.
- The lectures and examples helped me understand difficult concepts.
- The instructor knows if the class is understanding him/her or not.
- The instructor makes efforts to be sure that the students are following the presentation.
- Classroom presentations are good.
- The instructor communicates recent advances in the field.
- The instructor presents origins of ideas and concepts.
- The instructor demonstrates extensive knowledge of the field.
- The instructor appears to know the subject matter well.
- The instructor is enthusiastic about the subject.
- The instructor seems to enjoy teaching.
- The instructor has interest and concern in the quality of his/her teaching.
- The instructor encourages class discussion.
- The instructor encouraged class participation.
- The instructor presents points of view other than his/her own.
- The instructor invites criticism of his/her own ideas.
- I felt free to ask questions and to express my opinions in class.
- The instructor is approachable and is sensitive to students' needs.
- The instructor is careful and precise in answering questions.
- The instructor is responsive to student questions in and out of class.
- The instructor responded to student questions in a helpful way.
- The instructor is professional and courteous in his or her treatment of students.
- The instructor treated students with respect.
- The instructor relates to students as individuals.
- The instructor is willing and available to help students.
- The instructor has a genuine interest in students.
- The instructor is valued for advice not directly related to the course.
- The instructor is punctual in meeting and conducting class.
- The instructor is available during office hours.
- The instructor was available and willing to help during office hours.
- The instructor is a dynamic and energetic person.
- The instructor has an interesting style of presentation.
- The instructor stimulates interest in course material.
- My interest in the subject area has been stimulated by this course.
- The instructor was intellectually stimulating and prompted me to think creatively and critically.
- The instructor makes an effort to minimize cheating/plagiarism

Table 1 (continued)

-
- This course is among the best I have had at this university.
 - This professor is among those from whom I have learned the most.
 - I learned a lot from this instructor.
 - The instructor is a very good teacher.
 - I would recommend this instructor to other students.
-

¹Questions are examples from SET used in randomly selected departments from five colleges in the authors' university. The sample was drawn by our Director of Institutional Research. The colleges and departments represented are the College of Arts, Media, and Communication (Journalism Department); the College of Business and Economics (Marketing Department); the College of Humanities (Philosophy Department); the College of Science and Mathematics (Biology Department); and the College of Social and Behavioral Science (Urban Studies and Planning Department). The number of questions asked by these individual departments on their SET ranges from 12 to 18.

cate what they liked best about the course, how they think the course could be improved, and the like. POR sometimes use a combination of standardized questions and narrative, but most POR emphasize narrative (Kohut, Burnap, & Yon, 2007; Seldin, 1999a, 1999b).

In this article, we examine the views about SET and POR expressed in past literature and report the results of a qualitative study that explored faculty members' views of both SET and POR. In addition, we look at whom university faculty members most perceive as the experts in evaluating teaching quality. Are they other faculty members? Or are they students?

Overview of Relevant Literature

Student Evaluations of Teaching (SET). The literature on SET often points to instructor concerns about student evaluations of teaching. Based on the literature, these largely center around two factors. The first is the questionable validity of SET. As educational institutions increasingly view students as consumers of education, they are also more influenced by reports of student satisfaction (Share, 1997). Unfortunately, reports of student satisfaction may be more based on expected grades (Olshavsky & Spreng, 1995), similarity of thinking styles (Betoret, 2007), and social factors (Blackhart, Peruche, DeWall, & Joiner, 2006) than on actual learning outcomes (Marks, 2000; Simpson & Siguaw, 2000). Expected grades frequently, although not always, correlate with positive ratings of professors, and rigor is often negatively related or believed to be negatively related to SET (Clayson, 2004; Clayson & Haley, 1990; Heckert, Latier, Ringwald-Burton, & Drazen, 2006; Sojka, Gupta, & Deeter-Schmelz, 2002). As educational institutions value SET in their assessments of instructor performance, faculty members may, therefore, feel tempted to lower standards and inflate grades (Churchill, 2006; Fram & Pearse, 2000; Share). With such concerns in mind, Wright (2006) suggested implementing a tracking system where the students of faculty members who had received low ratings could be randomly sampled and then interviewed in depth to discover what really went on in the class. He argued that such a system might protect skilled and effective instructors whose low ratings largely reflect that they are demanding teachers.

Second, faculty members are apprehensive about the emphasis given SET in the retention, tenure, and promotion (RTP) process (Simpson & Siguaw, 2000).

It is argued that SET often reflects perceptions and impressions based on instructor personality and likeability, and on course workload and difficulty, rather than on assessments based on teaching effectiveness and actual learning (Blackhart et al., 2006; Clayson & Haley, 1990; Clayson & Sheffet, 2006; Marks, 2000). Indeed most questions listed in Table 1 focus on such perceptions. Clayson and Sheffet found student perceptions of instructors' personality and evaluation of instruction to be strongly related and concluded that students associate personality with instructional effectiveness.

Thus some faculty members distrust SET because they view SET scores as little more than measures of a faculty member's popularity (Chandler, 1978; Clayson & Sheffet, 2006) or even as an opportunity for students to exact revenge on a rigorous instructor (Wright, 2006). Because of concerns about whether student evaluations of teaching truly reflect an instructor's effectiveness in facilitating student learning, the emphasis on SET in retention, tenure, and promotion (RTP) decisions is disquieting.

McKeachie (1997) in a review of literature agreed that student ratings are valid measures of teaching, but that contextual variables such as grading leniency may influence the level of ratings. As such, he argued that the problem lies in a lack of sophistication among personnel committees who use student ratings. Thus more attention should be directed toward methods of ensuring more careful and thoughtful use.

Similarly, Greenwald (1997) concluded that much disagreement exists about the validity and usefulness of SET. Considerable concern revolves around construct validity, the degree to which SET measure what they purport to measure. The knowledge that students gain from a course should be a direct function of teaching effectiveness, yet SET often emphasize other factors that have little to do with actual learning (Clayson & Sheffet, 2006; Lang, 2007; Marks, 2000). Chonko, Tanner, and Davis (2002) found that the top six expectations of professors cited by students were that the instructor is interesting, helps students, communicates well, is easy to talk to, has a good personality, and is kind. Far fewer students cited characteristics such as whether the professor is concerned about whether students learn, is challenging, and is knowledgeable.

Considerable research on the concept of teaching shows teaching to be multidimensional and calls for instructor evaluation across a variety of dimensions (Feldman, 1989; Marsh, 1991; Seldin, 1999a, 1999b; Smart, Kelley, & Conant, 2003). Some of these dimensions such as the ability to facilitate problem-solving, to foster creativity, and to sharpen critical thinking are infrequently addressed in SET instruments (Seldin). However, these are typically key learning objectives in courses emphasizing active learning pedagogies. Faculty members who favor active learning pedagogies such as real-world experiential projects sometimes also lament that evaluations are administered too early in the term before students can fully appreciate the value of the experience (Gaidis & Andrews, 1990).

There also are some differences in how students and faculty generally view the dimensions of effective teaching. Student and faculty views of one another depend on how much they agree on the characteristics of excellent instructors (Goldstein & Benassi, 2006). Interviews with students reveal the attributes most

associated by students with outstanding teaching to be rapport, delivery, fairness, knowledge and credibility, organization, and preparation (Faranda & Clarke, 2004). Studies have found generally high correlations between students' and faculty members' evaluations of teaching, but also significant differences between the weights placed on the various dimensions. In addition, some dimensions valued by faculty members such as facilitating the achievement of key learning objectives and modeling rigorous thinking may be unrecognized by students (Buskist, Sikorski, Buckley, & Saville, 2002; Schaeffer, Epting, Zinn, & Buskist, 2003).

Peer Observation Reports (POR). For these and other reasons, a growing number of institutions of higher learning now require evaluations by faculty peers based on in-class observations, that is, peer observation reports (POR). Seldin (1999b) reported that the use of faculty peer evaluation nearly tripled between 1978 and 1998. The increase was partly in response to regulatory and political action calling for increased accountability in higher education, and/or the requirements of accrediting agencies similarly demanding accountability for academic performance. DeZure (1999) suggested that peer observations are more valid than SET for evaluating such learning-oriented criteria as substantive content, pedagogical content, and ethical standards of practice. Kohut et al. (2007) found that a majority of faculty respondents in their study (both faculty members who evaluate and faculty members who are evaluated) regard peer evaluations as valid and useful means of evaluating teaching effectiveness. However, the acceptance of POR has not increased as rapidly as predicted, and some regard it as involving a highly subjective process (DeZure; Kohut et al.; Osborne, 1998; Yon, Burnap, & Kohut, 2002). For example, POR is more widely used in the United Kingdom than in the United States (Kohut et al.).

Seldin (1999a) concluded that teaching evaluations in general have often failed on at least two counts. First, they frequently provide inadequate motivation for faculty members to improve teaching. Second, they largely fail to distinguish between poor, adequate, and good teaching. These failures are attributed to a variety of causes. For example, many evaluation programs, although conceived of as mechanisms to improve the quality of teaching and to support faculty development, are in practice used primarily for administrative purposes in the retention, tenure, and promotion process (DeZure, 1999; Osborne, 1998). As such, faculty members may react to evaluations as carrying an implicit threat of punitive consequences, as opposed to valuing evaluation as a source of feedback for continual improvement (Seldin).

According to Yon et al. (2002), many university instructors see POR as less than optimal evidence of effective teaching, and they are uneasy about the use of POR as a critical part of the promotion and tenure dossier. Several authors (compare Keig, 2000) have concluded that teaching is more likely to improve when evaluation programs are expressly designed for instructional improvement (formative evaluation) rather than through the retention, tenure, and promotion process (summative evaluation). In addition, instructors are sometimes concerned that isolated classroom visits may not provide a full picture of teaching quality and that students are better equipped to assess a holistic experience of the course. This factor has led to interest in the use of so-called "mystery students" (Douglas & Douglas, 2006). Similar in concept to mystery

shoppers, mystery students may be professional teachers playing the role of full-time students to evaluate the performance of instructors in the university.

Some concerns about POR are about the potential for contamination by personal biases and unprofessional behavior. Biases might reflect the prior reputation of the faculty member and a preference for one teaching style or philosophical orientation over another (Kreber, 2002). Morrison, Sweeney, and Heffernan (2006) in a recent review of the literature on learning styles and teaching effectiveness commented that there is controversy about the efficacy of using a single teaching style perhaps appropriate to most but not all students, versus using multiple teaching styles to increase teaching effectiveness for a wider group. They concluded that instructors should build skills across all styles as students can be trained to develop versatile learning styles. Biases may also reflect stereotyping based on gender, race, English as a second language, and so forth (Centra & Gaubatz, 2000; DeZure, 1999; Martin, 1984). A less obvious bias from POR may emanate from resentment over salary inequities, especially salary compression and salary inversion where senior faculty members evaluate more recently hired and less experienced faculty members who earn higher incomes (Lamb & Moates, 1999).

Jai and Lackritz (2001) enumerated some of the more common concerns about POR. These include beliefs that the characteristics of good teaching are too elusive to measure and that evaluations allow for bias toward one teaching style. Faculty members also are concerned about lack of training on the part of evaluators, the potential for atypical performance on the day of evaluation, that classroom visitations are intrusive and awkward, and that POR carry an implicit threat. There is also the possibility that because peers feel a professional responsibility to provide feedback for improvement and attempt to offer constructive comments, these comments may be interpreted as negative and critical, thereby eroding a sense of collegiality. Studies from Webster (1990) and Lammers and Kirchner (1985) found that on scale items for which ratings differed across peer evaluations, student evaluations, and self-evaluations, peers consistently gave the lowest ratings when evaluating teaching performance.

Evaluations of Teaching and Academic Freedom

Finally, concern exists that both SET and POR may be used as tools for exercising administrative power and political/institutional control over faculty, and that the imposition of such evaluation threatens academic freedom. Using Foucault's (1977, 1980) ideas on discipline and power, Wicks (2004) conceptualized how individual instructors lose their academic freedom and argued that the academic freedom of untenured faculty is rendered particularly sensitive and vulnerable to measurement and classification that can be used to rationalize rule-like administrative decisions. As argued by Wicks, "There is a side to tenure that is not widely discussed—one that places faculty under close scrutiny, discourages academic freedom and increases managerial control" (p. 625). Similarly, Haskell (1997a, 1997b) argued that the widespread use of SET, particularly in salary decisions and in retention, tenure, and promotion (RTP) decisions, infringes on the instructional responsibilities of faculty by providing a control mechanism over curriculum, course content, grading, standards, and teaching methodology.

Summary

The literature on both student evaluations of teaching (SET) and peer observation reports (POR) offers serious concerns about their use. Concerns about SET largely focus on questions about validity and apprehension over the common emphasis on SET in the retention, tenure, and promotion process. The growing use of POR largely stems from these concerns. It is believed that peer observations may be more valid than SET for evaluating substantive and pedagogical content. On the other hand, POR are also criticized as being vulnerable to personal biases and subjectivity. Other issues about POR include lack of training on the part of evaluators and the potential for atypical performance on the day of evaluation. Last, both measures have potential for abuse and may be improperly used to limit academic freedom.

Method

The study reported here is an exploratory inquiry into the perceptions of university faculty members about student evaluations of teaching (SET) and peer observation reports (POR). Most of the existing studies related to this subject have focused either just on SET or just on POR, or they have used a survey method with predetermined response options. We chose to use a qualitative method in order to explore new ideas and to allow for a broader interpretation of individual perceptions of who are the real experts in evaluating teaching effectiveness.

Interviews were conducted with eight professors, a number found often to be sufficient when conducting in-depth interviews (McCracken, 1988). In formulating the research design, we decided to conduct a minimum of eight interviews and then conduct additional interviews if necessary to the point where the information revealed became redundant. In conducting the interviews, we found considerable redundancy as early as the sixth interview. We decided to stop at the eighth interview as little new information was being added.

Informants included both junior and senior faculty members and instructors teaching in both undergraduate and graduate programs. The informants had held positions at research- and comprehensive/teaching-oriented institutions, at large public and private universities, and at smaller private colleges and universities. Sampling was purposeful to obtain diversity of experience levels, academic rank, and perspectives (McCracken, 1988; Strauss & Corbin, 1990). Pseudonyms are used. A brief profile of each informant is provided in Table 2.

All interviews were conducted by one author, tape-recorded with permission, and transcribed word for word. Interviews lasted an average of about an hour (ranging from about 45 minutes to about 90 minutes). Seven of the eight interviews were conducted face to face, and the eighth (with Lea) was conducted by phone due to distance. In all cases, the peer evaluations discussed refer to evaluations of teaching by tenured colleagues in the same academic department. Without exception, tenured faculty informants (Ed, Sue, Jan, Amy) had been evaluated by peers and students as part of the retention, tenure, and promotion (RTP) process prior to tenure, and this continued at least yearly following tenure while under consideration for promotion from associate to full professor. After their promotions, the two full professors (Ed and Sue) had

Table 2
Informant Profiles

<i>Pseudonym</i>	<i>Profile</i>
Ed	Tenured full professor in the field of marketing, with nearly 3 decades teaching experience in a large, public comprehensive university in the U.S. west. Teaches in both undergraduate and graduate programs. Prior teaching experience at a research-oriented university in the U.S. Midwest. Approximate age: Late 50s.
Sue	Tenured full professor in the fields of marketing and management at a large, comprehensive public university in the US west. Nearly two decades total teaching experience, including lecturer positions at a small, private liberal arts university and a large, public research university in the US west. Teaches in both undergraduate and graduate programs, with primary emphasis at the undergraduate level. Approximate age: Mid 40s.
Lea	Clinical professor in the fields of management and marketing at a poly-technic university in the US northeast. Nearly three decades total teaching experience. Prior teaching experience includes a tenure track position at a small, private liberal arts college on the US east coast; and lecturer positions at large, private and public research and comprehensive universities in the U.S. west. Has taught in both undergraduate and graduate programs, with primary emphasis at the graduate level. Approximate age: Late 40s.
Jan	Tenured associate professor in the field of marketing with 9 years in tenure track and tenured positions at a large, public comprehensive university on the US west coast; and approximately 4 years previous lecturer positions at large, public comprehensive universities in the US west. Has taught in both undergraduate and graduate programs, with primary emphasis at the undergraduate level. Approximate age: Late 40s.
Ann	Tenure track assistant professor in 3rd year at a large, comprehensive public university on the US west coast. Has nearly 2 decades prior teaching experience in the field of marketing in a tenure track position (large private research university in the US west) and in multiple visiting scholar and lecturer positions (at several large and small public and private research institutions in the US northeast, on the US east coast, and in Europe). Teaches in both undergraduate and graduate programs. Approximate age: Early 40s.
Amy	Newly promoted and tenured associate professor in the field of finance. In 7th year at a large, public comprehensive university in the US west. Has taught in both undergraduate and graduate programs, with primary emphasis at the undergraduate level. Approximate age: Late 30s.
Liz	Tenure track assistant professor in the field of law. In 5th year at a large, public university on the US west coast. Has also taught undergraduate management courses. Approximate age: Late 30s.
Dan	Tenure track assistant professor in the field of marketing. In 5th year at a large, public comprehensive university on the US west coast. Previous teaching experience as a lecturer at both a large, private research-oriented university, and at a small, private liberal arts university in the US west. Teaches in undergraduate and graduate programs. Approximate age: Early 40s.

been evaluated at least yearly by students and every five years by peers as part of the post-tenure review process. The current assistant professors (Ann, Dan, Liz) were evaluated every term by both students and peers. The clinical professor (Lea) and every informant who had held non-tenure track lecturer positions had been evaluated every term by students while in these positions. Most lecturers were evaluated by peers as well. All who are of senior rank (associate

or full) had evaluated their peers following tenure and promotion. Those who were recently tenured (Jan and Amy) commented on the sudden transition from being evaluated to being an evaluator.

The interviews began with the author who served as interviewer briefly explaining that the purpose of the study was to gain insight into faculty perceptions of SET and POR. Informants were then asked to discuss their experiences with SET and POR, particularly focusing on what they viewed as benefits, limitations, and problems. Interviews were minimally structured so as to allow informants to broach topics in their own ways. Following the initial general question and discussion, if the topics had not already been covered, follow-up questions asked informants: "Which type of teaching evaluation, if either, do you believe better assesses the quality of teaching?" "What do you believe college students look for when evaluating teaching, and what do you believe faculty peers look for in evaluating teaching?" Finally, all informants were asked to respond to whom they believed to be the real experts in judging the quality of instruction—faculty peers or students. Informants were asked to elaborate on their answers and explain why they felt as they did.

The transcribed interviews were analyzed as an iterative process of coding, categorizing, and abstracting data as outlined in guides for conducting qualitative and interpretive research in general, and in-depth interviews in particular (McCracken, 1988; Miles & Huberman, 1984; Strauss & Corbin, 1990). Transcripts were read and coded at an individual level, then subjected to comparison and additional coding and categorization in the light of subsequent interviews. Themes were identified through this iterative whole-to-part and part-to-whole process. We continued analysis and discussion among ourselves through several iterations until consensus was reached. Last, conclusions were submitted to informants for validation.

Results

Several broad themes emerged from the interviews. The first pertains to whether faculty peers or students were the experts in assessing teaching quality. The second and third are about the advantages and disadvantages of student evaluations. The final themes examine the advantages and disadvantages of faculty peer evaluations. Each broad theme is associated with more specific issues as summarized in Table 3 and discussed below.

Who are the Experts: Students or Faculty Members?

Informants generally considered faculty members to be the experts, but acknowledged that students also provided useful assessments of teaching. Informants noted that faculty were experts in their fields and also may have given considerable thought to what constitutes effective teaching and pedagogy. On the other hand, students were the recipients of instruction, and therefore might bring another type of expertise.

Faculty members are the experts. Seven of the eight informants flatly stated that overall, faculty peers were the experts in evaluating the quality of teaching. Ed, a tenured full professor with nearly three decades of teaching experience, said, "I'd say the faculty, the peers, are the experts and the students are not." Sue, a tenured full professor with nearly two decades of teaching experience, expressed concern that students currently enrolled in a course did not yet

Table 3
Major Themes and Subthemes

Who are the Experts—Students or Faculty Members?

- Faculty members are the experts
- Faculty members are the experts in evaluating content, students in evaluating delivery
- Students are experts at assessing their own response to a class
- Faculty members are the experts, but motivated students can also be experts
- Student expertise can only be captured by asking the right questions

Advantages of Student Evaluation

- Students know best what was effective for their learning
- Students observe the whole class
- Student samples are larger

Disadvantages of Student Evaluation

- Student evaluation of teaching is affected by grades
- Pressure to please students
- Based on entertainment value
- Intentional manipulation

Advantages of Faculty Peer Evaluation

- Peers bring professional knowledge and experience
- More objective and rational than student evaluation
- Helpful for faculty development
- Need to separate evaluation and development

Disadvantages of Faculty Peer Evaluation

- Unintentional bias
 - Intentional bias
 - Atypical performance
-

have the perspective to evaluate the course. She went on to say that many aspects of learning were not appreciated or understood by students until they progressed to more advanced courses where it was apparent that they were building on knowledge and skills learned previously. Further, some aspects of learning might not be appreciated until put into practice at work. This is one reason why she values the assessments of alumni more than assessments of current students. She concluded, "I don't think the students are the best indicator of good teaching. Not at the time when we do evaluations."

Similarly, Jan (a tenured associate professor with more than a decade of teaching experience) and Ann (an untenured assistant professor but with nearly two decades of full-time teaching at various institutions) both said that while they were taking a course, students were often unaware of how much they were learning. Both these instructors and others like Sue and Lea (a clinical professor with nearly three decades of teaching experience at several institutions) said that this was particularly true when an instructor maintained a high level of rigor and the students were required to put forth substantial effort, yet expected that they would not earn the grade they wanted. Lea and Sue were both strongly convinced that rigor could negatively affect teaching evaluations, but that students often appreciated it later when they realized that they had learned something of value.

Faculty members are the experts in evaluating content, students in evaluating delivery. More specifically, informants typically (seven out of eight) considered

faculty peers to be the experts in assessing course content, whereas they felt that students were more qualified to assess the quality of delivery. Ed recalled an example to illustrate his belief that students were not experts in evaluating content:

There was a study where they had two professors or presenters come into a class and the first presenter was very factual and had good content but he was very dry in his delivery and then the students rated him. The second professor was void of content but he had lots of enthusiasm, a very interesting character, and the students rated him much higher. And so that kind of summarizes the way I view it.

These informants uniformly expressed concern that students taking a course did not yet have a complete basis for judging content. They noted that students did not yet know what the content of the course should be. Further, they could not compare how one instructor taught the course with how another would teach it or evaluate the appropriateness of the pedagogy. For example, according to Ann,

The students aren't experts in evaluating the quality of the content. They're just not in a position to do that. They're learning it for the first time and they're taking the professor's word for the fact that this is the most important stuff ... and the correct way to interpret it, [and] this is the most up-to-date information. They just take our word for it, so they can't assess the quality of the content. Our colleagues are in a better position to do that.

On the other hand, these same informants felt that because they were experiencing the class as learners, students did have expertise in evaluating teaching delivery. For example, Dan, a tenure track assistant professor in his fifth year of full-time teaching, said, "Students are the experts [at] assessing teaching delivery. They have more recent and more intense exposure to a range of teaching styles than any faculty peer would." Sue felt that students could evaluate clarity, "[They can see] either you're able to explain it or you're not." She observed that students as learners knew if the material was presented to them understandably.

Students are experts at assessing their own response to a class. Six of the informants opined that students knew whether they felt engaged and whether they felt that they were learning. Jan pointed out, "Students are the experts on how the class affected them. Were they interested? Did they become interested? Was the instructor helpful? Did the instructor help to make things clear?" Similarly, Sue said that students knew, "Either I can understand you, or I can't." Thus, she concluded, students are experts at accessing their own responses to a particular faculty member's style of teaching.

Faculty members are the experts, but motivated students can also be experts. Four informants qualified their views that faculty members were the experts, saying that especially motivated students also could be experts. Amy (a recently tenured associate professor with approximately seven years teaching experience) said, "If there is a really hard working student, following step-by-step [my instructions in the course] then their input would be the most valuable because they know [which] areas are not good enough." Ed added, "It depends on the student. Some students are very concerned about the content, learning."

Thus although perhaps not all students in a course are considered experts, these faculty members do value the assessments of students who are conscientious, motivated, and fully engaged.

Student expertise can only be captured by asking the right questions. Finally, two of the informants offered the opinion that although students did have sufficient expertise to judge some aspects of teaching, SET instruments were not always designed to capture their insights effectively. Lea (a clinical professor with nearly two decades teaching experience) said, "If students are given an instrument that asks appropriate questions, students can comment on things like: Was the professor prepared? Was the professor excited about the material? They can clearly comment on what happened in class." She and others felt, however, that SET instruments too often encouraged students to focus away from learning. Lea expressed the view that SET questions frequently overemphasized such factors as how charismatic, entertaining, witty, kind, or lenient the professor was.

Advantages of Student Evaluation

Students know best what was effective for their learning. The most commonly mentioned advantage of SET among this group of informants, six of whom commented on it, was that students could evaluate what worked and what did not work for themselves as learners. Informants commented that SET thus provided useful direct information that POR could not provide. Ann, currently a tenure-track assistant professor who had previously held another tenure-track position and several nontenure-track visiting and lecturer positions at research-oriented and comprehensive private and public universities, explained that in encountering a new group of students,

I always think of the students as the target market and I am always thinking about [how I can] tailor my message to the market. So, what the student evaluations tell me is whether I actually had a good indication of what my target market was [and] did [my teaching] resonate with them.

In contrast, she felt that faculty peers, "may or may not be able to tell ... whether it is resonating."

Jan said, "Student evaluations have the advantage that they can measure how students reacted to the class and to the instructor as a person facilitating their learning." She went on to note that when SET were administered to an entire class, they could provide insight into how the typical student responded to the course and provided a distribution of responses to each item included on the instrument.

Students observe the whole class. Five informants pointed out that whereas POR were typically limited to the observation of one class period and that one class period could be atypical, students observed the class over the entire term. Ed explained, "Students get to see the whole picture. They get to see every meeting. They get to see how the class is organized every time." Dan said, "Students see what we do day after day, in a much more comprehensive way than peer evaluation." Three informants (Ed, Sue, and Jan, all senior faculty members who conduct peer observations) pointed out that a faculty member being evaluated might take care to be particularly well prepared with interesting material and class activities on the day of the peer evaluation. However, Jan

also pointed out that the faculty member might be less spontaneous and engaging because he or she felt nervous under the pressure of being evaluated. In either case, it was felt that a peer evaluator might be unable to provide a valid, holistic appraisal of teaching effectiveness.

Student samples are larger. Five of the informants also cited as an advantage of SET the relatively large numbers of students who evaluate an instructor. Whereas POR typically involve one observer, SET reports involve an entire class or at least all those attending class on the day of the evaluation. Thus SET in the aggregate may be less prone to individual biases. Jan pointed out that SET allow instructors to see the distribution of responses among students, whereas a peer evaluator's opinion could be more idiosyncratic. Liz, a tenure-track assistant professor in her fifth year, said:

The student who has an axe to grind against the professor could take it out [on him/her], but the fact is that in the average class [there are] about 40 students who are responding to this.... That's going to wash out and you're going to have also the medians and the averages.

Disadvantages of Student Evaluation

Student evaluation of teaching is affected by grades. The most common concern expressed about SET by all eight informants was the influence of grading and grade expectations. Some mentioned studies that had found correlations between SET and individual grades or grade distributions. Lea reflected, "Students tend, unless the evaluation instrument is very well designed, to answer on the basis of 'did I like them?' And 'did I like them?' has largely to do with 'am I going to get an A from them?'" Jan said, "Students tend to relate [their evaluation] to the grade." She said that faculty members might feel tempted to grade more leniently, especially before tenure, because they felt pressured to improve or maintain their student evaluations. She speculated that this might be a root cause of grade inflation. Ann offered that students might view grading and responses to teaching evaluations as reciprocal, where students who expected to receive a high grade gave the instructor high ratings on the evaluation instrument, and students receiving lower grades gave lower ratings.

Pressure to please students. Closely related to the concern that evaluations were affected by grades, four of the informants stated that because SET were used in the retention, tenure, and promotion process, instructors felt pressured to please students or make them feel good rather than giving an honest assessment of their academic performance. For example, Jan said,

I am quite certain that instructors do feel pressured to please the students. That is good if it translates into being committed, interesting, clear, helpful.... But it is bad if it translates into pandering, grade inflation, lowering standards, not feeling free to express warranted critique of written work, not feeling free to express warranted critique of lazy or sloppy thinking in class discussions, not feeling free to express warranted critique of lack of preparedness.... That impedes learning.

Dan offered the alternative perspective that the desire to please students, "may be good because students generally do not work as hard if they are irritated with an instructor. And if they do not work as hard, they will not learn

as much.” In contrast, Lea expressed the prevalent perspective that SET, “can place professors at a disadvantage in the evaluation process [when they] want to push or challenge students past their comfort zones.”

Based on entertainment value. Four informants criticized SET for rewarding the entertainment value of the course or the professor’s teaching style more than the amount of learning. These informants believed that peers were more likely to evaluate the substantive content of a course, whether or not the content was delivered in an entertaining style. For example, Ann said,

I think the entertainment value is obviously weighted very heavily in the evaluations, and so I know that people who are very content-heavy in their lectures, who have a hard time also bringing in some sort of entertainment value to their lectures, are going to get knocked [down] by the students even though those are really very important lectures.

These informants believed that professors who were witty and colorful were more likely to receive high teaching evaluations, even if little learning took place. They all acknowledged that an entertaining style of presentation could enhance learning; however, concern was expressed over the potential of SET to reward instructors who emphasized form with little content.

Intentional manipulation. One informant expressed concern that instructors could intentionally manipulate student evaluations. Ed explained, “[Instructors] can do things before the evaluations.... There are cases where the professor has given a party [or] where they say, ‘now I’ll offer extra credit where I didn’t before.’” In his view, this represented real potential to compromise the validity of SET, and he believed that faculty peers were typically more objective than students, or at least less susceptible to such overt manipulations.

Advantages of Faculty Peer Evaluation

Peers bring professional knowledge and experience. Seven of the eight informants spoke of specific areas of faculty expertise as an advantage of faculty peer evaluations. They felt that faculty expertise in evaluating teaching effectiveness was particularly salient when departments had identified agreed-on learning objectives or pedagogies. They felt that the ability to evaluate whether such learning objectives were being achieved or whether appropriate pedagogies were being effectively used was simply beyond the expertise of most students, but that faculty peers were sensitized to such norms. Further, they expressed that faculty peers were experts in their fields and could judge better than students the substantive content and the currency of material covered in a class. Last, they expressed the opinion that faculty peers could offer constructive feedback based on their own experience of what worked. For example, Sue said,

Faculty peers know, or should know, what your department wants you to be doing. So they know what sort of standards you should be trying to achieve. They would [also] be more cognizant of the fact that there are different learning styles, and hence different teaching styles are appropriate.... A student isn’t going to be thinking about those things.

Dan explained, “Experienced peers know teaching techniques and things that they or other faculty members have tried that can help improve my teaching. They can say, ‘This is what I tried that helped my students understand this

difficult topic.” Ed pointed out that faculty peers were knowledgeable about course content:

The advantage of [peer evaluations] is that you have a senior faculty member who has taught the same class, who knows what the content should be and can see whether that content is being delivered. The students don’t have that ability. They just know what looks good to them.

More objective and rational than student evaluation. Six informants saw POR as more objective and rational than SET and cited this as an important advantage. For example, Lea said, “Faculty do evaluations coming at it in a more objective way, [more] professional.” Ann reflected, “A peer can see beyond emotion [and is] much more rational about the evaluation [so there is] a focus much more on assisting the recipient.” She explained that, for example, a student might evaluate a professor based largely on his or her affect toward the instructor’s personality or style, or give the professor roughly the same grade as she or he earned on the last exam or paper. These informants felt that evaluations by faculty peers were less likely to be colored by such emotional responses.

Helpful for faculty development. Four informants most positive toward POR generally cited its usefulness as a tool for faculty professional development. They offered that POR provided a fresh perspective on one’s teaching. According to Ann,

Students don’t give feedback that helps in the development of a faculty member in terms of their teaching style. [SET] can help in terms of what students like and don’t like [but] faculty members are much better [by comparison] in giving feedback that helps in terms of teaching style.

Jan characterized POR as:

The perspective of someone who has been in the trenches themselves.... A good peer evaluator would have empathy for both the students and the instructor, and give the instructor some ideas for how to possibly handle some situations better.

Dan pointed out that a more experienced faculty member may be able to offer advice that can help a junior colleague concerned about how to teach a particularly difficult or confusing concept or module.

Need to separate evaluation and development. Although informants generally valued POR both for development and for evaluation, Sue and Ann both talked of the desire for more emphasis on development separate from the retention, tenure, and promotion process. For example, Sue said,

For RTP it’s important to find out what peers think, and for faculty development it’s helpful to get other people’s perspectives ... but it would be nice if those weren’t always counted ... in the RTP process, meaning it’s nice if it’s not formal.

These two informants expressed a desire to separate evaluation from possible punitive consequences to seek out peer evaluation freely as a tool for gaining insight into how to improve teaching quality. Ann described with appreciation a productive experience she had at a university that offered this option.

Disadvantages of Faculty Peer Evaluation

Unintentional bias. Although informants said that POR were typically more objective than SET, serious concerns were expressed about the potential for bias and unfairness in the faculty peer review process. Six informants pointed out that POR were vulnerable to bias, although this bias might be unintentional. For example, Amy said, "The limitation is each professor has their [own favored] approach, their own point of view. They lose the perspective that there is some other way, some other approach." Similarly, Sue said, "If [the peer is] not open to the fact that there are different styles ... if you don't do something exactly the way they do it, and therefore you're wrong, that would be a disadvantage." Jan pointed out that with the growing emphasis on active learning pedagogies, a faculty member delivering a more traditional lecture might be given a less positive evaluation based on the preferences of the evaluator rather than on the quality or content of the lecture.

Intentional bias. Five informants also cautioned that intentional bias did sometimes exist and could seriously contaminate the POR process. They speculated or recalled from an experience they had witnessed that personal biases and conflicts could overpower objectivity, and POR could be used as a means of castigation or for retaliation. Liz spoke from personal experience, saying,

Peer evaluations can be used by the wrong people as tools of retaliation. [Another faculty member with whom I had a conflict] gave me a peer evaluation that was wholly unfair. [The faculty member] was saying things happened in the classroom that didn't take place. Luckily I had student witnesses and their notes that basically conflicted with [the evaluator's] version.

Ann pointed out that intentional bias can also unfairly favor a faculty member:

It can be a tool of retaliation [and] it can also be a tool used to promote someone who doesn't necessarily deserve it. So, in other words, peer evaluation in the RTP process can be steeped in the very office politics that the RTP process should rise above. So it can really muddy the waters, I think, for RTP.

Atypical performance. One junior and two senior faculty members pointed out that classroom performance under observation was sometimes atypical. Ed talked of a potential problem associated with giving advance notice: "The faculty member knows [the classroom observation] is going to occur and when it is going to occur, so [he or she] can put forth a special effort to make something look good during that one session." By contrast, Jan pointed out that observation was stressful and recalled being, "so worried about making a mistake [I was] less spontaneous and less effective than usual." In either case, the session under observation might be uncharacteristic of the faculty member's usual performance.

Discussion

The perceptions of informants interviewed in this study raise concerns about reliance on a single source of information as evidence of teaching effectiveness. Informants' responses indicate that student evaluations of teaching (SET) and peer observation reports (POR) seem to provide complementary information. In general, informants viewed POR as a better source of information for faculty development and for assessing course content and pedagogy, which agrees

with the concerns in the literature. On the other hand, informants believed that SET provided a better assessment of the quality of what goes on throughout the entire term and in interactions with students, extending the findings of the literature in this area.

Are peers or students the experts in evaluating teaching? An overwhelming majority of those interviewed (seven of eight informants) felt that faculty peers were the experts, especially for evaluation of content. On the other hand, it was also suggested that the number of observers and quantity of observations in SET could offset the inherent disadvantages of individual student evaluations.

The results suggest that POR is perceived as superior to SET for evaluating the substance and currency of subject matter in a class. Further, constructive feedback from the evaluator can offer a fresh perspective as well as useful feedback on teaching methods. Such feedback is useful for faculty development purposes (Lam, 2006). Further, peer evaluators are well qualified to review syllabi, assignments, examinations, and other teaching materials to ensure that departmental standards are upheld (North, 1999). They are more aware of the value of innovative pedagogies and of important educational objectives such as helping students develop critical thinking and problem solving skills (Buskist et al., 2002).

Although informants expressed several concerns about SET, they also valued SET. Informants suggested that SET could provide a valuable perspective on teaching effectiveness from the point of view of the learner. Informants said that students were best equipped to provide feedback on the quality of faculty-student interactions, the clarity of instruction, resonance and interest, and perceived fairness and workload. On the other hand, informants expressed the concern that students were less capable of evaluating the aspects of teaching that occur outside the classroom such as preparation and maintenance of currency and may not be attuned to the importance of achieving desired learning outcomes or using appropriate pedagogical techniques (Cashin, 1999; Friedman, Rodriguez, & McComb 2001; Schaeffer et al., 2003).

In order to address the perceived shortcomings of POR and SET, we present the following suggestions. These are derived from a synthesis of the literature, the critical recommendations from our informants, and from our own experiences with SET and professional development through POR.

Engage the faculty in developing the POR program. As indicated by our review of the literature and by the concerns expressed among our informants, considerable uneasiness and distrust appears to be associated with POR. Thus we suggest that for a program of peer evaluation to succeed, faculty members who both evaluate and are evaluated should feel ownership of the process (Van Dyne & Pierce, 2004). Those who will be evaluated should feel that they have more to gain than to lose from peer evaluation and should view it as an opportunity to obtain valuable feedback to improve teaching. To achieve these ends, the purposes of peer evaluation should be clearly defined, and faculty members should be involved in developing the peer evaluation program and should also control its implementation (Peel, 2005). The formative benefits of evaluation should be reinforced on an ongoing basis, and when possible separated from administrative retention, tenure, and promotion decisions. Kohut et al. (2007) found that both faculty members who serve as peer ob-

servers and faculty members who are observed valued pre- and post-observation meetings as a way for faculty to participate in the process and contribute guidelines for their academic units.

Train the evaluators. Interviews indicated that those being evaluated sometimes question the capability and objectivity of the evaluator. Further, faculty members who serve as peer evaluators are sometimes reluctant or feel unprepared to assess other faculty members' performance (Kohut et al., 2007). In our study Jan, who had recently been tenured and taken on the role of evaluator, discussed this at some length.

Both situations can be ameliorated with training. We suggest that training should provide clarity on the purposes of peer evaluation and on common pitfalls to avoid. Seldin (1999a) suggests that training programs should include such components as what to look for, the mechanics of the program, how the results will be used, how teaching evaluation leads to teaching improvement, and the responsibilities of evaluators. Similarly, we suggest that students might be better informed on how to complete SET and more encouraged to offer practical and objective critiques rather than merely indicate their satisfaction with the instructors' personality or the difficulty of the course and the instructors' grading. If students perceive that their comments will be directly used to develop future teaching improvements, they may be more inclined to give thoughtful and constructive feedback.

Assure fairness and confidentiality. Informants expressed concern over both unintentional and intentional bias. It is essential that faculty peer evaluation programs used for retention, tenure, and promotion decisions comply with all relevant employment law, including ensuring protection from discriminatory and retaliatory practices (Lamb & Moates, 1999; Seldin, 1999a). However, even with clearly communicated standards and effective training, unfairness may exist. Thus the process should be monitored and corrective action taken when personal bias or injustice is evident (Cosh, 2002).

Build a trusting and constructive climate. Perhaps most important, Higgerson (1999) proposed that building a climate of collegiality and mutual respect was essential to an effective POR program. In the absence of a collegial climate, almost any evaluative effort may lead to further distrust and dissent (Shortland, 2004). POR will be viewed at best as a regulatory annoyance and at worst as a punitive tool. On the other hand, evaluative comments will be more favorably received and formative benefits more likely realized in an already constructive and collegial environment (Kohut et al., 2007).

Conclusions

A limitation of the qualitative method used in this study is that the results are not generalizable. In-depth interviews were used as an exploratory tool, and our findings provide the groundwork for further study. We recommend that future research use more quantitative techniques to determine the extent to which the views of faculty informants in this study about SET and POR are held among faculty members.

Although revealing some interesting perspectives, this study is only the beginning of what could be an ongoing inquiry into student versus peer evaluations of teaching effectiveness. This study found that although the faculty informants generally perceive faculty peer evaluators to be experts in

evaluating teaching, these same faculty members do not always regard peer evaluations as more accurate than student evaluations. In fact in some cases, and in evaluating some aspects of teaching, POR are viewed as less accurate because, unlike students, faculty peers do not experience the class day to day as learners. It is important to note that this conclusion is at least partly drawn from informants' observations of their own skill levels in evaluating others. Other measures may reveal even less positive perceptions about faculty peer evaluations of teaching. In addition, understanding the prevailing attitudes among faculty members, students, and administrators may foster better acceptance of both student and peer evaluations in appropriate contexts.

Our findings suggest that both instructors and students can be experts, and the relative consequence of each group's expertise depends on what is being evaluated and the purposes for which the evaluations are collected. This is consistent with findings from earlier research. A key contribution of this study is that the findings underline the importance of using a composite evaluation of teaching effectiveness, acknowledging the expertise of both students and faculty peers. As indicated by the responses of our informants, the focus and emphasis of SET and POS are different and complementary in that both have advantages and both have shortcomings. Thus overreliance on one versus the other will lead to evaluations of teaching effectiveness that are at best incomplete.

Considering the advantages and disadvantages of SET and POR raises the issue of standards and expectations. A problem with both student and peer evaluation is that there is no objective standard against which evaluators judge teaching. Even highly motivated students can only base their evaluations on the teaching they have received in the past, which can vary substantially depending on the college or university and the individual student's experience and progress toward a degree. For example, students in small colleges may encounter fewer faculty members and a narrower range of faculty teaching styles than do students studying in large universities. Similarly, graduate students and students nearing the completion of their undergraduate degrees have encountered more instructors and probably a greater diversity of teaching styles and skills than have most freshmen and sophomores. Faculty peers, especially if they are recently tenured and just starting to serve as peer evaluators, may have an even more limited recent basis for comparison of one instructor with another. On the other hand, the more closely faculty members are held to universal standards, the more academic freedom would be threatened.

We recommend that future research compare SET and POR with objective assessment standards (e.g., measures of student achievement over time on embedded measures) to determine the relative effectiveness of student versus peer evaluation of teaching. In addition, the use of professional evaluators (e.g., business practitioners in business education, experienced elementary school-teachers in teacher education, engineering practitioners in engineering education) may bring a different perspective on the criteria and measurement of teaching effectiveness as compared with either SET or POR.

Another issue suggested from the perceptions of our informants is that student evaluations are effective when they are done internally in a systematic

and reliable way. The perceived validity of SET among our informants largely stems from the number and frequency of student observations. That is, an entire class evaluates an instructor over the course of an entire semester or term, culminating in one SET report. By contrast, the evaluations posted on Internet rating sites (e.g., ratemyprofessor.com, pickaprof.com) may consist of nothing more than responses from a few disgruntled students who received low grades. Although these sites do generally contain both positive and negative evaluations, students motivated to go on these sites on their own time to post evaluations are often outliers in one way or another. They may be particularly pleased, particularly disgruntled, or particularly interested in voicing their opinions. Such rating sites rarely capture the evaluations of the majority of students who fall into none of these categories.

One implication of our findings may be that as students are voluntarily going to Internet rating sites to post their opinions, and as universities are now required to allow instructors' grading distributions to be published on such sites (e.g., pickaprof.com has recently won court cases based on the Public Records Act against US universities that attempted to withhold such information), it may be in the best interest of everyone for universities to allow SET records to be published. One advantage of SET identified in our study is that they represent the views of the entire class. Because of this, published SET would provide more accurate information for potential students choosing courses and instructors than do the ratings of just a handful of students reflecting particularly negative or positive points of view, as is typically found on private online sites.

References

- Betoret, F.D. (2007). The influence of students' and teachers' thinking styles on student course satisfaction and on their learning process. *Educational Psychology, 27*, 219-234.
- Blackhart, G.C., Peruche, B.M., DeWall, C.N., & Joiner, Jr., T.E. (2006). Factors influencing teaching evaluations in higher education. *Teaching of Psychology, 33*(1), 37-39.
- Buskist, W., Sikorski, J., Buckley T., & Saville, B.K. (2002). Elements of master teaching. In S.F. Davis & W. Buskist (Eds.), *The teaching of psychology: Essays in honor of Wilbert J. McKeachie and Charles L. Brewer* (pp. 27-39). Mahwah, NJ: Erlbaum.
- Cashin, W.E. (1999). Student ratings of teaching: Uses and misuses. In P. Seldin (Ed.), *Changing practices in evaluating teaching* (pp. 25-44). Bolton, MA: Anker.
- Centra, J.A., & Gaubatz, N.B. (2000). Is there gender bias in student evaluations of teaching? *Journal of Higher Education, 70*(1), 17-33.
- Chandler, T.A. (1978). The questionable status of student evaluations of teaching. *Teaching of Psychology, 5*(3), 150-152.
- Chonko, L.P., Tanner, J.F., & Davis, R. (2002). What are they thinking? Students' expectations and self-assessments. *Journal of Education for Business, 77*(5), 271-281.
- Churchill, L. (2006). Professor Goodgrade: Or how I learned to stop worrying and give lots of A's. *Chronicle of Higher Education, 52*, 25.
- Clayson, D.E. (2004). A test of the reciprocity effect in the student evaluation of instructors in marketing classes. *Marketing Education Review, 14*(2), 11-36.
- Clayson, D.E., & Haley, D.A. (1990). Student evaluations in marketing: What is actually being measured? *Journal of Marketing Education, 12*(3), 9-17.
- Clayson, D.E., & Sheffet, M.J. (2006). Personality and the student evaluation of teaching. *Journal of Marketing Education, 28*(2), 149-160.
- Cosh, J. (2002). Peer observation in higher education—A reflective approach. *Innovations in Education and Training International, 35*(2), 171-176.
- DeZure, D. (1999). Evaluating teaching through peer classroom observation. In P. Seldin (Ed.), *Changing practices in evaluating teaching* (pp. 70-96). Bolton, MA: Anker.
- Douglas, A., & Douglas, J. (2006). Campus spies? Using mystery students to evaluate university performance. *Educational Research, 48*(1), 111-119.

- Faranda, W.T., & Clarke III, I. (2004). Student observations of outstanding teaching: Implications for marketing educators. *Journal of Marketing Education*, 26(3), 271-281.
- Feldman, K.A. (1989). The association between student ratings of specific instructional dimensions and student achievement: Refining and extending the synthesis of data from multisection validity studies. *Research in Higher Education*, 30(6), 583-645.
- Foucault, M. (1977). *Discipline and punish: The birth of the prison* (A. Sheridan, Trans.). New York: Pantheon Books.
- Foucault, M. (1980). *Power/knowledge: Selected interviews and other writings 1972-77*. (C. Gordon, Ed., C. Gordon, L. Marshall, J. Mepham, & K. Soper, Trans.). New York: Pantheon Books.
- Fram, E.H., & Pearse, R. (2000). "Tough love" teaching generates student hostility. *College Teaching*, 48(2), 42-43.
- Friedman, P., Rodriguez, F., & McComb, J. (2001). Why students do and do not attend classes. *College Teaching*, 49(4), 124-133.
- Gaidis, W.C., & Andrews, J.C. (1990). Management of experiential learning projects in marketing coursework. *Journal of Marketing Education*, 12(2), 49-60.
- Goldstein, G.S., & Benassi, V.A. (2006). Students' and instructors' beliefs about excellent lecturers and discussion leaders. *Research in Higher Education*, 47, 685-707.
- Greenwald, A.G. (1997). Validity concerns and usefulness of student ratings of instruction. *American Psychologist*, 52, 1182-1186.
- Haskell, R.E. (1997a). Academic freedom, promotion, reappointments, tenure and the administrative use of student evaluation of faculty (SEF) (Part IV): Analysis and implications of view from the court in relation to academic freedom, standards, and quality instruction. *Educational Policy Archives*, 5(21), Retrieved June 16, 2007, from: <http://epaa.asu.edu/epaa/v5n21.html>
- Haskell, R.E. (1997b). Academic freedom, tenure, and student evaluation of faculty: Galloping polls in the 21st century. *Educational Policy Archives*, 5(6), Retrieved June 16, 2007, from: <http://epaa.asu.edu/epaa/v5n6.html>
- Heckert, T.M., Latier, A., Ringwald-Burton, A., & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to "buy" better evaluations through lenient grading? *College Student Journal*, 40, 588-596.
- Higgerson, M.L. (1999). Building a climate conducive to effective teaching evaluation. In P. Seldin (Ed.), *Changing practices in evaluating teaching* (pp. 194-212). Bolton, MA: Anker.
- Jai, G., & Lackritz, J.R. (2001). Peer evaluation in the classroom: A check for sex and race/ethnicity effects. *Journal of Education for Business*, 76, 274-282.
- Keig, L. (2000). Formative peer review of teaching: Attitudes of faculty at liberal arts colleges toward colleague's assessment. *Journal of Personnel Evaluation in Education*, 14(1), 67-88.
- Kohut, G.F., Burnap, C., & Yon, M.G. (2007). Peer observation of teaching: Perceptions of the observer and the observed. *College Teaching*, 55(1), 19-25.
- Kreber, C. (2002). Teaching excellence, teaching expertise, and the scholarship of teaching. *Innovative Higher Education*, 27(1), 5-23.
- Lam, T.C. (2006). Group problem-solving among teachers: A case study of how to improve a colleague's teaching. *Social Psychology of Education*, 9, 273-299.
- Lamb, S.W., & Moates, W.H. (1999). A model to access compression inequities in faculty salaries. *Public Personnel Management*, 28, 689-700.
- Lammers, H.B., & Kirchner, D.F. (1985). The construct validity of teaching behavior evaluation methods: A multitrait-multimethod analysis. *Journal of Marketing Education*, 7(2), 35-44.
- Lang, J.M. (2007). Did you learn anything? Students are very accurate judges of the most important question we can ask them about their classroom experiences. *Chronicle of Higher Education*, 53(March 9), 27.
- Marks, R.B. (2000). Determinants of student evaluations of global measures of instructor and course value. *Journal of Marketing Education*, 22(2), 108-119.
- Marsh, H.W. (1991). Multidimensional students' evaluation of teaching effectiveness: A test of alternative higher-order structures. *Journal of Educational Psychology*, 83, 285-296.
- Martin, E. (1984). Power and authority in the classroom: Sexist stereotypes in teaching evaluations. *Signs*, 9, 482-492.
- McCracken, G. (1988). *The long interview*. Newbury Park, CA: Sage.
- McKeachie, W.J. (1997). Student ratings: The validity of use. *American Psychologist*, 52, 1218-1225.
- Miles, M.B., & Huberman, A.M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage.

- Morrison, M., Sweeney, A., & Heffernan, T. (2006). Karn's learning styles and learning effectiveness: A rejoinder. *Journal of Marketing Education, 28*(1), 64-68.
- North, J.D. (1999). Administrative courage to evaluate the complexities of teaching. In P. Seldin (Ed.), *Changing practices in evaluating teaching* (pp. 183-189). Bolton, MA: Anker.
- Olshavsky, R.W., & Spreng, R.A. (1995). Consumer satisfaction and students: Some pitfalls of being customer driven. *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior, 8*(2), 69-77.
- Osborne, J.L. (1998). Integrating student and peer evaluation of teaching. *College Teaching, 46*(1), 136-139.
- Peel, D. (2005). Peer observation as a transformatory tool? *Teaching in Higher Education, 10*, 489-504.
- Schaeffer, G., Epting, K., Zinn, T., & Buskist, W. (2003). Student and faculty perceptions of effective teaching: A successful replication. *Teaching of Psychology, 30*(2), 133-136.
- Seldin, P. (1999a). Building successful teaching evaluation programs. In P. Seldin (Ed.), *Changing practices in evaluating teaching* (pp. 213-242). Bolton, MA: Anker.
- Seldin, P. (1999b). Current practices—Good and bad—Nationally. In P. Seldin (Ed.), *Changing practices in evaluating teaching* (pp. 1-24). Bolton, MA: Anker.
- Share, C.E. (1997). Implications of considering students as consumers. *College Teaching, 45*(4), 122-123.
- Shortland, S. (2004). Peer observation: A tool for staff development or compliance? *Journal of Further and Higher Education, 28*(2), 219-228.
- Simpson, P.M., & Siguaw, J.A. (2000). Student evaluations of teaching: An exploratory study of the faculty response. *Journal of Marketing Education, 22*(3), 199-213.
- Smart, D.T., Kelley, C.A., & Conant, J.S. (2003). Mastering the art of teaching: Pursuing excellence in a new millennium. *Journal of Marketing Education, 25*(1), 71-78.
- Sojka, J., Gupta, A.K., & Deeter-Schmelz, D.R. (2002). Student and faculty perceptions of student evaluations of teaching. *College Teaching, 50*(2), 44-49.
- Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, CA: Sage.
- Van Dyne, L., & Pierce, J.L. (2004). Psychological ownership and feelings of possession: Three field studies predicting employee attitudes and organizational citizenship behavior. *Journal of Organizational Behavior, 25*, 439-459.
- Webster, C. (1990). Evaluation of marketing professors: A comparison of student, peer, and self-evaluations. *Journal of Marketing Education, 12*(1), 11-17.
- Wicks, D. (2004). The institution of tenure: Freedom or discipline? *Management Decision, 42*, 619-627.
- Wright, R.E. (2006). Student evaluations of faculty: Concerns raised in the literature, and possible solutions. *College Student Journal, 40*, 417-422
- Yon, M., Burnap, C., & Kohut, G. (2002). Evidence of effective teaching: Perceptions of peer reviewers. *College Teaching, 50*, 104-110.